Casualty Actuarial Society E-Forum, Fall 2008

Including the 2008 Reserves Call Papers

2008 FALL E-FORUM

Table of Contents

2008 Reserves Call Papers

Additional Papers

2008 RESERVES CALL PAPER PROGRAM

In July 2007 the Casualty Actuarial Society (CAS) and the CAS Committee on Reserves (CASCOR) extended a call for discussion papers on reserving topics. Some of the 17 accepted papers will be presented at the Casualty Loss Reserve Seminar CLRS scheduled for September 18-19, 2008 in Washington D.C.

CAS Committee on Reserves

Thomas A. Ryan, *Chairperson*
Mark R. Shapland, *Vice Chairperson*

Ron Fowler	Thomas R. Kolde	Vladimir Shander
John F. Gibson	ZhenZhen Lai	Stephen James Talley
Aaron M. Halpert	Dale F. Ogden	Mark S. Wenger
Gloria A. Huberman	Susan R. Pino	Ernest I. Wilson

Cheri Widowski, CAS Staff Liaison

E-Forum Committee

Glenn M. Walker, *Chairperson*

Mark A. Florenz	Darci Z. Noonan	Yingjie Zhang
Karl Goring	Zongli Sun	
Dennis L. Lange	Windrie Wong	

Elizabeth A. Smith, CAS Staff Liaison

For information on submitting a paper to the *E-Forum*, visit http://www.casact.org/pubs/forum/.

Manually Adjustable Link Ratio Model for Reserving

Emmanuel T. Bardis, FCAS, MAAA, Ph.D., Ali Majidi, Ph.D., Aktuar (DAV)
and Daniel M. Murphy, FCAS, MAAA

Abstract: The chain ladder method is very popular in General/Property-Casualty Insurance actuarial circles. Mack [1] expanded the deterministic algorithm to include calculations for the variance of the chain ladder projections. The assumptions underlying the chain ladder method are important in regards to the appropriateness of the deterministic projections; they are even more important in regards to the appropriateness of the stochastic results. The purpose of this paper is to introduce more statistical rigor to this popular method and help close the link between practice and statistical theory. We will discuss residual analysis and other statistical measures as they apply to the chain ladder method so that the appropriateness of its deterministic and stochastic results can be objectively measured based on statistically rigorous principles. We will also show how the regression approach of Murphy [2] can be expanded so that link ratios "selected judgmentally" can be seen as conforming to an underlying statistical model.

Keywords: chain ladder; selection; residuals; Mack; Murphy

1. INTRODUCTION

A big part of the actuarial research in the last two decades is dedicated to reserving. While many statistical methods have been dedicated to this problem, none of them is broadly accepted by the practitioners. The aim of this paper is to reduce, or even to close, the gap between practice and theory by embedding this practice into a theoretical flexible framework. The most popular method to solve the central problem of reserving, namely to estimate an "expected value for the outstanding payments=Best Estimate," is the chain ladder method. This is the reason of the popularity of the analysis of Mack [1], where the standard chain-ladder approach is discussed. Murphy [2] considers the more general question of "loss development method," where the chain ladder method is treated as a special case of a more general linear regression approach. Zenwirth [3] calls this family the "extended link ratio family," he criticizes its prediction power and suggests the "probability trend family" instead. However Zenwirth's approach is not consistent with the traditional chain ladder method and the user input associated with this method. The incorporation of user judgement is a typical Bayesian problem, and the approach suggested from Verall [4] is a theoretical rigorous way to tackle the inflexibility of the previous methods. The necessity of the MCMC algorithm (Markov Chain Monte Carlo) in this method makes Verall's approach hard to describe and the basic assumptions of prior distributions for the link ratios are not easily verified.

The purpose of this paper is to present an appropriate model, which

1. Is compatible with the way practitioners implement the chain ladder method and

2. Provides a statistical framework that will help test the underlying assumptions of the chain ladder method (for example for approval of an internal model[1] in Solvency II-context, or the use of benchmarks for the reserving exercise).

In the first section we will propose a model that is built around the regression interpretation of the chain ladder method similar to Murphy [2]. It turns out that a flexible formulation of the chain ladder method along the lines of a regression model satisfies the above-mentioned requirements. Furthermore we will demonstrate how this embedded statistical process can be used to test the appropriateness of the "actuarial selected link ratios" both visually and statistically. Finally we will suggest how to proceed if the approach taken is not appropriate and demonstrate with an example.

2. THE LINK RATIO APPROACH

We start with the usual notation, where the observed cumulative paid losses are denoted by the set $\boldsymbol{D} = \{\boldsymbol{C}_{ij} \mid 1 \le \boldsymbol{i} \le \boldsymbol{I}, 1 \le \boldsymbol{j} \le \boldsymbol{I} + 1 - \boldsymbol{i}\}$. A regression model equivalent to the chain ladder method is

$$\boldsymbol{C}_{ik+1} = \boldsymbol{f}_k \boldsymbol{C}_{ik} + \sigma_k \varepsilon_{i,k} \boldsymbol{C}_{i,k}^{\alpha_k/2} \tag{1}$$

$$\varepsilon_{i,k} \sim \aleph(0,1), 1 \le \boldsymbol{i} \le \boldsymbol{I}, 1 \le \boldsymbol{k} \le \boldsymbol{I} + 1 - \boldsymbol{i} \tag{2}$$

Thereby the set $\{\varepsilon_{ik} \mid 1 \le \boldsymbol{i} \le \boldsymbol{I}, 1 \le \boldsymbol{k} \le \boldsymbol{I} + 1 - \boldsymbol{i}\}$ is assumed to be "noise" or independent identical distributed (i.i.d.) normal[2] random variables with mean 0 and standard deviation 1. Making explicit the implicit assumption of the error term is crucial for assessing the appropriateness of a model because it provides a data set of residuals for model testing. Under these assumptions the least square estimate of the link ratio, given the set of observations $\boldsymbol{D}$, can be calculated through weighted averages of the observed link ratios:

The optimal solution of model (1), (2) is specified by the parameters $(\hat{f}, \hat{\sigma}, \hat{\alpha})$ (the "model specification") where the solution for the values of the αs is discussed below.

2.1 Chain Ladder

The model introduced in Mack [1] is a special case of the model (1), (2) with $\alpha_k=1$, $k=1,\ldots,I$. Mack noted in this model the minimum variance estimator

---

[1] "Proposal for a DIRECTIVE OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on the taking-up and pursuit of the business of Insurance and Reinsurance" http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:52007PC0361:EN:NOT

[2] The normality assumption is made to assure that the chain ladder link ratios correspond to ML estimators. Other distributions can be assumed as well, but that might lead to an ML solution other than the least squares solution.

$$\widehat{f}_k = LR_k(1) = \sum_{i=1}^{n-k} \frac{C_{i,k}}{\sum_{j=1}^{n-k} C_{j,k}} \cdot \frac{C_{i,k+1}}{C_{i,k}} = \frac{\sum_{i=1}^{n-k} C_{i,k+1}}{\sum_{i=1}^{n-k} C_{i,k}},$$

and derived estimators for uncertainty, popularly known as "The Mack Formula." In other words, if we specify a "variance assumption" by selecting the alpha parameter, then the link ratios in this model as well as the uncertainty of the estimators are also selected. This *model* embeds, by making these extensions, the traditional chain ladder *method* in a statistical framework.

Hereby it is important to distinguish between a model and a method. A model is a mathematical description of an observation, phenomenon, etc. and produces "best-fitted" parameters based on the underlying data characteristics. A method, on the other hand, is an algorithm that makes certain assumptions and produces estimates based on a number of predetermined steps. Thus a method can always be used to calculate some estimates, whereas a model is based on assumptions that need to be tested, before the model is used. The traditional chain ladder *method* is "consistent" with many stochastic models that have been created around it, such as the Mack/Murphy Model or the over-dispersed Poisson model. By "consistent" we mean that, given the model that is appropriate for the data on hand, the chain ladder method is a reasonable algorithm to produce reserve estimates that are similar to the estimates of these models. However, actuaries are used to *selecting* link ratio judgmentally because estimated link ratios by averaging methods can be inappropriate in cases when the stochastic component of the loss generation process is made complex by the influence of many unknown and unobserved parameters. An experienced actuary recognizes, for example, trends in the triangles and adjusts the link ratios manually, or uses benchmark pattern instead. There is no doubt that such a manual extension of the model makes sense, but no matter how experienced an actuary is, the appropriateness of his or her selection is always open to question. The model framework of this paper can be used to answer that question with more objective statistical tests.[3]

2.2 Residuals and Model Selection

In the traditional world, actuaries' methods and selections are defended by their expertise and experience. However, mathematical and graphical tools can provide more objective ways to defend their selections and to communicate their answers. One of the most important diagnosis and validation tools are residuals, which are in general the difference between a "data set" and its "formulaic representation." In the chain ladder case, the formulaic representation of the data is given by the specifications of the model parameters.

---

[3] Furthermore, we mention here that the residuals are often used to simulate the distribution of the stochastic reserving process through the bootstrapping approach. The core of the bootstrapping method is the "independent identical" assumption in (2). The bootstrapping results will be wrong if this assumption is violated.

If we reformulate (1) with respect to $\boldsymbol{\varepsilon_{i,k}}$, we obtain

$$\boldsymbol{\varepsilon_{i,k}} = (\boldsymbol{C_{ik+1}} - \boldsymbol{f_k C_{ik}}) / (\boldsymbol{\sigma_k C_{i,k}^{\alpha_k/2}}).$$

This residual assumption can be validated with the data set.

We define the "corresponding" residuals of a model specification $(\hat{\boldsymbol{f}}, \hat{\sigma}, \hat{\alpha})$ by

$$r_{i,k} := r_{i,k}(\hat{f}, \hat{\sigma}, \hat{\alpha}) := (C_{ik+1} - \hat{f}_k C_{ik}) / (\hat{\sigma}_k C_{i,k}^{\hat{\alpha}_k/2}) \tag{4}$$

We start by selecting the parameters in this model and proposing a certain estimate, which corresponds to a hypothesis for the future liabilities that leads to an estimate for the reserves. The question is now, how confident are we in that estimate? Taking (2) and (3) together our chosen estimates need to fulfill the hypothesis

"The data set $\{\boldsymbol{r_{i,k}}\}$ looks like noise."

Although we have a subjective feeling for a data set looking like noise, we could hardly test it without further clarifications. However the hypothesis "i.i.d. normal distributed" can be tested through visual tests (e.g., QQ-Plots) as well as statistical tests (e.g., Shapiro-Francia-test for normality [5]).

Now one can raise the question: What should we do if the test fails? We change the link ratios manually. Of course this is not new. Actuaries have always selected link ratios manually by employing experience, judgment, benchmarks, etc.

Assuming we manually change the link ratios, the next question is: Is the new set of link ratios more appropriate than those selected initially?

In the next sections we describe an approach to answer these questions and show how to use the approach to fine-tune the selected link ratios in a controlled work flow way.

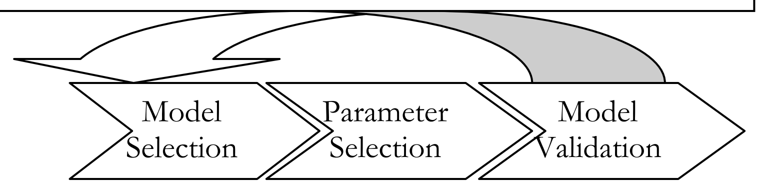
3. SELECTED LINK RATIO MODEL

Consider the regression approach (1) to the chain ladder method. The problem with the common actuarial practice is that when the selected link ratios are not the volume weighted average, then they are not consistent with the best linear unbiased estimators calculated by the statistical models employed in stochastic reserving exercises. In particular non-volume-weighted-average selected link ratios are not proper estimators for f_k according to Mack's model, and his associated uncertainty estimators employing such selections will be incorrect.[4] A matter of a greater concern though is that the residual definition is not valid for the new model and thus the selected model cannot be tested.

[4] In Mack's 1999 paper he expanded his formulas to incorporate simple averages in addition to weighted averages.

In the remainder of the paper we close this gap in a sense that for each "reasonably" selected link ratio set we provide a statistical model which has this set of selected link ratios as its best linear unbiased regression estimators. Using this tool, we are now able to incorporate a statistical work flow cycle into the reserving process:

Figure 1: Actuarial validation loop



This diagram shows of course only the work flow assuming that the data is appropriate. However one major part of the reserving exercise is reviewing the underlying data. We will see that the residuals can help the actuary identify outliers and trends.

As actuaries select, evaluate, and re-select link ratios, they are implicitly reformulating the model (1) by "selecting" a different α parameter each time. This correspondence is established by the following two theorems that prove the existence of the α parameters that solve model (1) for selected link ratios that are reasonable. By reasonable selected link ratios we mean selected link ratios within the range produced from the various average link ratios based on the empirical data.

3.1 Theorem (Link Ratio Function)

We consider for a given triangle the corresponding link ratio function as in (3) and denote the set of all *reasonable* link ratios with $\mathrm{LR}_{\boldsymbol{k}}(\Re) := \{\mathrm{LR}_{\boldsymbol{k}}(\alpha) \mid \alpha \in \Re\}$ and $i_{\min,k}, i_{\max,k}$ be the index of $\min\{C_{j,k}, j < n-k\}, \max\{C_{j,k}, j < n-k\}$ respectively. Then

1. If $c, d \in LR_k(\Re)$, then the whole interval $[c,d] \subseteq LR_k(\Re)$
2. $\mathrm{LR}_{\boldsymbol{k}}(\alpha) \to \boldsymbol{F}_{\boldsymbol{i}_{\min},\boldsymbol{k}}, (\alpha \to \infty)$
3. $\mathrm{LR}_{\boldsymbol{k}}(\alpha) \to \boldsymbol{F}_{\boldsymbol{i}_{\max},\boldsymbol{k}}, (\alpha \to -\infty)$
4. In particular, every value between the straight average link ratio, the weighted average link ratio and the link ratios corresponding to the minimum and maximum weight $\min\{C_{j,k}, j < n-k\}, \max\{C_{j,k}, j < n-k\}$ respectively, is reasonable.

3.2 Theorem (Existence)

Let $\{\boldsymbol{h_k}; \boldsymbol{k} \leq \boldsymbol{n}-2\}$ be a set of reasonable link ratios (as defined in 3.1) with $\boldsymbol{h_k} \in \mathrm{LR}_{\boldsymbol{k}}(\Re)$, $\boldsymbol{k} \leq \boldsymbol{n}-2$. Then for each *k* there is at least one α such that $\boldsymbol{h_k}$ is the ML-estimator of (1). We define

$$\hat{\alpha}_k := \max(\min\{\alpha > 0 \mid h_k = \mathrm{LR}(\alpha)\}, \max\{\alpha \leq 0 \mid h_k = \mathrm{LR}(\alpha)\}).$$

Then $\hat{\alpha}_k$ is well defined and can be calculated using a solver.[5] In other words among all possible α we take the one with smallest absolute value, and in cases, where two possible α have exactly the same absolute value, we choose the positive.

The proofs of both theorems are relegated to the appendix.

The condition $k \leq n$-2 is necessary because for the last development period ($k=n$-1) a regression type of approach is not useful as there is only one observation.

Remark 1:

In the original chain ladder method modeled in Mack (1993) the standard deviations of payments of all development periods is assumed to be proportional to the square root of payments of the previous development year. But why is it the square root, and why should this hold for all development years? Theorem 3.2 Theorem (Existence) relaxes this requirement. It shows that even with judgmentally selected link ratios an underlying statistical model exists such that the selected link ratios are the optimal parameters.[6]

Although assumptions cannot be tested, residuals can, which enables us to find the appropriate chain ladder model that is consistent with the actuary's link ratio selections. This underlines the thought that models offer "proposals" to understand the data structure. To cite George Box: "Essentially, all models are wrong, but some are useful."

4. EXAMPLES

Example 1:

We first consider the following triangle, which is discussed in Mack (1993) and in Zehnwirth (2004). The weighted averages link ratios are shown below:

---

[5] For example the Newton-Algorithm with starting point 0.

[6] In fact in some cases there can be more than one $\hat{\alpha}_k$ for the same link ratio. In other words, it is possible to have more than one standard deviation assumption associated with the same link ratio.

Table 1

	5,012	8,269	10,907	11,805	13,539	16,181	18,009	18,608	18,662	18,834
	106	4,285	5,396	10,666	13,782	15,599	15,496	16,169	16,704	
	3,410	8,992	13,873	16,141	18,735	22,214	22,863	23,466		
	5,655	11,555	15,766	21,266	23,425	26,083	27,067			
	1,092	9,565	15,836	22,169	25,955	26,180				
	1,513	6,445	11,702	12,935	15,852					
	557	4,020	10,946	12,314						
	1,351	6,947	13,112							
	3,133	5,395								
	2,063									
Simple Average	8.206	1.696	1.315	1.183	1.127	1.043	1.034	1.018	1.009	
Weighted Average	2.999	1.624	1.271	1.172	1.113	1.042	1.033	1.017	1.009	

Although this triangle is quite well understood, we try to "analyze" it again.

First we declare our *goal*, which is to find a model, which describes our data with a certain confidence.

- <u>Model Selection:</u> We start with the link ratio model, which means that we believe

$$C_{ik+1} = f_k C_{ik} + \sigma_k \varepsilon_{i,k} C_{i,k}^{\alpha_k/2}.$$

- <u>Parameter Selection</u>: This means in our case, that we choose a set of link ratios and *calculate* the corresponding variance assumption. We start here with the simple averages.

- <u>Model Validation</u>: Now we need to test the corresponding residuals.

Table 2

-0.5313	-0.7949	-0.7322	-0.5395	0.9132	1.3861	-0.1275	-0.7071
2.6108	-0.9210	2.0882	1.6351	0.0653	-0.9937	1.0576	0.7071
-0.4513	-0.3229	-0.4763	-0.3326	0.7867	-0.2809	-0.9301	
-0.4994	-0.6992	0.1083	-1.2187	-0.1807	-0.1115		
0.0448	-0.0850	0.2693	-0.1818	-1.5844			
-0.3198	0.2526	-0.6596	0.6376				
-0.0801	2.1662	-0.5977					
-0.2483	0.4040						
-0.5254							

The following plot graphs the residuals along the accident-year dimension and helps the practitioner to identify the existence, or absence, of any trends. The graph below suggests that the residuals are, for the most part, random.

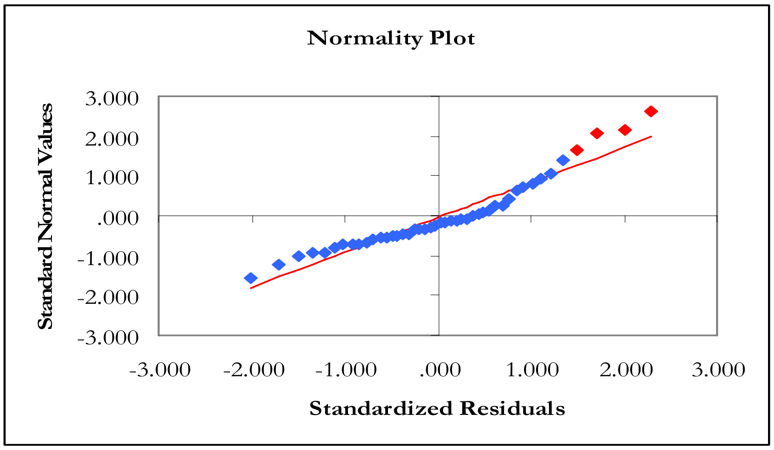


Figure 2 The residuals for the first example with the selected simple average link ratios against the quantiles of the normal distribution (red line)

Additionally we could test the data in several different other ways to make sure we are confident about the "noise hypothesis." In particular the Shapiro-Francia *P*-Value is 2.6%, which suggests that the assumption of normality of the residuals is rejected at the 5.0% confidence level. This means we would need to go back to one of the previous steps.

- Model (Re)Selection: With an exception of a few outliers, the model was acceptable, so we might still stay with the same model.
- Parameter (Re)Selection: Obviously the first few link ratio "produces" outliers, so we might change the first three selected link ratios to be the volume-weighted ones. That means we would select:

Selection	2.999	1.624	1.271	1.183	1.127	1.043	1.034	1.018	1.009
alpha	1.000	1.000	1.000	2.000	2.000	2.000	2.000	2.000	

- Model Validation: The Shapiro Francia test delivers a *P*-Value of 12.0%, so dependent on our level of statistical confidence we could accept this model, the selected parameters (and the corresponding "best estimate" reserves, the standard deviation, etc.). By comparing Figure 2 and Figure 3, we see that the selected link ratio set is a much better approximation of the normal distribution than the simple average link ratios.

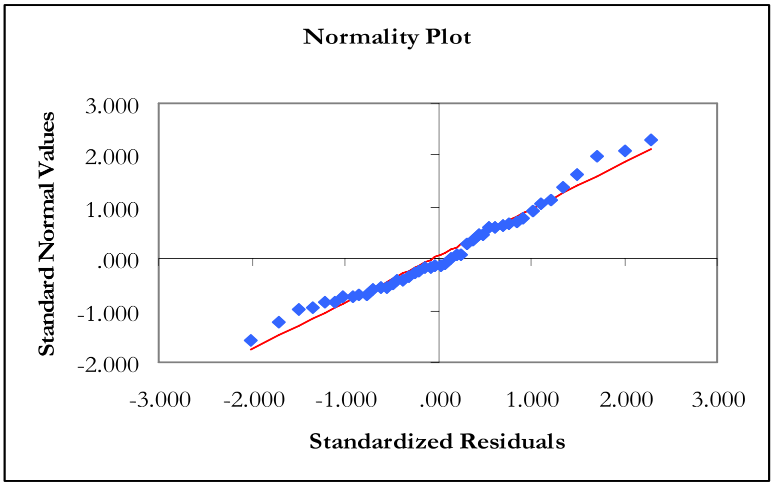


Figure 3 The residuals for the first example with the selected link ratios against the quantiles of the normal distribution (red line)

The Chain ladder link ratios, based on the volume weighted averages, deliver a *P*-Value of 23.4% for the Shapiro Francia test.

2.999	1.624	1.271	1.172	1.113	1.042	1.033	1.017	1.009

In other words the volume weighted link ratios are easily acceptable with our 5% level of confidence, but this demonstrates again that many models are "similarly wrong," but good enough for this task. We have chosen this well known-example to demonstrate the different steps in Figure 1.

Example 2:

We consider now the following triangle:

Table 3

29	97	216	388	580	764	930	1,119	1,322	1,526	1,657	1,720	1,739	1,748	1,752
30	102	227	403	631	849	1,046	1,270	1,518	1,703	1,820	1,877	1,894	1,901	
35	107	234	451	723	984	1,221	1,496	1,714	1,880	1,987	2,037	2,056		
34	112	268	526	850	1,162	1,447	1,689	1,888	2,037	2,134	2,178			
34	123	308	622	1,014	1,393	1,648	1,869	2,048	2,181	2,265				
42	152	373	745	1,216	1,555	1,786	1,984	2,145	2,265					
49	185	449	898	1,322	1,630	1,839	2,019	2,163						
58	217	537	939	1,319	1,594	1,779	1,938							
70	262	550	917	1,262	1,510	1,679								
88	261	518	846	1,154	1,379									
76	235	466	755	1,033										
68	207	411	673											
58	185	372												
53	167													

- Model Selection: We consider again the link ratio model.
- Parameter Selection: Before selecting the parameters, we might want to look at the link ratios and probably try the "latest year averages" because of the possible trend in the most recent calendar years.

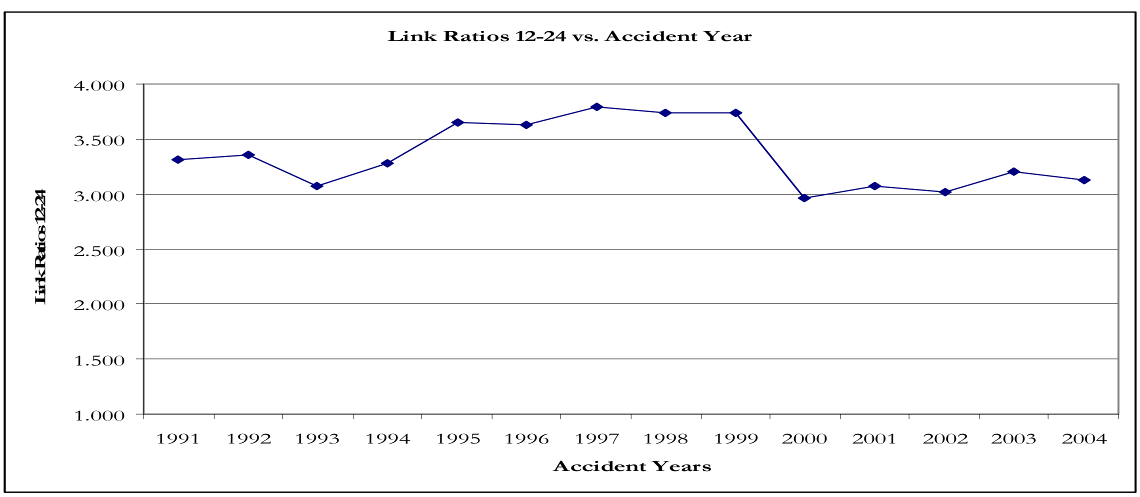


Figure 4 The link ratios for the first development period

- Model Validation: The following tables show the selected link ratios and the corresponding weights:

Chain Ladder
VW All Years

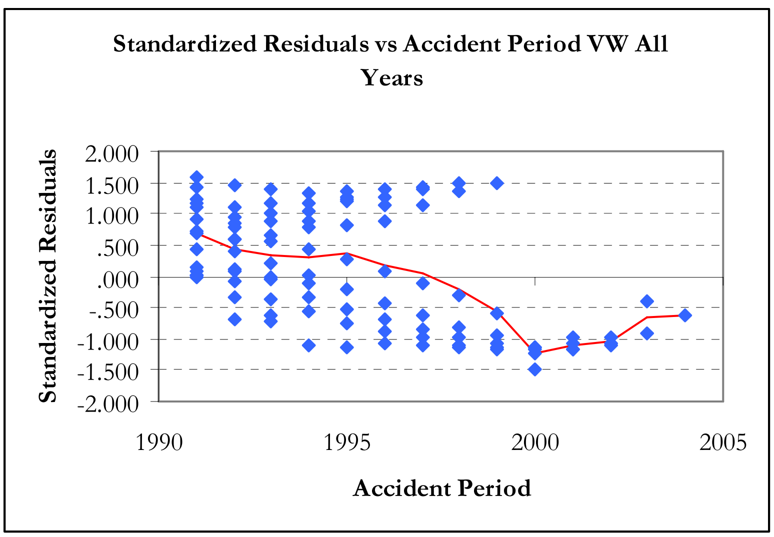
Link ratio	3.325	2.196	1.791	1.483	1.273	1.169	1.144	1.118	1.090	1.057	1.028	1.010	1.004	1.002
alpha	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
***P* Value**	0.0054%													

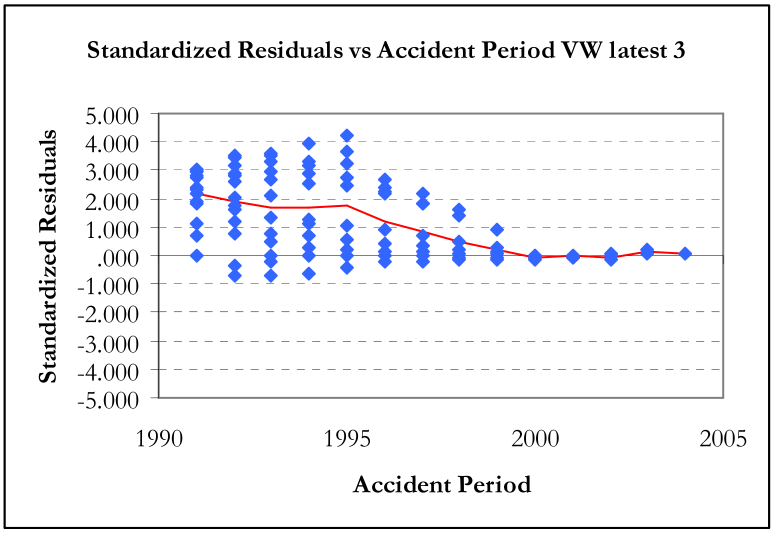
VW Latest 5

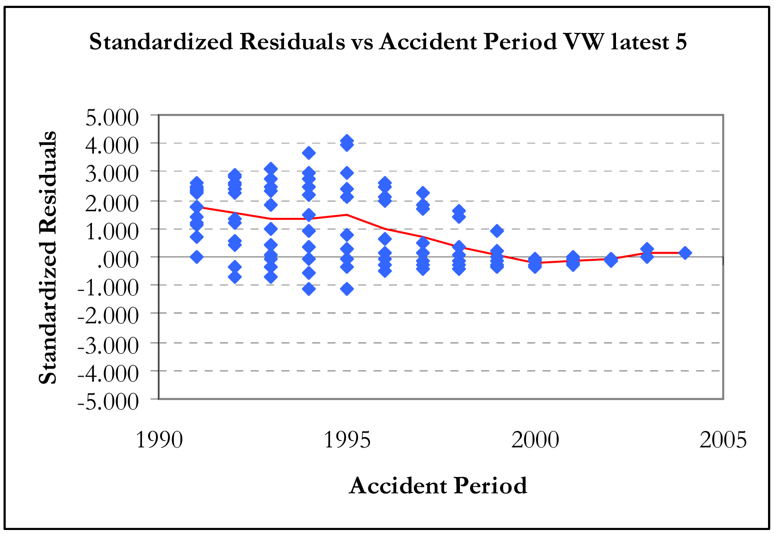
Link ratio	3.063	2.015	1.664	1.398	1.222	1.137	1.118	1.099	1.081	1.057	1.028	1.010	1.004	1.002
alpha	6.430	5.076	4.719	4.300	4.065	3.948	3.892	3.854	3.727	1.000	1.000	1.000	1.000	1.000
***P* Value**	0.0058%													

VW Latest 3

Link ratio	3.111	1.994	1.630	1.370	1.200	1.119	1.099	1.082	1.066	1.047	1.025	1.010	1.004	1.002
alpha	6.170	5.274	4.697	4.433	4.178	4.044	3.967	3.903	3.713	3.538	3.404	1.000	1.000	1.000
***P* Value**	0.0016%													







The test of normality rejected the assumption for all three types of selected link ratios. After these three loops of trying different levels of diagnosis, we might reconsider the model.

- Model Selection: We might now consider a more complex model, for example: $C_{ik+1} = g_k + f_k C_{ik} + \sigma_k \varepsilon_{i,k} C_{i,k}^{\alpha_k/2}$. For this model we refer the reader to Murphy [2].

The data might be even too complex for this model, but we demonstrate here the controlled way of actuarial work, which, of course, needs actuarial judgment, but also statistical tools to quantify the level of confidence for objective communication and assurance of quality (for example, for approval of an internal model in Solvency II-context).

5. CONCLUSION AND FURTHER RESEARCH

As already mentioned before, an alternative approach to ours would be the Bayesian approach, which means one could define a priori for the α_k and derive the a posteriori distribution for the variance assumption.

We have shown how to use the more flexible regression model (1) to reproduce the results of the traditional chain ladder methodology, which offers both consistency with the actuarial reserving work flow and statistical diagnostic tools. It is now quite obvious that the recursive formula of Mack/Murphy for the overall reserve uncertainty can be adapted to the selected link ratio model. In addition to that, a similar approach for the uncertainty of the BF method or Cape Cod method seems to be straight forward. We mention here also that any kind of bootstrapping can be done using the tested residuals. As we mentioned before, for bootstrapping purposes the residuals should be tested to assure proper results.

Even though the approach we introduced here is much more flexible than just employing average link ratios, there are many cases, where the model is not capable of modeling the structure in an appropriate way (such that the residuals looks like noise). In these cases, taking a more complicated method with more prediction power is necessary. The most natural way of making another step towards flexibility is to use the regression model of Murphy [2] with an intercept.

6. APPENDIX

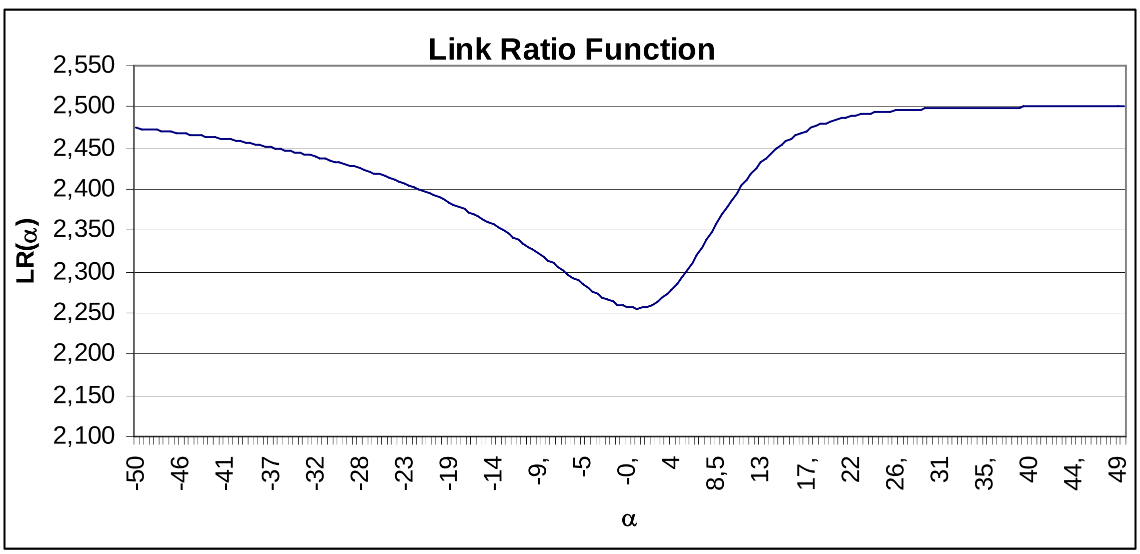
Proof of Theorem 3.1 (Link Ratio Function)

1. If $\mathrm{LR}_k : \Re \to \Re$ is a differentiable function and in particular continuous, its range is an interval in the set of real numbers.
2. We first note for arbitrary α that $\sum_{j=1}^{n-k} w_{j,k}^{\alpha} = 1$. Without loss of generality we assume $\boldsymbol{C}_{\boldsymbol{i}_{\min},\boldsymbol{k}} < \boldsymbol{C}_{\boldsymbol{j},\boldsymbol{k}}, (\boldsymbol{j} \le \boldsymbol{n} - \boldsymbol{k})$. It is now sufficient to prove $\boldsymbol{w}_{\boldsymbol{i}_{\min},\boldsymbol{k}}^{\alpha} \to 1$ as $\alpha \to \infty$. This can be seen by rewriting the weight
$$\boldsymbol{w}_{\boldsymbol{i}_{\min},\boldsymbol{k}}^{\alpha} = \boldsymbol{C}_{\boldsymbol{i}_{\min},\boldsymbol{k}}^{2-\alpha} / \sum_{\boldsymbol{j}=1}^{\boldsymbol{n}-\boldsymbol{k}} \boldsymbol{C}_{\boldsymbol{j},\boldsymbol{k}}^{2-\alpha} = \boldsymbol{C}_{\boldsymbol{i}_{\min},\boldsymbol{k}}^{2} / \sum_{\boldsymbol{j}=1}^{\boldsymbol{n}-\boldsymbol{k}} \boldsymbol{C}_{\boldsymbol{j},\boldsymbol{k}}^{2} \cdot (\boldsymbol{C}_{\boldsymbol{i}_{\min},\boldsymbol{k}} / \boldsymbol{C}_{\boldsymbol{j},\boldsymbol{k}})^{\alpha} .$$
Obviously all $\cdot(\boldsymbol{C}_{\boldsymbol{i}_{\min},\boldsymbol{k}} / \boldsymbol{C}_{\boldsymbol{j},\boldsymbol{k}}) < 1, \boldsymbol{j} \ne \boldsymbol{i}_{\min}$, thus all terms converge to 0 except for $\boldsymbol{j} = \boldsymbol{i}_{\min}$, so that we see $\sum_{\boldsymbol{j}=1}^{\boldsymbol{n}-\boldsymbol{k}} \boldsymbol{C}_{\boldsymbol{j},\boldsymbol{k}}^{2} \cdot (\boldsymbol{C}_{\boldsymbol{i}_{\min},\boldsymbol{k}} / \boldsymbol{C}_{\boldsymbol{j},\boldsymbol{k}})^{\alpha} \to \boldsymbol{C}_{\boldsymbol{i}_{\min},\boldsymbol{k}}^{2}$ as $\alpha \to \infty$.
3. Similar to 2, we can deduce $\boldsymbol{w}_{\boldsymbol{i}_{\max},\boldsymbol{k}}^{\alpha} \to 1$ as $\alpha \to -\infty$.
4. The weighted average and the simple average correspond to $\mathrm{LR}_{\boldsymbol{k}}(2), \mathrm{LR}_{\boldsymbol{k}}(1)$, respectively. This, with 1 above, proves the theorem.

The following example illustrates the function $\mathrm{LR}_k(\alpha)$ with an example, where $\boldsymbol{F}_{i_{\min},k} = \boldsymbol{F}_{i_{\max},k} = 2.5$. This is a case, where for all link ratios, except for the minimum for $\alpha = 0$, there are two different variance assumptions, which lead to the same link ratio. Also the infinitesimal behavior of the function is stated in the following graph.

Table 4: Link Ratio Example

	152	380	**2.5000**
	185	449	2.4270
	217	537	2.4747
	250	550	2.2000
	262	655	**2.5000**
	235	466	1.9830
	207	411	1.9855
	185	372	2.0108
Simple Average:		α=2	**2.2601**
VW Average:		α=1	**2.2563**
quadratic		α=0	**2.2559**



Proof of Theorem 3.2

Using Theorem 3.1 we observe that the set $\{\alpha \in \Re \mid h_k = \mathrm{LR}(\alpha)\}$ is not empty. Furthermore we note that $h_k = \mathrm{LR}(\alpha) \Leftrightarrow (h_k \cdot \sum_{j=1}^{n-k} C_{j,k}^{2-\alpha} - \sum_{i=1}^{n-k} C_{i,k+1} C_{i,k}^{1-\alpha}) = 0$, which can be solved with an appropriate numerical solver algorithm.

Consider again the example in Table 4. Then we get two solutions for the link ratio 2.400: -21.4 and 10.7, thus we set the variance estimation to max(-21.4, 10.7)=10.7.

7. REFERENCES

[1] Mack, Thomas, "Distribution-Free Calculation of the Standard Error of Chain Ladder Reserve Estimates," *ASTIN Bulletin* 1993, 23(2):213-225.
[2] Murphy, Daniel, "Unbiased Loss Development Factors," *Proceedings of the Casualty Actuarial Society* 1994, 81:154-222.
[3] Zenwirth, Ben and Glenn Barnett, "Best Estimate Reserving," *Proceedings of the Casualty Actuarial Society* 2000, 87(167):245-321.
[4] Verall, Richard J., "A Bayesian generalized linear model for the Bornhuetter-Ferguson method of claims reserving," *North American Actuarial Journal* 2004, 8(3):67-89.
[5] Royston, P., "A pocket-calculator algorithm for the Shapiro-Francia test for normality: An application to medicine," *Statistics in Medicine* 1993, 12(2):181-184

BIOGRAPHIES

Emmanuel T. Bardis is an actuarial consultant in the Boston office of Towers Perrin. He received the 2007 Hachemeister Prize for "Considerations Regarding Standards of Materiality in Estimates of Outstanding Liabilities," a paper he co-authored. Emmanuel received his Ph.D. in mathematics from the University of Notre Dame.

Ali Majidi is a mathematical consultant with Integrated Risk Management of Munich Re. He worked as an actuarial consultant with Towers Perrin before. Ali holds a Ph.D. in mathematical statistics and has done research in the field of nonparametric regression.

Daniel M. Murphy is a senior consultant in the San Francisco office of Towers Perrin. Formerly the chief actuary of Argonaut Insurance Company, he won the Woodward-Fondiller award for his paper "Unbiased Loss Development Factors." Dan received his masters in statistics from the University of Illinois.

Meaningful Intervals

Glen Barnett, Ph.D, David Odell, Ph.D, and Ben Zehnwirth, Ph.D, AIA, AIAA[1]

1 Professorial Visiting Fellow, School of Actuarial Studies, Australian School of Business, University of NSW

Abstract: Reserve ranges and risk capital requirements can be related to statistical interval estimates. While not all sources of uncertainty are readily incorporated into an interval estimate, such intervals give a lower bound on the size of the required interval. We discuss the calculation of interval estimates, for both the estimate of the mean and for the liability process itself, show how to tell if the model is a reasonable description of the data and show that when it is not, the interval estimates may sometimes be disastrously wrong.

Many practitioners are now using probabilistic versions of standard actuarial techniques, sometimes employing quite sophisticated tools in their estimation. However, none of these developments avoid the need for stringent checking of the suitability of model assumptions, a necessity that is often overlooked.

We discuss some of the statistical models underlying a variety of standard methods, construct a number of diagnostics for model assessment for several models and discuss how the underlying ideas carry over to many other methods for the estimation of liabilities. These tools are easy to implement and use. They allow practitioners to use the corresponding models with greater confidence, and gain additional information about the triangle. This information can have important consequences for the insurer.

We illustrate that some popular approaches—the Mack chain ladder, the quasi-Poisson GLM—and consequently predictions based on them (both bootstrapped and otherwise) have structure not present in real triangles, and don't describe some features of the data. Consequently their *associated intervals fail to have the desired properties*.

We point out that many aspects of the reserving problem and the structure of real data lead us to model on the log scale. We briefly describe the Probabilistic Trend Family (PTF) models and its extension to the multivariate case (MPTF) and show that these model families can capture the patterns in real data and produce more reasonable prediction intervals.

INTRODUCTION

It is important to distinguish between variability and uncertainty.

Variability is the effect of chance and a function of the system. Additional data points don't reduce the process variability.

Uncertainty is a lack of knowledge about the parameters that characterize the physical system that is being modeled. This may be reducible with additional information.

While separate concepts, variability and uncertainty are not completely unrelated—in general, the uncertainty of a parameter estimate will be related to the variability.

An interval for the mean is a form of confidence interval, based on the associated parameter uncertainty (and possibly including other sources of uncertainty).

An interval for a future payment (a prediction interval) must incorporate both the variability of the process and the uncertainty in the mean.

A sum or even a linear combination of future payments will similarly incorporate process variability for each term, parameter uncertainty for each term, and parameter covariances between each pair of terms. If the model includes correlations between observations, process covariance will also come into the prediction intervals.

The basic properties of confidence intervals and prediction intervals in regression models are presented in many standard statistical textbooks. See, for example, Wackerly, Mendenhall, and Schaeffer (2002). Many of the same principles encountered in the regression context apply more generally and form a useful basis for extending the discussion to intervals in the context of loss reserving and also calculation of risk capital requirements.

SECTION 1: CONFIDENCE INTERVALS AND PREDICTION INTERVALS

Consider the following simple motivating example.

Suppose a fair coin is tossed 100 times and we count the number of heads (X). To draw a parallel with insurance, imagine you pay a dollar for each head.

The mean number of heads (the mean of X) is 50. The standard deviation of X is 5. The binomial probability of each possible outcome of X (0, 1, 2,…100) is known. There is no uncertainty about the coin's mean, its variability, nor any of the probabilities associated with each outcome.

A 100% confidence interval for the mean is [50, 50]. However, the probability that X is equal to the mean of its distribution, 50, is approximately 0.08. A 95% prediction interval for the outcome X is [40, 60]. This prediction interval cannot be shortened without reducing the coverage probability. The inherent variability in the outcomes is termed process variability.

Suppose now that we do not know the true probability of a head, p, perhaps because the coin, or the method by which it is tossed is in some way not fair. Before the coin is tossed 100 times, a preliminary observation is made: it is tossed 20 times to get an estimate of the probability p.

Let's say that 10 heads are observed. Now the estimate of the probability of a head in one toss (10/20 = 0.5) is just that—an estimate. It is uncertain.

We can create a confidence interval (CI) for p and also for the mean number of heads in 100 tosses ($100p$). The CI is an interval around the estimated mean, 50. This confidence interval is not the same as a prediction interval for the *outcome from 100 tosses*.

The risk you're exposed to is the risk of the process, not the mean of the process (you don't pay the mean). Hence, even when the model is known, adequate risk capital is derived from the process variability. However, estimated process variability is insufficient—because the parameters are unknown. Our estimates are not equal to the true values, so we must *also* account for parameter uncertainty.

A prediction interval includes process variability and is therefore wider than a confidence interval. A confidence interval is an interval for a parameter, which is a constant (though unknown), while a prediction interval is for a random variable. The liability on a line of business is a random variable, not a parameter.

Predictive variance (the average variation between the predicted value and the actual outcome) is the sum of process variance and the variance of the parameter estimate (parameter uncertainty).

A 95% prediction interval for the number of heads in 100 tosses will be wider than [40, 60] (which accounts for process variability alone). For example, it might be, say [35, 65]. This interval could only be reduced to at best [40, 60] through additional sampling to reduce the parameter uncertainty. But you cannot make a 95% prediction interval narrower than [40, 60] in this circumstance—only the parameter uncertainty can be reduced.

Consider another situation with the same mean. A fair roulette wheel, numbered 0, 1, 2, 3,....., 100 is spun only once, and let R be the random variable that represents the outcome. The mean of R is 50, and its standard deviation is about 29. There is no uncertainty about the variability in the outcome. The probability that $R=50$ is 1/101.

A 100% *confidence interval* for the mean is [50, 50], as it was for the fair coin (no parameter uncertainty). A 95% *prediction* interval for the outcome R is [3, 98] (there are several such intervals of equal width).

Each process (fair coin and balanced roulette wheel) has the same mean, or if you like the same "best estimate." But the wheel requires more risk capital.

Reserve Ranges

In some countries, a "range" for reserving relates to uncertainty in the mean. As we have seen, a confidence interval can be produced by considering the uncertainty in the mean arising from parameter uncertainty. In some other countries, reserve calculations will incorporate process uncertainty and consider predictive distributions. In some cases a one-sided interval may be required. More sophisticated approaches can formally incorporate several additional sources of uncertainty into either kind of interval, but consideration of these is beyond the scope of this paper. When some important components of uncertainty are not formally introduced into the calculations, the upper ends of ranges implied by the statistical intervals (whether confidence intervals or prediction intervals) would be lower bounds on the required endpoints (and conversely where lower ends of ranges are required).

Risk Capital

It is important to recognize that the insurer is exposed to the loss process itself, not its mean. That risk includes process risk, and in order to hold risk capital adequate to cover the risk faced, process risk must be included in calculations. This implies that prediction intervals, rather than confidence intervals, are the appropriate starting point.

Reliance on Assumptions

When the assumptions of the model are reasonable, the derived interval estimates will be suitable inputs to reserving and risk-capital calculations. Conversely, when model assumptions are not met, derived intervals may have nothing like the desired properties. It is important to see that the model for the past is a reasonable description, and that the model for the future contains any relevant information about how that may change.

It is unfortunately the case that many models that have been used in loss reserving frequently fail to describe the data. We illustrate this problem by predicting the final calendar year (without using it in the estimation) and showing that the prediction intervals don't behave as they should if the model was adequate. It would seem that being able to predict the past is at least a minimum requirement for a model. If we cannot predict the past, on what basis can we assert we are able to predict the future?

SECTION 2: A BASIC PREDICTION PROBLEM

The simplest prediction problem illustrates many of the issues. Consider the following example.

We have n observations $Y_{1,} \ldots Y_n \sim F(\mu,\sigma^2)$.

Equivalently, $Y_i = \mu + \varepsilon_i \quad \varepsilon_i \sim F(0, \sigma^2)$.

While for sufficiently large samples, the distribution may not be critical for a confidence interval, it becomes important in the case of prediction intervals, since the observation being predicted will come from that distribution. In order to deal with prediction, we either need to make a distributional assumption (and check it), or use some methodology—such as the bootstrap—that allows us to deal (at least approximately) with the distribution.

Now we want to forecast another observation, Y_{n+1} ($= \mu + \varepsilon_{n+1}$).

So we have:

$$\hat{Y}_{n+1} = \hat{\mu} + \hat{\varepsilon}_{n+1} \quad (\hat{\varepsilon}_{n+1} \text{ is the forecast of the error term})$$

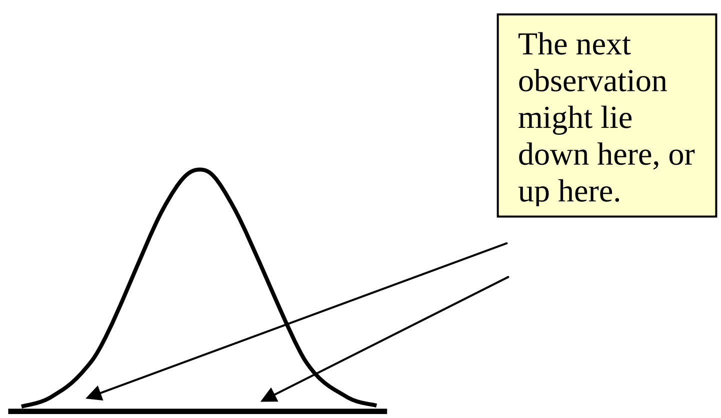
$$= \hat{\mu} + 0.$$

That is, our best estimate of the next observation is exactly equal to our current estimate of the mean.

Case (i): The mean, μ, is known:

$$\hat{Y}_{n+1} = \mu + 0.$$

This is equivalent to the case where we knew the coin was fair. Even though we know everything about the process, in predicting Y_{n+1}, we are predicting a random quantity.



The variance of the prediction of Y_{n+1} is:

$\mathrm{Var}(\mu) + \mathrm{Var}(\varepsilon_{n+1}) = 0 + \sigma^2.$

It is important to remember that the *risk to the business* is not simply from the uncertainty in the mean—for example, the value at risk is related to the amount you will actually pay, not its mean.

Even when the mean is known exactly, there is still underlying process uncertainty (with 100 tosses of a fair coin, might get 46 heads or 57 heads).

It doesn't really make sense to talk about a mean (or any other aspect of the distribution) in the absence of a probabilistic model—otherwise what distribution is it the mean *of*? Without some kind of model we cannot even be clear what we're talking about.

With loss modeling, you design a model to describe what's going on with the data. Assumptions need to be explicit so that you can check that the distribution is consonant with the data.

In our motivating example, you would not want to use the coin model if your data was actually coming from the roulette wheel.

Given a suitable model, in practice we simply won't know the mean or other relevant parameters—we only have a sample to tell us about them. (For simplicity we confine our attention to the mean in this discussion.)

Case (ii): The mean, μ, is unknown.

The best estimate for the next observation is still the mean, but we now have to estimate it. This estimate is based on past values and is not exactly equal to the actual underlying population mean — even with a perfect model due to the effects of random variation, the estimate will differ from the underlying value.

While the estimate is uncertain, we can obtain an estimate of the uncertainty, if the model is a good description. So we will have an estimate of the mean and we'll also be able to get a confidence interval for the mean.

This interval is designed so that if we were able to re-run history (re-toss our coin, respin our roulette wheel), many times, the intervals we generate will include the unknown mean a given fraction of the time.

If the model doesn't describe the data, however, the confidence interval may not have anything close to required probability coverage.

Confidence interval for the mean of the coin model

Again, we count how many heads in 100 tosses, but we have a small sample (20 tosses) with which to assess the probability of a head. As mentioned before, let's assume that we observe 10 heads (to keep the mean prediction unchanged).

Obtaining 10 heads in 20 tosses yields $\hat{p} = ½$. $\hat{\mu} = 100.\hat{p} = 50$.

$\mathrm{Var}(\hat{\mu}) = 100^2\ \mathrm{Var}(\hat{p}) = 100^2\ ½.\ ½.\ /20 = 11.18^2$.

An approximate 95% CI for the mean, μ is

$\hat{\mu} \pm 1.96\ \mathrm{s.e.}(\hat{\mu}) \approx (29, 71)$.

Note that the interval here can be based on a normal approximation, due to the central limit theorem. (If the distribution is sufficiently skewed or heavy tailed, the sample may need to be larger for the normal approximation to be reasonable, but in the case of the binomial with p not too far from 0.5, a sample of 20 should be plenty).

Now we again want to look at a *prediction interval*, but here with an estimated mean.

We want to predict a random outcome where we don't know the mean. (In this example we assume the variance is known. In many practical cases uncertainty in the variance does not greatly alter the limits of intervals.)

To understand your business you need to understand the actual risk of the process, not just the risk in the estimate of the mean.

Let's revisit the simple model $Y_{n+1} = \mu + \varepsilon_{n+1}$:

$$\hat{Y}_{n+1} = \hat{\mu} + \hat{\varepsilon}_{n+1}\ .$$

Now μ is unknown. So $\hat{Y}_{n+1} = \hat{\mu} + 0$.

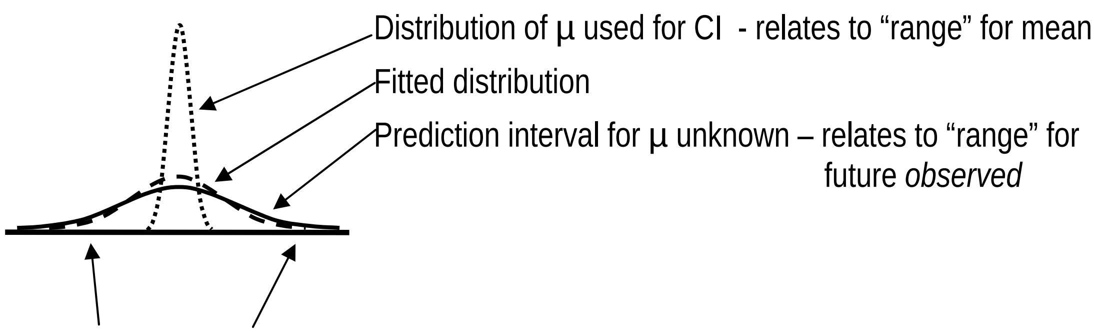
Variance of forecast

$$= \mathrm{Var}\ (\hat{\mu}) + \mathrm{Var}(\varepsilon_{n+1})$$

$$= \sigma^2/n \quad + \quad \sigma^2\ .$$

(In practice, σ^2 is replaced by its estimate, of course. We are ignoring the parameter uncertainty in the variance for the present discussion.)

Now imagine that the distribution, F, is normal. The next observation might lie



down here, or up here. (While we've assumed normality here, the issues in the diagram apply more widely.)

In the coin experiment, the predictive distribution is approximately normal.

Y_{n+1} is the number of heads on our next run of 100 tosses. Its predictive variance is

$\mathrm{Var}(Y_{n+1} \mid \mu = \hat{\mu}) + \mathrm{Var}(\hat{\mu}) = 100\ \tfrac{1}{2}.\ \tfrac{1}{2} + 100^2\ \tfrac{1}{2}.\ \tfrac{1}{2}.\ /20 = 12.25^2$

So an approximate 95% CI for the forecast Y_{n+1} is

$\hat{Y}_{n+1} \pm 1.96\ \mathrm{s.e.}(\hat{Y}_{n+1}) \approx (26{,}74)$.

An alternative way to look at the predictive variance

Here we derive the predictive variance as the variance of the prediction error.

Prediction error = $Y_{n+1} - \hat{Y}_{n+1}$.

Predictive variance = Var(prediction error)

$= \mathrm{Var}(Y_{n+1} - \hat{Y}_{n+1})$

$= \mathrm{Var}(Y_{n+1}) + \mathrm{Var}(\hat{Y}_{n+1}) - 2\,\mathrm{Cov}(Y_{n+1}, \hat{Y}_{n+1})$

$= \mathrm{Var}(Y_{n+1}) + \mathrm{Var}(\hat{Y}_{n+1}) - 0$

$= \mathrm{Var}(Y_{n+1}) + \mathrm{Var}(\hat{Y}_{n+1})$

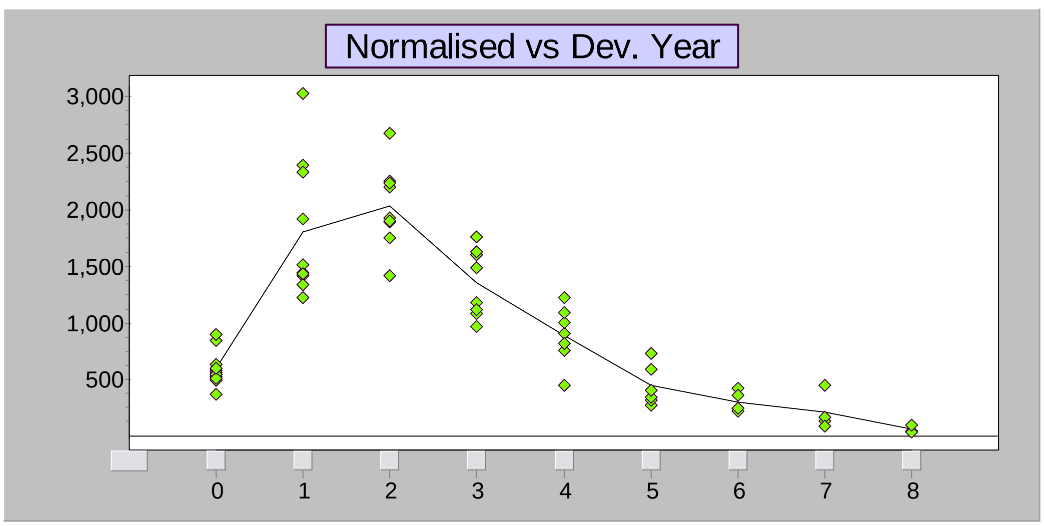
= process variability + parameter uncertainty.

Note that this result only relies on the fact that Cov $(Y_{n+1}, \hat{Y}_{n+1}) = 0$, which, for example, occurs when observations are independent (since then the forecast $\hat{Y}_{n+1}$ is a function of only past observations, while Y_{n+1} is a future observation, which will be independent of it) and the result follows.

With more complex models the calculation of the variance of the estimate of the mean is more complicated, but the principle remains the same.

Confidence intervals vs. Prediction intervals—a basic loss example

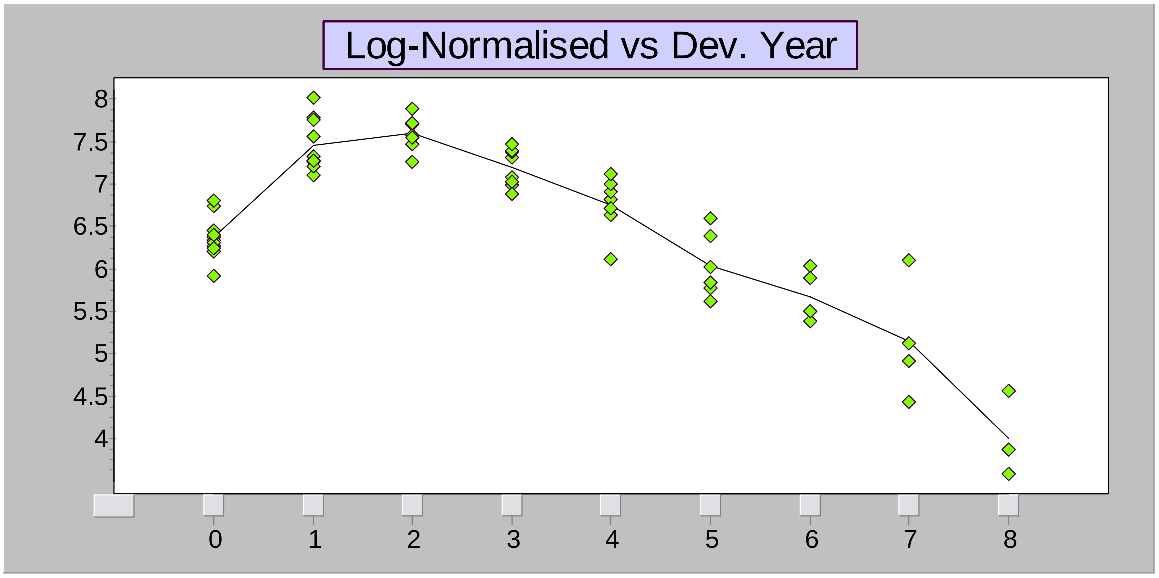
Let's look at some real long-tail data that has been inflation-adjusted and then normalized for a measure of exposure. This is the CTP data that was analyzed in Barnett and Zehnwith (2000).



We see a clear runoff pattern against development year. In this instance the trends in the accident-year and calendar-year directions sufficiently small that we can ignore them for illustrative purposes.

Note that in the figure above, the points have a tendency to "clump" just below the mean and be more spread out above the mean—the normalized data is skewed to the right.

On the log-scale this skewness disappears, and the variance is pretty stable across years. The skewness is removed so the values appear much more symmetric about center and the spread looks fairly constant.



Consider a single development—say DY 3:

AY	Normalized	Log
1	1,489	7.306
2	1,606	7.381
3	1,087	6.991
4	1,628	7.395
5	1,178	7.072
6	1,118	7.019
7	1,761	7.474
8	972	6.879

The maximum likelihood estimates of μ and σ are 7.190 and 0.210, respectively.

(NB: the MLE of σ is s_n, the standard deviation with the n denominator, not the more common s_{n-1}).

Assuming a random sample from a process with constant mean, we would predict the mean for next value as 7.190. However, without some indication of its accuracy, this is not very helpful.

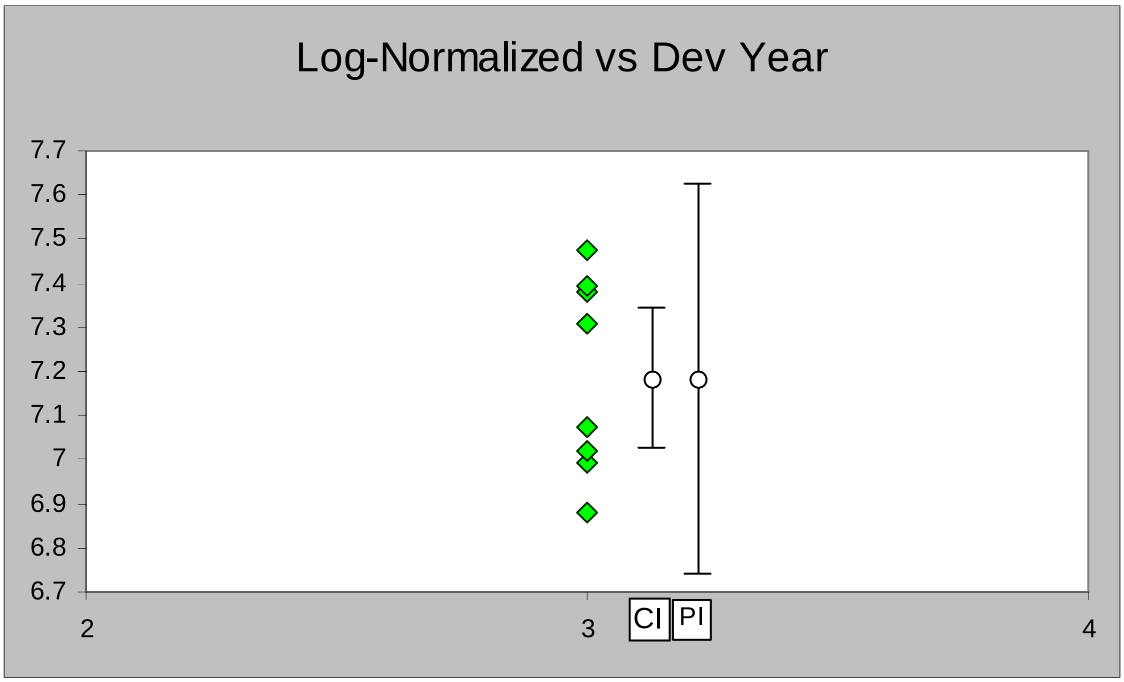
A 95% confidence interval for μ is: (7.034, 7.345).

Prediction:

Recall that the *predictive* variance is $\mathrm{Var}(\hat{\mu}) + \mathrm{Var}(\varepsilon_{n+1})$

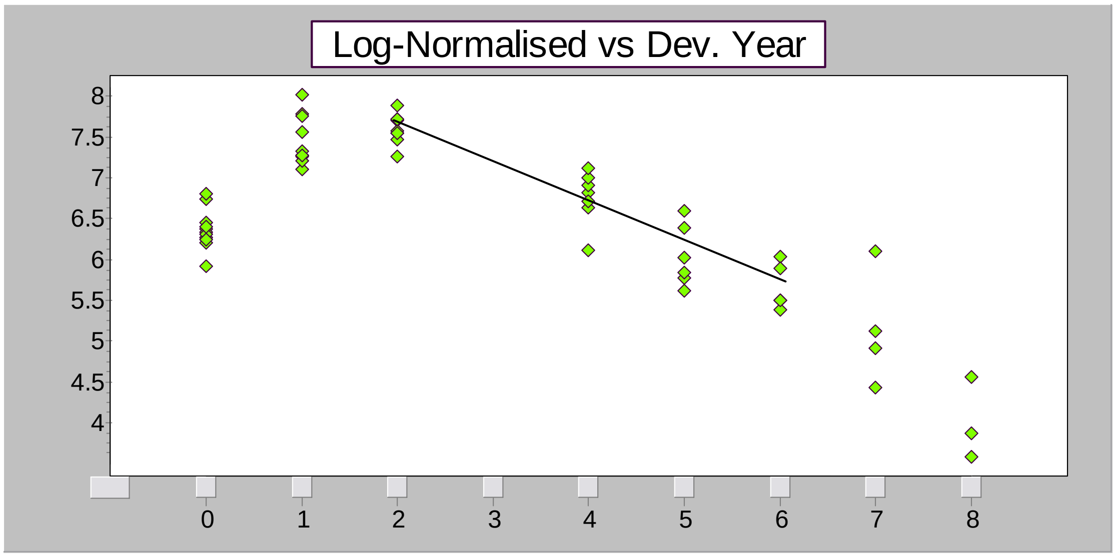
(= parameter uncertainty + process var).

A 95% *prediction interval for* Y_{n+1} is (6.75,7.63). See the figure below.



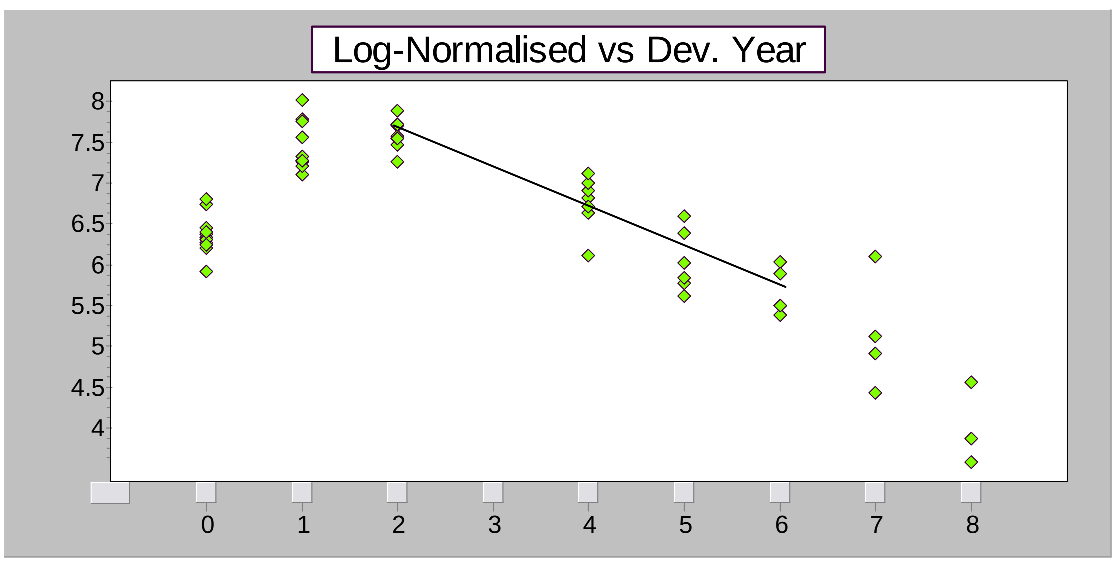
The intervals here are fairly wide. More data reduces parameter uncertainty (e.g., more than 20 tosses of the coin in the earlier trial would make the intervals smaller). In some cases you can go back and get more loss data and eventually you'll have another year of data. However, process variability doesn't reduce with more data—it's an aspect of the process. We can measure it more accurately, but the thing we're measuring is not changing.

As we can see in the figure below, nearby developments are related: if DY 3 was all missing, you could take a fair guess at where it was.



So in this case we *do* have more data!

To take full advantage of this, we need a model to relate the development years. Even just fitting line through DY2-4 has a reasonably large effect on the width of the *confidence interval* (the grey bars shift inward, to the black bars).



However, it only changes the *prediction interval* by ~2%—so calculated VaR hardly changes.

Note that so far this prediction interval is on the *log* scale. To take a *prediction* interval back to the normalized-dollar scale, we just back-transform the endpoints of the prediction interval. To produce a *confidence interval* for the *mean on the normalized-dollar* scale is harder. We *can't* just backtransform the limits on the confidence interval—that's going to give an interval for the *median*, not the mean. However, we can scale the interval for the median to produce an interval for the mean.

Further, to convert the interval for these scaled dollars to original dollars, we need to re-scale the interval for the inflation and exposures.

There are some companies around for whom (for some lines of business) the process variance is very large—some have a coefficient of variation near 0.6 (so the standard deviation is > 60% of the mean). That's just a feature of the data. You may not be able to control it, but you sure need to *know* it.

Why take logs?

Taking logs tends to stabilize variance. Multiplicative effects (percentage changes, including economic effects such as inflation) become additive. Exponential growth or decay becomes linear. Skewness is usually eliminated. Distributions tend to look near normal, making least squares reasonable. Using logs is a familiar way of dealing with many of these issues—indeed, it's standard in many parts of finance.

Note that for these benefits to work, we have to take logs of *incremental* figures (such as incremental paid), rather than cumulative paid or incurred losses. For example, inflation in the past period affects payments now, but not past payments—so cumulatives (which are also present in incurred figures) will contain a mix of payments across past rates.

SECTION 3: DIAGNOSTIC DISPLAYS FOR CHAIN LADDER MODELS

In this section we consider two models that reproduce chain ladder forecasts, the regression model (Mack model, Mack, 1993) and the quasi-Poisson GLM (Hachemeister and Stanard, 1975, and apparently independently by Renshaw and Verrall, 1994).

Many common regression diagnostics for model adequacy relate to analysis of residuals, particularly residual plots. In many cases these work very well for examining many aspects of model adequacy. When it comes to assessing predictive ability, the focus should, where possible, shift to examining the ability to predict data not used in the estimation. In a regression context, a subset of the data is held aside and predicted from the remainder. Generally the subset is selected at random from the original data. However, in our case, we cannot completely ignore the time-series structure and the fact that we're predicting outside the range of the data. Our prediction is always of future calendar time. Consequently the subsets that can be held aside and assessed for predictive ability are those at the most recent time periods.

This is common in analysis of time series. For example, models are sometimes selected so as to minimize one-step-ahead prediction errors. See, for example, Chatfield (2000).

Out of sample predictive testing

The critical question for a model being used for prediction is whether the estimated model can predict outside the sample used in the estimation. Since the triangle is a time series, where a new diagonal is observed at each calendar period, prediction (unlike predictions for a model without a time dimension) is of calendar periods after the observed data. To do out-of-sample tests of predictions, it is therefore important to retain a subset of the most recent calendar periods of observations for post-sample predictive testing. We refer to this post-sample-predictive testing as *model validation* (note that some other authors use the term to mean various other things, often related to checking the usefulness or appropriateness of a model).

Imagine we have data up to time t. We use only data up to time t–k to estimate the model and predict the next k periods (in our case, calendar periods), so that we can compare the ability of the model to predict actual observations not used in the estimation. We can, for example, compute the prediction errors (or validation residuals), the difference between observed and predicted in the validation period. If these prediction errors are divided by the predictive standard error, the resulting standardized validation residuals can be plotted against time (calendar period most importantly, and also accident and development period), and against predicted values, (as well as against any other likely predictor), in similar fashion to ordinary residual plots. Indeed, the within-sample residuals and "post-sample" predictive errors (validation residuals) can be combined into a single display.

One step ahead prediction errors are related to validation residuals, but at each calendar time step only the next calendar period is predicted; then the next period of data is brought in and another period is predicted.

In the case of ratio models such as the chain-ladder, prediction is only possible within the range of accident and development years used in estimation, so out of sample prediction cannot be done for all observations left out of the estimation. The use of one step ahead prediction errors maximizes the number of out-of-sample cases that have predictions. Further, when reserving, the liability for the next calendar period is generally a large portion of the total liability, and the liability estimated will typically be updated once it is observed; this makes one-step-ahead prediction errors a particularly useful criterion for model evaluation when dealing with ratio models like the chain ladder.

For a discussion of the use of out of sample prediction errors and in particular one-step-ahead prediction errors in time series, see Chatfield (2000), chapter 6.

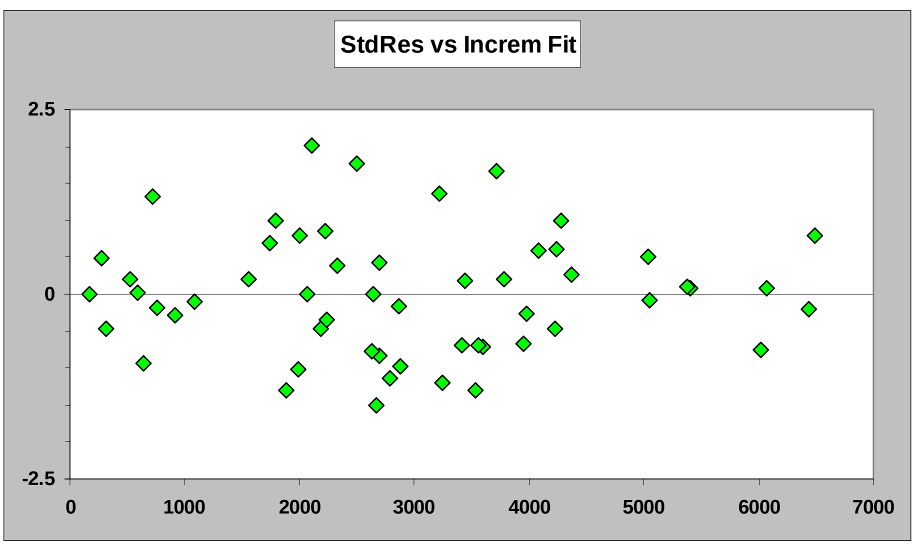
For many models, the patterns in residual plots when compared with the patterns in validation residuals or one step ahead prediction errors appear quite similar. In this circumstance, ordinary residual plots will generally be sufficient for identifying model inadequacy.

Critically, in the case of the Poisson and quasi-Poisson GLM that reproduce the chain ladder, the prediction errors and the residuals *do* show different patterns.

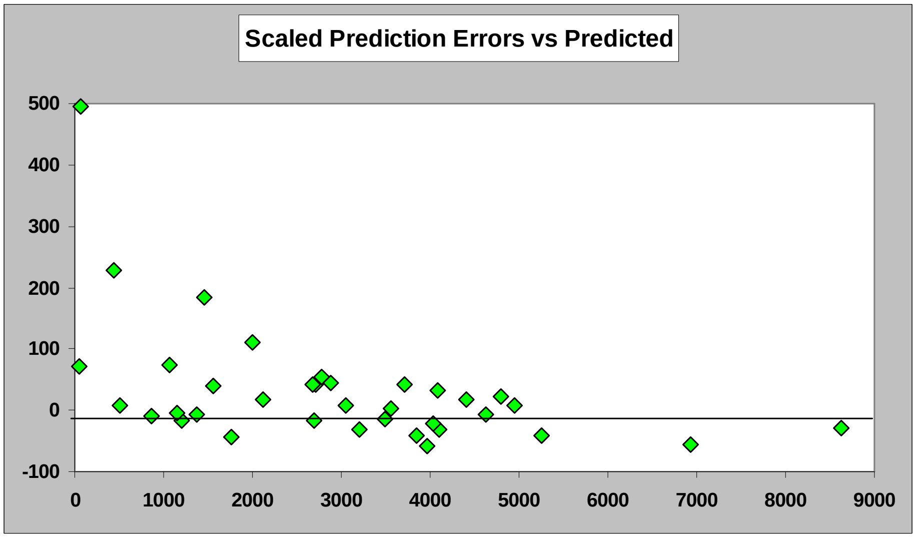
Illustration:

This data was used in Mack (1994). The data are incurred losses for automatic facultative business in general liability, taken from the Reinsurance Association of America's Historical Loss Development Study.

If we fit a quasi- (or overdispersed) Poisson GLM and plot standardized residuals against fitted values, the plot appears to show little pattern:

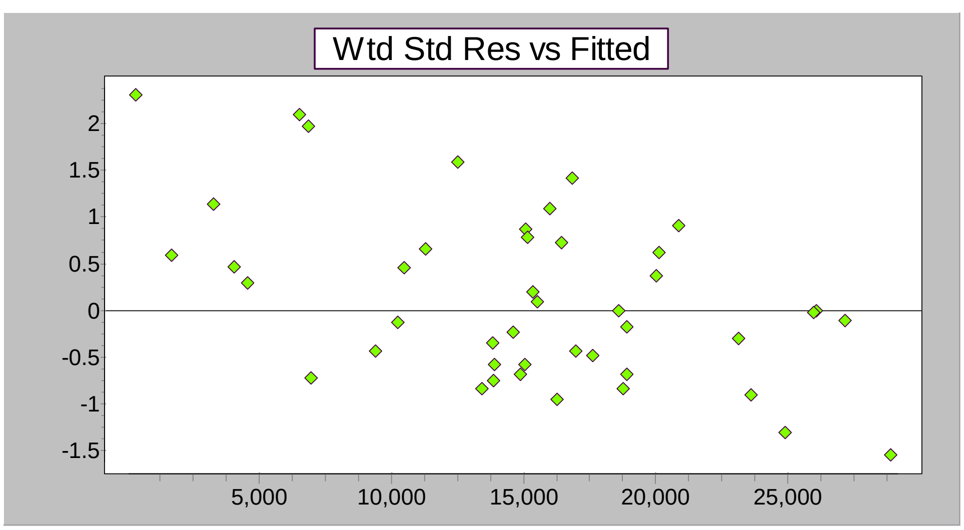


However, if we plot *one step ahead prediction errors* (scaled by dividing the prediction errors by $\hat{\mu}^{1/2}$) against predicted values, we *do* see a distinct pattern of mostly positive prediction errors for small predictions with a downward trend toward more negative prediction errors for large predictions:



Prediction errors above have not been standardized to have unit variance. The underlying quasi-Poisson scale parameter would have a different estimate for each calendar-year prediction; it was felt that the additional noise from separate scaling would not improve the ability of this diagnostic to show model deficiencies. On the other hand, using a common estimate across all the calendar periods would simply alter the scale on the right-hand side without changing the plot at all, and has the disadvantage that for many predictions you'd have to scale them using "future" information. On the whole it seems prudent to avoid the scaling issue for this display, but as a diagnostic tool, such scaling is not a major issue.

This problem of quite different patterns for prediction errors and residuals does not generally occur with the Mack formulation of the chain ladder, where ordinary residuals are sufficient to identify this problem:



As noted in Barnett and Zehnwirth (2000), this downward trend is caused by a simple failure of the ratio assumption—it is not true that $E(Y|X) = \beta X$, as would be true of any model where the next cumulative is assumed to be (on average) a multiple of the previous one. (For this data, the relationship between a cumulative and the previous cumulative does not go through the origin.)

The above plot is against cumulatives because in the Mack formulation, that's what is being predicted. (Note that the quasi-Poisson GLM residuals vs. cumulative fitted rather than incremental fitted still looks flat.)

Why is the problem obvious in the residuals for the Mack version of the chain ladder model, but not in the plots of GLM residuals vs. fitted (either incremental or cumulative)?

Even though the two models share the same prediction function, the *fitted values* of the two models are quite different.

On the cumulative scale, if X is the most recent cumulative (on the last diagonal) and Y is the next (future) one, both models have the prediction-function $E(Y|X) = \beta X$.

However, *within* the data, while the Mack model uses the same form for the fit—$E(Y|X) = \beta X$, the GLM does not—you can write it as $E(Y) = \beta E(X)$, which seems similar enough that it might be imagined it would not make much difference, but the right-hand side involves "future" values not available to predictions. This allows the fit to "shift" itself to compensate, so you can't see the problem in the fits. However, the out-of-sample prediction function is the same as for the Mack formulation, and so the predictions from the GLM suffers from *exactly* the same problem—once you forecast future values, you're assuming $E(Y|X) = \beta X$ for the future—and this assumption needs to be checked! It cannot be assessed using a *within-data* (residual) analysis of the GLM; it needs to be assessed by checking the actual assumption, whether via use of prediction residuals, or by checking Mack residuals.

Adequate model assessment of quasi-Poisson GLMs therefore *requires* the use of some form of out-of-sample prediction, and because of the structure of the chain ladder, this assessment seems to be best done with one-step-ahead prediction errors. For many other models, such as the Mack model, this would be useful but not as critical, since we can identify the problem even in the residuals.

SECTION 4: THE BOOTSTRAP

The bootstrap is, at heart, a way to obtain an approximate sampling distribution for a statistic (and hence, if required, produce a confidence interval). Where that statistic is a suitable estimator for a population parameter of interest, the bootstrap enables inferences about that parameter. In the case of simple situations the bootstrap is very simple in form, but more complex situations can also be dealt with. The bootstrap can be modified in order to produce a predictive distribution (and hence, if required, prediction intervals).

It is predictive distributions that are generally of prime interest to insurers (because they pay the outcome of the process, not its mean). The bootstrap has become quite popular in reserving in recent years.

The bootstrap does not require the user to assume a distribution for the data. Instead, sampling distributions are obtained by resampling the data.

However, the bootstrap certainly does not avoid the need for assumptions, nor for checking those assumptions. The bootstrap is far from a cure-all. It suffers from essentially the same problems as finding predictive distributions and sampling distributions of statistics by any other means. These problems are exacerbated by the time-series nature of the forecasting problem—because reserving requires prediction into never-before-observed calendar periods, model inadequacy in the calendar-year direction becomes a critical problem. In particular, the most popular actuarial techniques—those most often used with the bootstrap—don't have any parameters in that direction, and are frequently mis-specified with respect to the behavior against calendar time. The bootstrap does not solve this problem.

Further, commonly used versions of the bootstrap can be sensitive to overparameterization—and overparameterization is a common problem with standard techniques.

A basic bootstrap introduction

The *bootstrap* was devised by Efron (1979), growing out of earlier work on the *jackknife*. He further developed it in a book (Efron, 1982), and various other papers. These days there are numerous books relating to the bootstrap, such as Efron and Tibshirani (1994). A good introduction to the basic bootstrap may be found in Moore et al. (2003); it can be obtained online.

The original form of the bootstrap is where the data itself is resampled, in order to get an approximation to the sampling distribution of some statistic of interest, so an inference can be made about a corresponding population statistic.

For example, in the context of a simple model $E(X_i) = \mu$, $i = 1, 2, \ldots, n$, where the Xs assumed to be independent, the population statistic of interest is the mean, μ, and the sampling statistic of interest would typically be the sample mean, $\bar{x}$.

Consequently, we estimate the population mean by the sample mean ($\hat{\mu} = \bar{x}$) —but how good is that estimate? If we were to collect many samples, how far would the sample means typically be from the population mean?

While that question could be answered if we could directly take many samples from the population, typically we cannot resample the original population again. If we assume a distribution, we could infer the behavior of the sample mean from the assumed distribution, and then check that the sample could reasonably have come from the assumed distribution.

(Note that rather than needing to assume an entire distribution, if the population variance were assumed known, we could compute the variance of the sample mean, and given a large enough sample, we might consider applying the central limit theorem (CLT) in order to produce an approximate interval for the population parameter, without further assumptions about the distributional form. There are many issues that arise. One such issue is whether or not the sample is large enough—the number of observations per parameter in reserving is often quite small. Indeed, many common techniques have some parameters whose estimates are based on only a single observation! Another issue is that to be able to apply the CLT we assumed a variance—if instead we estimate the variance, then the inference about the mean depends on the distribution again. As the sample sizes become large enough that we may apply Slutsky's theorem, then for example a *t*-statistic is asymptotically normal, even though in small samples the *t*-statistic *only has a t-distribution if the data were normal*. Lastly, and perhaps most importantly when we want a *predictive distribution*, the CLT generally cannot help.)

In the case of bootstrapping, the *sample* is itself resampled, and then from that, inferences about the behavior of samples from the population are made on the basis of those resamples. The empirical distribution of the original sample is taken as the best estimate of the population distribution.

In the simple example above, we repeatedly draw samples of size n (with replacement) from the original sample, and compute the distribution of the statistic (the sample mean) of each resample. Not all of the original sample will be present in the resample—on average a little under 2/3 of the original observations will appear, and the rest will be repetitions of values already in the sample. A few observations may appear more than twice.

The standard error, the bias and even the distribution of an estimator about the population value can be approximated using these resamples, by replacing the population distribution, F by the empirical distribution F_n.

For more complex models, this direct resampling approach may not be suitable. For example, in a regression model, there is a difficulty with resampling the responses directly, since they will typically have *different* means.

For regression models, one approach is to keep all the predictors with each observation and sample them together. That is, if $\underline{X}$ is a matrix of predictors (sometimes called a *design matrix*) and $\underline{y}$ is a data-vector, for the multiple regression model $\underline{Y} = \underline{X}\,\beta + \underline{\varepsilon}$, then the rows of the augmented design matrix $[X|y]$ are resampled. (This is particularly useful when the Xs are thought of as random.)

A similar approach can be used when computing multivariate statistics, such as correlations.

Another approach is to resample the *residuals* from the model. The residuals are estimates of the error term, and in many models the errors (or in some cases, scaled errors) at least share a common mean and variance. The bootstrap in this case assumes more than that—they should have a common distribution (in some applications this assumption is violated).

In this case (with the assumption of equal variance), after fitting the model and estimating the parameters, the residuals from the model are computed: $e_i = y_i - \hat{y}_i$, and then the residuals are resampled as if they were the data.

Then a new sample is generated from the resampled residuals by adding them to the fitted values, and the model is fitted to the new bootstrap sample. The procedure is repeated many times.

Forms of this *residual resampling* bootstrap have been used almost exclusively in reserving, even when the other form of the bootstrap could be used.

If the model is correct, appropriately implemented residual resampling works. If it is incorrect, the resampling scheme will be affected by it, some more than others, though in general the size of the difference in predicted variance is small. More sophisticated versions of this kind of resampling scheme, such as the second bootstrap procedure in Pinheiro et al. (2003) can reduce the impact of model misspecification when the prediction is, as is common for regression models, within the range of the data. However, the underlying problem of amplification of unfitted calendar-year effects remains, as we shall see.

For the examples in this paper we use a slightly augmented version of *Sampler 2* given in Pinheiro et al.—the prediction errors are added to the predictions to yield bootstrap-simulated predictive values, so that we can directly find the proportion of the bootstrap predictive distribution below the actual values in one-step-ahead predictions.

In the case of reserving, the special structure of the problem means that while often we predict inside the range of observed accident years, and usually also within the range of observed development years, we are always projecting *outside* the range of observed calendar years—precisely the direction in which the models corresponding to most standard techniques are inadequate.

As a number of authors have noted, the chain ladder models the data using a two-way cross-classification scheme (that is, like a two-way main-effects ANOVA model in a log-link). As discussed in Barnett et al. (2005), this is an unsuitable approach in the accident and development direction, but the issues in the calendar direction are even more problematic. Even the more sophisticated approaches to residual resampling can fail on the reserving problem if the model is unsuitable.

Assessing bootstrap prediction intervals

When calculating predictive distributions with the bootstrap, we can in similar fashion make plots of standardized prediction errors against predicted values and against calendar years. Of course, since the prediction *errors* are the same, the only change would be a difference in the amount by which each prediction error is scaled (since we have bootstrap standard errors in place of asymptotic standard errors from an assumed model); the broad pattern will not change, however, so the plot based on asymptotic results are useful prior to performing the bootstrap.

Since we can produce the entire predictive distribution via the bootstrap, we can evaluate the percentiles of the omitted observations from their bootstrapped predictive distributions—if the model is suitable, the data should be reasonably close to "random" percentiles from the predictive distribution. This further information will be of particular interest for the most recent calendar periods (since the ability of the model to predict recent periods gives our best available indication if there is any hope for it in the immediate future—if your model cannot predict last year you cannot have a great deal of confidence in its ability to predict *next* year).

We could look at a visual diagnostic, such as the set of predictive distributions with the position of each value marked on it, though it may be desirable to look at all of them together on a single plot, if the scale can be rendered so that enough detail can be gleaned from each individual component. It may be necessary to "summarize" the distribution somewhat in order to see where the values lie (for example, indicating 10^{th}, 25^{th}, 50^{th}, 75^{th}, and 90^{th} percentiles, rather than showing the entire bootstrap density). In order to more readily compare values it may help to standardize by subtracting the mean and dividing by the standard deviation, though in many cases, if the means don't vary over too many standard deviations, simply looking at the original predicted values on (whether on the original scale or on a log scale) may be sufficient—sometimes a little judgment is required as to which plot will be most informative.

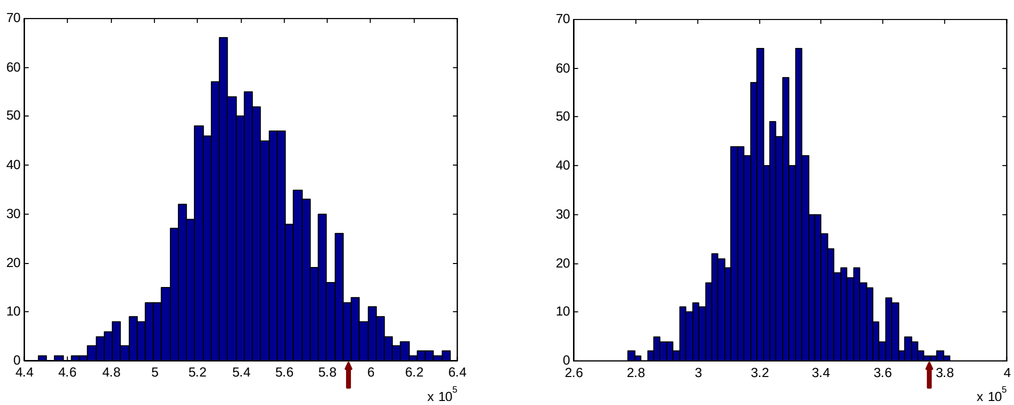
SECTION 5: EXAMPLES

Example 1

ABC data is workers compensation data for a large company. This data was analyzed in some detail in Barnett and Zehnwirth (2000).

In this example we actually use the bootstrap predictions discussed in the basic bootstrap introduction above, based on the second algorithm from Pinheiro et al. (2003). Below are the predictive distributions for the first two values (after DY0) for the last diagonal, for a quasi-Poisson GLM fitted to the data prior to the final calendar year, which was omitted. The brown arrows mark the *actual* observation that the predictive distribution is attempting to predict.

ABC Predictive distribution for last diagonal - histograms

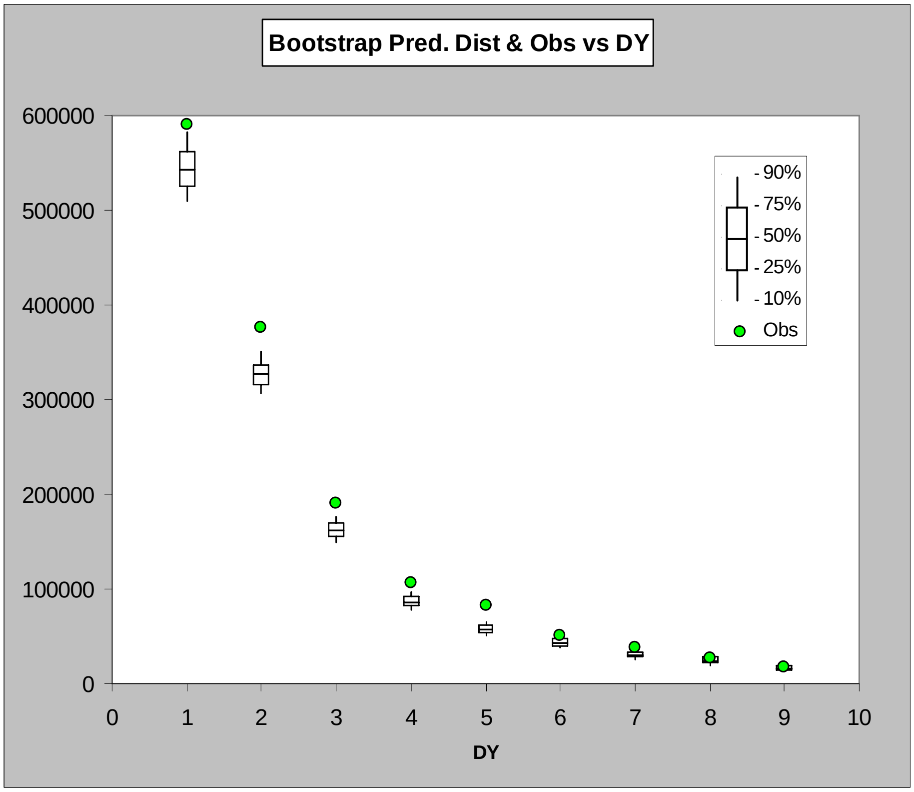


Delay 1 **Delay 2**

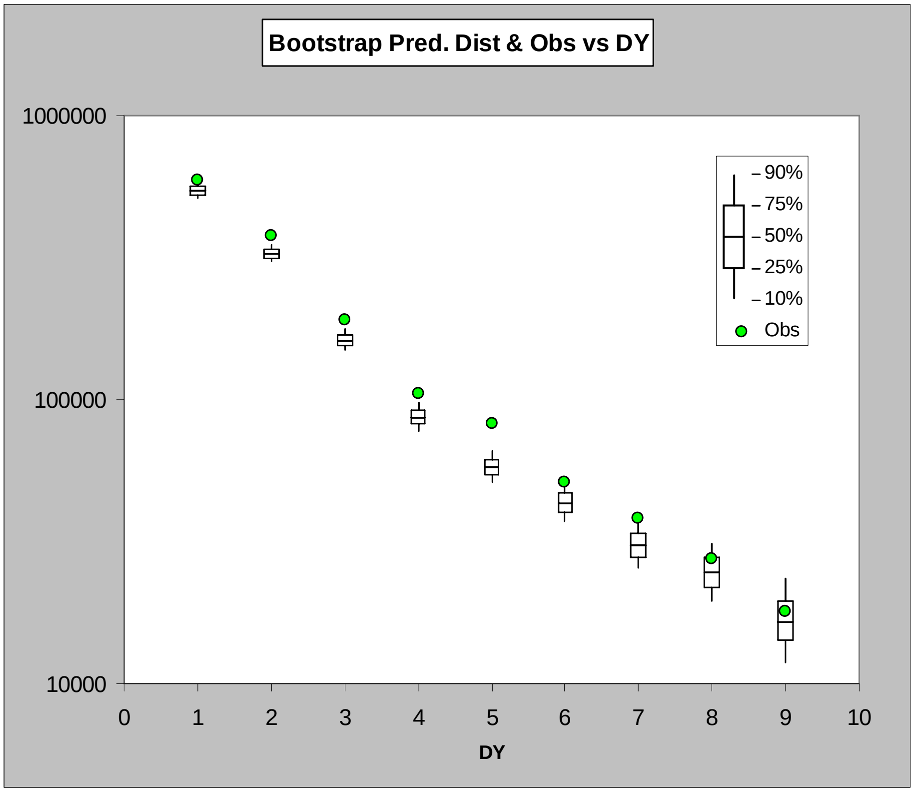
For the two distributions shown, the observed values sit fairly high. For a single observation, this might happens by chance, even with an appropriate model, of course.

The runoff decreases sharply for this triangle, so most of this information in the histograms would be lost if we looked at them on a single plot. Consequently, for a more detailed examination, the bootstrap results are reduced to a five-number summary (in the form of a boxplot) of the percentiles:

ABC Predictive distribution for last diagonal—box and whisker plots



The actual payments for the first seven development periods are all very high, but it's a little hard to see the details in the last few periods. Let's look at them on the log-scale:



Now we can see that in all cases the observations sit above the median of the predictive distribution, and all but the last two are above the upper quartile.

Below is a summary table of the bootstrap distribution for the final calendar year:

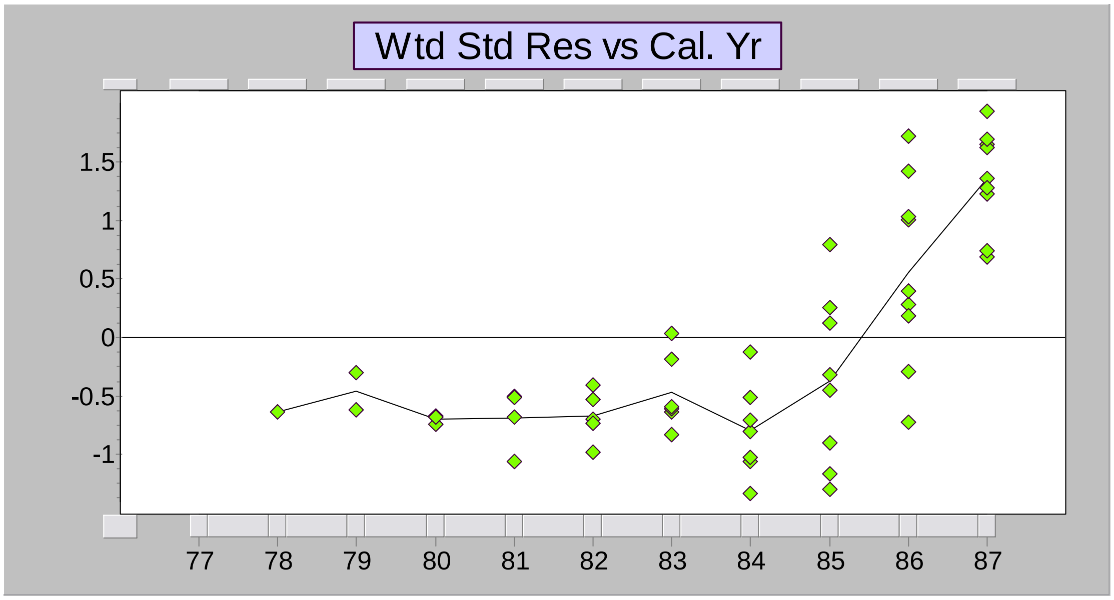
ABC: Bootstrap Predictive distributions for last calendar year

DY	Actual	10%	25%	50%	75%	90%	% ≤obs
0	496200						
1	590400	509620	525430	542150	562070	583270	93.9
2	375400	306580	315890	326600	337060	351080	99.6
3	190400	148750	155240	161520	169110	176290	98.9
4	105600	77760	81850	86220	91330	97340	99.2
5	82400	51050	54270	57740	61590	65730	100
6	51000	37440	40300	43360	46950	51380	89.3
7	38000	25490	27940	30770	33680	37110	92
8	27400	19430	21920	24540	27840	30970	72.9
9	18000	11930	14210	16460	19630	23450	63.9
10	12200						

So what's going on? Why is this predicting so badly?

We would see via one-step-ahead prediction errors that there's a problem with the assumption of no calendar-period trend; alternatively, as we noted earlier, we simply can look at residuals from a

Mack-style model and get a similar impression:

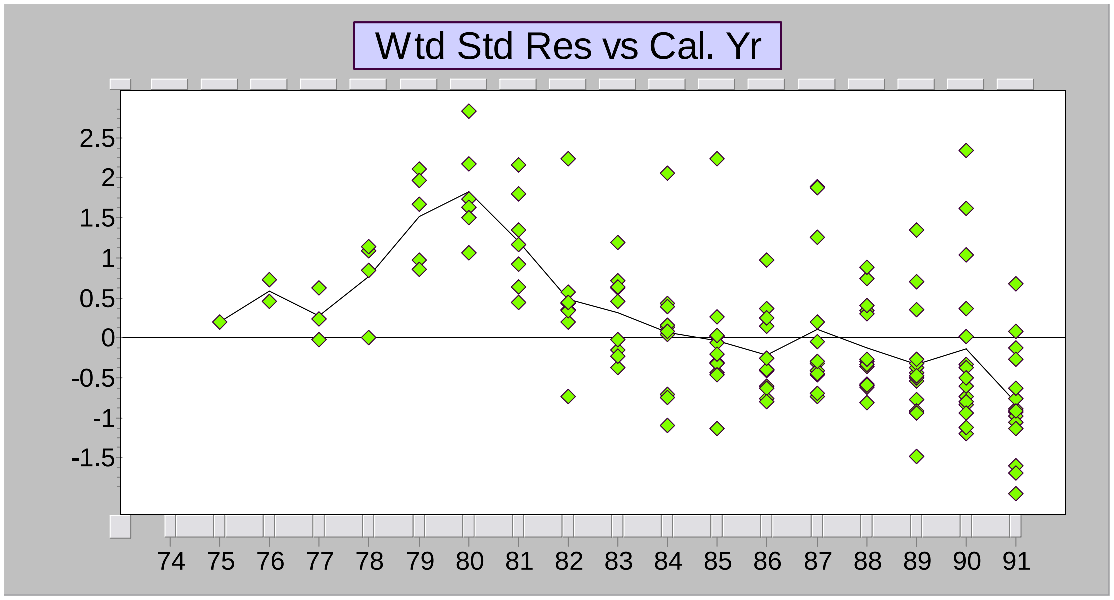


There is a strong trend-change in the calendar-year direction. Consequently, predictions of the last calendar year will be too low. One major difficulty with the common use of the chain ladder in the absence of careful consideration of the remaining calendar-period trend is that there is no opportunity to apply proper judgment of the future trends in this direction. The practitioner lacks a context for seeking all the information relevant to scenarios for future behavior.

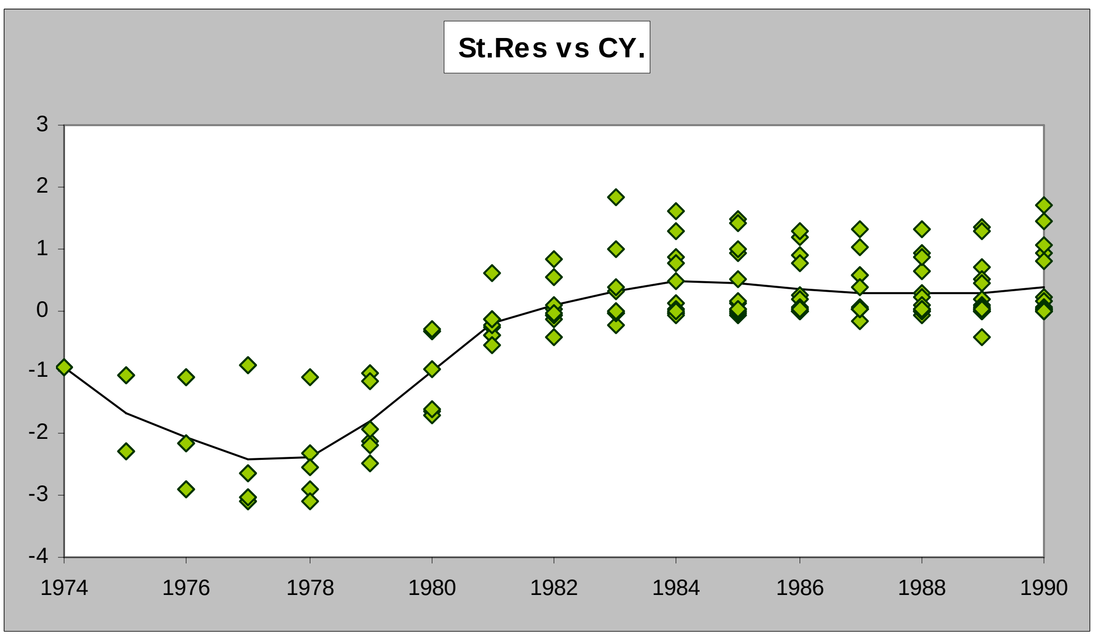
Example 2—LR high

As we have seen, we can look at diagnostics, which would have allowed us to assess *before we try to produce bootstrap prediction intervals* whether we should proceed.

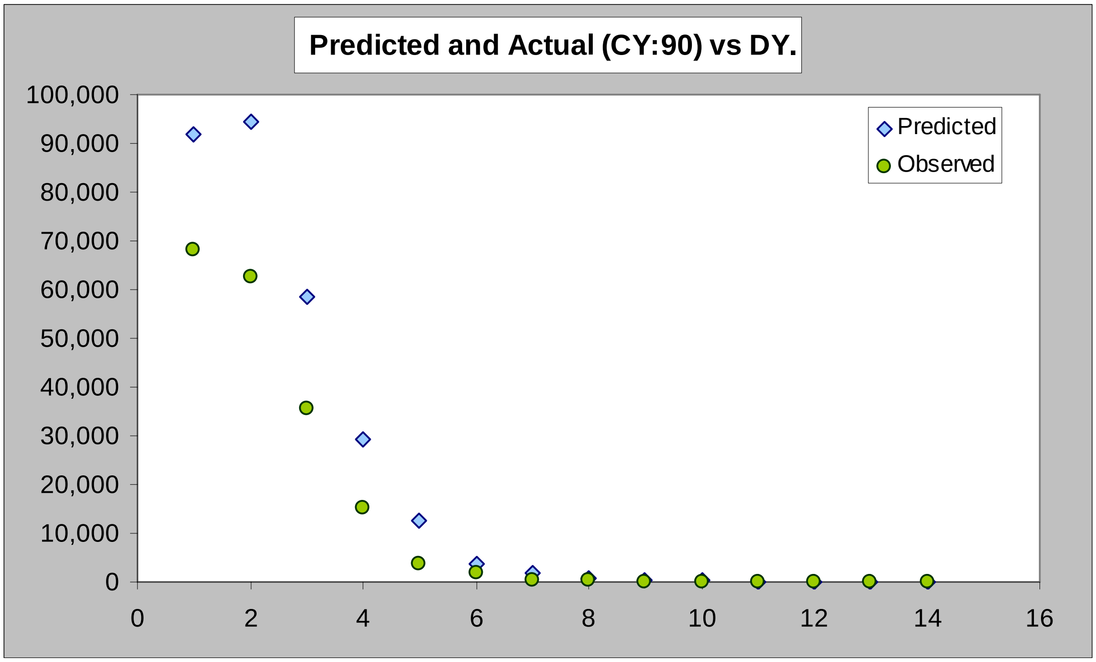
Here are the standardized residuals vs. calendar years from a Mack-style chain ladder fit. As you can see, there's a lot of structure.



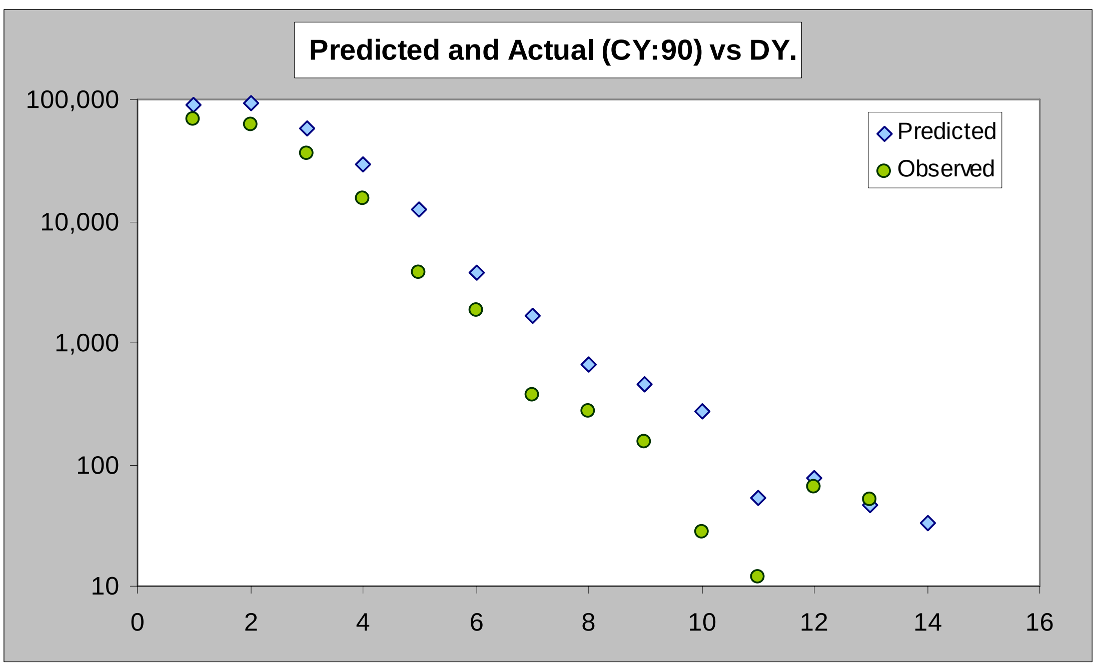
There's also structure in the quasi-Poisson GLM formulation of the chain ladder— residuals show there are strong trend changes in the calendar-year direction:



However, as we described before, this residual plot gives the incorrect impression that the GLM is underpredicting. This impression is incorrect, as we see by looking at the validation (one step ahead predictions) for the last year:

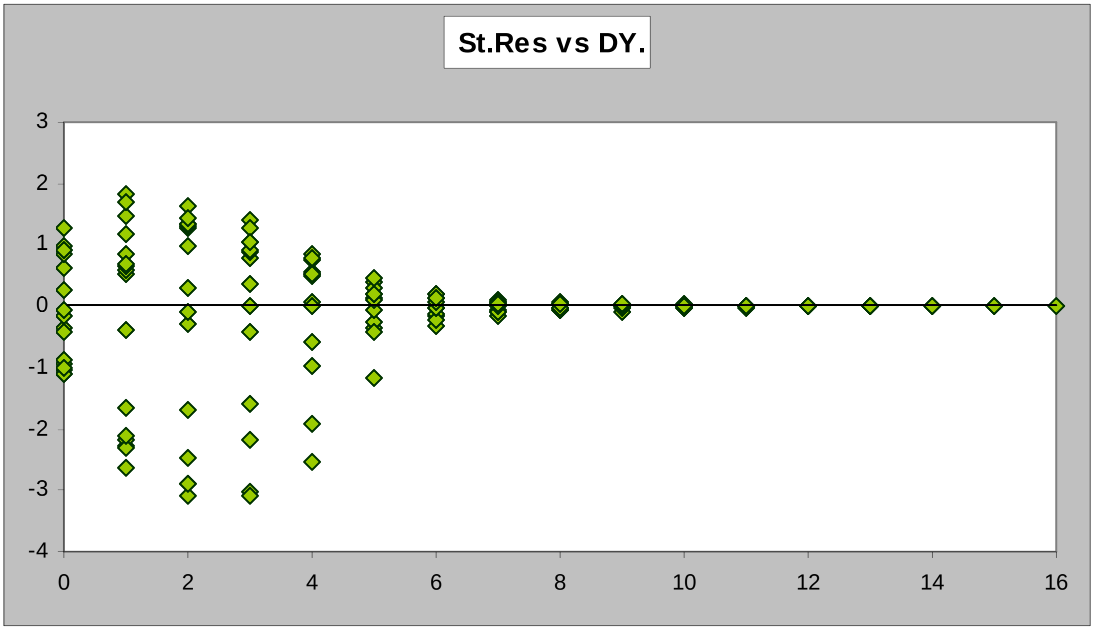


It's a little hard to see detail over on the right, so let's look at the same plot on the log scale:



The Mack-model residual plot gave a good indication of the predictive performance of the chain ladder (bootstrapped or not) for both the Mack model *and* the quasi- (overdispersed) Poisson GLM. It's always a good idea to validate the last calendar year (look at one-step-ahead prediction errors), but a quick approximation of the performance is usually given by examining residuals from a Mack-chain ladder model.

A further problem with the GLM is revealed by the plot of residuals vs. development year. The assumed variance function does not reflect what's in the data—and hence the prediction intervals cannot be correct:



Example 3

The next example has been widely used in the literature relating to the chain ladder. Indeed, Pinheiro et al. (2003) referred to it as a "benchmark for claims reserving models." The data come from Taylor and Ashe (1983).

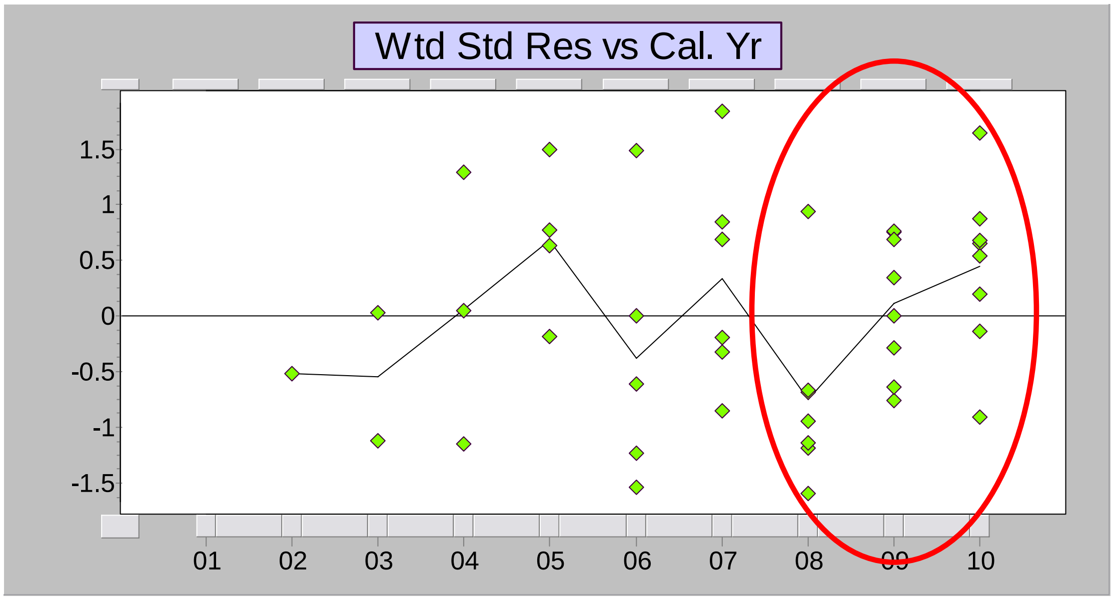
Here are the bootstrap predictive means and s.d.s for the last diagonal (i.e., with that data not used in the estimation) for a quasi-Poisson GLM, and the actual payments for comparison:

DY:	1	2	3	4	5	6	7	8
CL pred	931994	1000686	1115232	482991	325851	443060	231680	309629
mean:	958887	1021227	1114169	490137	328453	452636	242346	327365
stdev:	452285	331706	318026	195225	156289	200080	152170	227644
actual:	986608	1443370	1063269	705960	470639	206286	280405	425046

Firstly, there is an apparent bias in the bootstrap means. The chain ladder predictions sit below the bootstrap means, indicating a bias. Since, the ML for a Poisson is unbiased, if the model is correct, these predictions should be unbiased. This doesn't *necessarily* indicate a bad predictive model, but is there anything going on?

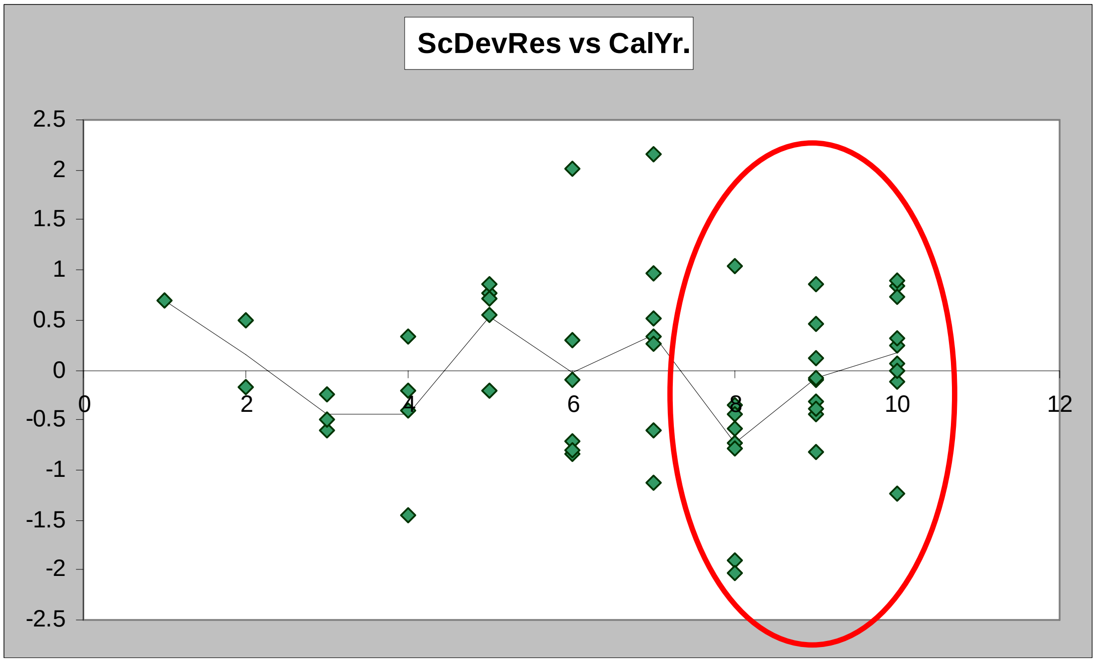
In fact there is, and we can see problems in residual plots.

Here is a plot of the residuals vs. calendar year from a Mack-type fit:



Strong calendar-period effects are evident in the last few years. The existence of a calendar-period effect was already noted by Taylor and Ashe in 1983 (who included the late calendar-year effect in some of their models), but it has been ignored by almost every author to consider this data since. If the trend were to continue for next year, the forecasts may be quite wrong. If we didn't examine the residuals, we may not even be aware that this problem is present.

Exactly the same effect appears when fitting a quasi-Poisson two-way cross classification with log-link:



There's a benefit in examining Mack residuals before fitting a quasi-Poisson GLM—the residuals are a little easier to produce, and the plot of residuals vs. fitted has more information about the predictive ability of the model.

Some other considerations

All chain ladder-reproducing models (including both the quasi-Poisson GLM and the Mack model) must assume that the variance of the losses is proportional to the mean (or they will necessarily fail to reproduce the chain ladder). This assumption is found to be rarely tenable in practice—as we saw in example 2—and for an obvious reason. While it can make sense with claim counts—for example, when the counts are higher on average they also tend to be more spread but with lower coefficient of variation. If the counts happen to be Poisson-distributed, the variance will be proportional to the mean (in fact equal to it). Heterogeneity or various forms of dependence in claim probabilities can make the Poisson untenable even for claim numbers. But with claim payments, the amount paid on each claim is itself a random variable, not a constant, and variable claim payments will make the variation increase faster than the mean. Simple variation in claim size (such as a constant percentage change, whether due to inflation effects or change in mix of business or any number of other effects) will make the variance increase as the square of the mean, while claim size effects that vary from policy to policy can make it increase still faster. Dependence in claim size effects across policies can make it increase faster again. Consequently the chain ladder assumption of variance proportional to mean must be viewed with a great deal of caution and carefully checked.

The chain ladder model is overparameterized. It assumes, for example, that there is *no* information in nearby development periods about the level of payments in a given development, yet the development generally follows a fairly smooth trend—indicating that there is information there, and that the trend could be described with few parameters. This overparameterization leads to unstable forecasts.

Finally, in respect of the bootstrap, the sample statistic may in some circumstances be very inefficient as an estimator of the corresponding population quantities. It would be prudent to check that it makes sense to use the estimator you have in mind for distributions that would plausibly describe the data.

SECTION 6: DEALING WITH OBSERVED MODEL INADEQUACY

As we have seen, the chain ladder models considered so far don't predict well with the data triangles we looked at—even though two of them are "standards" for illustrating ratio models.

Calendar-period trend changes

Pattern in the plot of residuals or prediction residuals vs. calendar period indicates calendar-period trend changes.

As we have seen, calendar-period trend changes do show up in real data. Further, because substantive changes don't generally occur frequently (such as every year or two), but more occasionally, changed rates may sometimes be expected to continue for some period (though it depends on the cause of the change—for example, with the triangle ABC, the cause of the calendar-period trend change was a known change in legislation, for which the higher identified rate was not expected to continue; in that case the projections of a ratio method will be too high after the rate drops back. It has sometimes been stated that ratio methods project at an average of past calendar-period rates, but in fact it is not the case that there is a single rate at which future observations are being inflated. Even if it were true, whether the new rate is to continue or discontinue, an average rate would be unsuitable.

Changing inflation can be modeled properly with calendar-year parameters. However, we must beware—the loglinear quasi-Poisson GLM cannot be readily modified in this way. That is, while it is possible to add calendar-year parameters to the GLM (it's no longer chain ladder, of course), the new model is demonstrably unsuitable for inflated payment data. Imagine a triangle with no inflation that otherwise meets the assumptions of the loglinear quasi-Poisson GLM. Now construct a new triangle from the old that has constant nonzero inflation in the last four calendar periods—say running at 10% per period. Note that Var($k.X$) = k^2.Var(X). The variance of the inflated observations increases as the square of the factor by which they are inflated. But the quasi-Poisson model requires that the variance increase proportional to the mean, so the model requires Var($k.X$) = k.Var(X). Since there will be a different factor ("k") for each calendar period in the inflating region, this model *cannot* be consistent with the data.

Trend in the one-step-ahead prediction residuals vs. predicted plot

When there are no changing calendar-period trends, trend in the plot of one-step-ahead prediction residuals vs. predicted values often occurs, and indicates a model inadequacy. In the case of the loglinear quasi-Poisson GLM, this pattern does not appear in the equivalent "within-data" plot—the plot of residuals vs. fitted values.

In the case of the Mack model, if it is present in the plot of prediction residuals, it will generally also be seen in the corresponding residuals vs. fitted plot. In the case of the Mack model, it implies the need for an intercept term (see Barnett and Zehnwirth, 2000, or Murphy, 1994). There does not appear to be a simple modification of the quasi-Poisson GLM that is able to deal with this form of model inadequacy. And ordinary residual plots don't reveal its presence.

Because of the frequent presence of superimposed inflation (claims inflation at different rates to economic inflation), it is necessary to model incremental values. We believe that a log-transform is frequently beneficial both from the point of view of linearizing inflation effects, linearizing trend in the late developments (reducing the number of parameters required), and for stabilizing the variance in terms of the mean.

The Probabilistic Trend Family of models

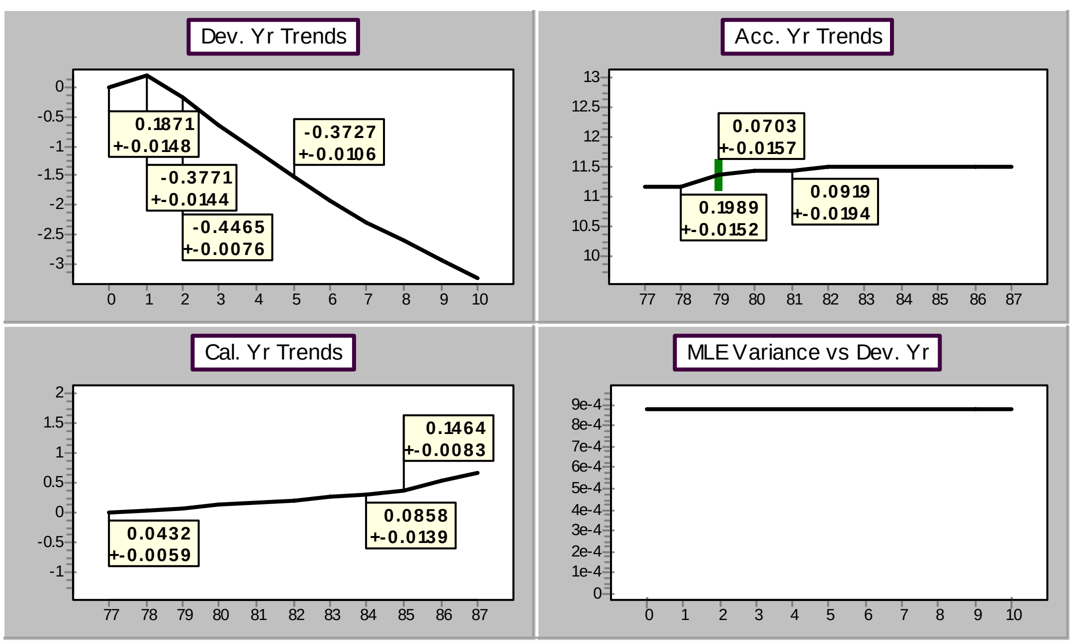
When models better describe the characteristics of the data, the prediction intervals tend to have the required properties (such as including near to the anticipated proportion of future observations).

The Probabilistic Trend Family (PTF) models consist of a model for the mean trends in the three directions of the triangle, and a model for the random variability about the trends. It is applied to log-incrementals, adjusted for exposures and economic inflation (where these are available). Many triangles are well described by a few parameters in the early developments (to capture the "run-off"), and where there are trend changes against the accident periods or calendar periods (indeed, in many cases the timing of these may be known in advance), parameters in those directions as well. Often the variability is constant on the log scale, though sometimes it exhibits a variance change against the development periods, requiring some modeling of the variance. The distribution of the inflation-adjusted and exposure normalized data is assumed to be independent lognormal—which implies normality on the log-scale. This assumption should be checked, but is in practice almost always a good description of the data. The observations are assumed to be independent.

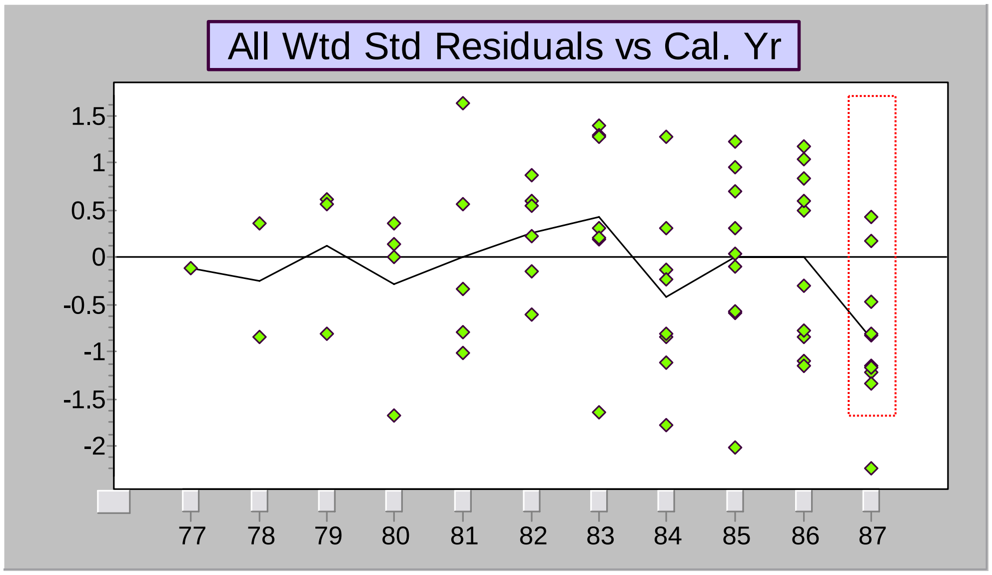
Accident-period parameters represent "levels," while development and calendar-period parameters describe linear trends (in the logs). The Probabilistic Trend Family is described in more detail in Barnett and Zehnwirth (2000).

Because of the simple form of these models, they may be represented pictorially by a decomposition of the model for the mean into trend changes in each direction and the model for the variability about it.

For example, the figure below shows a display of a reasonable model for the ABC data of Example 1. (This model includes treating the observation at 1982 delay five as an outlier and giving it zero weight in the estimation.)

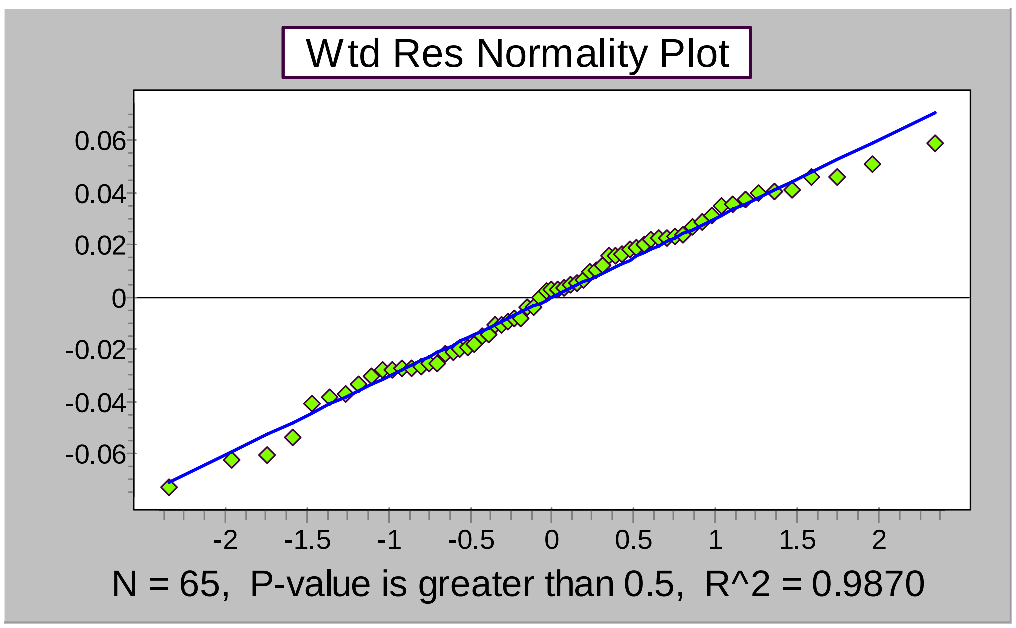


Examination of diagnostic plots indicate that the model is a reasonable description of the data. For example, the next figure shows residuals, with prediction residuals for the final diagonal, against calendar year. The final calendar period is reasonably well predicted by the model, even though those observations are not included in the estimation.



Standardized residuals and one step ahead prediction errors with approximate 90% confidence interval for the predictions.

The final plot clearly shows that the lognormal assumption seems reasonable.



Residuals against expected normal scores for all years

Multivariate PTF

Where related triangles are being analyzed, such as different subsets of a line of business (excess of loss layers, different territories, or different claim types), different lines within a single company, or related lines across several business units, it is essential to be able to model the related triangles together. It may be that there are related trend changes across triangles and the errors about the models may be correlated. The PTF models may be extended by incorporating correlated error terms and the possibility of related parameters or changes in parameters. When the models are the same (in terms of where trend changes are located), these form generalized least squares (GLS) models. Where parameters are unrelated, these are seemingly unrelated regression (SUR) models.

These models are especially useful for calculation of diversification effects, for example, in risk capital calculations or for reserving.

By providing an adequate description of the ABC data, the identified model from the Probabilistic Trend Family is able to predict the final year (in the sense that the observations are reasonably consistent with the predictive distribution, as indicated by the validation residuals). As long as the model used to predict the future (which will be informed by the model for the past) is valid, confidence intervals and predictive intervals should have close to the right coverage probabilities, making them suitable inputs to the determination of the relevant ranges.

SECTION 7: CONCLUSIONS

Prediction intervals are important components of risk capital calculations, but such intervals rely on the model's predictive assumptions. We frequently find that for commonly used models, those predictive assumptions are violated, and we find that the models often don't predict the most recent data well.

Consequently, when fitting a quasi-Poisson GLM, it's important to check the one-step-ahead prediction errors in order to see how it performs as a predictive model—the residuals against fitted values don't show you the problems. Alternatively, the Mack residuals can be useful approximate diagnostic tools for the *predictive* assumptions of the quasi-Poisson GLM.

Predictive diagnostics should also be looked at before bootstrapping a model and once a bootstrap has been done, you should also validate at least the last year—that is, examine whether the actual values from the last calendar year could plausibly have come from the bootstrap predictive distribution standing a year earlier.

The use of the bootstrap does not remove the need to check assumptions relating to the appropriateness of the model. Indeed, it is clear that there's a critical need to check the assumptions. Use of the bootstrap does not avoid the fact that chain-ladder type models have no simple descriptors of features in the data. We show in several examples that there is much remaining structure in the residuals.

If it is the predictive behavior that is of interest, prediction errors are appropriate tools to use in standard diagnostics, and they can be analyzed in the same way as residuals are for models where prediction is within the range of the data.

Checking the model when utilizing the bootstrap technique is achieved in much the same way as it is for any other model—via diagnostics—but they should be predictive diagnostics selected with a clear understanding of the problem, the model and the way in which the bootstrap works.

REFERENCES

[1] Ashe, F., "An essay at measuring the variance of estimates of outstanding claim payments," *ASTIN Bulletin* 1986, 16S:99–113.

[2] Barnett, G. and B. Zehnwirth, "Best Estimates for Reserves," *Proceedings of the Casualty Actuarial Society* 2000*,* 87(167):245-321.

[3] England P.D. and R.J. Verrall, "Stochastic Claims Reserving in General Insurance," *British Actuarial Journal* 2002, 8(3):443-518.

[4] Barnett, G., B. Zehnwirth, and E. Dubossarsky, "When can accident years be regarded as development years?" *Proceedings of the Casualty Actuarial Society* 2005, 92(177):239-256, www.casact.org/pubs/proceed/proceed05/05249.pdf

[5] Chatfield, C., *Time-Series Forecasting* (Boca Raton, Florida: Chapman and Hall/CRC Press, 2000).

[6] Efron, B. "Bootstrap methods: another look at the jackknife," *Annals of Statistics* 1979, 7:1-26.

[7] Efron, B. *The Jacknife, the Bootstrap and Other Resampling Plans,* vol. 38*,* (Philadelphia, Pa., Society for Industrial and Applied Mathematics, 1982).

[8] Efron, B. and R. Tibshirani, *An Introduction to the Bootstrap* (New York: Chapman and Hall,. 1994).

[9] England, P.D. and R. J. Verrall, "Analytic and bootstrap estimates of prediction errors in claims reserving," *Insurance: Mathematics and Economics* 1999, 25(3):281-293.

[10] Hachemeister, C. and J. Stanard, "IBNR Claims Count Estimation with Static Lag Functions," Presented at the 1975 Spring Meeting of the Casualty Actuarial Society.

[11] Mack, T. "Measuring the variability of chain ladder reserve estimates," *Casualty Actuarial Society Forum*, Spring 1994, vol. 1:101-182.

[12] Mack, T., "Which stochastic model is underlying the chain ladder method?" *Insurance Mathematics and Economics* 1994, 15(2-3):133-138.

[13] Moore, D. S., G.P. McCabe, W.M. Duckworth, and S.L. Sclove, "Bootstrap Methods and Permutation Tests," companion chapter 18 in *The Practice of Business Statistics*, (New York: W. H. Freeman, 2003), http://bcs.whfreeman.com/pbs/cat_140/chap18.pdf.

[14] Murphy, D., "Unbiased Loss Development Factors," *Proceedings of the Casualty Actuarial Society* 1994, 81(154):154–222.

[15] Pinheiro, P.J.R., J. M. Andrade e Silva, and M. L. Centeno, "Bootstrap Methodology in Claim Reserving," *Journal of Risk and Insurance* 2003, 70(4):701-714

[16] Reinsurance Association of America, *Historical Loss Development Study*, (Washington: RAA, 1991).

[17] Renshaw, A.E. and R. J. Verrall, "A stochastic model underlying the chain ladder technique," *Proceedings of the XXV ASTIN Colloquium* 1994*,* Cannes, France.

[18] Stanard, J.N. "A Simulation Test of Prediction Errors of Loss Reserve Estimation Techniques," *Proceedings of the Casualty Actuarial Society* 1985, 72(137):124-148.

[19] Taylor, G.C. and F.R. Ashe, "Second Moments of Estimates of Outstanding Claims," *Journal of Econometrics* 1983, 23(1)37-61.

Modeling with the Multivariate Probabilistic Trend Family

Glen Barnett, Ph.D. and Ben Zehnwirth, Ph.D., AIA, AIAA[1]

1 Professorial Visiting Fellow, School of Actuarial Studies, Australian School of Business, University of NSW

Abstract: This paper motivates the benefits of modeling trends, volatility and correlations through a study of real data triangles.

We show that a model that is demonstrably unable to forecast the recent past of the historical triangle cannot be expected to tell us anything useful about the future of the same process. Naturally, the same basic model fitted by applying a more sophisticated tool will suffer the same fate. The use of GLMs, bootstrapping, or Bayesian statistics cannot avoid the basic defects of traditional methods.

With traditional techniques the parameters (e.g. age-to-age factors) are a function of the data. By contrast, in the Probabilistic Trend Family (PTF) modeling framework the model design as well as the parameters are a function of the data. We illustrate PTF modeling (e.g., Barnett and Zehnwirth, 2000) on a variety of real triangles.

The PTF modeling framework is extended to the simultaneous modeling of multiple triangles. The multivariate modeling framework (MPTF), apart from describing the volatility in each triangle also describes correlations between them in two different ways.

The MPTF modeling framework can be used in a number of innovative ways yielding useful information about the risk characteristics of the business. There are important implications for economic capital calculations and optimal retention. In order to compute economic capital for reserve risk and underwriting risk the correlations between lines of business need to be known. To assess the correlations accurately (whether from related trends or correlated errors), a model for each line that describes the trend structure and the volatility about the trend structure needs first to be identified.

SECTION 1: RATIO MODELS AND THE CHAIN LADDER

The intuitive basis of ratio models is simply that it is expected that if a known cumulative paid loss (or an incurred loss) in one period is high, that the corresponding figure for the next period will also tend to be high, and if it is low, it will subsequently tend to be low. The *ratio assumption* holds that the next value is expected to be in direct proportion to the current known figure.

In what follows, we will drop the parenthetical reference to incurred losses, but the subsequent discussion will (with possibly minor but obvious alterations) continue to apply.

Notation

We label the accident years in the triangle $i = 1, \ldots n$, and the calendar years similarly, $t = 1, \ldots n$. For ease of notation, development year is taken as the delay between accident and payment year ($j = t - i$), and hence $j = 0, 1, 2, \ldots, n{-}1$. Let the incremental amount be p_{ij} , and let the cumulatives total paid or incurred to date in an accident year be c_{ij}. Further, let y_{ij} be the logarithms of incrementals (possibly after adjusting for economic inflation or exposures).

Ratios

If C_{ij} is the figure for accident period i, development period j then an obvious way to record that intuitive expectation is

$$\mathrm{E}(C_{ij} \mid C_{i,j-1} = c) = \beta_j c\ . \tag{1}$$

There are a variety of ways to estimate the ratio parameters, β_j. Different estimators will correspond to different implicit assumptions about the remainder of the model.

On a worldwide basis, ratio methods continue to be very popular, and the chain ladder is undoubtedly the most popular of the ratio methods in use. The standard chain ladder estimators correspond to assuming that the variance of the C_{ij} is proportional to the mean (in the sense that the estimators and subsequent forecasts are optimal under that assumption; otherwise better estimators exist).

These assumptions underly chain ladder forecasts. If we consider a particular accident year, i, and j-1 is the last observed development in that year, the chain ladder forecast of C_{ij} (the first future cumulative payment in that year) follows equation (1).

There are a number of models that reproduce the chain ladder. The approach of Mack (1993) and Murphy (1994) reproduces chain ladder forecasts by combining equation (1) with the variance-proportional-to-mean assumption. It is possible to derive predictive means and variances from that.

The other widely used model that reproduces the chain ladder forecasts is the (quasi-) Poisson model with log-link and linear predictor corresponding to a two-way main-effects ANOVA. This Poisson model was introduced to actuaries by Hachemeister and Stanard (1975). The approach became popular in the 1990s after the paper by Renshaw and Verrall (1994). This model is now popularly referred to as the overdispersed Poisson model (ODP) – however the term "overdispersed" can be incorrect because the scale parameter ϕ, may (with some choices of scale, such as may occur if the data are measured in millions or billions of dollars) easily be less than 1, and therefore underdispersed. The term quasi-Poisson is more appropriate, since it applies irrespective of the size of the scale parameter.

Within the data, this model has a different structure for the mean, which corresponds essentially to the assumption:

$$\mathrm{E}(C_{ij}) = \beta_j \mathrm{E}(C_{i,j-1})\ . \tag{2}$$

However, the *forecasting function* of this model has the same structure as (1) – as indeed it must in order to reproduce the chain ladder forecasts. For example, just as with the Mack model, for a given accident year, i, and j-1 is the last observed development, the chain ladder forecast of C_{ij} follows equation (1).

SECTION 2: ASSESSING THE CHAIN LADDER ASSUMPTIONS

2.1 Chain ladder assumptions and diagnostics

Since the primary interest for actuaries lies in the forecasting performance of these models (rather than the estimates within the data), it is crucial to test that the structure described by model (1) for ratio models in general; it is still necessary to assess within-data fit (such as via residual plots), but it is also necessary to assess the appropriateness of the prediction equation if the two differ. This is discussed in detail in Barnett, Zehnwirth, and Odell (2008), where one-step-ahead prediction errors are used to discover inadequacies of the *predictive performance* of the quasi-Poisson GLM, and it is shown that ordinary residuals do not indicate problems with the predictive properties of this model.

For present purposes, it suffices to approximate the predictive performance of the quasi-Poisson GLM by looking at displays of the residuals from the Mack model as an adjunct to the residual analysis of the GLM fit. Mack residuals are a good approximation because the one-step-ahead prediction errors and the Mack residuals are based on the same estimates – but in the case of the quasi-Poisson GLM, on fewer observations (the Mack residuals use data from all available accident years for its residuals, while the GLM one-step-ahead predictions don't use data from accident years later than the observation under consideration). In practice, simply using the Mack residuals is an effective way to assess the predictive performance of the two-way quasi-Poisson GLM with log-link.

Consequently, for the present paper, unless otherwise noted, we will use residuals from a Mack-style model (modified as indicated below) as an indicator of predictive performance of both of the popular chain-ladder-reproducing models.

ELRF models

In Barnett and Zehnwirth (2000), we expanded the framework of Murphy (who added intercepts) to construct diagnostic tests of ratio models, including the chain ladder, constructing the Extended Link Ratio Family of models (*ELRF*).

We then analyzed a number of real data triangles using standard regression diagnostics in the ELRF framework and showed that the assumptions of ratio models failed for all of the data sets analyzed. In this paper, we have carried out a small study using such diagnostics on randomly chosen triangles (in section 3) to assess the suitability of the chain ladder for a variety of different triangles.

2.2 Further issues with the chain ladder

Barnett, Zehnwirth, and Dubossarsky (2005) describe the incremental chain ladder forecasts are the same whether cumulation and ratios are done across development years or down accident years. This property – that development and accident years are interchangeable – has some worrying implications for all chain ladder-reproducing models.

These include:

(i) The fact that the accident and development directions are not distinguished by the model, despite the fact that we know them to behave quite differently;

(ii) being fully parameterized (implicit or explicit) in both those directions – in spite of the fact that the development years and accident years may both be described with only a few parameters;

(iii) Both directions exhibit clear relationships within themselves (though of different kinds), but the model doesn't attempt to utilize this information.

A second issue occurs in the case of the quasi-Poisson GLM, which we don't believe has been noted elsewhere. It is possible to add calendar-period parameters to the quasi-Poisson model, and it has been suggested that this would be suitable for dealing with changing inflation. However, this is not the case, as can be seen by a simple argument.

Imagine that the portfolio has been stable for many years, the shape of the runoff has remained unchanged, and that superimposed inflation has been zero. Then, for the most recent calendar year, superimposed inflation at 10% occurs. Clearly, the mean for each observation is 10% higher than the corresponding observation a year earlier, and – since $\text{Var}(cX) = c^2\,\text{Var}(X)$ – the standard deviation is also 10% higher. Yet if we model the calendar period changes with parameters in the quasi-Poisson model, the model asserts that the standard deviation will just under 5% higher. The model description of the variance is incorrect. Consequently, prediction intervals – including bootstrap prediction intervals – based on this model will be too narrow.

Note that if the inflation is random, rather than constant, the induced variance will increase still faster.

The bootstrap and Bayesian methods

Refinements such as the bootstrap and the use of Bayesian methods don't alter the adequacy or lack or adequacy of the mean and variance components of a given model – if the mean and variance of the data are not well described without them, the mean and variance will not be will described with them (unless those components of the model itself are altered).

SECTION 3: THE CHAIN LADDER EXPERIMENT

Design

A random sample of 25 complete triangles from Schedule P data was taken. Either the Paid or Incurred triangle was used (60% chance of selecting paid, 40% chance of selecting incurred).

For each such triangle, the suitability of the ratio (chain ladder-like) assumptions were tested according to a number of criteria

Criteria

There were four primary diagnostic criteria on which the chain ladder assumptions were assessed.

1) Do the lines relating incremental to previous cumulative need an intercept?

2) Are the correlations between incremental and previous cumulative "small?"

3) Do the residuals exhibit substantial changing trends against calendar periods?

4) Do the residuals have trends against fitted values?

If the answer to any of these questions is a clear yes, the chain ladder model does not hold.

In addition, there was a fifth criterion:

5) After removing a linear trend down the accident periods, are remaining linear correlations between incremental and previous cumulative "small?"

The fifth criterion is to see whether there is an alternative explanation for a correlation between an incremental and the previous cumulative due to increases running down the accident years (which could be caused by increasing exposures or inflation). By fitting a simple linear accident trend, the chain ladder fails this criterion if for a substantive number of years the ratio is no longer a significant predictor.

Because of the small sample sizes, the assessment is not always obvious graphically, so we perform this test in the regression by looking at the ratio fitted last to see if it still has explanatory power.

The first two questions are assessed individually for pairs of adjacent developments (individual regressions), while the latter two questions are assessed globally (across all residuals). The final question is related to question 2 and again assessed on adjacent pairs of developments individually.

When assessment is performed on individual pairs of developments, only the first six pairs of developments are considered (there are too few points after that for a reliable assessment).

1) *Zero intercept:* Least squares line on graph of incremental vs. previous cumulative has intercept that is plausibly near origin; 68% of triangles failed on many years (3-6), (all triangles failed at least some years).

2) *Incremental significantly correlated with previous cumulative*: 52% of triangles fail on many years (3-6), 68% fail on several years.

3) *No CY trend changes*: 64% of triangles display strong CY trends, 28% display moderate CY trends, and 8% display weak indication of (plausibly random) CY trends.

4) *No trend in residual vs. fitted*: 32% show strong trends vs. fitted (linear or quadratic), 28% show moderate trends vs. fitted. 40% show little trend vs. fitted values.

Overall, 96% of triangles fail substantively (strong residual trends or fail on many years for the first pair of criteria) on at least one of these four criteria. Four percent show at least partial success for the chain ladder assumptions.

5) *Relationship not just a proxy for accident trends:* 84% substantively fail (that is, the relationship between y and x is substantively accounted for by a simple increasing or decreasing trend in the accident period direction, which indicates that exposures or inflation should be brought into the model). A further 4% partially fail (relationship remains for a few periods). In 4% the relationship for several years is strongly quadratic (a failure of the chain ladder assumption of linearity). In 8% there is remaining linear correlation between incremental and previous cumulative after accounting for a simple AY trend in incrementals.

Some of these criteria are plainly subjective (such as those based on appearance of diagnostic plots), but those effects classified as "strong" have deficiencies that are quite plain (though there's no clear dividing line between strong and moderate effects, so some people may well classify some of the moderate effects as strong or vice-versa); there is also room in the moderate/weak borderline for disagreement. Nevertheless, even by extremely generous criteria, still only a fraction (far less than a quarter) of the triangles considered could be even remotely regarded as suitable for the chain ladder.

If the results of this small study were to hold more generally, it seems that relatively few triangles will satisfy the basic assumptions about the mean for ratio models. Note that there are actually further assumptions required, such as variance assumptions – for the chain ladder, the assumption is that the variance is proportional to the mean. We have seen in other examples that these assumptions are not always tenable, so the proportion of triangles for which the model assumptions are plausible may be substantially lower than the present study might suggest.

It seems the only prudent course is to assume that ratio models do not describe a triangle, unless it has been clearly established that the model is a plausible description of the data.

SECTION 4: PROBABILISTIC TREND FAMILY MODELS

The problems that are regularly identified in the ratio models lead us to try to solve them.

i) We need to be able to deal with calendar-year trends (such as superimposed inflation) *and* to be able deal properly with the effect that multiplicative effects such as inflation have on the variance. This is one of many aspects that lead us to move from considering merely a log-link (as in the quasi-Poisson model) to a model with variance proportional to the square of the mean (of course, variance assumptions should be checked for reasonableness, as with any other assumption).

ii) The intercept term in the ELRF models is often necessary, but once calendar- and accident-year trends have been modeled, even in a very crude way, ratios rarely contribute anything further. This suggests that we should model the development year levels (corresponding to intercept terms) as well as effects in the calendar- and accident-years' directions.

iii) The overparameterization issue suggests that we should take advantage of the smoothness typically seen in the development period direction by relating the levels. Similarly, the periods of constant inflation or stable accident year levels tend to be fairly long-lived and we should be able to take advantage of that fact. (Further, in the full version of the model it is possible to relate/smooth accident years in a more extensive way than presented here.)

4.1 The basic model

All of these issues together point us toward what we call the Probabilistic Trend Family of models (*PTF*), which has the suggested features. It models incremental payments on the log scale. This allows inflation to be modeled appropriately (inflation impacts current and future payments, not past payments; consequently, cumulative payments and incurred are unsuited to models dealing with inflation). These may be normalized for a measure of exposure and adjusted for inflation before taking logs. The model described here may be fitted in a regression package (even in Excel), but predictive distributions of aggregates is somewhat involved.

The PTF approach models data as four components:

$$\text{data} = \text{development trend} + \text{accident trend} + \text{calendar trend} + \text{random.} \tag{4.1}$$

The first three components model the mean and the final component models the variability about the mean.

(i) *development trend*: The structure for modeling the runoff in the development year direction is as follows:

$$\delta_j = \gamma_1 + \gamma_2 + \ldots + \gamma_j \ . \tag{4.2}$$

The gamma parameters (γ_j) represent the shift in mean between the previous development and the current one. This allows models to adapt to shifts in level.

Setting gamma parameters to be equal allows the efficient modeling of constant trends. In a particular case, the fitted development year trend (runoff), $\hat{\delta}_p$, looks like this:

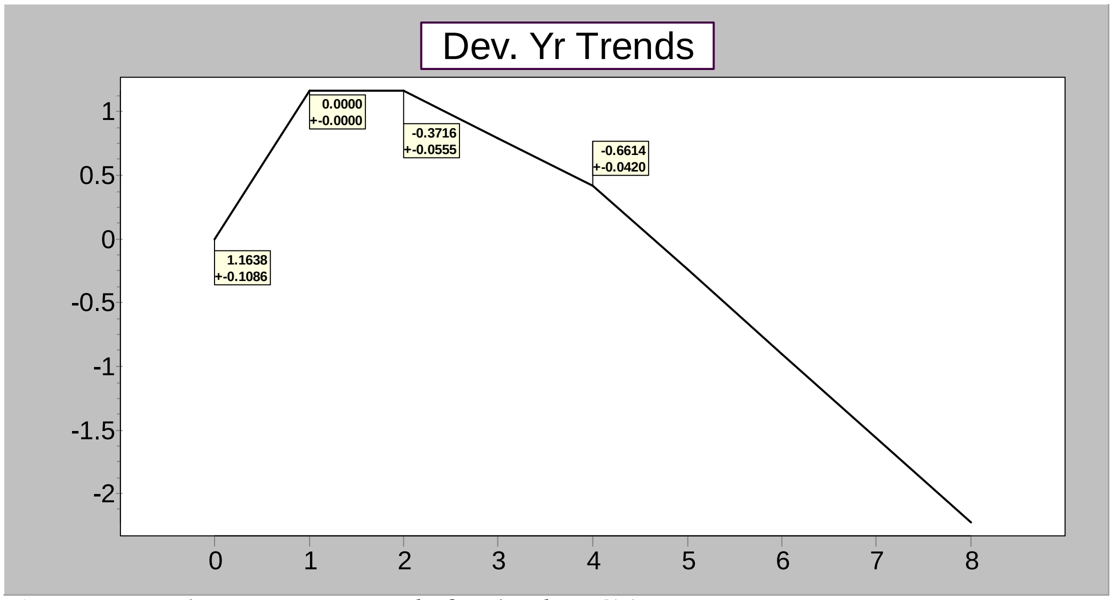


Figure 1: Development year trends for the data CTP

The level in development year zero is represented by a parameter alpha (α), and the gamma parameters represent (percentage) shifts to each development period after that.

For a single accident year, the average log(payment) in development j is:

$$E(y_j) = \alpha + \gamma_1 + \gamma_2 + \ldots + \gamma_j \ . \tag{4.3}$$

In the case where there are no fitted trends in the mean in the other directions (accident year, calendar year), the model for the mean for the entire array can be of the same form:

$$E(y_{ij}) = \alpha + \gamma_1 + \gamma_2 + \ldots + \gamma_j \ . \tag{4.4}$$

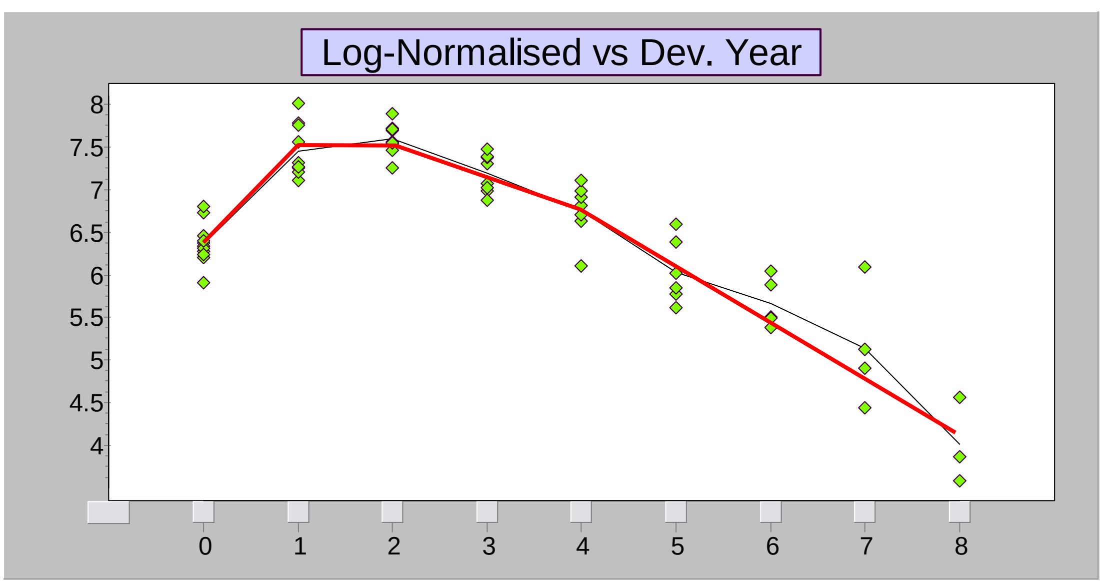


Figure 2: CTP data vs. development year, showing the fitted trend of model type (4.4)

(ii) *Accident trend*: Changing accident-year trends are possible by allowing the level for the accident year to change. The level of development year 0 in accident year *i* is α_i. After fitting calendar trends, the accident period trends are often fairly stable for long periods of time. Consequently, consecutive α- (or level-) *parameters* may be set to be equal.

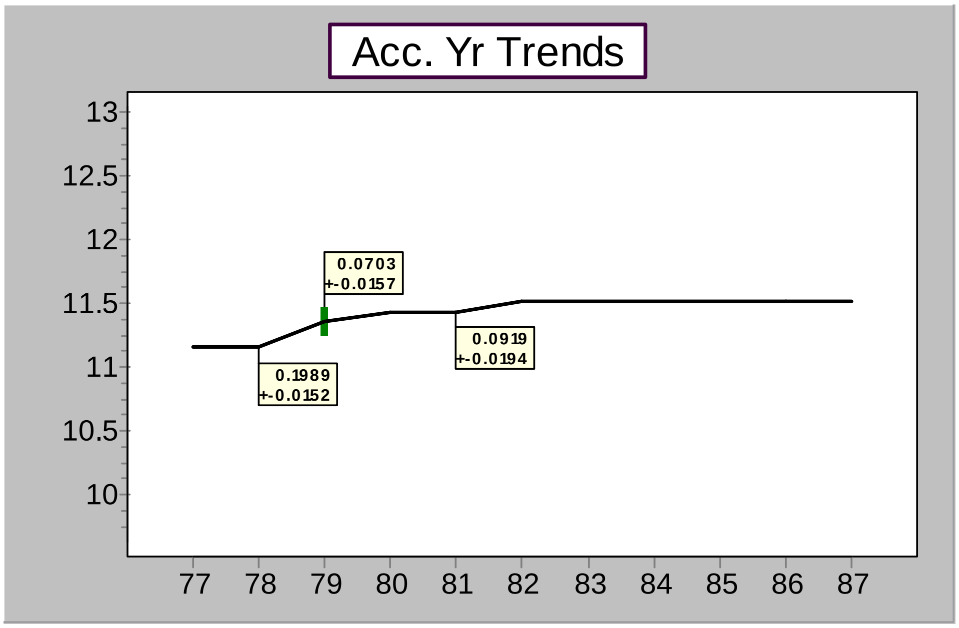


Figure 3: Accident year trends for the triangle ABC. Flat line segments show where consecutive parameters are equal.

Note that we can't show the fit to the data in this direction (because of effects in the other direction – development year effects are almost always present), unless we remove the effects of the other directions from the data (partial residuals).

Further, it sometimes makes sense to allow non-adjacent accident years to have the same level. For example, it is desirable when there has been a temporary shift in the level of paid losses that then returns to the original level.

(iii) *Calendar trend*: In real triangles we frequently find periods of stable, constant, or near-constant inflation, with sometimes abrupt changes. This can be modeled in similar fashion to the development period trend:

$$\kappa_t = \iota_2 + \iota_3 + \ldots + \iota_t \ . \tag{4.5}$$

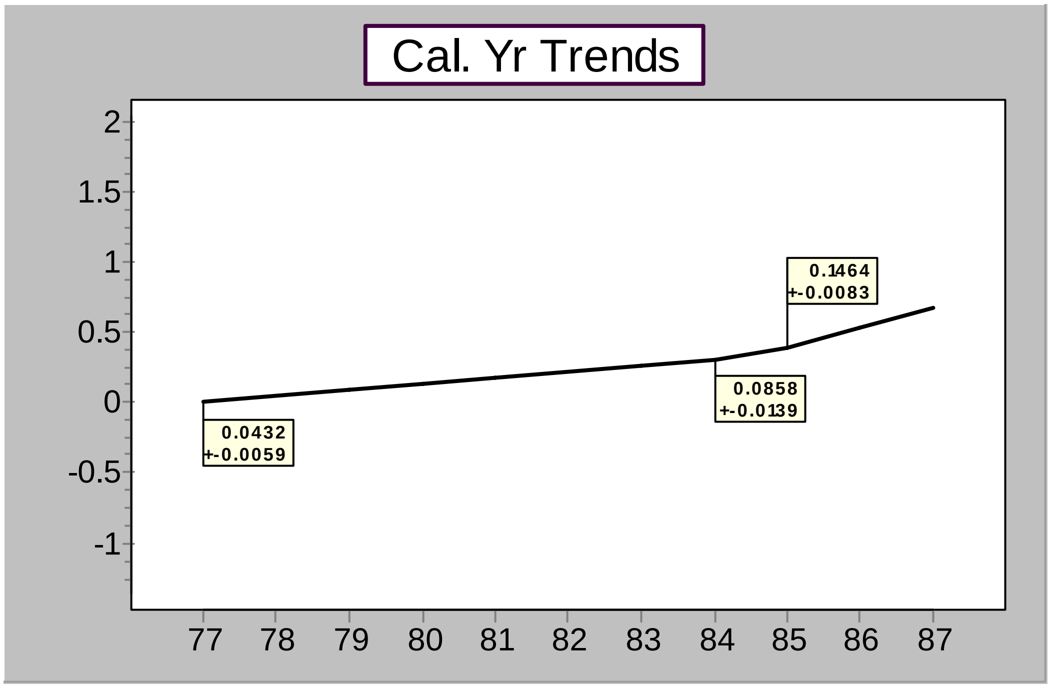


Figure 4: Calendar-year trends for the triangle ABC

Multicollinearity and the projection of trends to adjacent directions

Note that while it might sometimes be of interest to also have linear trends in the accident year direction (corresponding to periods constant growth in the accident year direction), it's not possible to simultaneously estimate trends in all years in all three directions at once, due to multicollinearity.

This is because of a fundamental property of triangles – trends in any of the directions project onto the adjacent direction(s). Consequently, it is necessary to restrict the potential to have a linear trend in one of the directions – in our case, the accident-year direction. A single linear trend in the accident year direction may be picked up by a trend in the calendar-year direction.

(iv) Random component:

The random component aims to describe everything of interest about the data not described by the mean. The distribution of the data about the mean (also called the distribution of the *error term*) is an essential part of the model. The errors are assumed to be independent, and by default the variance is taken to be constant on the log-scale (constant coefficient of variation).

Combining the components

This leaves the basic form of the Probabilistic Trend Family of models as:

$$y_{ij} = \alpha_i + \delta_j + \kappa_{i+j} + \varepsilon_{ij} \;; \quad \varepsilon_{ij} \sim (0, \sigma^2),$$

where each of the terms describing the mean has a particular structure in order to allow for the tendency for the trends or levels to be stable. This equation is of the form *data = accident trend + development trend + calendar trend + random*, as with equation 4.1. Each component is parameterized so as to be able to provide parsimonious descriptions of the trends typically seen in real data.

An assumed error distribution is not required in order to estimate parameters, but estimation is via least squares, which is optimal in the case of normality. In order to calculate explicit predictive distributions, we make the assumptions that the ε_{ij} are normal (that is, that the original p_{ij} are lognormal), but this assumption should be checked (since if it's implausible, the predictive distributions will not have the desired coverage probabilities). So we explicitly assume that ε_{ij} are normal, and so the original data are assumed to be lognormal.

We summarize this model in terms of four pictures, representing the four components. This gives an instant visual understanding of what is going on in the model. When combined with residual plots, the practitioner is able to rapidly assess both the components of the model for the mean and the characteristics of the data about the mean. For example, here is a reasonable model for a particular set of data (ABC, analyzed in detail in Barnett and Zehnwirth, 2000):

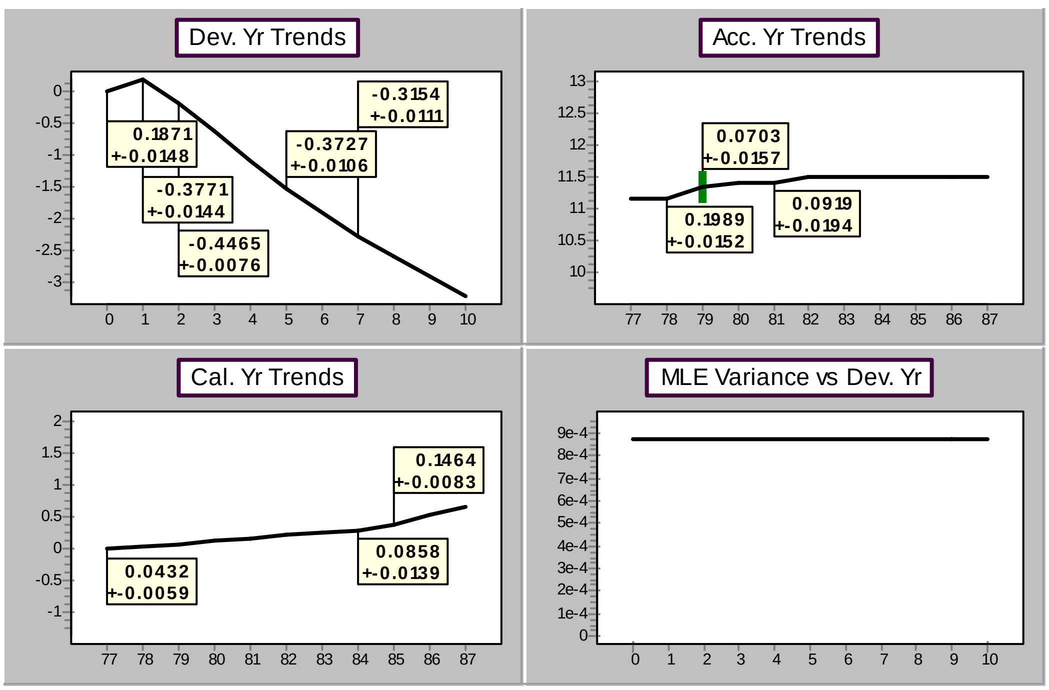


Figure 5a: Model trends in the three directions and fitted variance.

The residuals from this model are below.

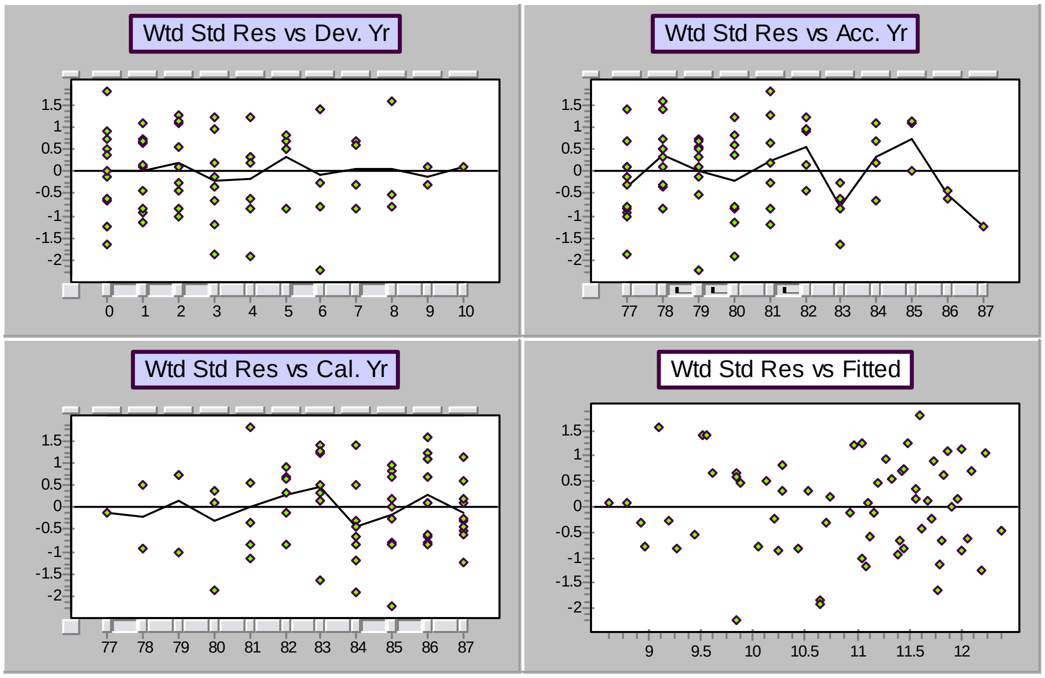


Figure 5b: Residuals vs. the three directions and against fitted values.

There is some pattern remaining in the accident-year direction that is better captured with a model that smoothes the accident-year trends rather than one that adds further discrete parameters; the modest remaining movements don't support separate parameters.

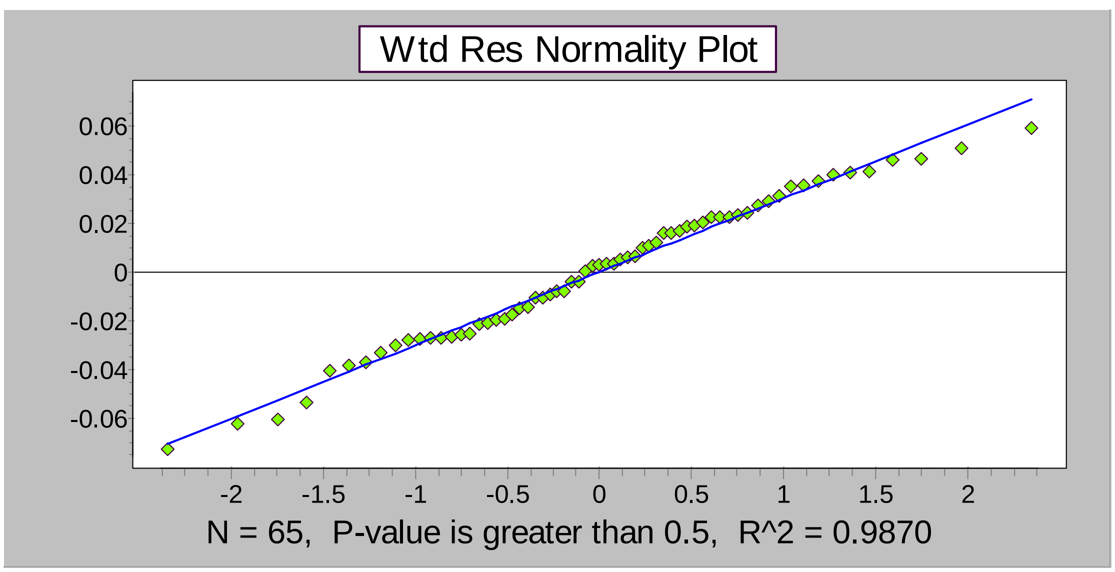


Figure 6: Assessment of the normality of the residuals. The lack of curvature indicates that the data are quite consistent with the assumption of normality.

4.2 Does this model family fit the data?

Being designed to describe features we regularly find in loss triangles, the model family includes many models that describe some triangles well – but each triangle is different, so a particular model should be chosen to describe the data at hand. Model critique via diagnostic displays and statistics should always be carefully assessed.

Indeed, the rich PTF model family includes in it a model whose linear predictor is identical in form to that of the quasi-Poisson GLM. Note that both the PTF models and the quasi-Poisson GLM are based on describing the mean on the log-scale, and that a PTF model with $\kappa_t = 0$ for all calendar years and with all α_i and γ_j different has an identical description of the mean (that is, as a two-way cross classification structure in logs) as a quasi-Poisson GLM. Further, as noted in section 2.2, the Mack-type chain ladder model also (necessarily) has parameters in the same places (though in that case, the accident-year parameters are implicit, being contained in the conditioning on the first development). Consequently, if there is a circumstance in which the PTF models fail to describe the mean, then a chain ladder-like model *is also guaranteed to fail*. On the other hand, it is frequently the case that a PTF model will succeed in describing the data where a ratio model fails.

Numerous examples are given, for example, in Barnett and Zehnwirth (2000), where triangles that are not well described by a chain ladder or other ratio model are well captured by a PTF model.

4.3 Extending the Basic Probabilistic Trend Family

Modeling changes in variance:

While the constant (log-scale) variance assumption is often tenable for many triangles, sometimes the variance is not constant. However, when it occurs, while the two are related, often the variance is seen to change more clearly with the development period than with the mean. When it is not reasonably constant, the variance is often relatively stable for a number of years, and so when necessary we allow the variance to change with development, while allowing adjacent variance parameters to be equal.

As mentioned before, a more complex form of smoothing (a generalization of exponential smoothing) is sometimes used (notably in the accident-year direction) in the PTF framework, but we will not pursue details of that since it requires more specialized algorithms than the simple regression methods required for the model described here to implement.

There are a variety of other ways in which this model can be extended. For example, the same structure on the mean could be used on count data in a GLM framework.

In the remainder of the paper we describe an extension of the PTF framework to the analysis of multiple triangles (such as different types of claim – asbestos vs. non-asbestos, different territories, several excess of loss layers, or different lines of business.

SECTION 5: MULTIVARIATE PTF MODELS

We relate PTF models by allowing for two different type of relationship between triangles:

(i) Payments may be correlated about their overall trends (if one is higher than average, the other may be higher than average).

(ii) The trends, or the differences in trends, may be related across triangles. For example, it is often the case with two related triangles that not only does superimposed inflation change in the same places in both, but the size of the change is similar in both triangles.

Where the individual models have the same *design* (the same pattern of observations with identical parameter structure), the combined model for both triangles taken together is a *generalized least squares* model (GLS). In this case, the parameter estimates are not affected by the correlations between triangles, and the calculations may be performed in two stages (estimation in individual triangles followed by estimation of the between triangle correlation).

While the correlations between triangles don't affect the estimates in GLS, they do affect the distribution of the forecast of the aggregate (or, indeed, the difference or even other functions if they occur).

More generally, the designs differ somewhat, and the combined model is a seemingly unrelated regression model (SUR). However, as noted in section 4.3, the models can actually get slightly more complicated than SUR.

As with the discussion of PTF models, we will concentrate on an SUR approach in this paper. Packages are available that can estimate SUR models, for example, the free statistical package *R* (http://cran.r-project.org/) has an extension package called *systemfit* which can fit SUR models.

SECTION 6: EXAMPLES

6.1 Relationships between two lines of business (LOB1/LOB3)

We consider a pair of triangles from two lines of business, which we label here as LOB1 and LOB3. They are full 10x10 triangles over the same period of time.

Here is a simple model fit for LOB1 without setting any non-consecutive parameters to be equal (though the pair of peaks against accident years suggests that it might be reasonable to pool information in that way).

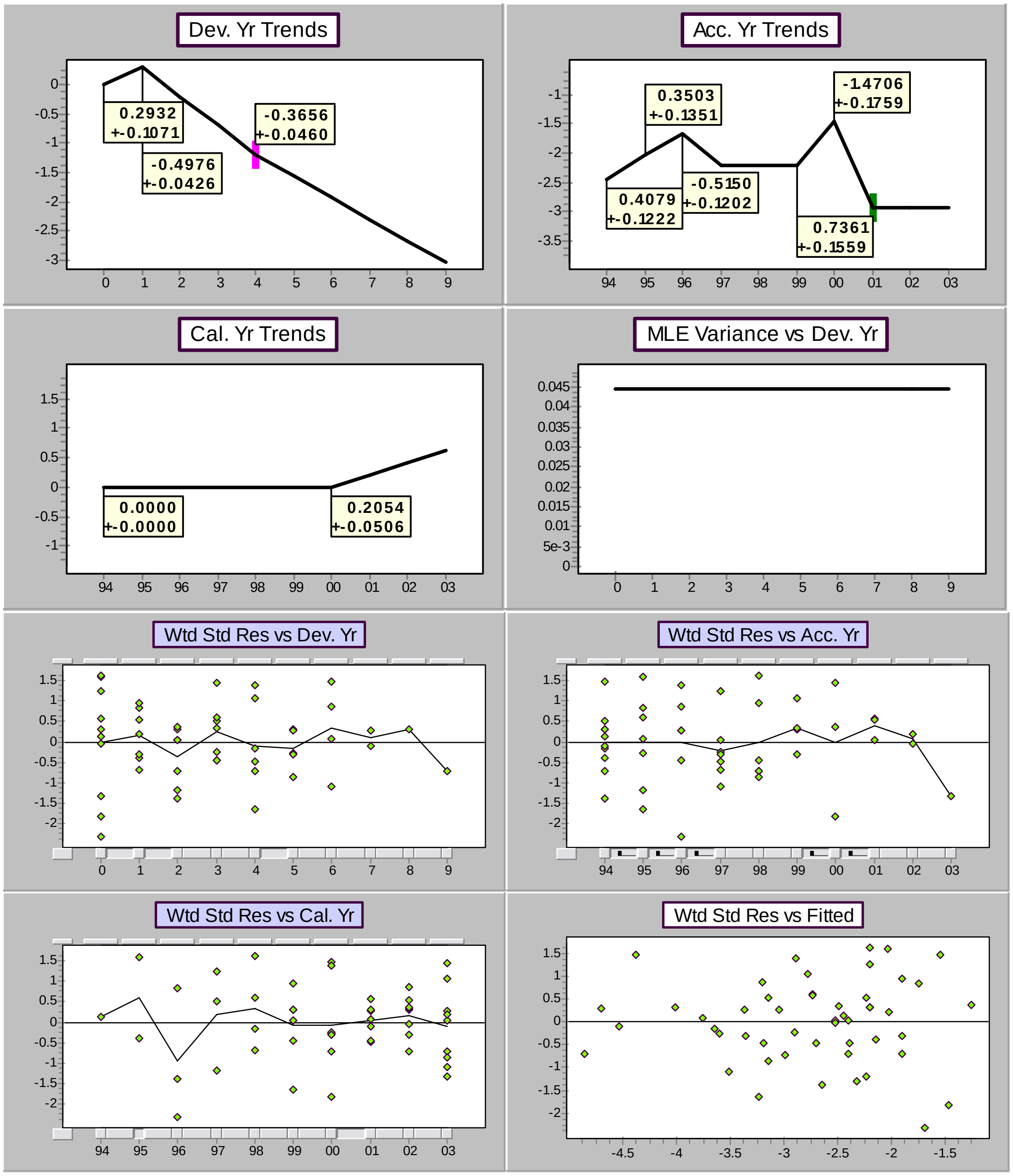


Figure 7: Model display and residuals for LOB1

We see this model provides a reasonable description of the data; it suggests that beginning in 2000, there was a dramatic (0.2054) increase in calendar-period trend (about 23% p.a.). While you should note that there was a substantial drop in the accident-year level in 2000, both the increase in calendar-year trend and the drop against accident year appear to be required for a reasonable description of the data.

It turns out that LOB3 has a major change in calendar- and accident-year trends that also occur in the same places as LOB1 (calendar-year 2000 and accident-years 2000 and 2001 respectively).

However, it is quite difficult to model LOB3 well, on its own. If we fit a reasonable model to the other directions, the accident year residuals appear as follows:

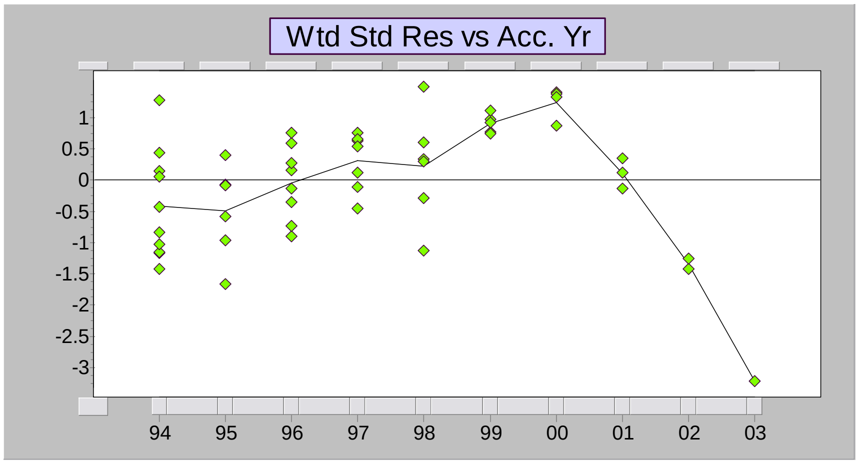


Figure 8: remaining trends in the accident year direction after removing trends in the other directions

The above residual display indicates a need to capture two linear trends against accident years. There are a variety of ways to model that with this data. One approach is to allow the accident-year levels to increase until the change, as with the following model:

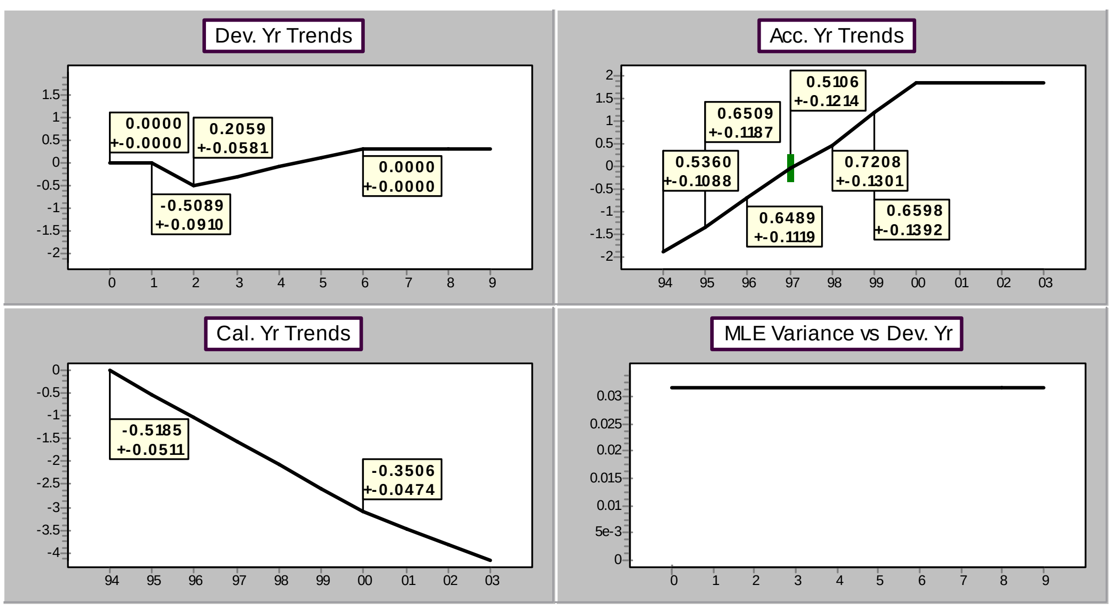


Figure 9a: An overparameterized model for LOB3

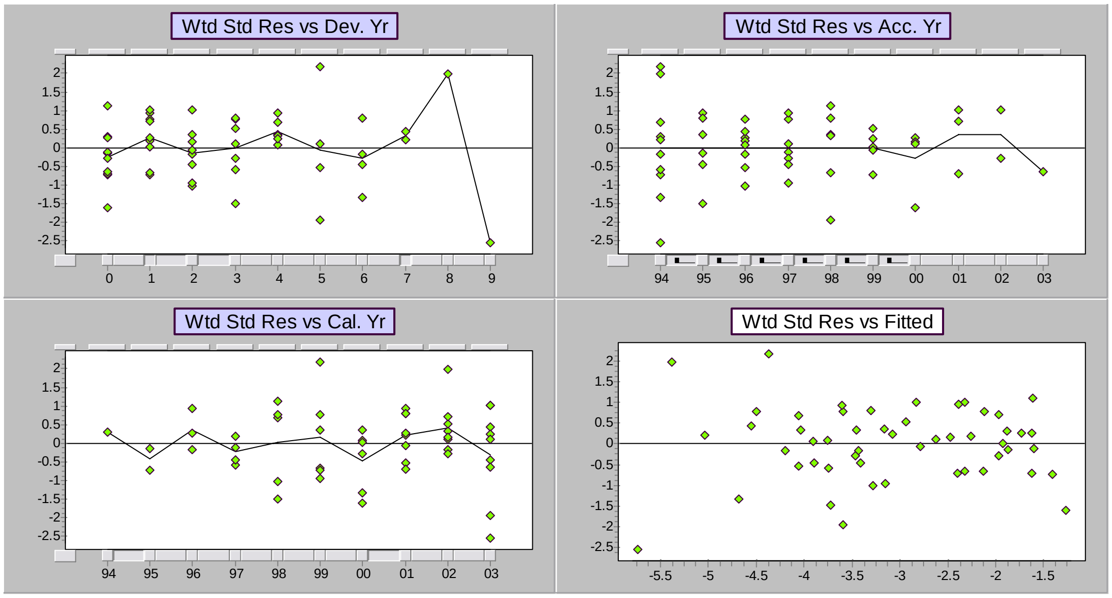


Figure 9b: Residuals from the above model.

This particular formulation of the model is somewhat overparameterized – though not nearly as much as the chain ladder, and it yields a substantially better description of the data. For example, see the residuals from a chain ladder fit in the calendar-year direction:

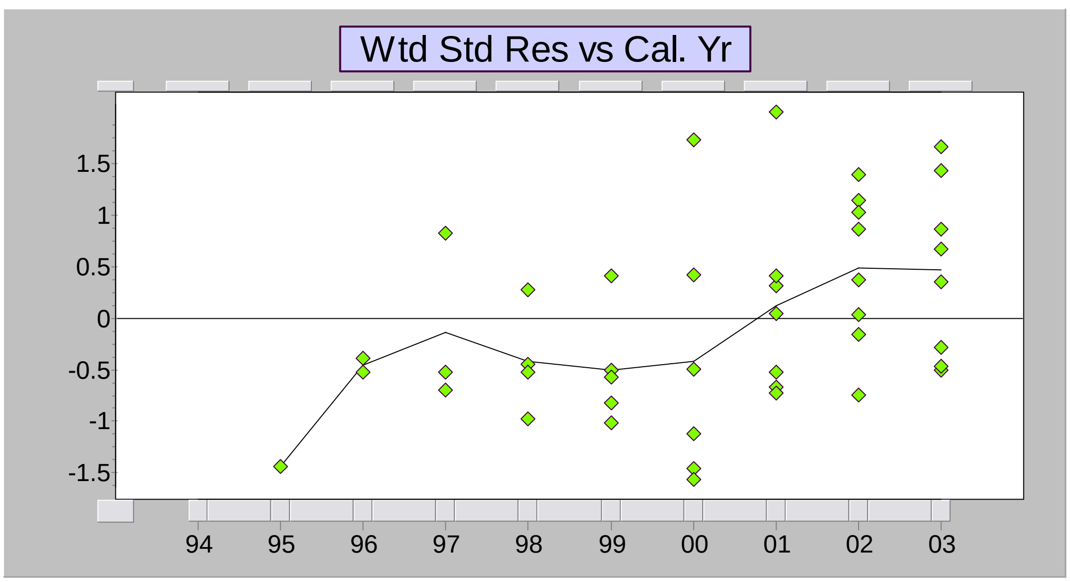


Figure 10: LOB3 Chain ladder residuals vs. calendar years

The graph shows distinct remaining calendar-year trends. Additionally, intercepts are necessary for several years (that is, ratios are inadequate). LOB1 shows similar uncaptured calendar-period trends when fitting ratios.

There is a certain amount of difficulty in obtaining simple descriptions of the combined effect of the accident- and calendar-period effects in LOB3. However, when the triangles are modeled together, it becomes more straightforward. Let us model both triangles together:

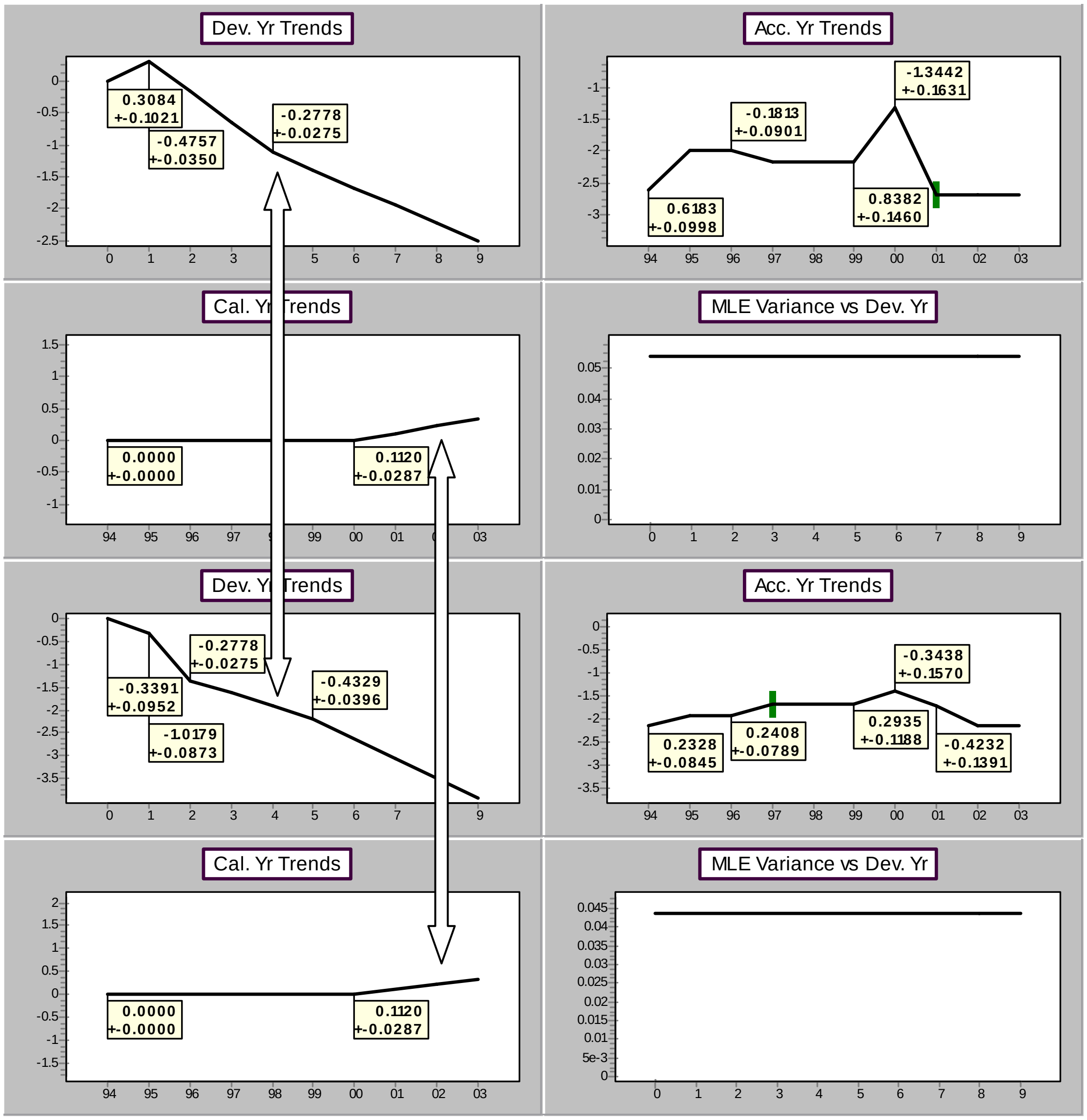


Figure 11: Composite Model for LOB1(top) and LOB3 (bottom). The arrows indicate equal trends.

In the model above, two development period trends (for overlapping periods of developments) and two concurrent calendar-period trends have been set to be equal. The accident-year level in the first and last accident years within LOB3 are also equal to each other.

The fit for each of the two lines of business has borrowed strength from the other..

This model has stabilized the estimates of trend very well, which has been particularly helpful for LOB3 – in fact (even when double-counting the common parameters), the number of parameters used to describe each array is smaller, yet the residuals are reasonable:

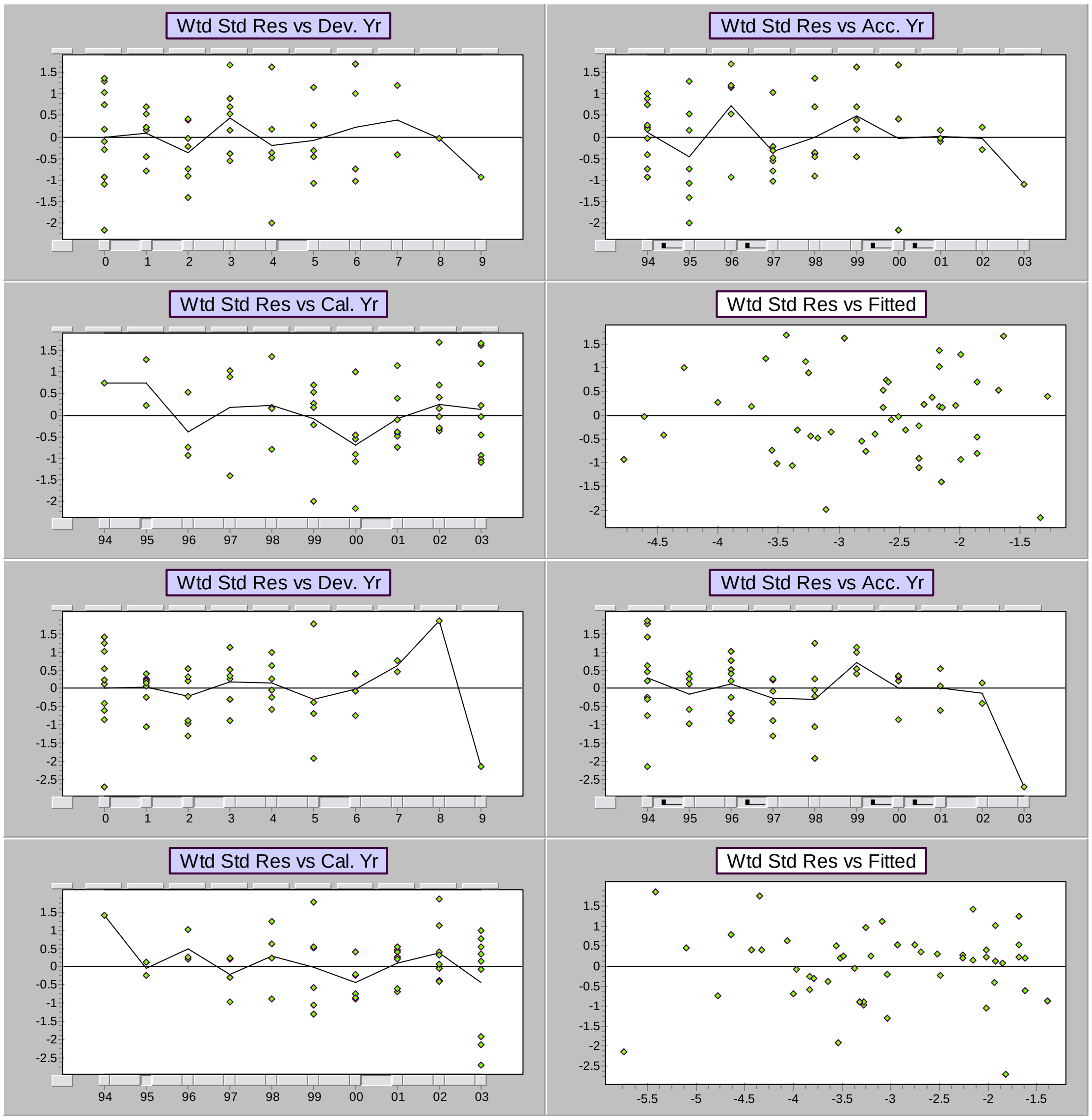


Figure 12: residuals for the composite model for LOB1 (top) and LOB3 (bottom)

Somewhat better results can be obtained if the smoothing of accident year trend estimates is used, but the mathematics is beyond the scope of this paper.

The estimated residual correlation between triangles is 0.39. This correlation can be seen in the moderate tendency for corresponding plots

Forecasts:

Note that the forecasts will be correlated in two ways:

(i) Parameter correlation: both the fact that two parameters will be shared (one development and one calendar period) between triangles and the fact the remaining parameters will be correlated across triangles contribute to the parameter correlation;

(ii) The model assumes that the data are correlated around the model. The estimate of the size of this correlation is roughly 0.4.

Let's examine a table of aggregate forecasts for the two lines and their sum.

Composite Model Forecast:		Accident-Year Summary:		1 Unit = $1,000
	Mean Reserve	Ultimate	Standard Dev.	CV of Reserve
LOB1	705,717	2,362,497	95,613	0.14
LOB3	1,182,458	4,560,860	165,913	0.14
Aggreg.	1,888,175	6,923,357	244,212	0.13

The CV of the aggregate is smaller than the CV of the individual lines, as we would anticipate.

By way of comparison, the figures for the individual models have slightly higher coefficients of variation; while the aggregate of the means for the models fitted individually is almost identical (around half a percent smaller), the standard deviations are substantially reduced because of the better use of information (reducing the contribution of parameter uncertainty to risk capital).

We can compute forecasts and standard deviations of aggregates, but in order to compute quantiles (such as for value at risk calculations or some risk capital calculations) it is necessary to simulate.

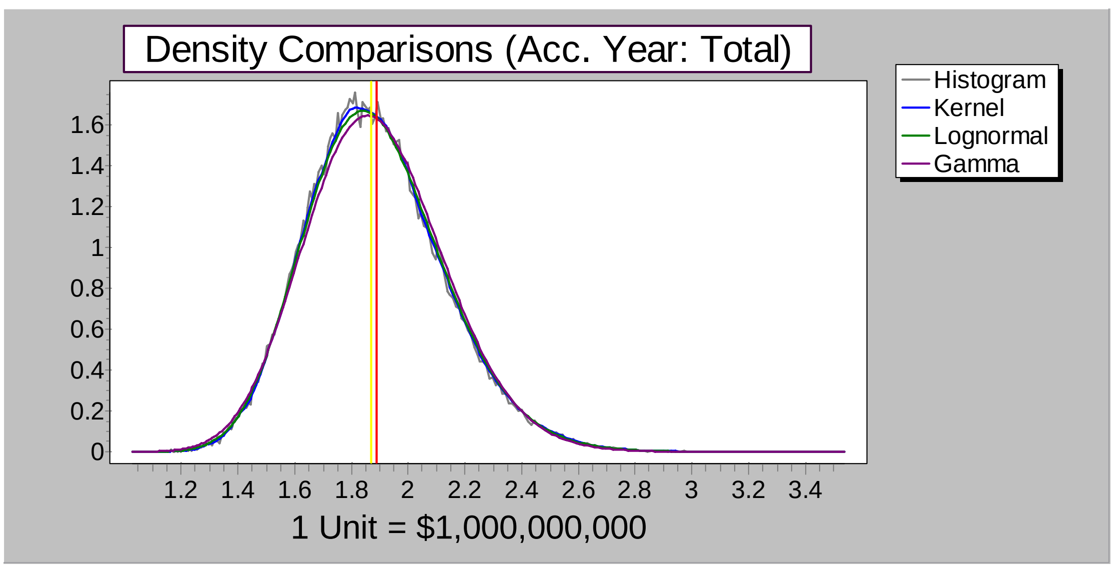


Figure 13: simulated distribution of aggregate of LOB1 and LOB3 based on 100,000 simulations. Rough grey line = simulated values, blue line is a kernel density estimate (smoothed simulated values), green curve = lognormal with the calculated mean and standard deviation, and purple = equivalent gamma density.

The PALD simulations enable the computation of quantiles, either directly from the simulated values or from one of the smooth curves.

LOB 1 LOB 3:Composite Reserve PALD Summary
Selected Quantile Statistics and Value at Risk (Acc. Year: Total) Unit = $1 Billion

%	Sample			Kernel			LogNormal			Gamma		
	Quantile	# S.D.s	VaR	Quantile	# S.D.s	VaR	Quantile	# S.D.s	VaR	Quantile	# S.D.s	VaR
90.0	2.205	1.298	0.317	2.208	1.308	0.319	2.209	1.312	0.321	2.207	1.306	0.319
95.0	2.313	1.738	**0.424**	2.316	1.751	0.428	2.315	1.746	0.426	2.307	1.715	0.419
98.0	2.443	2.271	0.555	2.446	2.284	0.558	2.440	2.259	0.552	2.423	2.190	0.535
99.0	2.532	2.638	0.644	2.536	2.653	0.648	2.527	2.616	0.639	2.502	2.515	0.614
99.5	2.616	2.981	0.728	2.619	2.993	0.731	2.609	2.953	0.721	2.576	2.818	0.688
99.9	2.796	3.717	0.908	2.805	3.755	0.917	2.788	3.685	0.900	2.734	3.462	0.845

Let's compare those results with corresponding results where the process correlations are set to zero:

LOB 1 LOB 3:Composite Reserve PALD Summary
Selected Quantile Statistics and Value at Risk (Acc. Year: Total) Unit = $1 Billion

%	Sample			Kernel			LogNormal			Gamma		
	Quantile	# S.D.s	VaR	Quantile	# S.D.s	VaR	Quantile	# S.D.s	VaR	Quantile	# S.D.s	VaR
90.0	2.246	1.311	0.292	2.248	1.322	0.294	2.246	1.310	0.291	2.244	1.303	0.290
95.0	2.342	1.744	**0.388**	2.345	1.756	0.391	2.340	1.735	0.386	2.334	1.707	0.380
98.0	2.457	2.262	0.503	2.460	2.274	0.506	2.451	2.235	0.497	2.438	2.174	0.483
99.0	2.537	2.618	0.582	2.541	2.639	0.587	2.528	2.581	0.574	2.509	2.492	0.554
99.5	2.611	2.951	0.656	2.617	2.978	0.662	2.601	2.907	0.647	2.575	2.789	0.620
99.9	2.793	3.772	0.839	2.796	3.783	0.841	2.757	3.610	0.803	2.714	3.417	0.760

Compare, for example, the VaR at the 95th percentile for the fitted model with the estimated correlation is $424 million, whereas if the two triangles are assumed to have zero process correlation, the VaR is $388 million. If process correlation is not taken into account, the VaR is $36 million (8.5%) too small. Similar effects are seen with TailVar, though the effect (in percentage terms) is a little larger at a given level (due to looking further into the tail). *Economic capital requirements are underestimated if the correlations between the predictions are not incorporated.*

6.2 Excess-of-loss type layers

In this example we consider the relationships between two layers net of reinsurance (losses limited to 1M and losses limited to 2M), and the layer corresponding to the difference between them, (1M excess of 1M). If the analysis was done at the scale of the original dollars, the three arrays would be linearly dependent (the first and third should add to the second), but on the log scale this is no longer the case.

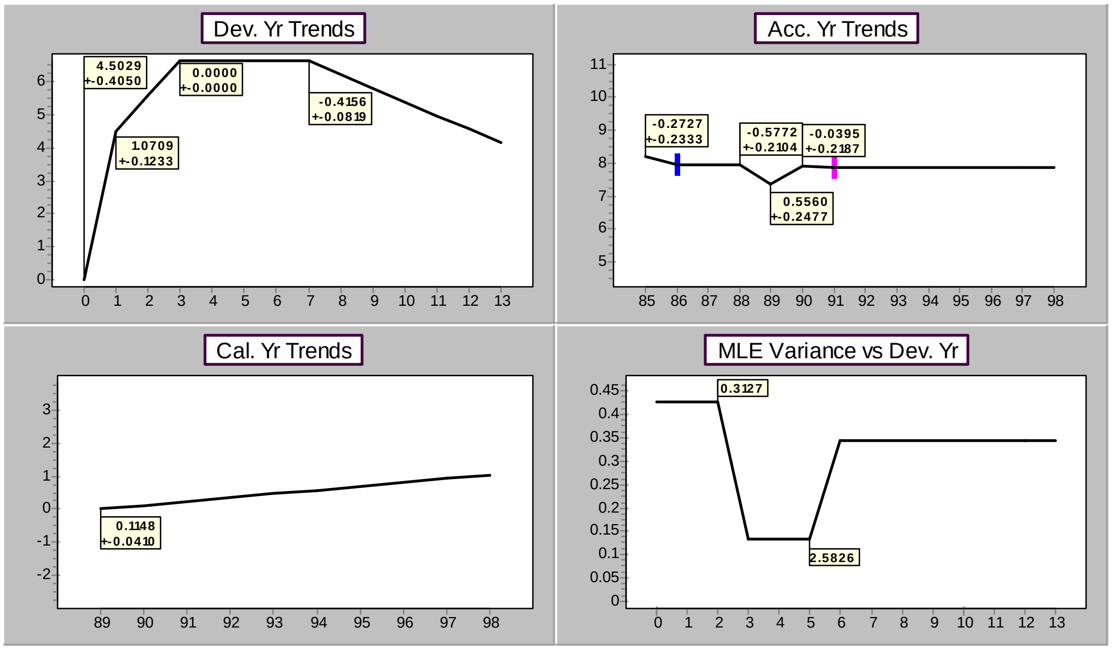


Figure 14a: model displays for layer 1M

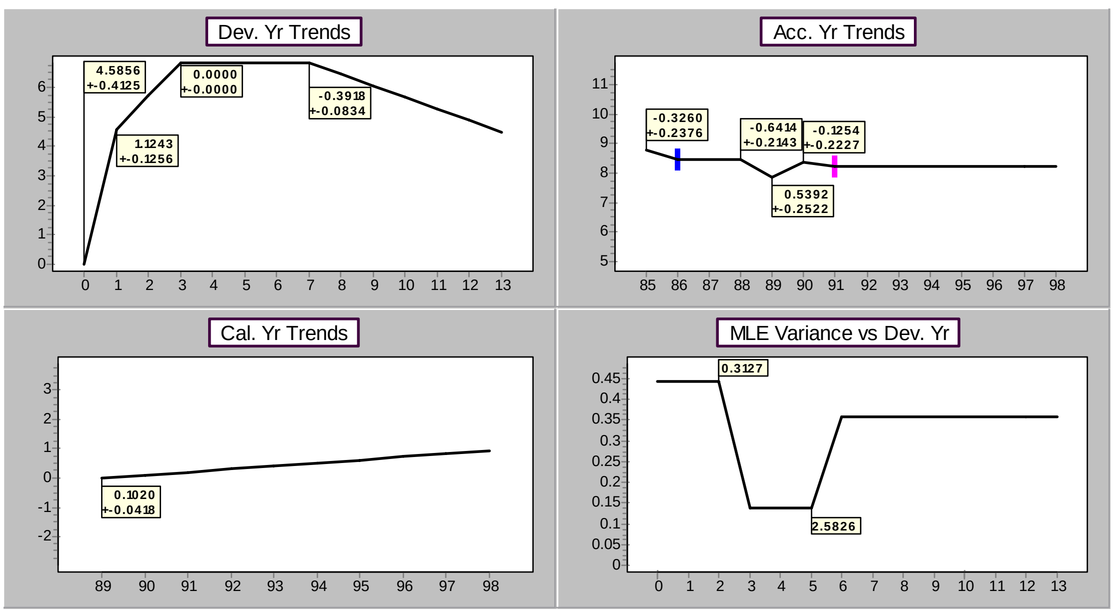


Figure 14b: model displays for layer 2M

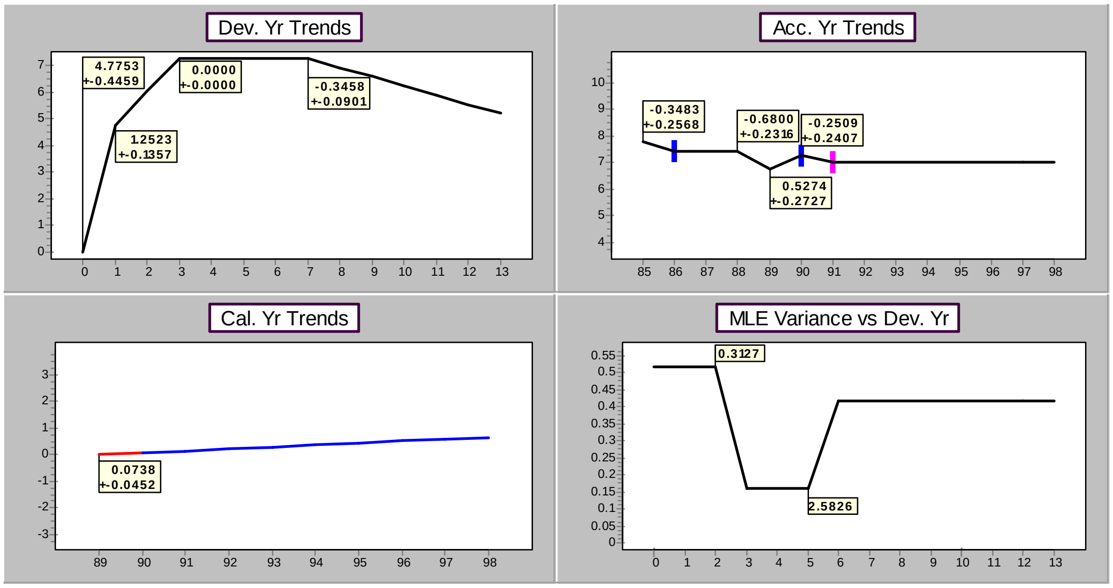


Figure 14c: model displays for layer 1M XS 1M

We see that the model chosen for the two ground-up layers and for the excess layer are quite similar, but the calendar-year trend for the excess layer (the third model plot) has a large standard error (there is high uncertainty about the underlying value) – indeed we can't be sure there's a non-zero trend there at all! On the other hand, if we're projecting that trend out into the future, we don't want to set it to zero; its estimate corresponds to an annual inflation of about 7.5% per annum.

It would be useful if we could borrow trend information from the ground up layers to help estimate some of the trends, in which case the estimate of the calendar-period trend may be better estimated.

As it turns out, a suitable combined multivariate model of the same form does exactly this. The *difference* in accident-year level from 1988 to 1989 is very close for the layer to 1M and for the next layer (and further, for the layer to 1M, the levels either side of the down-and-up bump are very close, though this has much less impact). If we borrow strength across the first two layers by setting that change in level to be equal, the standard error on the calendar-period trend comes down substantially, and the estimate is now extremely close to zero (0.0026), or about a quarter of a percent. With this model, it makes little difference to the forecasts whether or not the estimate of superimposed inflation for that layer is set to zero.

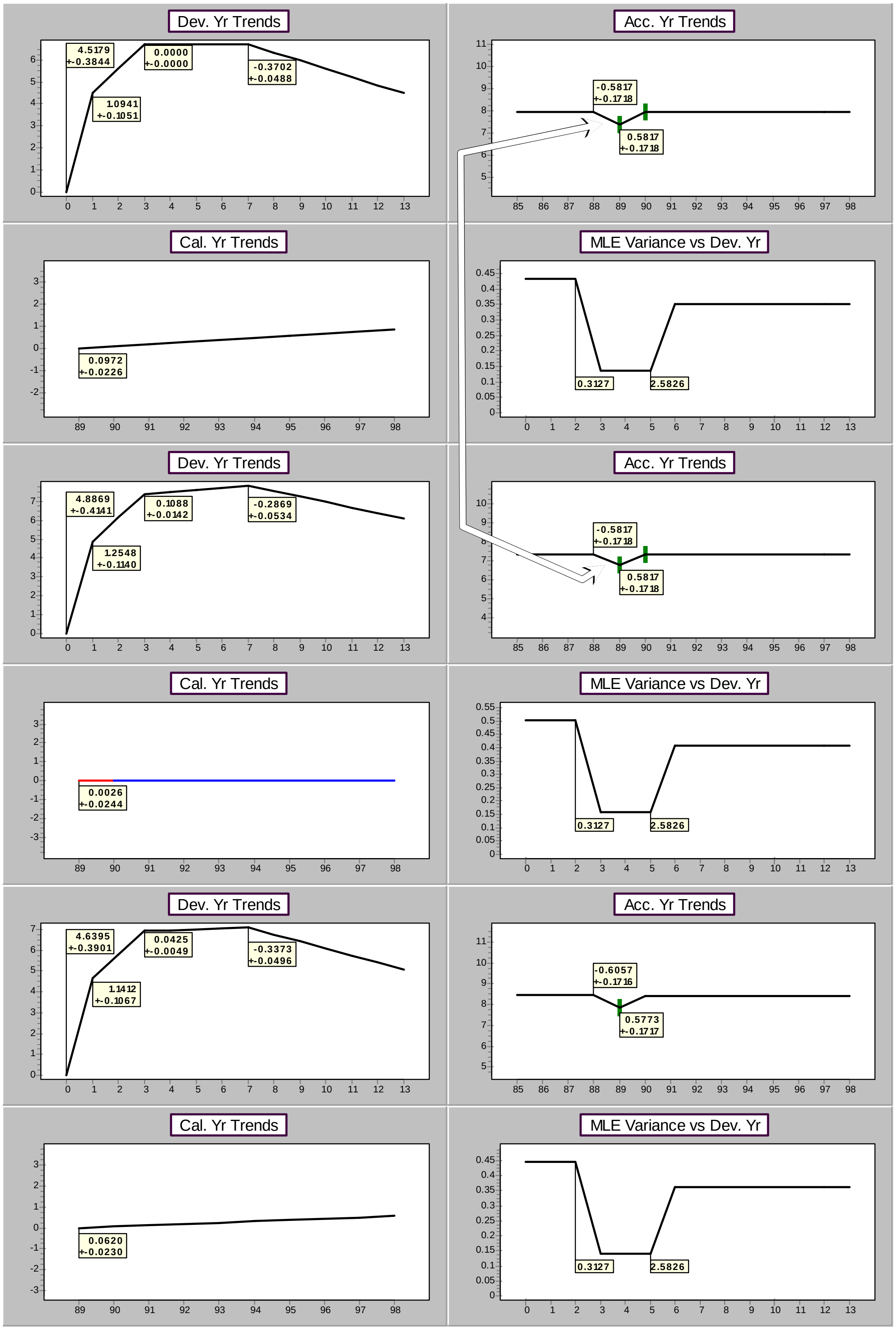
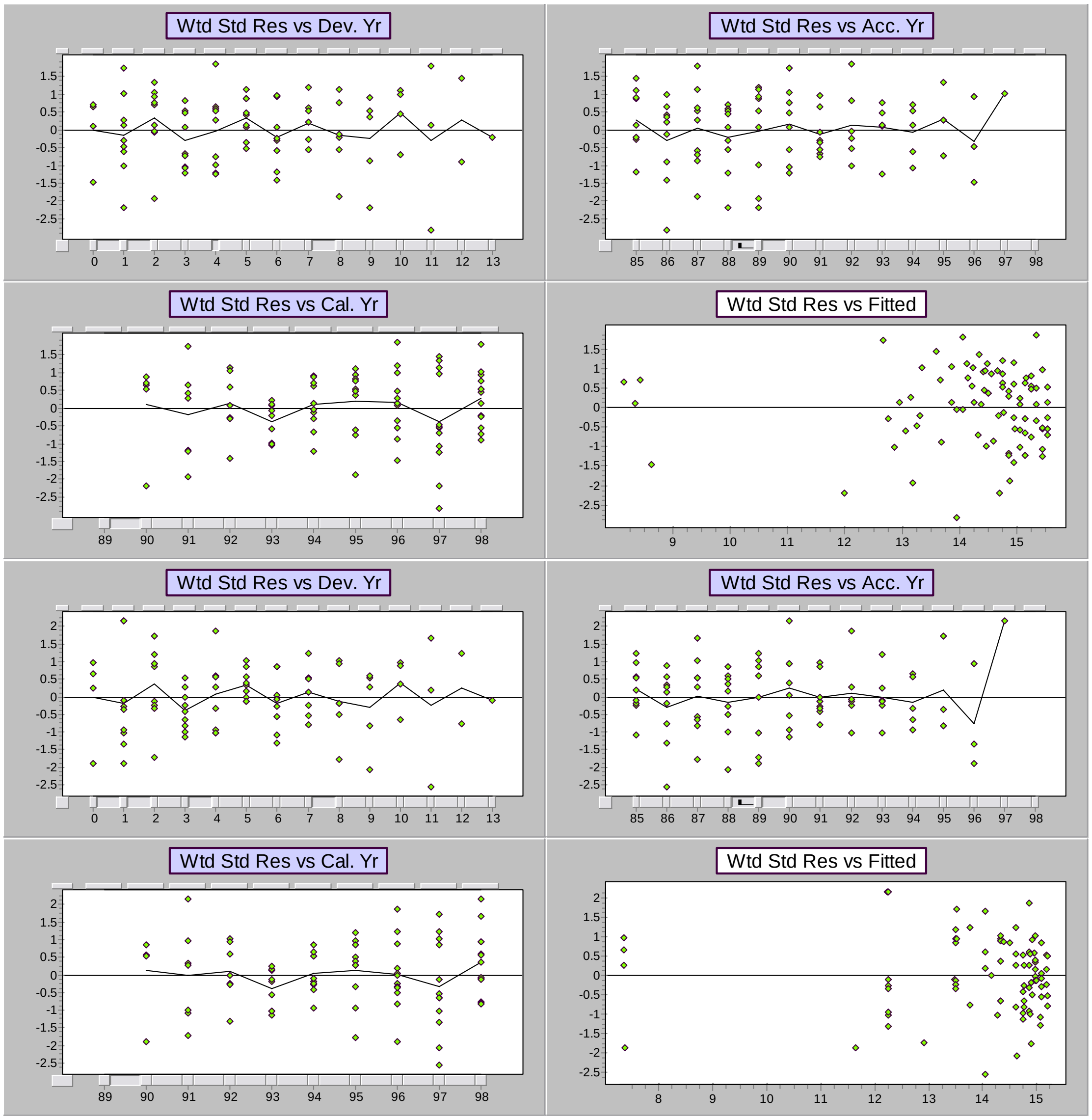


Figure 15: Model displays for combined model for layers 1M, 2M, and 1MXS1M. The arrow indicates a shift in accident-year level that is set equal across layers.

Here are the residuals from the above model



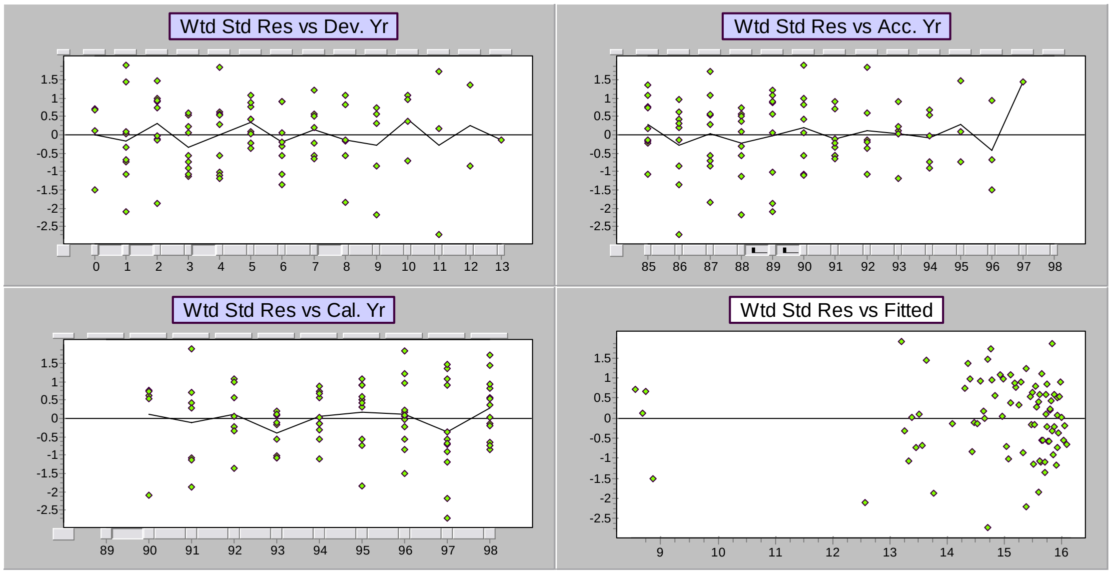


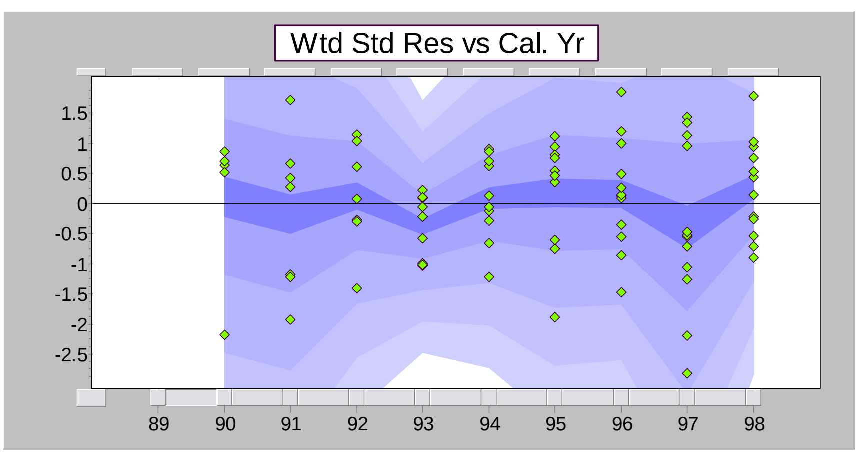
Figure 16: Residuals from the fitted composite model

The residuals indicate a reasonable fit for this model. The residuals do tend to "move together (perhaps not surprisingly!), and this is reflected in the correlations, which are very high.

Correlations between triangles:

	1M:PL(I)	1MXS1M	All 2M
1M:	1	0.960	0.995
1MXS1M:	0.960	1	0.982
2M:	0.995	0.982	1

These correlations can be seen by comparing corresponding residual displays across triangles. For example, in the residual displays below, the blue contours show multiples of standard deviations either side of the means by year.



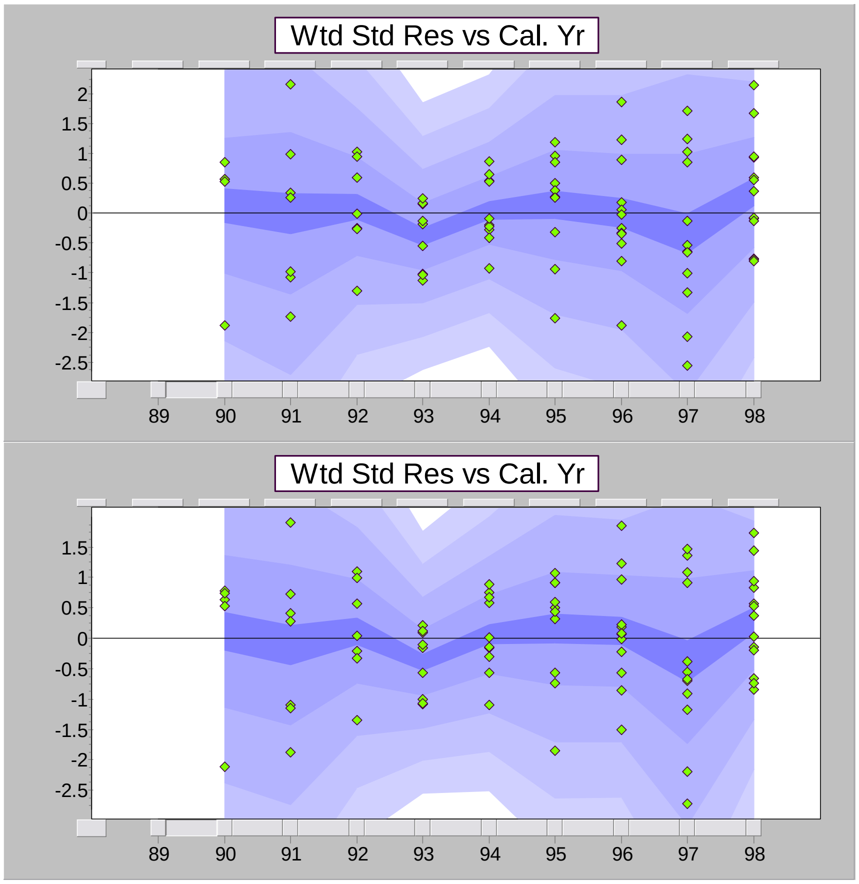


Figure 17: Residuals vs. calendar year for 1M, 1MXS1M, and 2M.

The "wiggles" in the blue lines are similar across each triangle, because of the high correlations between corresponding points. The tendency of the residuals to move up and down together is their correlation.

Here is a summary of the results from the forecasts with this model:

1M 1Mxs1M 2M:Composite Reserve Forecast Summaries:
Accident Year Summary 1 Unit = $1M

				CV	
	Mean Reserve	Ultimate	Standard Dev.	Reserve	Ultimate
1M	454	692	92	0.20	0.13
1MXS1M	244	421	47	0.21	0.11
2M	677	1,113	138	0.20	0.12
1M+1MXS1M	678	1,114	138	0.20	0.12

Note that adding 1M and 1MXS1M gives essentially the same answer as 2M, as we would wish. This consistency is preserved across a variety of reasonable models for this data.

The CV for the 2M and for the smaller layer 1M is almost the same.

Moving from retaining losses below 2M to losses below 1M (reinsuring 1M XS 1M) doesn't help the cedant's CV at all!

6.3 Gross vs. Net data

The following data is gross and net of Excess of loss-type reinsurance. Analysis of gross and net data together can help to assess the value of the reinsurance, and aid the design of reinsurance better suited to the direct insurer's needs.

Here is a display of a good fitted model for the gross data.

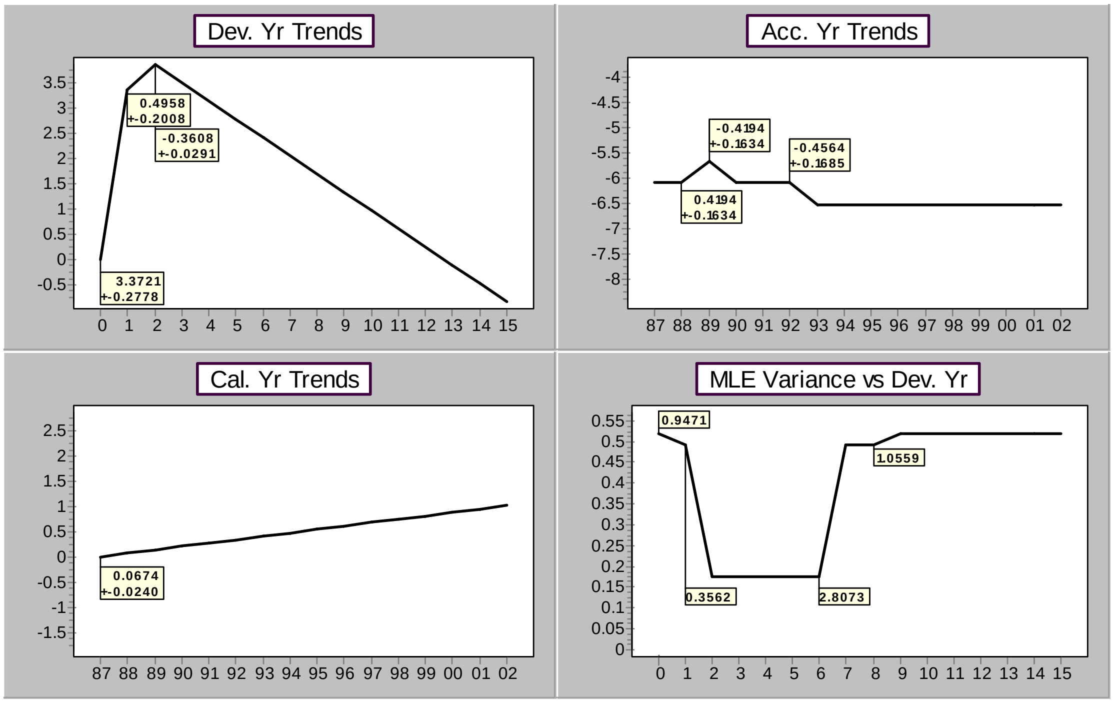


Figure 18: Model display for the Gross data.

There is smoothing of accident-period levels between 88 and 89 and between 92 and 93. The accident-year level for the 1988 (and earlier) mean is set equal to the 1990-1992 mean. Further note that the model has the level of variance of observations at developments 2-6 substantially lower than for the other years (heteroskedasticity). There's a stable trend of around 7% per annum ($e^{0.0674}$–1) in the calendar-year direction.

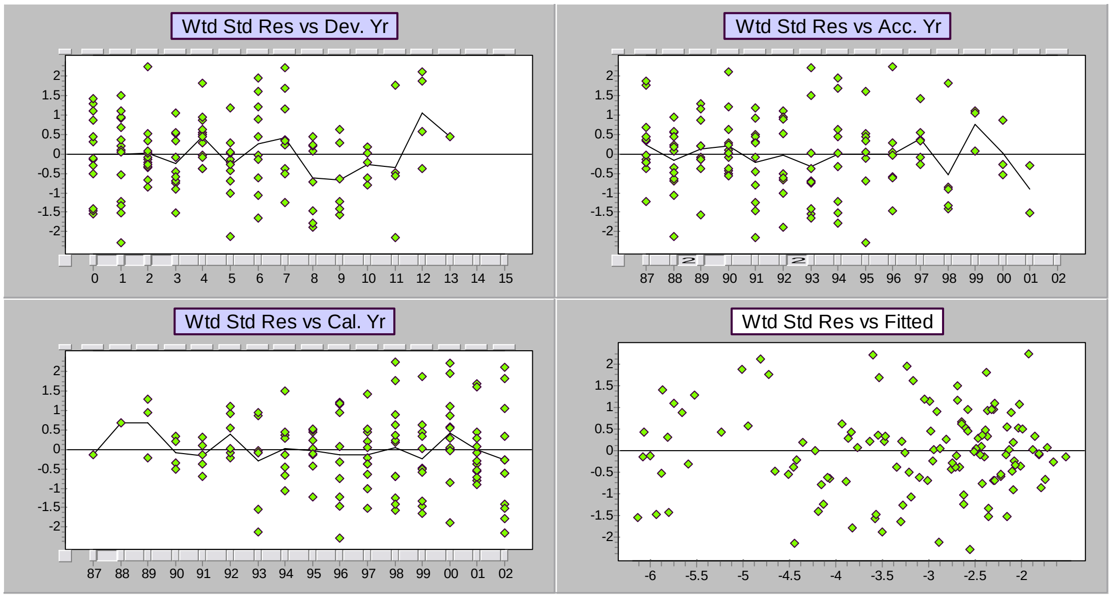


Figure 19: Residual plots for the model. The fit to the calendar- and accident-year directions is good; there is some lack of fit in the tail (where payments are very low).

Now let's model the Net data:

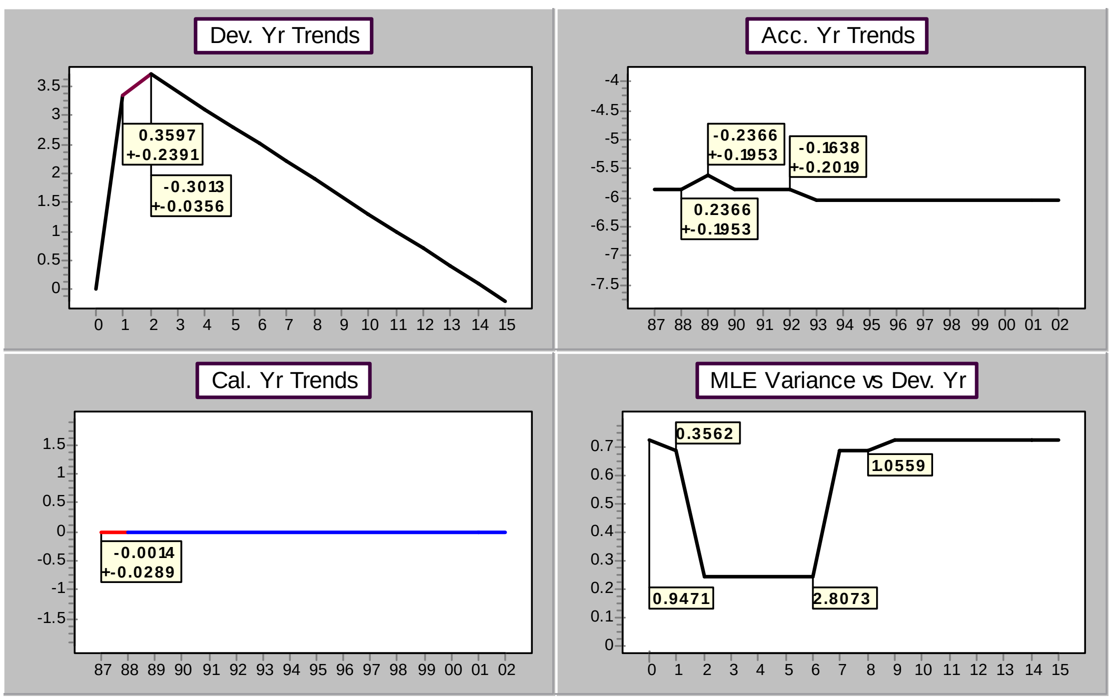


Figure 20a: Model display for the Net data

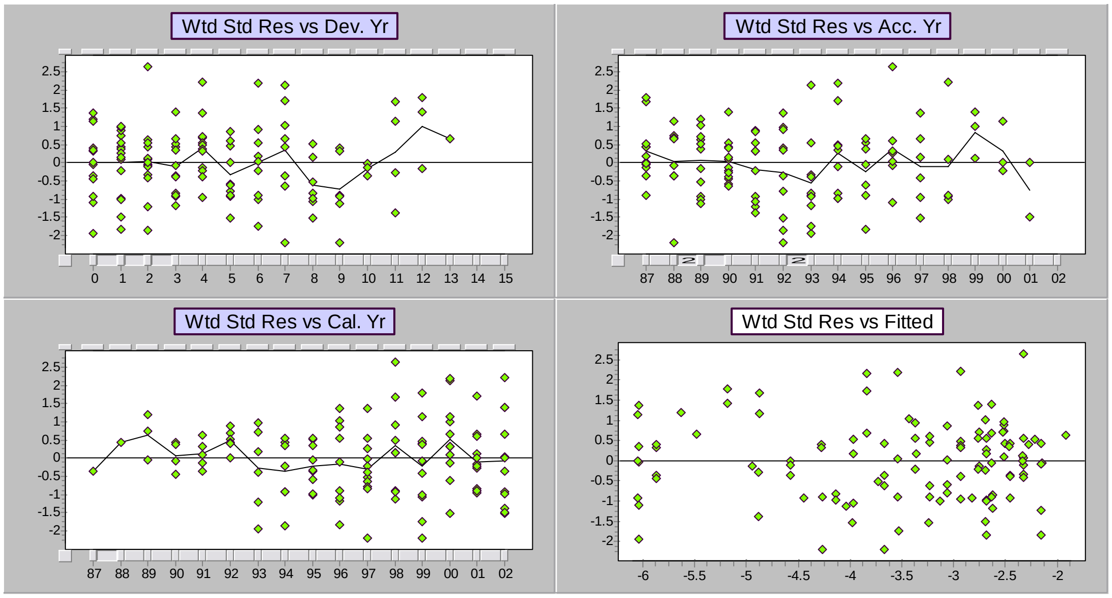


Figure 20b: Residuals for the Net data

Here, the model is quite similar to the gross, but as we see, the estimated calendar-year trend is essentially zero (0.14% with a standard error of about 3%). The fit is again, reasonably good.

The original insurer seems to be entirely ceding the growth (calendar-period inflation). While not particularly surprising, this is important information for both the cedant and the reinsurer.

In Figure 21, below, the graph shows the relationship between residuals for the two triangles. The correlation between the residuals of these models is quite high (this is common with reinsurance data), at around 0.81, and the normality assumption is reasonable.

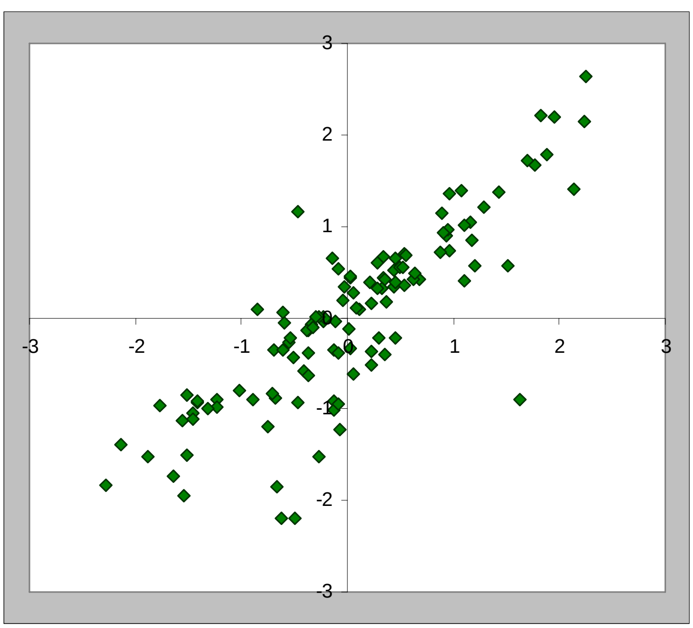


Figure 21*:* Plot of corresponding residuals from the two models, Net vs. Gross

Combined model:

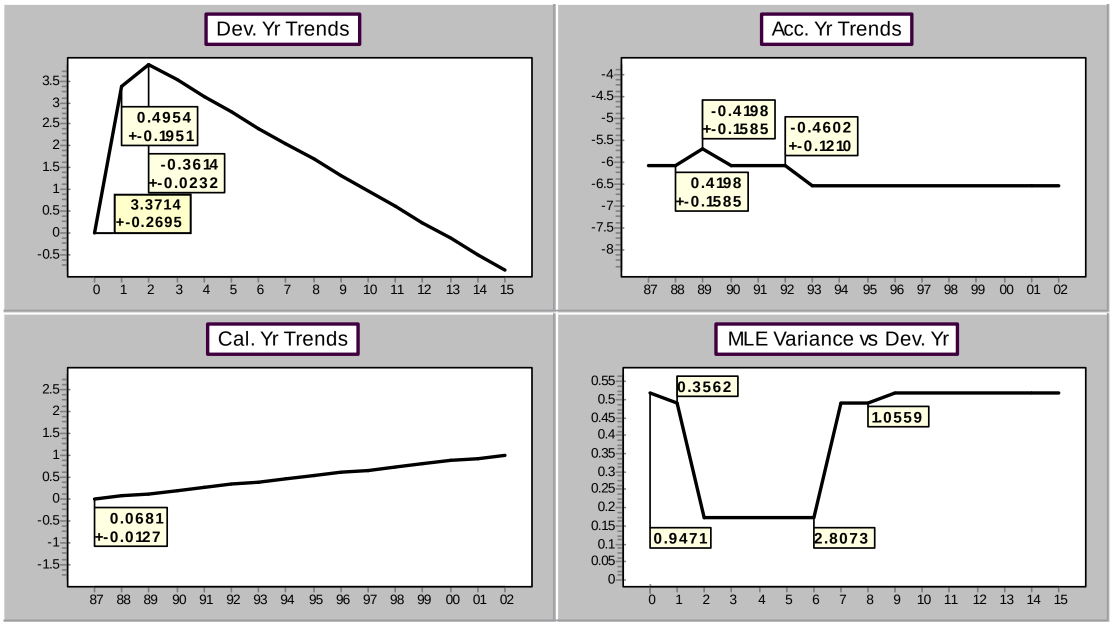


Figure 22: Model display for the combined model for the Gross data

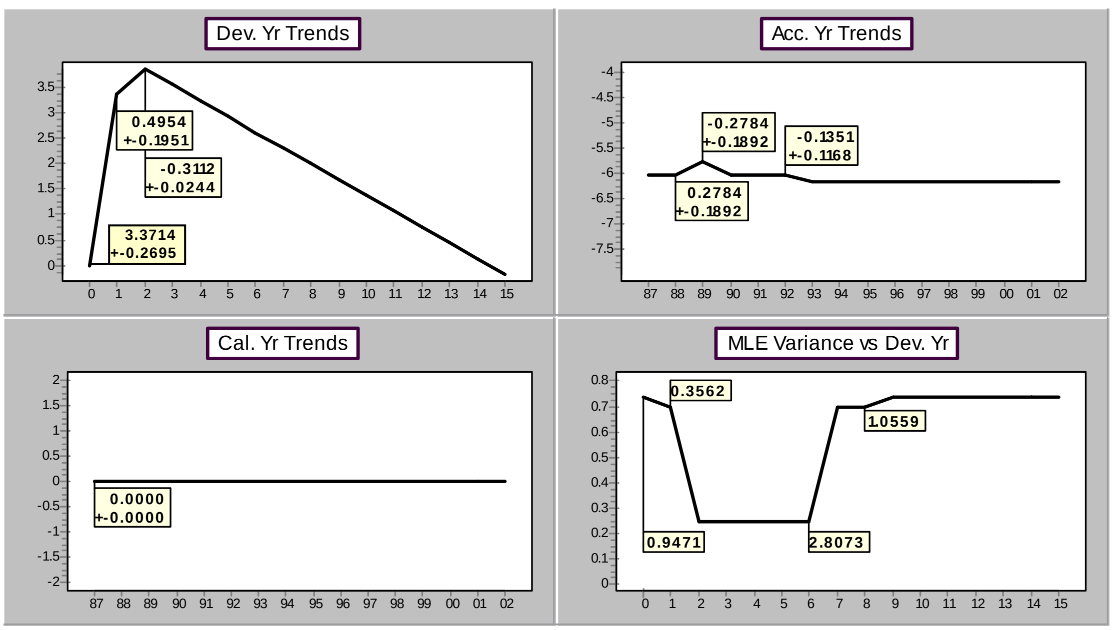


Figure 23: Model display of combined model for Net data

The first two development period trends had high uncertainties and have been set to be equal (they're effectively percentage changes in level across development periods).

Correlation in residuals is a little higher at around 83%, but otherwise the models are similar. The estimated trend for the calendar years for the Net data became even smaller (and less uncertain), and was set to zero (though it would make little difference if it was retained in this case, just as with the previous example).

Forecasting this combined model yields an interesting result:

Comp N & G: Composite: Accident Year Summary

	Mean Reserve	Ultimate	Standard Dev.	CV Reserve	CV Ultimate
:FAC ENG G:	101,705	216,641	15,450	0.15	0.07
:FAC ENG N:	52,821	134,323	8,725	0.17	0.06

The CV for the Net data (0.17) is slightly higher than for the Gross (0.15)! Even taking into account the fact that some of the trends will be less certain for the Net data (which can pull up the CV), this seems to suggest that the reinsurance is not achieving the goal of reducing the riskiness for the cedant.

DISCUSSION

These studies have some very important consequences for capital risk charges for both claims liability and underwriting. The forecast standard errors considered here include process variability, which is a major component of the risk, and which we cannot reduce below its inherent level; we can, however, reduce parameter uncertainty.

The fact that (percentage) changes in trend are often closely related across related triangles means that we can estimate critical components, such as calendar-year trends with less uncertainty – and so predictive distributions can become more concentrated. At the same time, we must consider the impact of the relationships between the triangles – relationships between parameters (including equal trends or trend changes) mean that the forecasts are correlated, and then there is the contribution of the correlation of the error term across triangles. These correlations contribute to the estimate of risk and so are critical to the estimation of economic capital.

Between some lines of business, after incorporating common trends, there's often little residual correlation, but as we saw in section 6.1, we can certainly get substantial correlations – neither close to independent nor close to very high dependence—and this can make a substantive difference to required capital. Consequently, using either the assumption of independence or the assumption that the dependence attains its upper bound could lead to either a radical under- or over- assessment of the required risk capital.

REFERENCES

[1] Ashe, F., "An essay at measuring the variance of estimates of outstanding claim payments," *ASTIN Bulletin* 1986, 16(S):S99-S113.

[2] Barnett, G. and B. Zehnwirth, "Best Estimates for Reserves," *Proceedings of the Casualty Actuarial Society* 2000, 87(167):245-321.

[3] Barnett, G., B. Zehnwirth, and E. Dubossarsky, "When Can Accident Years Be Regarded As Development Years?" *Proceedings of the Casualty Actuarial Society* 2005, 92(177):239-256.

[4] Barnett, G., B. Zehnwirth and D. Odell, "Meaningful Intervals," Paper submitted to the CAS 2008 Reserves Call Paper Program, Casualty Actuarial Society *E-Forum*, Fall 2008, http://www.casact.org/pubs/forum/08fforum/.

[5] England, P. D. and R. J., Verrall, "Predictive Distributions of Outstanding Liabilities In General Insurance," *Annals of Actuarial Science* 2006, 1(2):221-270.

[6] Hachemeister, C. A. and J. N. Stanard, "IBNR Claims Count Estimation with Static Lag Functions," Paper presented at the 1975 Spring Meeting of the Casualty Actuarial Society.

[7] Mack, T., "Distribution-free Calculation of the Standard Error of Chain Ladder Reserve Estimates," *ASTIN Bulletin* 1993, 23(2):213-225.

[8] Murphy, D., "Unbiased Loss Development Factors," *Proceedings of the Casualty Actuarial Society* 1994, 81(154):154–222.

[9] Renshaw, A. E. and R. J. Verrall, "A Stochastic Model Underlying the Chain-Ladder Technique," *British Actuarial Journal* 1998, 4(4):903-923.

Reserving with Incomplete Exposure Information

David R. Clark, FCAS, MAAA

Abstract. This paper outlines a reserving method that allows the actuary to use exposure information, such as onlevel premium, even if that information is only available for a limited number of years. The method is a simple blending of methods already in wide use, but can be shown to be based on a common underlying statistical model. The paper provides an overview of the Over-Dispersed Poisson model, and how it relates to Multiplicative LDF, Cape Cod, and Bornhuetter-Ferguson methods.

Motivation. The reserving actuary may have reliable exposure information (e.g., onlevel premium) for only a few recent years of data, rather than for the full historical period for which reserves need to be set.

Method. This incomplete exposure information can still be used, by implementing a hybrid reserving method equivalent to the Cape Cod method for the recent years and the Multiplicative LDF method for older years.

Results. We show how common reserving methods can be derived from a single statistical model, and then show how these methods are best combined when partial information is available.

Conclusions. This is a practical solution to the problem of stabilizing loss projections for recent accident years, incorporating available rate change information, and being responsive to actual loss emergence.

Keywords. Reserving, GLM, Chain ladder, Cape Cod, Bornhuetter-Ferguson.

1. INTRODUCTION

The purpose of this paper is to outline a method for estimating a stable reserve for immature years on long-tailed lines of business.

In order to bring more stability to these reserve estimates, it is helpful to bring in an exposure base that is proportional to expected loss by year. Optimally, this exposure base would be something like payroll or sales, but more commonly only historical premium is available. Historical premium is not directly applicable because of significant changes in rate adequacy over time – a phenomenon called the "insurance cycle." Instead we need to adjust the historical premium to an "onlevel premium" basis that is truly proportional to the expected losses by year. Unfortunately, the rate level indices required to make this adjustment may only be available for a limited number of years.

We will propose that even this limited information can be used in the reserve review in a straightforward way, with a method that is a combination of the Multiplicative LDF (a.k.a. Chain ladder) and Cape Cod (a.k.a. Stanard-Bühlmann) methods to form a single unified method. This unified method can be shown to be a best use of the available data and to be consistent with the other methods because they are all relying on the same underlying statistical model.

1.1 Research Context

There have been several past papers surveying statistical models applied to insurance loss development (payment or reporting) patterns. Recent examples are the CAS Working Party on Reserve Variability (2005), and the classification paper by Schmidt (2006). This prior research has aided greatly in viewing the loss development phenomenon from a statistical viewpoint; and showing connections between various models.

1.2 Objective

We will not intend to break new ground from a theoretical standpoint. Instead, we will build on the theory already established and draw some important practical implications. Specifically, we will show how best to incorporate limited exposure information into a reserve review in a consistent manner. By grounding this method in sound theory, we can show how it is consistent with current models and how it is an improvement over some popular techniques such as the Bornhuetter-Ferguson method.

What is new in this paper is the demonstration that a single unified method, which combines a Multiplicative LDF for older years and Cape Cod for more recent years, is built upon a single statistical model. The result is that limited exposure information can be incorporated for the years in which it is available.

1.3 Outline

The remainder of the paper proceeds as follows.

Section 2.1 will provide a description of the reserving problem faced for long-tailed business. We will introduce a numerical example to illustrate this problem.

Section 2.2 will give some basic definitions to set the groundwork for addressing the problem.

Section 3.1 will describe the Over-Dispersed Poisson (ODP) model as the basic structure underlying all of the methods to be discussed.

Section 3.2 will look at three methods in common use; and how they relate to the ODP model.

Section 3.3 will look at a unified method that combines the Multiplicative LDF and Cape Cod methods to incorporate limited exposure information.

Section 4 gives further discussion of practical issues of the Unified method, including issues in creating an appropriate exposure index.

2. PRELIMINARIES: THE RESERVING PROBLEM

We now proceed to give a more detailed description of the reserving problem to be addressed.

2.1 A Realistic Example

You are a reserving actuary reviewing the medical malpractice line of business. You will be working with an eight-year development triangle of cumulative paid loss data as shown below.[1]

Cumulative Paid Loss Triangle

AY	12	24	36	48	60	72	84	96
1999	257	1,143	2,402	3,478	4,456	5,080	5,284	5,481
2000	266	1,167	2,604	3,897	4,522	5,299	5,464	
2001	347	1,400	2,839	3,984	5,131	5,427		
2002	279	1,186	2,450	3,858	4,417			
2003	245	992	2,508	3,047				
2004	220	1,269	1,714					
2005	214	829						
2006	215							

This data shows a development pattern in which relatively little loss is paid in the first year. As a benchmark, you calculate standard chain ladder development factors, which confirm that only about 5.4% (=1/18.520 as shown below) of the loss would be paid as of the first twelve months—

assuming that there is no tail beyond the eighth year. Based on this, accident year 2006 seems too immature to expect the loss development method to yield a reliable result.

Development Factors (age-to-age link ratios)

AY	12-24	24-36	36-48	48-60	60-72	72-84	84-96	96-Ult
1999	4.447	2.101	1.448	1.281	1.140	1.040	1.037	
2000	4.387	2.231	1.497	1.160	1.172	1.031		
2001	4.035	2.028	1.403	1.288	1.058			
2002	4.251	2.066	1.575	1.145				
2003	4.049	2.528	1.215					
2004	5.768	1.351						
2005	3.874							
Wtd Avg	4.369	2.028	1.427	1.217	1.120	1.036	1.037	
LDF	18.520	4.239	2.090	1.465	1.203	1.074	1.037	1.000

In the past, reserves for immature years were often set using the Bornhuetter-Ferguson method, with a plan loss ratio used as the a priori expected value. However, in researching old reserve reviews, you have found that the plan loss ratio has consistently been set at about a 60% ELR, plus or minus a few points. By contrast, the actual experience has displayed a long-term cyclical pattern with a much wider range of loss ratios.

AY	Earned Premium	Latest Diagonal	LDF	Ultimate Loss	Loss Ratio
1999	5,400	5,481	1.000	5,481	101.5%
2000	5,900	5,464	1.037	5,668	96.1%
2001	6,500	5,427	1.074	5,829	89.7%
2002	8,500	4,417	1.203	5,315	62.5%
2003	10,200	3,047	1.465	4,464	43.8%
2004	11,000	1,714	2.090	3,582	32.6%
2005	11,300	829	4.239	3,514	31.1%
2006	11,500	215	18.520	3,982	34.6%
Total	70,300	26,594		37,835	53.8%

The 60% ELR might have been right for some periods (as Lewis Carroll observed: even a stopped clock is right twice a day…), but in general it has not proved to be an accurate number.

[1] This triangle is based on a section of industrywide medical malpractice data, but has been modified. The example is intended to be realistic, if somewhat better behaved than most accounts, but should not be used for any purpose other

Instead we have some evidence that the improving loss ratios from 1999 to 2006 were due in large part to significant rate increases. We know that this information should be used in the analysis, but unfortunately we only have a reliable monitor for rate changes starting in 2002.

What do we do?

2.2 Laying the Groundwork for a Solution

Before giving a detailed explanation of the models available to us and a proposed solution to the example above, it is worth carefully defining some key concepts.[2]

Model = A mathematical or empirical representation of a specified phenomenon

Method = A systematic procedure for estimating the unpaid claims

The "Model" is a mathematical description of the form of the world that we are analyzing, though with simplifying assumptions, such as the assumption that all accident years have the same expected loss development pattern.

The "Method" is the step-by-step procedure, or algorithm, that a person will follow to get from the original data to a final numerical result. In our insurance example above, we applied the chain ladder method to calculate our ultimate loss ratios.

Some may ask: why bother defining a model at all? Why not just select a method that seems reasonable and leave it there? There are three reasons:

1) A model gives criteria for deciding which of several possible methods is the "best" one (e.g., criteria of unbiasedness and minimum variance).

2) A model forces us to make all of our assumptions explicit so that they can be tested (e.g., with residual plots and goodness-of-fit criteria).

3) A model provides the theory for creating ranges around our reserve estimate (either

---

than illustrating the ideas in this paper.

[2] These two definitions come from Actuarial Standard of Practice No. 43; see Shapland (2007) for a more rigorous definition of these terms.

standard deviation or percentile distributions).

The 2005 CAS Working Party on Reserve Variability gives a more complete explanation of these reasons for creating a model. For the present paper, the primary purpose of introducing the mathematical model will be to show the "family relationship" of the methods presented.

Two more concepts need to be introduced before we proceed with our model.

Over-Parameterization = when we have too few data points relative to the number of model parameters

Model Constraints = user-supplied information that sets parameters, or relationships between parameters, rather than having them estimated from the data

The concept of over-parameterization is sometimes referred to as over-fitting or responding to the noise in the data rather than the signal. This can be a significant problem in the loss reserving context where we are working with data summarized into the triangle format. Constraining the model parameters is one way of reducing the instability from over-parameterizing and will be key to understanding the differences in the methods that we discuss below.

3. A FAMILY OF RESERVING MODELS AND METHODS

We now turn to a model that provides a framework for all of the familiar reserving methods. It points to a useful solution to our particular problem.

3.1 The Over-Dispersed Poisson (ODP) Model

The model presented here is derived from the theory of generalized linear models (GLM). GLM theory is an expansion of the theory of linear regression that allows for a broader category of error distributions beyond the normal Gaussian distribution, and also allows for the linear relationship of independent variables to be transformed by a "link function" in predicting the dependent variable.[3]

The structure of our model will be a multiplicative combination of accident year (y) and development period (d) factors. The dependent variable that we are attempting to fit will be the incremental loss for a given accident year in a given development period, and will be denoted $c_{y,d}$. For our example, this will be referred to as incremental paid, but the theory could be equally applied to reported data.

$$E(c_{y,d}) = \mu_{y,d} = v_y \cdot ELR \cdot \beta_d \tag{3.1.1}$$

Within this formula, the parameter v_y is an exposure or volume measure by accident year that is proportional to ultimate loss. This can be thought of as onlevel premium, though Section 4 of this paper will give a more detailed discussion as to how to create the measure. The ELR is an expected loss ratio, which represents the ratio of expected ultimate loss to the exposure measure. Because the exposures v_y already vary by accident year in proportion to expected loss, we only need a single value for ELR. The last parameter β_d is the development period relativity and may be thought of as the percent paid during a given calendar year.

This type of multiplicative combination of independent parameters indicates a log-link within GLM. That is, we would need to take logarithms of each side of the equation in order to transform the problem into a linear form.

Next, we will assume that the expected variance of an actual point from the expected value is in proportion to the expected value. The variance-to-mean ratio is represented as a dispersion parameter ϕ.

$$\mathrm{Var}(c_{y,d}) = \phi \cdot E(c_{y,d}) = \phi \cdot \mu_{y,d} \tag{3.1.2}$$

The GLM framework makes use of distributions within the exponential family for the error

[3] See Mildenhall (1999) for a good introduction to GLM in general, or Renshaw & Verrall (1998) for the GLM directly corresponding to this reserving application.

function. The assumption that the variance is proportional to the mean uniquely identifies the distribution as Poisson. The Poisson distribution is defined on the positive integers, $\{0,1,2,3,\cdots\}$, with variance equal to its mean, but this is generalized to the over-dispersed Poisson (ODP) model to be defined on multiples of the dispersion parameter, $\{0\phi,1\phi,2\phi,3\phi,\cdots\}$.[4]

With this model defined, the maximum likelihood estimates for the parameters can be found. We can actually do this by maximizing the quasi-log-likelihood (QLL) function,[5] a simplified version of the log-likelihood that does not depend on the dispersion parameter ϕ.

$$QLL \quad = \quad \sum_{y=1}^{n} \sum_{d=1}^{n+1-y} \left\{ \ln\left(\mu_{y,d}\right) \cdot c_{y,d} - \mu_{y,d} \right\} \quad = \quad \sum_{y=1}^{n} \sum_{d=1}^{n+1-y} \ln\left(v_y \cdot \mathrm{ELR} \cdot \beta_d\right) \cdot c_{y,d} - v_y \cdot \mathrm{ELR} \cdot \beta_d \tag{3.1.3}$$

We maximize the quasi-log-likelihood by solving for the parameters that set all of the derivatives equal to zero. For example:

$$\frac{\partial\, QLL}{\partial\, \beta_d} \quad = \quad 0 \qquad \forall\, d \tag{3.1.4}$$

Taking these derivatives guarantees that totals of the fitted losses in each column (development age) are equal to the actual losses. The model may therefore be described as unbiased.[6]

$$\sum_{y=1}^{n+1-d} c_{y,d} \quad = \quad \sum_{y=1}^{n+1-d} v_y \cdot \mathrm{ELR} \cdot \beta_d \qquad \forall\, d \tag{3.1.5}$$

---

[4] Venter (2007) prefers to call this the Poisson-constant-severity (PCS) model rather than ODP, because it can be interpreted as a collective risk model in which the number of claims follows a Poisson distribution, and every claim amount is the same value. However, there is no need to force this interpretation; we can simply view it as a discretized aggregate loss model for a given mean and variance.

[5] See Renshaw and Verrall (1998) for the full detail on this. They also note "We find it easiest to retain the assumption that the data have a Poisson distribution at the moment, although in all that follows in this section it is only the form of the likelihood which is important."

[6] The unbiasedness of row and column parameters as seen in "balancing" their totals may be familiar from the problem of classification ratemaking as described in Mildenhall (1999). More rigorously, we define unbiasedness as a characteristic of an estimator whose expected value is equal to the expected value of the random variable. That is, $E\left(\sum_{y=1}^{n+1-d} c_{y,d}\right) = E\left(\sum_{y=1}^{n+1-d} v_y \cdot \mathrm{E\hat{L}R} \cdot \hat{\beta}_d\right)$. This means that the total of the fitted values corresponding to the observed payments will be unbiased; this does not mean that the estimated reserve for the future periods will also be unbiased (cf. Taylor 2003).

Depending on the further constraints on v_y and ELR, certain row totals will also have fitted values that equal the actual values. Our choice of reserving method will depend on how we define these constraints.

3.2 Common Methods – Based on Constraining the ODP Model

Having defined the basic ODP model, we proceed to show how it is related to three familiar reserving methods.

3.2.1 The Multiplicative LDF Method

We begin with a fully unconstrained model, for which we assume that the vector of exposure measures is not available and must be estimated from the data in the development triangle. The exposure values v_y and ELR are therefore considered parameters to be estimated by the model. We start by defining:

$$\alpha_y = v_y \cdot \text{ELR} \tag{3.2.1}$$

Then we need to add a fitting criterion that the derivative of the QLL with respect to each α_y is set equal to zero.

$$\frac{\partial \, \text{QLL}}{\partial \, \alpha_y} = 0 \qquad \forall \, y \tag{3.2.2}$$

Taking these derivatives guarantees that the row totals of fitted and actual values are equal for every accident year.

$$\sum_{d=1}^{n+1-y} c_{y,d} = \sum_{d=1}^{n+1-y} \alpha_y \cdot \beta_d \qquad \forall \, y \tag{3.2.3}$$

An easy way of estimating the α and β parameters for this model is to use the chain-ladder method of loss development factors. The parameters α_y represent the ultimate loss by year; the parameters β_d are a function of the weighted average LDFs.

AY *y*	Ult. Loss (alpha)	LDF	% of Ult	Incremental % (beta)
1999	5,481	1.000	100.00%	3.59%
2000	5,668	1.037	96.41%	3.31%
2001	5,829	1.074	93.10%	10.00%
2002	5,315	1.203	83.10%	14.84%
2003	4,464	1.465	68.26%	20.41%
2004	3,582	2.090	47.85%	24.26%
2005	3,514	4.239	23.59%	18.19%
2006	3,982	18.520	5.40%	5.40%
			Total of Betas:	100.00%

A simple inspection of the actual and fitted incremental triangles will confirm that both the row and column totals are equal.

Actual Incremental Payments

AY	0-12	12-24	24-36	36-48	48-60	60-72	72-84	84-96
1999	257	886	1,259	1,076	978	624	204	197
2000	266	901	1,437	1,293	625	777	165	
2001	347	1,053	1,439	1,145	1,147	296		
2002	279	907	1,264	1,408	559			
2003	245	747	1,516	539				
2004	220	1,049	445					
2005	214	615						
2006	215							
Total:	2,043	6,158	7,360	5,461	3,309	1,697	369	197

Fitted Incremental Payments

AY	0-12	12-24	24-36	36-48	48-60	60-72	72-84	84-96
1999	296	997	1,330	1,119	814	548	181	197
2000	306	1,031	1,375	1,157	841	566	188	
2001	315	1,060	1,414	1,190	865	583		
2002	287	967	1,289	1,085	789			
2003	241	812	1,083	911				
2004	193	652	869					
2005	190	639						
2006	215							
Total:	2,043	6,158	7,360	5,461	3,309	1,697	369	197

An important observation from this exercise is that we have set the tail factor at age 96 months equal to 1.000. That is, we are assuming that there is no further development beyond the eighth year. In fact, this is merely done by convention – we can include any tail factor that we would like beyond the eighth year. We include a tail factor by dividing all of our β parameters by the selected 96-ultimate LDF, and then also multiplying all of the α parameters by the same amount. The cross-product will produce fitted values equal to the model above.

What this tells us is that the fully unconstrained model provides us with no information about development beyond the periods in the historical data.[7]

A second observation from this unconstrained model is that, while we usually think of it in multiplicative terms, it can equivalently be considered an additive model:

Ultimate Loss	=	(Paid Loss) × LDF
Ultimate Loss	=	(Paid Loss) + (Expected Ultimate)×(1-1/LDF)
	where	Expected Ultimate = (Paid Loss)×LDF

A final observation is that our example includes 36 actual data points, but those 36 data points are estimating 15 parameters (eight accident year factors plus seven development factors). This gives us few data points per parameter and, therefore, should be described as an over-parameterized model.

3.2.2 The Cape Cod Method

As noted above, the fully unconstrained model that produces the chain-ladder method has a problem with over-parameterization. We therefore move to a model that adds more constraints, by introducing an exposure measure that forces a relationship between the accident year ultimates.

$$\frac{\text{Expected Ultimate Loss in Year } i}{\text{Expected Ultimate Loss in Year } j} = \frac{\alpha_i}{\alpha_j} = \frac{v_i}{v_j} \qquad \forall\, i, j \tag{3.2.4}$$

Because the exposure or volume measures are supplied by the user, we only need to estimate the parameter ELR instead of the full vector of α_y. The Maximum Likelihood Estimator (MLE) for

---

[7] One way to fit a tail factor to the data is to constrain the model by assuming that all of the β s follow a known development pattern form. This is the model outlined in Clark (2003), but will not be addressed here.

the ELR is found by setting the derivative of the quasi-log-likelihood function (QLL) equal to zero.

$$\frac{\partial \text{ QLL}}{\partial \text{ ELR}} = \frac{\partial \sum_{y=1}^{n} \sum_{d=1}^{n+1-y} \left\{ \ln\left(v_y \cdot \text{ELR} \cdot \beta_d\right) \cdot c_{y,d} - v_y \cdot \text{ELR} \cdot \beta_d \right\}}{\partial \text{ ELR}} = 0 \tag{3.2.5}$$

This criterion results in a requirement that the sum of all the losses in the entire triangle is the same for fitted and actual values.

$$\sum_{y=1}^{n} \sum_{d=1}^{n+1-y} c_{y,d} = \sum_{y=1}^{n} \sum_{d=1}^{n+1-y} v_y \cdot \text{ELR} \cdot \beta_d \tag{3.2.6}$$

This does not add anything to our MLE criteria, since we had already required that column totals would be equal.

The method for estimating model parameters is:

1) Estimate an incremental loss ratio IncrLR_d for each development period:

$$\text{IncrLR}_d = \frac{\sum_{y=1}^{n+1-d} c_{y,d}}{\sum_{y=1}^{n+1-d} v_y} \qquad \forall\, d \tag{3.2.7}$$

2) Set the ELR as the sum of the incremental loss ratios:

$$\text{ELR} = \sum_{d=1}^{n} \text{IncrLR}_d \tag{3.2.8}$$

3) Set the development pattern parameters such that $\sum_{d=1}^{n} \beta_d = 1$:

$$\beta_d = \frac{\text{IncrLR}_d}{\text{ELR}} = \frac{\sum_{y=1}^{n+1-d} c_{y,d}}{\sum_{y=1}^{n+1-d} v_y \cdot \text{ELR}} \qquad \forall\, d \tag{3.2.9}$$

With this procedure, we accomplish the goal of having all of the column totals for the fitted triangle match those of the actual triangle; therefore the results are the maximum likelihood estimates.

We have been again assuming that there is no "tail" beyond the last age represented in the triangle. As with the Multiplicative LDF method, this is only by convention, and we can introduce any tail factor we wish by re-scaling the β and ELR parameters.

$$\beta_d \rightarrow \frac{\beta_d}{\mathrm{LDF}_n} \qquad \text{so that} \qquad \mathrm{LDF}_n = \frac{1}{\sum_{d=1}^{n} \beta_d} \tag{3.2.10}$$

The original ELR is then multiplied by the selected tail LDF_n to produce a final ELR.

$$\mathrm{ELR} = \left(\mathit{Original}\ \mathrm{ELR}\right) \cdot \mathrm{LDF}_n \tag{3.2.11}$$

The key concept to note is that the ELR and tail LDF_n are interdependent. If we change one of them, then the other will also need to change. This concept will be critical when we examine the Bornhuetter-Ferguson method.

In order to perform these calculations, we must first create an exposure index covering all of the accident years in the experience period. We saw above that the ultimate loss ratios were not constant by year, and so we cannot assume that historical premium is a good measure of exposure. We will instead make use of an onlevel factor to adjust for changes in rate adequacy. This way we can create a surrogate exposure base.

AY y	Earned Premium	Onlevel Factor	Exposures v_y
1999	5,400	2.200	11,880
2000	5,900	2.050	12,095
2001	6,500	1.850	12,025
2002	8,500	1.400	11,900
2003	10,200	1.200	12,240
2004	11,000	1.100	12,100
2005	11,300	1.050	11,865
2006	11,500	1.050	12,075
Total	70,300		96,180

The exposures v_y are estimated as the historical earned premium times the onlevel factor. These exposures are now assumed to be proportional to the ultimate expected losses by accident year and can be used in formula 3.2.7 to estimate the preliminary development parameters.

Actual Incremental Payments divided by Exposure

AY	0-12	12-24	24-36	36-48	48-60	60-72	72-84	84-96
1999	2.16%	7.46%	10.60%	9.06%	8.23%	5.25%	1.72%	1.66%
2000	2.20%	7.45%	11.88%	10.69%	5.17%	6.42%	1.36%	
2001	2.89%	8.76%	11.97%	9.52%	9.54%	2.46%		
2002	2.34%	7.62%	10.62%	11.83%	4.70%			
2003	2.00%	6.10%	12.39%	4.40%				
2004	1.82%	8.67%	3.68%					
2005	1.80%	5.18%						
2006	1.78%							
IncrLR:	2.12%	7.32%	10.19%	9.08%	6.91%	4.71%	1.54%	1.66%
Cumul:	2.12%	9.45%	19.63%	28.71%	35.62%	40.34%	41.88%	43.53%
Beta	**4.88%**	**16.82%**	**23.40%**	**20.86%**	**15.87%**	**10.83%**	**3.54%**	**3.81%**
Cumul:	4.88%	21.70%	45.10%	65.96%	81.83%	92.66%	96.19%	100.00%
LDF	20.495	4.609	2.217	1.516	1.222	1.079	1.040	1.000

These numbers are calculated additively rather than via chain ladder link ratios but the calculations are still very straightforward. The ELR to onlevel premium is calculated directly as 43.53% by summing the preliminary incremental loss ratios.

We can also calculate an LDF from the β s. However, this development pattern is not exactly equal to that produced by the chain ladder method. The key reason for this difference is that we are now making use of more information. For example, the 2006 year has loss as of 12 months of $215, which would not affect the chain ladder calculation (no link ratio is calculated from the 2006 year), whereas it does affect the result for the constrained model.

The next step is to use these parameters to project the ultimate losses by year. This is done with an additive formula.

$$\text{Ultimate Loss} \quad = \quad \text{(Paid Loss)} + \text{(Expected Ultimate)}\times(1\text{-}1/\text{LDF})$$

$$\text{where} \quad \text{Expected Ultimate} = \text{Exposure} \times \text{ELR}$$

AY y	Exposures v_y	ELR	Expected Ultimate	LDF	IBNR% 1-1/LDF	Latest Diagonal	Final Ultimate	Final L / R
1999	11,880	43.53%	5,172	1.000	0.00%	5,481	5,481	46.14%
2000	12,095	43.53%	5,265	1.040	3.81%	5,464	5,665	46.83%
2001	12,025	43.53%	5,235	1.079	7.34%	5,427	5,811	48.33%
2002	11,900	43.53%	5,181	1.222	18.17%	4,417	5,358	45.03%
2003	12,240	43.53%	5,329	1.516	34.04%	3,047	4,861	39.71%
2004	12,100	43.53%	5,268	2.217	54.90%	1,714	4,606	38.07%
2005	11,865	43.53%	5,165	4.609	78.30%	829	4,874	41.08%
2006	12,075	43.53%	5,257	20.495	95.12%	215	5,215	43.19%
Total	96,180	43.53%	41,871			26,594	41,871	43.53%

We may note that this is the same calculation that is often thought of as the Bornhuetter-Ferguson method, except that the ELR has been estimated from the data rather than from some a priori input.

This method can be equivalently applied by showing the ELR as the ratio of the latest diagonal of loss divided by the exposure corresponding to the expected loss-to-date. This is the format typically seen in the Cape Cod method, as shown below.

AY y	Exposures v_y	LDF	Expos / LDF	Latest Diagonal	Ultimate L / R	
1999	11,880	1.000	11,880	5,481	46.14%	
2000	12,095	1.040	11,634	5,464	46.96%	
2001	12,025	1.079	11,142	5,427	48.71%	
2002	11,900	1.222	9,737	4,417	45.36%	
2003	12,240	1.516	8,073	3,047	37.74%	
2004	12,100	2.217	5,457	1,714	31.41%	
2005	11,865	4.609	2,574	829	32.20%	
2006	12,075	20.495	589	215	36.49%	
Total	96,180		61,088	26,594	43.53%	= 26,594 / 61,088

This result is significant because it derives from the same underlying ODP model as we used for the Multiplicative LDF method. The only difference is that we have added a constraint that forces a certain behavior in the expected ultimate losses.

As with the Multiplicative LDF method, this Cape Cod method tells us nothing about development beyond the eight years in the historical data. We can again introduce a tail factor to change all of our β parameters, with an exact offsetting change to the ELR.

3.2.3 The Bornhuetter-Ferguson (BF) Method

As noted in the previous section, the Cape Cod method looks very much like a traditional Bornhuetter-Ferguson (BF) method, except that in the Cape Cod method the ELR is estimated from the data itself instead of being supplied by the analyst.

The BF method was originally created as a means of enforcing stability in the IBNR loss reserve estimate. As was stated in the original 1972 paper:

> The decision as to whether to develop the reserve as a direct function of case incurred losses or as a function of expected losses turns on the expected volatility of the data. If the data are extremely thin, the presence or absence of several large losses will impact greatly on the IBNR reserves if the reserve is a function of the case incurred.

This original quote implies an either/or decision: the IBNR reserve is either a function of case incurred losses or a function of expected losses. The GLM framework allows us to incorporate both sources of information in a single consistent model. We will start with the more general model, which incorporates the ELR into a GLM, and then move on to how the BF method is traditionally applied in practice.

For our example, let us suppose that the analyst has selected a 50% ELR for use in the BF method. To calculate the β parameters in this constrained model, we perform the same calculation as we used in the Cape Cod method, except that the denominator is the exposures times our selected 50% ELR.

$$\beta_d = \frac{\sum_{y=1}^{n+1-d} c_{y,d}}{\sum_{y=1}^{n+1-d} v_y \cdot \text{ELR}} \qquad \forall\, d \tag{3.2.12}$$

The form shown in formula 3.2.12 is the same as the pattern recommended in Mack (2006) as most consistent with the BF method.

Actual Incremental Payments divided by Exposure times ELR of 50%

AY	0-12	12-24	24-36	36-48	48-60	60-72	72-84	84-96
1999	4.33%	14.92%	21.20%	18.11%	16.46%	10.51%	3.43%	3.32%
2000	4.40%	14.90%	23.76%	21.38%	10.33%	12.85%	2.73%	
2001	5.77%	17.51%	23.93%	19.04%	19.08%	4.92%		
2002	4.69%	15.24%	21.24%	23.66%	9.39%			
2003	4.00%	12.21%	24.77%	8.81%				
2004	3.64%	17.34%	7.36%					
2005	3.61%	10.37%						
2006	3.56%							
Beta	**4.25%**	**14.64%**	**20.38%**	**18.16%**	**13.82%**	**9.43%**	**3.08%**	**3.32%**
Cumul:	4.25%	18.89%	39.27%	57.43%	71.25%	80.67%	83.75%	87.07%
LDF	23.539	5.293	2.547	1.741	1.404	1.240	1.194	1.149

The β parameters will all be in the same proportion to those estimated for the Cape Cod method. However, we no longer have the freedom to introduce a tail factor to go from the 96-month age to ultimate. Instead, the data and our selected ELR have forced a tail factor upon us (again formula 3.2.10).

$$\text{LDF}_n = \frac{1}{\sum_{d=1}^{n} \beta_d} \tag{3.2.10}$$

AY y	Exposures v_y	ELR	Expected Ultimate	LDF	IBNR% 1-1/LDF	Latest Diagonal	Final Ultimate	Final L / R
1999	11,880	50.00%	5,940	1.149	12.93%	5,481	6,249	52.60%
2000	12,095	50.00%	6,048	1.194	16.25%	5,464	6,447	53.30%
2001	12,025	50.00%	6,013	1.240	19.33%	5,427	6,589	54.79%
2002	11,900	50.00%	5,950	1.404	28.75%	4,417	6,128	51.49%
2003	12,240	50.00%	6,120	1.741	42.57%	3,047	5,652	46.18%
2004	12,100	50.00%	6,050	2.547	60.73%	1,714	5,388	44.53%
2005	11,865	50.00%	5,933	5.293	81.11%	829	5,641	47.54%
2006	12,075	50.00%	6,038	23.539	95.75%	215	5,996	49.66%
Total	96,180	50.00%	48,090			26,594	48,090	50.00%

This format is the same as for the Cape Cod method, except that the ELR has been fixed by the model user. We may again note that the final ultimate loss ratio (relative to onlevel premium) is equal to the selected ELR.

In this BF example, the selection of the 50% ELR results in an implied tail factor of 1.149. We could have used the Cape Cod method instead, including a 1.149 tail factor, and produced the same results as the BF method. The two methods are algebraically equivalent: either the ELR determines the tail factor or the tail factor determines the ELR.

Method	**Values Supplied by User**	**Estimated Parameters**
Multiplicative LDF	LDF_n	$v_1, v_2, v_3, v_4, v_5, v_6, v_7, v_8$ $\beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7, \beta_8$ ELR
Cape Cod	$v_1, v_2, v_3, v_4, v_5, v_6, v_7, v_8$ LDF_n	$\beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7, \beta_8$ ELR
Bornhuetter-Ferguson as a GLM	$v_1, v_2, v_3, v_4, v_5, v_6, v_7, v_8$ ELR	$\beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7, \beta_8$ LDF_n

There may be objections at this point that we are not presenting the traditional BF method as found in the original 1972 paper. In that paper, the development pattern (including the tail factor) is selected prior to and independent of the ELR; the ELR implied by the data is ignored and implicitly overwritten by user.

When model parameters are overwritten by the user, bias is introduced: the fitted values for the triangle will no longer balance to the actual values. This bias may remain unrecognized because the model assumptions underlying the BF selections are never made explicit and are therefore left untested.

For this reason, we seek a method that keeps the stability of the traditional BF method, but is more responsive to the loss experience by balancing to the historical paid loss values.

3.3 A Unified Method

Having reviewed the three traditional methods used in a reserve review, we may note some limitations in each.

- The Multiplicative LDF method is clearly over-parameterized.
- The Cape Cod method is attractive but requires an exposure for every AY.
- The traditional Bornhuetter-Ferguson method involves user-intervention, making it less responsive to the actual loss experience.

Given these limitations, the most attractive option would be the Cape Cod method. Unfortunately, we may not have the full data to implement it. This is where a combination or unified method becomes most useful.

We begin by slightly modifying our original model to have the ELR apply to a subset of years. For example, the most recent four years may be grouped together under the same ELR, with the older years being estimated separately. We begin again with the general model.

$$E(c_{y,d}) \quad = \quad \mu_{y,d} \quad = \quad v_y \cdot \mathrm{ELR} \cdot \beta_d \tag{3.3.1}$$

The key concept is that the exposure values, v_y, are not available for the older years and so must be estimated in the model just as was done for the Multiplicative LDF method. We define a group of years, g, in which the exposures are available as containing the indices for the more recent years 2003-2006: $g = \{5, 6, 7, 8\}$.

If all of the years are part of the group, $g = \{1, 2, \cdots, n\}$, then the Unified method is equivalent to the Cape Cod method. On the other extreme, if only the most recent year is included in the group, $g = \{n\}$, then the Unified method is equivalent to the Multiplicative LDF method.

To solve for the Maximum Likelihood Estimates (MLE) of this model, we again have the condition that the fitted column totals must equal the actual column totals. We also have a condition that the sum of all the rows in the subset of years, *g*, must balance between fitted and actual values. This can be written using an indicator function, $\delta(y \in g)$, which is equal to unity for years in the group and zero otherwise.

$$\sum_{y=1}^{n}\sum_{d=1}^{n+1-y} c_{y,d} \cdot \delta(y \in g) \quad = \quad \sum_{y=1}^{n}\sum_{d=1}^{n+1-y} v_y \cdot \mathrm{ELR} \cdot \beta_d \cdot \delta(y \in g) \tag{3.3.2}$$

For our eight-year example, this implies:

$$\sum_{d=1}^{9-y} c_{y,d} \quad = \quad \sum_{d=1}^{9-y} \hat{v}_y \cdot \mathrm{ELR} \cdot \beta_d \qquad \text{for } y = \{1, 2, 3, 4\} \tag{3.3.3}$$

$$\sum_{y=5}^{8}\sum_{d=1}^{9-y} c_{y,d} \quad = \quad \sum_{y=5}^{8}\sum_{d=1}^{9-y} v_y \cdot \widehat{\mathrm{ELR}} \cdot \beta_d$$

This method requires an iteration to solve for the maximum likelihood values, but it is not difficult. The iteration finds the values for the β_d and ELR parameters such that the column total and the grouped-row totals match the actual values.

The result is the "best" model in that it uses all of the available information, produces an unbiased fit, and satisfies the maximum likelihood criteria.

The concept of the Unified method may sound abstract at first, but a numerical example will show that the application is actually quite simple.[8]

We first assume that the rate adequacy index is only available for the second half of the experience period. The exposures for the earlier years are just placeholders and do not affect the

---

[8] For an alternative discussion of this approach to reducing the number of parameters, see Venter (2007). While he is reducing the number of development period parameters rather than the number of accident year parameters, the technique is the same.

final result.

AY y	Earned Premium	Onlevel Factor	Exposures v_y	
1999	5,400	na	5,400	Separate Years
2000	5,900	na	5,900	
2001	6,500	na	6,500	
2002	8,500	na	8,500	
2003	10,200	1.200	12,240	Grouped Years
2004	11,000	1.100	12,100	
2005	11,300	1.050	11,865	
2006	11,500	1.050	12,075	
Total	70,300			

The results of the Unified method can be displayed in the same format as was used for the other methods.[9] The difference in this final version is that the ELR is the same for the recent years but different for the earlier years. The expected loss for the earlier years is simply the result from the Multiplicative LDF method.

AY y	Exposures v_y	ELR	Expected Loss	LDF	IBNR% 1-1/LDF	Latest Diagonal	Final Ultimate	Final L / R
1999	5,400			1.000	0.00%	5,481	5,481	101.50%
2000	5,900			1.037	3.59%	5,464	5,668	96.06%
2001	6,500			1.074	6.90%	5,427	5,829	89.68%
2002	8,500			1.203	16.90%	4,417	5,315	62.53%
2003	12,240	33.14%	4,057	1.465	31.74%	3,047	4,335	35.41%
2004	12,100	33.14%	4,011	2.104	52.47%	1,714	3,818	31.56%
2005	11,865	33.14%	3,933	4.293	76.71%	829	3,846	32.41%
2006	12,075	33.14%	4,002	18.745	94.67%	215	4,004	33.16%
Total	74,580	na	na			26,594	38,296	51.35%
2003-2006	48,280	33.14%	16,002	2.757		5,805	16,002	33.14%

As can be seen in this example, the Unified method is the Multiplicative LDF applied to the old years and the Cape Cod applied to the more recent years. In order for this to be the maximum likelihood estimate, the development pattern and the ELR should be calculated simultaneously. This

[9] The results shown require a numerical iteration to find the MLE parameters, so the reader can verify that the numbers satisfy the balance for row and column totals but cannot easily re-derive the parameters. A practical compromise is given in section 4.1.

requirement can be relaxed in practice, by having the analyst separately select the loss development pattern (see Section 4.1).

This example also shows that the group of years for which the exposure base is available can be treated as a unit with an average LDF applied multiplicatively. The average LDF is calculated as a harmonic average using the exposures as weights.

$$2.757 = \frac{48{,}280}{\left(\frac{12{,}240}{1.465}+\frac{12{,}100}{2.104}+\frac{11{,}865}{4.293}+\frac{12{,}075}{18.745}\right)}$$

We may summarize the relationship of this Unified method to the Multiplicative LDF and Cape Cod cases in the following chart.

Method	Values Supplied by User	Estimated Parameters
Multiplicative LDF	LDF_n	$v_1, v_2, v_3, v_4, v_5, v_6, v_7, v_8$ $\beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7, \beta_8$ ELR
Cape Cod	$v_1, v_2, v_3, v_4, v_5, v_6, v_7, v_8$ LDF_n	$\beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7, \beta_8$ ELR
Unified	v_5, v_6, v_7, v_8 LDF_n	v_1, v_2, v_3, v_4 $\beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7, \beta_8$ ELR

4. PRACTICAL ISSUES FOR THE "UNIFIED" METHOD

Having outlined the general approach for applying a Unified method that combines Multiplicative LDF and Cape Cod methods, we now wish to address two practical issues.

4.1 Separating the Selection of the Development Pattern

As was noted in the description of the theory underlying the Unified method, it is necessary that the development pattern (viewed either as β s or LDFs) must be estimated simultaneously with the other parameters in order to have the maximum likelihood estimate for the reserves. This may be unrealistic in practice, because the reserving actuary will often choose to smooth out the development pattern by removing outlier points or giving more weight to more recent diagonals.

All of these methods allow this step to be done separately. What results is a model that is further constrained by the selection of the β parameters. For example, the Multiplicative LDF method now seeks to find the "best" (MLE) α parameters, representing ultimate losses by accident year, given a selected development pattern. Within the ODP model, the maximum likelihood estimate is found by applying the selected LDF to the latest diagonal of the cumulative loss triangle. Likewise, for the Unified method, we simply apply the same method as outlined in section 3.3, using the selected LDFs.

Having selected a loss pattern of β parameters, either from the triangle or from external information, we apply this to the latest diagonal of account data: multiplicatively for the old years and additively (via Cape Cod) for the more recent years. This is equivalent to a GLM with the β parameters constrained by the user and the ELR fit via MLE.

Method	**Values Supplied by User**	**Estimated Parameters**
Multiplicative LDF	LDF_n	$v_1, v_2, v_3, v_4, v_5, v_6, v_7, v_8$ $\beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7, \beta_8$ ELR
Cape Cod	$v_1, v_2, v_3, v_4, v_5, v_6, v_7, v_8$ LDF_n	$\beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7, \beta_8$ ELR
Unified	v_5, v_6, v_7, v_8 LDF_n	v_1, v_2, v_3, v_4 $\beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7, \beta_8$ ELR
Unified method with development pattern selected by user.	v_5, v_6, v_7, v_8 $\beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7, \beta_8$ LDF_n	v_1, v_2, v_3, v_4 ELR

Because of this result, we can also interpret the Unified method as a purely multiplicative approach. In our example, we have grouped the latest four years together to apply the Cape Cod method. The ultimate for that group of years can also be calculated by applying a single average LDF to the four-year block.

$$\text{LDF}_{\text{Group}} = \frac{v_{2003} + v_{2004} + v_{2005} + v_{2006}}{\dfrac{v_{2003}}{\text{LDF}_{48}} + \dfrac{v_{2004}}{\text{LDF}_{36}} + \dfrac{v_{2005}}{\text{LDF}_{24}} + \dfrac{v_{2006}}{\text{LDF}_{12}}} \tag{4.1.1}$$

In other words, we take a weighted harmonic average of the development patterns for each year in the block, using the exposures as the weights. This average factor then is applied to the four-year block itself. The IBNR can be allocated back down to the individual years using the same Cape Cod method.

The averaging approach accomplishes the same stabilizing goal that is the reason that many people now use the Bornhuetter-Ferguson method, but it better responds to the actual experience.

We should also note that this concept is not original with this paper. This averaging method is the same as would be used if you had a development pattern from accident quarters (AQ) and needed to estimate an accident year (AY) development factor. You would perform an average as below.

$$\text{LDF}_{AY\_12} = \frac{4}{\dfrac{1}{\text{LDF}_{AQ\_3}} + \dfrac{1}{\text{LDF}_{AQ\_6}} + \dfrac{1}{\text{LDF}_{AQ\_9}} + \dfrac{1}{\text{LDF}_{AQ\_12}}} \tag{4.1.2}$$

This is the same as our Unified group, with the assumption that exposures are uniform across quarters.

4.2 Creating the Exposure Index

A second practical problem is the need to create an appropriate onlevel factor. As stated previously, the resulting exposure measure v_y should be proportional to the expected loss for accident year "*y*."

The starting point for this calculation should be changes in the underlying pricing, including the key components:

- Changes to base rates and increased limits factors
- Changes to discretionary pricing modifications

- Changes to terms and conditions (e.g., removal of exclusions)
- Enforcement of underwriting standards (e.g., correct classifications, audits)

We want to adjust for these components so as to remove the effects of the "insurance cycle."

The second component for the onlevel factor is an adjustment for inflation trend. That is, we want to have each year's premium adjusted to a common rate level, but at the loss cost level for the specific year. We do this by adjusting the premium to a projected future level, reflecting rate changes and increases due to exposure inflation. That adjusted premium is then de-trended based on loss inflation.

AY *y*	Earned Premium	Rate Onlevel Factor	Exposure Trend at 3.0%	Onlevel Premium	Loss Trend at 6.0%	Exposures v_y	Final Onlevel Factor
A	B	C	D	E=B*C*D	F	G=E/F	H=G/B
1999	5,400	2.690	1.230	17,863	1.504	11,880	2.200
2000	5,900	2.435	1.194	17,157	1.419	12,095	2.050
2001	6,500	2.136	1.159	16,092	1.338	12,025	1.850
2002	8,500	1.570	1.126	15,023	1.262	11,900	1.400
2003	10,200	1.308	1.093	14,578	1.191	12,240	1.200
2004	11,000	1.165	1.061	13,596	1.124	12,100	1.100
2005	11,300	1.081	1.030	12,577	1.060	11,865	1.050
2006	11,500	1.050	1.000	12,075	1.000	12,075	1.050

Because this index involves estimates of inflation, as well as components of price adequacy that may be difficult to quantify, it is not an easy task to estimate it reliably for a long historical period. This is a practical argument for the Unified method to be applied rather than Cape Cod method.

5. CONCLUSIONS

This paper has outlined a "Unified" reserving method that is a combination of familiar Multiplicative LDF and Cape Cod methods. This Unified method allows the reserving actuary to make use of exposure information even if it can only be compiled for a few recent periods. This Unified method is based on the same statistical model that is common to both of these other methods.

This Unified method achieves the goal of stabilizing the reserves for immature periods, while also being more responsive to the actual loss payments than the traditional Bornhuetter-Ferguson method.

Acknowledgment

The author gratefully acknowledges the helpful reviews by Denise Ambrogio, Bobby Hancock, Gerhard Quarg, Simone Rechtsteiner, Ira Robbin, Jim Sandor, and Chuck Thayer. They offered valuable and constructive criticism even where they did not fully agree with all of the conclusions of this paper.

5. REFERENCES

[1] Bornhuetter, Ronald L., and R.E. Ferguson, "The Actuary and IBNR," *PCAS* LIX, 1972, 181-195.

[2] Clark, David R., "LDF Curve-Fitting and Stochastic Reserving: A Maximum Likelihood Approach," *CAS Forum*, Fall 2003, 41-91.

[3] Mack, Thomas and Gary Venter, "A Comparison of Models that Reproduce Chain Ladder Reserve Estimates," *ASTIN Colloquium International Actuarial Association - Brussels, Belgium* 1999: Toyko.

[4] Mack, Thomas, "Parameter Estimation for Bornhuetter/Ferguson," *CAS Forum*, Fall 2006, 141-157.

[5] Mildenhall, Stephen J., "A Systematic Relationship Between Minimum Bias and Generalized Linear Models," *PCAS* LXXXVI, 1999, 393-487.

[6] Renshaw, Arthur E., and Richard Verrall, "A Stochastic Model Underlying the Chain-Ladder Technique," *British Actuarial Journal* 4, 1998, 903-923.

[7] Schmidt, Klaus D., "Methods and Models of Loss Reserving Based on Run-Off Triangles: A Unifying Survey," *CAS Forum*, Fall 2006, 269-317.

[8] Shapland, Mark, "Loss Reserve Estimates: A Statistical Approach for Determining 'Reasonableness'," *Variance*, Vol. 1, No. 1, Spring 2007, 120-148.

[9] Taylor, Greg, "Chain Ladder Bias," *ASTIN Bulletin*, Vol. 33, 2003, 313-330.

[10] Venter, Gary, "Refining Reserve Runoff Ranges," *CAS E-Forum*, Summer 2007, http://www.casact.org/pubs/forum/07sforum/07s-venter2.pdf.

[11] Verrall, R.J. and P.D. England, "Comments on 'A comparison of stochastic models that reproduce chain ladder reserve estimates'," *Insurance: Mathematics and Economics* 26, 2000, 109-111.

[12] CAS Research Working Party on Quantifying Variability in Reserve Estimates, "The Analysis and Estimation of Loss & ALAE Variability: A Summary Report." *CAS Forum*, Fall 2005, 29-146.

Abbreviations:

BF	Bornhuetter-Ferguson
CL	Chain ladder
GLM	Generalized Linear Model
IBNR	Incurred But Not Reported loss (all loss beyond the amounts in the historical triangle)
LDF	Loss Development Factor, also known as an "age-to-ultimate" factor
ODP	Over-Dispersed Poisson

Reserving Model Notation:

$c_{y,d}$	Actual incremental losses in accident year "y" and development period "d"
β_d	Parameter for development period "d"; can be thought of as the percent of ultimate loss paid during a given development period
ELR	Expected Loss Ratio
v_y	Exposures or volume measure for accident year "y"; can be thought of as onlevel premium
ϕ	Dispersion parameter for ODP model, ϕ = ratio of variance to mean

Biography of the Author

Dave Clark is part of the Actuarial Research and Modeling team for Munich Reinsurance America, Inc. His paper "LDF Curve-Fitting and Stochastic Reserving: A Maximum Likelihood Approach" received the 2003 Reserves Call Paper Prize.

Property-Liability Insurance Loss Reserve Ranges Based on Economic Value

Stephen P. D'Arcy FCAS, Ph.D., Alfred Au, and Liang Zhang

Abstract

A variety of methods to measure the variability of property-liability loss reserves have been developed to meet the requirements of regulators, rating agencies and management. These methods focus on nominal, undiscounted reserves, in line with statutory reserve requirements. Recently, though, there has been a trend to consider the fair value, or economic value, of loss reserves. Insurance regulators worldwide are starting to consider the economic value of loss reserves, which reflects how much needs to be set aside today to settle these claims, instead of focusing on nominal values. If insurers switch to economic values for loss reserves, then reserve variability would need to be calculated on this basis as well. This approach will add considerable complexity to reserve variability calculations. This paper combines loss reserve variability and economic valuation. Loss reserve ranges are calculated on a nominal and economic basis for a simplified insurer to illustrate the key variables that impact loss reserve variability. Nominal interest rate and inflation volatility, interest rate-inflation correlation, and the relationship between claim cost and general inflation are key factors that affect economic loss reserve variability. Actuaries will need to focus on measuring these values accurately if insurers adopt economic valuation of loss reserves.

The model used in this project is available for download at:
http://www.business.uiuc.edu/~s-darcy/papers/LossReserveRangeModelv2.xls.

Keywords: loss reserve ranges, economic value, stochastic simulation

1. INTRODUCTION

Traditional loss reserving approaches in the property-liability field produced a single point estimate value. Although no one truly expects losses to develop at exactly the stated value, the focus was on a single value for reserves that did not reflect the uncertainty inherent in the process. As the use of stochastic models in the insurance industry grew, for Dynamic Financial Analysis (DFA), for Asset Liability Management (ALM) and other advanced financial techniques, loss reserve variability became an important issue. McClenahan (2003) describes the history of interest in reserve variability and loss reserve ranges. Hettinger (2006) surveys the different approaches used to establish reserve ranges. The CAS Working Party on Quantifying Variability in Reserve Estimates (2005) provides a detailed description of the issue of reserve variability, including an extensive bibliography and set of issues that still need to be addressed. The conclusions of this Working Party are that, despite extensive research on this area to date, there is no clear consensus within the actuarial profession as to the appropriate approach for measuring this uncertainty, and that much additional work needs to be done in this area. All of the approaches described in this report and suggestions for future research focus on measuring uncertainty in statutory loss reserves. Given recent attention to fair value insurance accounting, future research should also focus on more accurate economic reserve ranges.

The use of nominal values for loss reserves is sometimes justified as providing a safety load, or risk margin, over the true (economic) value of the reserves. However, risk margins determined in this way would fluctuate with interest rates and vary by loss payout patterns. A more appropriate approach, which is beyond the scope of this research, would be to establish risk margins based on the risks inherent in the reserve estimation process, such as determining the risk margin based on the difference between the expected economic value and a level such as the 75th percentile value.

The Financial Accounting Standards Board (FASB) and the International Accounting Standards Board (IASB) have proposed an alternative approach to valuing insurance liabilities, including loss reserves. This approach, termed Fair Value, proposes that loss reserves in financial reports be set at a level that reflects the value that would exist if these liabilities were sold to another party in an arms length transaction. The relative infrequency with which these exchanges actually take place, and the confidentiality surrounding most trades that do occur, make this approach to valuation more of a theoretical exercise than a practical one, at least in the current environment. However, Fair Value would reflect the time value of money, so the trend would be, if these proposals are implemented, to set loss reserves at their economic rather than nominal values. The issues involved, and financial implications, in Fair Value accounting are covered extensively in the Casualty Actuarial Society report, Fair Value of P&C Liabilities: Practical Implications (2004). However, despite the comprehensive nature of the papers included in this report, little attention is paid to the impact the use of Fair Value accounting would have on loss reserve ranges. If reserves are to be calculated on a fair value basis, then reserve ranges should also be based on this approach as well.

A final impetus for this project is the recent criticism of the casualty actuarial profession over inaccurate loss reserves and the profession's response to these attacks. A Standard & Poor's report (2003) blamed the reserve shortfalls the industry reported in 2002 and 2003 on actuarial "naivety or knavery." The actuarial profession responded strongly to this criticism, both with information and with investigation (Miller, 2004). The Casualty Actuarial Society formed a task force to address the issues of actuarial credibility. The report of this Task Force (2005) included the recommendation that actuarial valuations include ranges to indicate the level of uncertainty in the reserving process and that additional work be done to clarify what the ranges indicate. Once again, the focus was on statutory loss reserve indications rather than the economic value.

The critical problem with setting reserve ranges based on nominal values is the impact of inflation on loss development. Based on relatively recent history (the 1970s) and current economic conditions (increasing international demand for raw materials, vulnerable oil supplies, the U.S. Federal Reserve's response to the subprime credit crisis), increasing inflation has to be accorded some probability of occurring in the future by any actuary calculating loss reserve ranges. As inflation will affect all lines of business simultaneously, the impact of sustained high inflation would be to cause significant adverse loss reserve development for property-liability insurers. Loss reserve ranges based on nominal values would therefore include the high values that would be caused by a significant rise in inflation. However, inflation and interest rates are closely related, as first observed by Irving Fisher (1930) and confirmed by economists consistently since. The loss reserves impacted by high inflation would most likely be accompanied by high interest rates, so the economic value of those reserves would not be that much higher than the economic value of the point estimate for reserves. Using economic values to determine reserve ranges could also lead to narrower ranges and provide a clearer estimate of the true financial impact of reserve uncertainty.

This project utilizes realistic stochastic models for interest rates, inflation, and loss development to determine loss reserve distributions and ranges on both a nominal and economic basis, draws a comparison between the two approaches and explains why the appropriate measure of uncertainty is based on the economic value. This work builds on prior work by Ahlgrim, D'Arcy, and Gorvett (2005) developing a financial scenario generator for the CAS and SOA as well as research on the interest sensitivity of loss reserves by D'Arcy and Gorvett (2000) and Ahlgrim, D'Arcy, and Gorvett (2004).

This study measures the uncertainty in loss reserving that is based on process risk, the inherent variability of a known stochastic process. In this analysis, both the distribution of losses and the parameters of the distributions are given. Thus, unlike actual loss reserving applications, there is no model risk or parameter risk. Setting loss reserves in practice involves more degrees of uncertainty, and would therefore lead to greater variability in the underlying distributions of ultimate losses and larger reserve ranges. This study is meant to illustrate the difference between nominal and economic ranges, and starting with specified loss distributions more clearly demonstrates this effect.

The remainder of the paper proceeds as follows. Section 2 reviews loss reserving methods. Section 3 discusses the importance of Asset Liability Management. Section 4 looks at current developments leading towards the use of economic values for loss reserving. Section 5 provides evidence of trends in inflation. Section 6 discusses the models used in the project. Section 7 address the parameters used in the models. Section 8 describes how the model is run. Section 9 summarizes the results. Section 10 concludes this paper.

2. REVIEW OF LOSS RESERVING METHODS

A primary responsibility of insurers is to ensure they have adequate capital to pay outstanding losses. Much research has been done on methods to evaluate and set these loss reserves. Berquist and Sherman (1977) and Wiser, Cockley, and Gardner (2001) provide excellent descriptions of the standard approaches used to obtain a point estimate for loss reserves. Loss reserve ranges became an issue in the past two decades, and has also been addressed in numerous papers. For example, Mack (1993) presented the chain-ladder estimates and ways to calculate the variance of the estimate. Murphy (1994) offered other variations of the chain-ladder method in a regression setting. Venter (2007) worked on improving the accuracy of these estimates and reducing the variances of the ranges. Other contributors to loss reserve estimates and discussions on the strengths and weaknesses of various evaluation models include Zehnwirth (1994), Narayan and Warthen (1997), Barnett and Zehnwirth (1998), Patel and Raws (1998), and Kirschner, Kerley, and Isaacs (2002). These works typically deal with nominal undiscounted value of loss reserves in line with statutory reserve requirements. Shapland (2003) explores the meaning of "reasonable" loss reserves, emphasizing the need for models to take into account the various risks involved along with "reasonable" assumptions. His paper points out that reasonableness is subject to many aspects, such as culture, guidelines, availability of information, and the audience; as such the paper concludes that more specific input is needed on what should be considered "reasonable" in the actuarial profession.

Traditional methods rely on imbedded historical inflation to produce the nominal reserves. Outstanding losses will be exposed to the impact of inflation until they are finally paid. If the inflation rate during the experience period has been high, loss severity will be projected to increase significantly generating large loss reserves. Similarly, after periods of low inflation, loss severity will be projected to increase more slowly, leading to lower loss reserves. Because inflation and interest rates are correlated, an insurer with an effective Asset Liability Management (ALM) strategy for dealing with interest rate risk can alleviate some of the impact of changing inflation.

There have been reserving techniques that attempt to isolate the inflationary component from the other effects, such as those proposed by Butsic (1981), Richards (1981), and Taylor (1977). Butsic investigated the effect of inflation upon incurred losses and loss reserves, as well as the inflation effect on investment income. For both increases and decreases in inflation, these components are found to vary proportionally. According to Butsic, as competitive pricing is dependent on a combination of both claim costs and investment income, insurers are to a large extent unaffected by unanticipated changes in inflation. Richards provides a simplified technique to evaluate the impact of inflation on loss reserves by factoring out inflation from historical loss data. Assumptions of future inflation can then be factored in to project possible values of future loss reserves. Under the Taylor separation method, loss cost is divided into two components, the stationary claim delay distribution and exogeneous inflation. This method assumes the inflation component affects all loss payments made in a given year to the same degree, regardless of the original accident year. Essentially, unpaid losses are not considered to be fixed in value over time but rather are fully sensitive to inflation. An alternative to this assumption is proposed by D'Arcy and Gorvett (2000), which allows loss reserves to gradually become "fixed" in value from the time of the loss to the time of settlement. Inflation would only affect the unpaid losses that have not yet become fixed in value. These two methods will be described in detail in the model section.

3. ASSET LIABILITY MANAGEMENT

Asset Liability Management (ALM) is a process in which organizations manage risk by considering the impact that an event would have on both their assets and their liabilities; risk is managed by using the offsetting effects to reduce aggregate risk to an acceptable level. For example, the fall of the dollar against the euro might increase the cost of claims an insurer would have to pay on business written in Europe. If the insurer held assets denominated in euros, then these would increase in value as the dollar fell, offsetting some, or all, of the increased claim costs. Although ALM can be used to deal with any type of financial risk, in practice most insurers focus on interest rate risk. In this context, if both assets and liabilities change by the same amount when interest rates rise or fall, there will be no interest rate risk for the firm. However, if they respond differently, the firm will be exposed to interest rate risk. Prior to the 1970s, mismatches between assets and liabilities were not a significant concern. Interest rates in the United States experienced only minor fluctuations, making any losses due to asset-liability mismatch insignificant. However the late 1970s and early 1980s were a period of high and volatile interest rates, making ALM a necessity for any viable financial institution. If interest rates increase, fixed income bonds decrease in value and the economic value (the discounted value of future loss payments) of the loss reserves decreases. The opposite occurs for both the assets and liabilities when interest rates decrease. Ahlgrim, D'Arcy and Gorvett (2004) provide a detailed analysis of the effective duration and convexity of liabilities, for property-liability insurers under stochastic interest rates that shows how assets can be invested to reduce the impact of interest rate risk.

Insurers can employ an ALM program to reduce the impact of inflation on loss reserves and maintain their surplus despite changing interest rates. This requires insurers with short effective duration liabilities to hold short-term assets. Some insurers invest in longer duration assets that offer higher yields. During periods of stable or declining interest rates, this approach will provide a higher return. However, when interest rates rise this strategy can be costly.[1] The effect of duration mismatching on loss reserves given expectations of future inflation volatility is a complicated issue, and is outside the scope of this paper. As will be shown later, the higher the correlation between nominal interest rates and inflation, such as in the 1970s, the more important and significant ALM's impact will be.

4. ECONOMIC VALUE OF LOSS RESERVES

Recent reports by the Financial Accounting Standards Board (FASB) and the International Accounting Standards Committee (IASC) have advocated Fair Value accounting measures. The American Academy of Actuaries established the Fair Value Task Force to address this issue. The fair value of a financial asset or liability is its market value, or the market value of a similar asset or liability plus some adjustments. If a market does not exist, the asset or liability should be discounted to its present value at an appropriate capitalization rate depending on the risk components it encompasses. The Fair Value report by AAA (2002) provides details on the valuation principles. The promotion of Fair Value accounting, which considers both risk and the time value of money, indicates a new trend towards economic valuation.

The trend towards economic or market-value based measurement of the balance sheet replacing existing accounting measures is also seen in the European Union, where solvency regulation is currently under reform. CEA (2007) describes how the new Solvency II project takes an integrated risk approach that will better account for the risks an insurer is exposed to than the current fixed standards under Solvency I. Solvency II introduces the use of a market-consistent valuation of assets and liabilities and market consistent reserve valuation, much like those proposed under Fair Value accounting in the United States.

---

[1] In late 2007 and early 2008, many banks suffered significant losses by following a similar mismatched strategy. They used off-balance sheet Structured Investment Vehicles (SIV) that invested in long-term bonds, often tied to subprime mortgages, but financed the investments with short-term debt. When the value of the assets fell and the credit markets froze up increasing short-term borrowing costs, the banks incurred significant losses which, in some cases, cost the CEOs their jobs (Hilsenrath, 2008).

Australian regulations have required ranges based on economic value since 1999. The value of the insurer's liabilities is generally assumed to be independent of the insurer's underlying assets. The Audit and Actuarial Reporting and Valuation (2006) and Institute of Actuaries of Australia Professional Standard 300 (2007) require loss reserves to be discounted by current observable market-based rates. These rates are based on characteristics of the future obligations, or derived from a yield of a replicating portfolio of low-risk securities. The study mentions that appropriate allowance can be made for future claim escalation from inflation and superimposed inflation (e.g., social or legal costs), but no clear methodology is provided as to how inflation should be taken into account.

Although there has been much discussion on the meaning of fair or economic value, both within and outside the United States, little attention has been given so far to the impact of economic value on loss reserve ranges. This paper ties together the loss reserve ranges with the economic values to show the relationship between loss reserve ranges on a nominal and economic basis and to illustrate some of the issues involved in calculating reserve ranges on economic values.

The economic value of an insurer's liabilities is determined by discounting expected future cash flows emanating from the liabilities by their appropriate discount rate. Butsic (1988) and D'Arcy (1987) explore discounting reserves using a risk-adjusted interest rate that reflects the risk inherent in the outstanding reserve. Girard (2002) evaluates this using the company's cost of capital. Actuarial Standard of Practice No. 20 addresses issues actuaries should consider in determining discounted loss reserves. This Standard suggests that possible discount factors could be the risk-free interest rate or the discount rate used in asset valuation.

5. TRENDS IN INFLATION - LEVEL AND VOLATILITY

Inflation as measured by the 12-month change in the Consumer Price Index has varied widely, from -11% to +20% over the period 1922 through 2007 (Figure 1). Since the adoption of Keynesian economic policies in developed countries following World War II, the general trend has been to avoid deflation at the cost of persistent inflation.[2] Rapid increases in oil prices in the 1970s and the early 21st century have increased inflation rates. The steady depreciation of the dollar in recent years has also put additional inflationary pressures on the U. S. economy. Recently, concern

[2] There is some disagreement over how much of an impact Keynesian economic policies have had on inflation patterns. The impact of open market bond purchases by the Federal Reserve, particularly during full employment periods, could have a more significant impact on inflation. Regardless, the United States has not experienced significant deflation since the 1950s, so that is the period used to determine the parameters calculated for the models in this work.

over the financial consequences of the subprime mortgage crisis and credit crunch has led the U. S. Federal Reserve to lower the discount rate to shield the economy from a housing slump and stabilize turbulence in the financial markets. Lowering interest rates is likely to lead to an increase in future inflation. Oil prices have risen sharply, the dollar has dropped to historical lows against the euro and gold prices have soared. Falling prices of long-term government debt after the recent rate drop suggests investors concern over inflation. Thus, the potential for inflation to increase must be incorporated into any financial forecast.

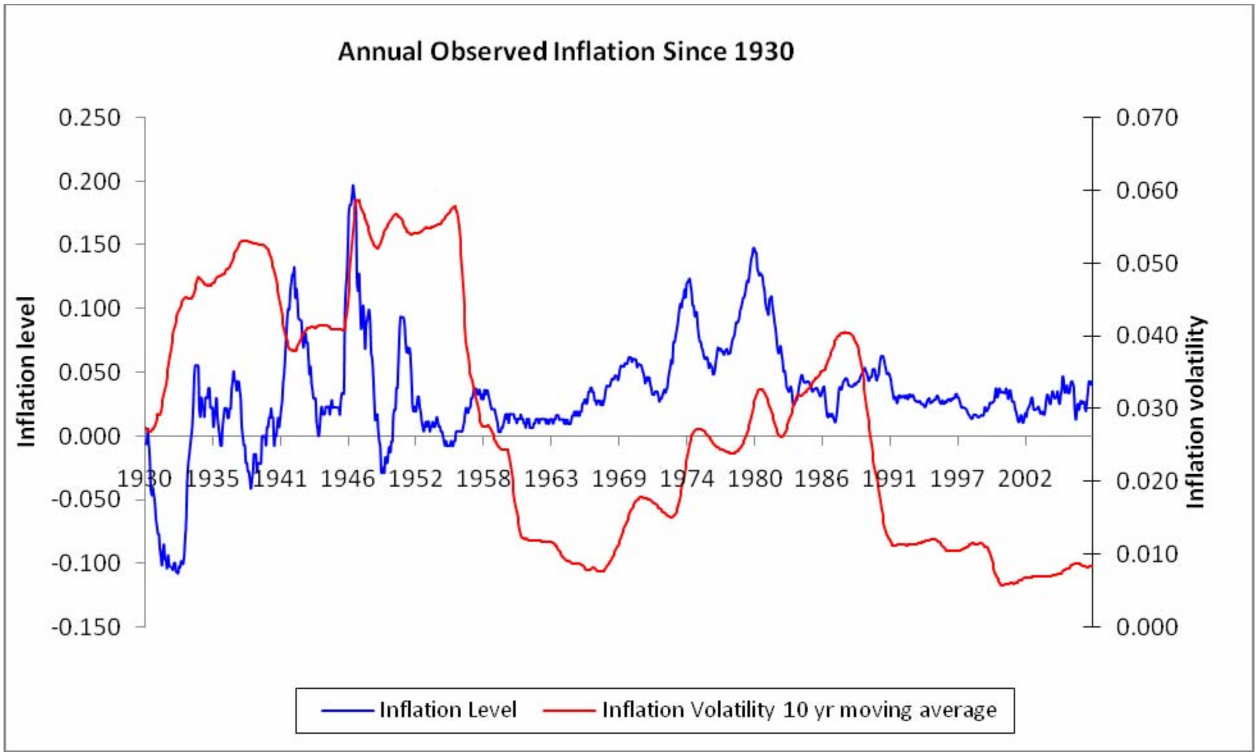


Figure 1

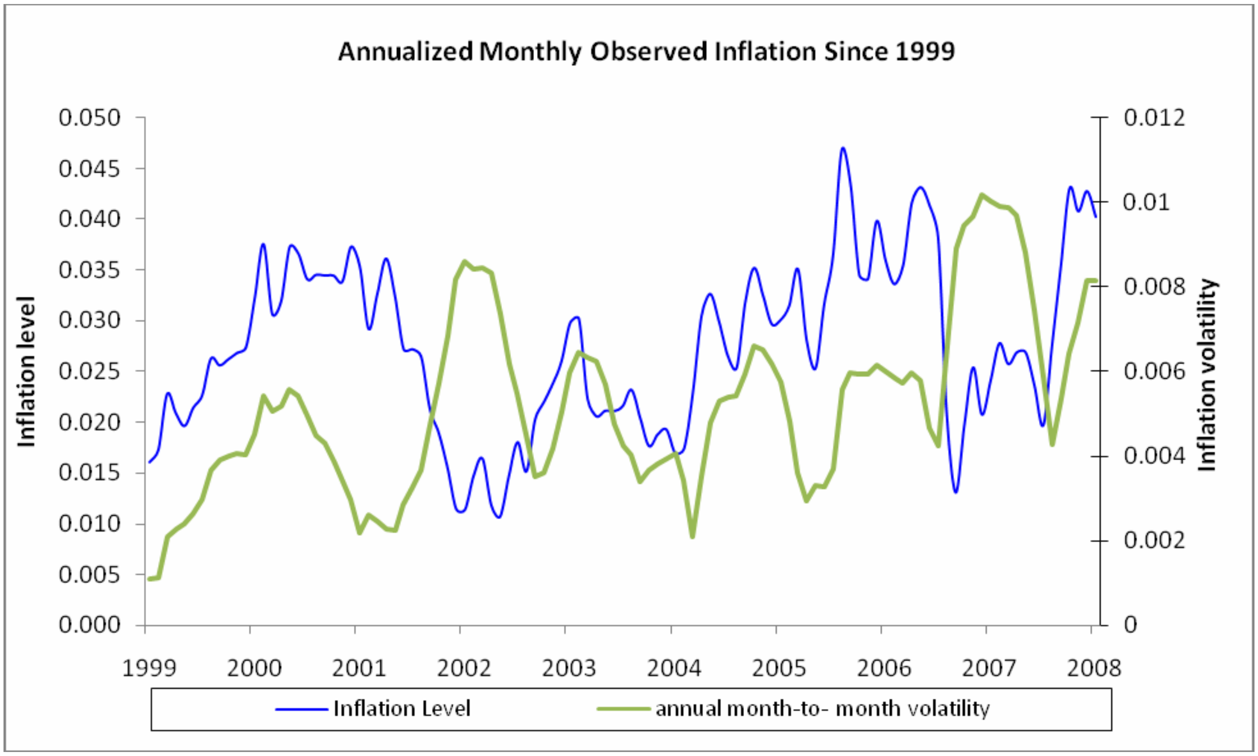


Figure 2

Figure 1 shows the inflation level and the inflation volatility (based on a ten-year moving average) since 1930. Note the periods of deflation that occurred during the Depression and right after World War II, and the inflation spikes of the 1940s, 1950s, 1970s, and 1980s. Inflation volatility has also experienced several spikes, most recently in the 1980s. For the last decade, volatility has been at historic lows. Figure 2 shows the same data from the past 10 years, where there appears to be a rise in both inflation and inflation volatility. On this graph, inflation volatility is shown on a year-by-year basis to show the recent volatility more clearly. With the current upward trend in inflation volatility, it is necessary to consider the possibility inflation volatility returning to the levels of the 1950s or the 1980s. Inflation volatility determines how accurately we are able to predict future inflation trends; the greater the volatility, the lower the ability to forecast future inflation, and thus, the greater uncertainty on its impact on loss reserves.

6. THE MODELS

The loss reserving model used in this research involves: a loss generation model for loss severity, a loss decay model for loss payout patterns, a two-factor Hull-White model for nominal interest rates, a Ornstein-Uhlenbeck model for inflation, adjustment for correlation between the nominal interest rate and inflation, adjustment for claims cost inflation, and a fixed claims model for the impact of inflation on unpaid claims. A sensitivity analysis worksheet is also built in to test the sensitivity of the parameters.

6.1 Loss generation model

The loss generation model generates aggregate claims based on the user's input of the number of claims, choice of distribution of the claim severity, and the mean and standard deviation of severity. The number of claims is assumed to be known. The severity of claims can follow a Normal, Log-normal or Pareto distribution.

6.2 Loss decay model

These losses can be settled either at a fixed time or at a rate based on a decay model over a number of years. If the claims are to be settled on a decaying basis, the decay model calculates the proportion of losses to be settled each year given a decay factor. For simplicity, loss severity is assumed to be independent of time to settlement. The decay model is of the following form:

$$X_{t+1} = (1-\alpha) * X_t \tag{6.1}$$

where X_i is the number of claims settled in year i, and α is the decay factor or the proportion of claims settled each year.

6.3 Nominal interest rate model

A two-factor Hull-White model is used to generate nominal interest rate paths. The Hull-White model uses a mean-reverting process with the short-term real interest rate reverting to a long-term real interest rate, which is itself stochastic and reverting to a long-term average level.

$$\begin{aligned} dr_t &= \kappa_r (l_t - r_t)dt + \sigma_r \sqrt{dt}\ dz_r \\ dl_t &= \kappa_l (\mu_r - l_t)dt + \sigma_l \sqrt{dt}\ dz_l \end{aligned} \tag{6.2}$$

where t is the time, r is the short-term rate, l is the long-term rate, κ is the mean reversion speed, μ is the average mean reversion level, dt is the time step, σ is the volatility, and dz is a Weiner process. This model allows for negative values, which do not typically occur for nominal interest rates. We impose a minimum short-term and long-term rate of 0% to adjust for this.

6.4 Inflation model

A one-factor Ornstein-Uhlenbeck model is used to generate inflation paths. The Ornstein-Uhlenbeck model uses a mean-reverting process with the current short-term inflation reverting to the long-term mean.

$$di_t = \kappa_r(\mu_r - i_t)dt + \sigma_r \sqrt{dt}\, dz_r \tag{6.3}$$

where t is the time, i is the current inflation, κ is the mean reversion speed, μ is the long-term inflation mean, dt is the time step, σ is the volatility, and dz is a Weiner process.

6.5 Correlated nominal and real interest rates

The short-term nominal interest rate and inflation rates are correlated through their random shock components. The random *dz* component is adjusted for a weighted average between a common correlated random component and an individual random component.

$$\begin{aligned} dz_{r,\,\text{nominal}} &= \rho dz_{\text{correlated}} + \sqrt{1-\rho^2}\,dz_{\text{nominal}} \\ dz_{r,\,\text{inflation}} &= \rho dz_{\text{correlated}} + \sqrt{1-\rho^2}\,dz_{\text{inflation}} \end{aligned} \tag{6.4}$$

where ρ is the correlation factor between the short-term interest rate and inflation rate, and *dz* are Weiner processes.

6.6 Masterson Claims Cost Index

Claim costs do not simply grow at the rate of inflation. The Masterson Claim Cost Index measures the rate at which claims costs are inflated over time by decomposing claim costs into its various components and inflating each part separately (Masterson, 1981; Masterson, 1987; Van Ark, 1996; Pecora, 2005). For this research, the Masterson Claim Cost Index is simplified to a linear projection of the general inflation rate.

6.7 Fixed claims model

Cash flows from unpaid claims are sensitive to inflation rate changes. Under the Taylor separation model (1977), any claim that has not been settled is subject to the full inflation in that year. If there is a car accident now and the claimant receives ongoing medical treatment for several years before the loss is settled, all medical costs are assumed to be impacted by inflation until the claim is paid. D'Arcy and Gorvett (2000) propose a model that reflects a different relationship between unpaid losses and inflation. Their model separates unpaid claims into portions that are "fixed" in value from those that are not. These fixed claims, once determined, will not be subject to future inflation while the remaining unfixed claims continue to be exposed to inflation. For example, medical treatment given over a period of time becomes fixed in value when the service is provided. If medical prices rise after some treatment has been provided, only future medical treatment will have this increased cost; medical treatment received before the price increase will have already been fixed. Any pain and suffering compensation is generally determined at a later date. This portion of the claim will likely continue to be affected by inflation until this claim is settled. As a result of only exposing partial loss segments to inflation, inflation's impact on the loss is greatly reduced. A representative function that displays these attributes is:

$$f(t) = k + \{(1-k-m)(t/T)^n\} \tag{6.5}$$

where $f(t)$ represents the proportion of the ultimate claims "fixed" at time t, k is the proportion of the claim that is fixed immediately, m is the proportion of the claim that will be fixed only when the claim is settled, and T is the time at which the claim is fully settled.

The model (6.5) can be divided into three cases by the value of the exponent n: the linear case $n = 1$, when claim value is fixed uniformly up to its ultimate settlement; the convex case $n > 1$, when the rate of fixing the value of a claim increases over time, and the concave case $n < 1$, when the rate of fixing the value of a claim increases quickly initially but slows down as time approaches the ultimate settlement date. The larger the n, the more closely the fixed claim model will resemble the Taylor model.

7. PARAMETERIZATION

Based on the ten-year loss development data of the auto insurance industry from A. M. Best's Aggregate and Averages over the period 1980-1996, approximately one-half of all remaining losses of the total loss value are settled each year up to the ultimate settlement year. Assuming loss severity to be independent of time of settlement, we use a decay factor $\alpha = 0.5$ for the number of claims settled each year. If loss severity is positively correlated with time of settlement, we would use a larger decay factor for the number of claims settled, but offset that by increasing the value of claims over time. Calculating the decay factor based on total loss value adjusts for the assumption that claims severity is independent of time to settlement.

Regressions were run against historical data to parameterize the Ornstein-Uhlenbeck inflation model and the two-factor Hull-White nominal interest rate model. These parameters are tabulated below:

Ornstein-Uhlenbeck inflation model			Two-factor Hull-White nominal interest rate model				
κ	μ	σ	κ_r	μ	σ_r	κ_l	σ_l
0.23	4.12%	1.90%	0.06	6.69%	1.55%	0.07	0.96%

The Fisher formula is an equilibrium statement that, on average, nominal rates and inflation are linked. Sarte (1998) has found that in an environment with stochastic inflation, the Fisher formula is still a reasonable approximation to its more complete counterpart in a dynamic endowment environment. It is important to note that inflation is sometimes a matter of government policy and the model should be adjusted to match the current economic situation.

The correlation between the three-month U. S. Treasury interest rates (the shortest securities issued) and percentage changes in the CPI index was determined for several periods as shown below. The relationship between inflation and interest rates hypothesized by Fisher applies to expected inflation and current interest rates. There is no reliable measure of expected inflation, so the actual inflation rate for a recent period is used here instead. The CPI is an estimate of a market basket of prices at a particular time; monthly changes include significant noise, as under or over-stated values in one month are adjusted the following month. This leads to the lowest values for the correlations. Inflation rates calculated based on three- and six-month CPI changes are more highly correlated with interest rates. The problem introduced by increasing the time period for determining the current inflation rate is that these rates may be less indicative of expected inflation. To run the model, we selected the one-month inflation value over the more recent time period, or 45%.

Correlation between 3-Month Treasury Bill Rate and Inflation

Years	1934-2007	1934-1970	1971-2007
One-Month Inflation Rates	0.241	0.007	0.459
Three-Month Inflation Rates	0.317	0.006	0.556
Six-Month Inflation Rates	0.364	0.011	0.615
Twelve-Month Inflation Rates	0.414	-0.007	0.684

Other values for this correlation are shown in the sensitivity tests. The Masterson claims cost index for auto insurance bodily injury from 1936 to 2004 was regressed against the historical inflation rate using a fixed intercept of 0. The slope of the regression increases over time indicating that claim costs have been increasing more than CPI inflation benchmarks. A slope of 1.6 was selected for this model; other values are illustrated in the sensitivity section.

For the fixed claim model, we are using the linear case, with the parameter for k (portion of claim fixed at inception of claim) of 0.15 as suggested in D'Arcy and Gorvett (2000), but the parameter for m (portion of the claim fixed at settlement) at 0.5. The sensitivity of these values is examined in a later section.

8. RUNNING THE MODEL

This model is available on the lead author's Web Site (http://www.business.uiuc.edu/~s-darcy/papers/LossReserveRangeModelv2.xls) and will also be made available through the CAS Web Site so any interested reader can run the model to reproduce the results here or test alternative parameters. The loss reserve model, which is designed in Microsoft Excel, begins with an input worksheet for the user to enter the parameters for each model used and the number of iterations to be made in the simulation. For each iteration, the model generates a loss distribution, a nominal interest rate path and an inflation path that are used to produce the nominal and economic loss ranges. An output worksheet collects the values from each iteration run and calculates the mean, standard deviation, and reserve ranges for both the nominal and economic value cases. The summary sheet collects these key statistics, the parameters used, and the number of iterations in the simulation in side-by-side columns for comparison.

The model is set to generate 1000 random log-normally distributed claims settled on a decaying basis over 10 years. The mean and standard deviation of the losses are arbitrarily set to 1000 and 500, respectively. The decay model then calculates the proportion of these claims settled at each time step up to the 10th year.

The generated losses are compounded at the inflation rate up to their time of settlement. This is the nominal, undiscounted value of losses that insurers are statutorily required to have as a reserve. The interest rate model generates cumulative interest rate paths corresponding to each time period up to settlement. The nominal values are then discounted back by this cumulative interest rate factor to obtain the economic value of losses.

For a simplified example, assume a single claim of \$1000 (based on the price level in effect when the loss occurred) is settled at the end of five years, and the annual nominal interest rate is 5%. Also assume that the inflation is equal to one-half of the nominal rate throughout the five years, i.e., $(1+5\%)^{0.5} - 1 = 0.0247$. The nominal value of the loss reserve would be $\$1000 * (1+2.47\%)^5 = \$1129.73$. This nominal value is discounted back by the interest rate over the five years to get the economic value $\$1129.73 * (1+5\%)^{-5} = \$885.17$. In economic terms, the amount that should be reserved for handling this loss in today's dollars is \$885.17. Now consider what would happen if interest rates changed by 200 basis points up or down. If the nominal rate is 7%, inflation will be $(1+7\%)^{0.5} - 1 = 3.44\%$, and the nominal value and economic value will be \$1184.30 and \$844.39, respectively. If the nominal rate is 3%, inflation will be $(1+3\%)^{0.5} - 1 = 1.49\%$, and the nominal value and economic value will be \$1076.70 and \$928.77, respectively. Thus, the nominal value range will be \$1129.73 - \$1076.70 = \$53.03, and the economic value range will be \$928.77 - \$885.17 = \$43.60. The economic value range is only 82% of the nominal value range. This is a simplified example illustrating three possible values of one claim, assuming inflation is proportional to the nominal rate. Under circumstances such as this, the reserve range based on economic values will be smaller than reserve ranges based on nominal values.

Now consider a book of 1000 such claims and allow inflation to vary independently of nominal rates. The average nominal and economic values of these 1000 claims are determined based on the interest rate and inflation paths generated for that simulation. This claims generation process is repeated for 10,000 simulations, with each simulation generating a different interest rate and inflation path for the 1000 claims of that iteration, and a distribution of nominal and economic loss reserves are generated. The mean, standard deviation, minimum, maximum, as well as the 5, 25, 75, and 95 percentile for both the nominal and economic loss ranges are determined and compared. A confidence interval ratio is computed by dividing the economic range confidence interval by the nominal range confidence interval for both a 50 percent (ranging from the 25th percentile to the 75th percentile) and 90 percent (ranging from the 5th percentile to the 95th percentile) confidence interval. These ratios will be used as an indicator of the difference in volatility between the economic loss ranges and nominal loss ranges.

9. RESULTS

To examine the effects of how the confidence interval is affected by changes in the assumptions, 10,000 simulations were run for each of the following cases. As the 50 percent and 90 percent confidence interval ratios turn out to be fairly close, only the 90 percent confidence interval ratios are shown here. The complete results are available from the authors. A monthly time step was chosen to provide a close approximation to continuous interest rate models, as inflation data are only available monthly.

9.1 Taylor Model versus Fixed Claim

The first example is based on running the model with the following assumptions: 1) monthly time step, 2) a correlation factor of 45% between the nominal interest rate and inflation, 3) claims inflation rate of 1.6 times the general inflation rate, 4) the Taylor separation model. This is Case A. Figure 3 shows the distributions for both the nominal and economic values; as would be expected, the economic values are lower than the nominal values; the economic reserve range turns out to be approximately 94% of the nominal loss reserve range. Discounting does not reduce the ranges much. The second example, Case B, incorporates the fixed loss model suggested by D'Arcy and Gorvett (2000). In this case there is a significant decrease in the standard deviation of the nominal and economic reserves because losses are only partially exposed to inflation throughout its time to settlement. (The portion of the claim that is fixed is no longer affected by future inflation.) In this case the confidence interval ratio (the economic range divided by the nominal range) is 102%. Discounting reserves reduces the level of the reserves, but not the range. We will treat Case B as the base case and examine additional changes in relationship to this case. The mean values, standard deviations, 5th and 95th percentiles and the 90% confidence intervals are for both nominal and economic values for Case A and Case B are shown on Table 1.

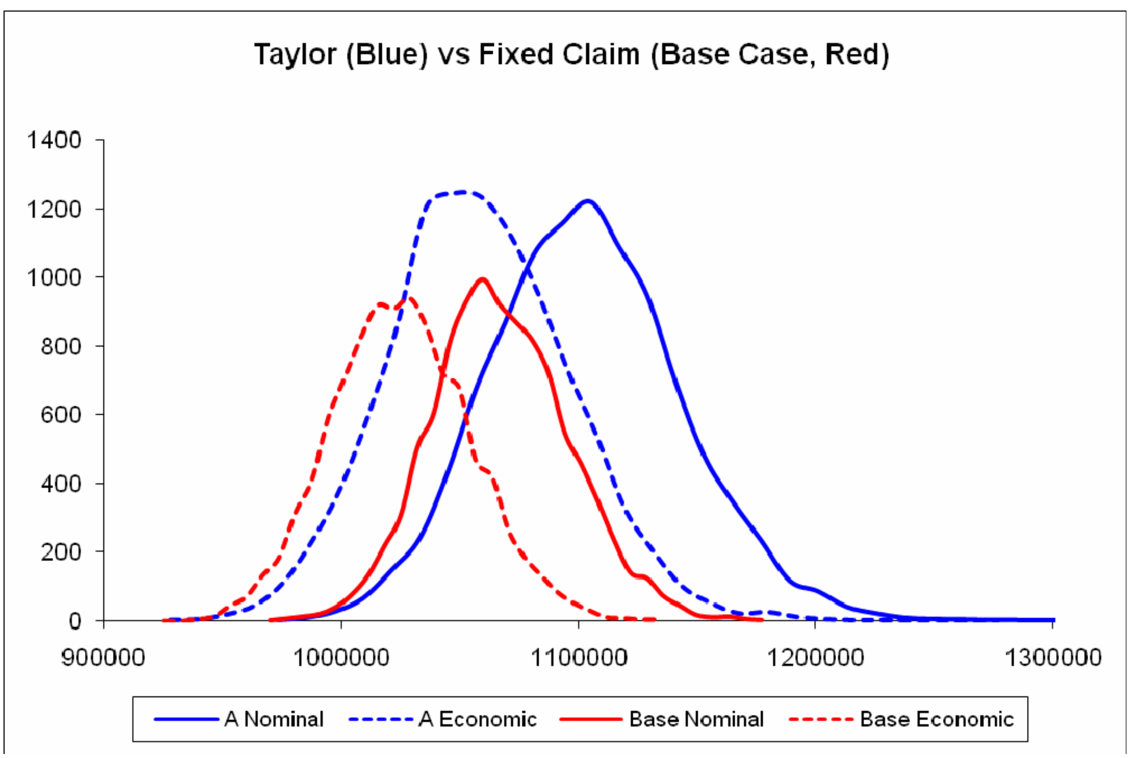


Figure 3

Case	Mean	Standard Deviation	Percentiles 5th	95th	90% Confidence Interval	Confidence Interval Ratio
A - nominal	1097016.132	40241.05642	1033427.88	1165584.07	132156.19	94.15%
A - economic	1052020.311	37895.67459	991242.95	1115669.17	124426.22	
B - nominal	1063890.967	28761.53405	1018485.77	1112663.88	94178.11	101.78%
B - economic	1021643.248	29192.93832	974956.77	1070811.09	95854.32	

Table 1

9.2 High Claims Cost Inflation

The relationship between claims inflation and the general inflation rate has varied widely over the period 1936 through 2004, but claims inflation is consistently higher than overall inflation. One reason for this is that medical costs are a major component of auto insurance claims and these have consistently outpaced general inflation. The third-party payer relationship also reduces resistance to cost increases, leading to higher inflation. Recently, the relationship between claims cost and inflation has increased significantly; between 2001 and 2004, auto bodily injury costs increased 1.9 times the general inflation rate. For Case C, the claim cost inflation factor will be 1.9 and the standard deviation of the nominal range will be increased 1.9 times the original inflation volatility. As the nominal range for loss reserves is the range of actual losses inflated by the claims cost, higher claims cost inflation will increase the nominal range and reduce the confidence interval ratio. The distributions for both the Base Case and Case C are shown on Figure 4, and the key metrics of Case C are shown on Table 2. For Case C the confidence interval ratio of the economic range to the nominal range drops to 97%.

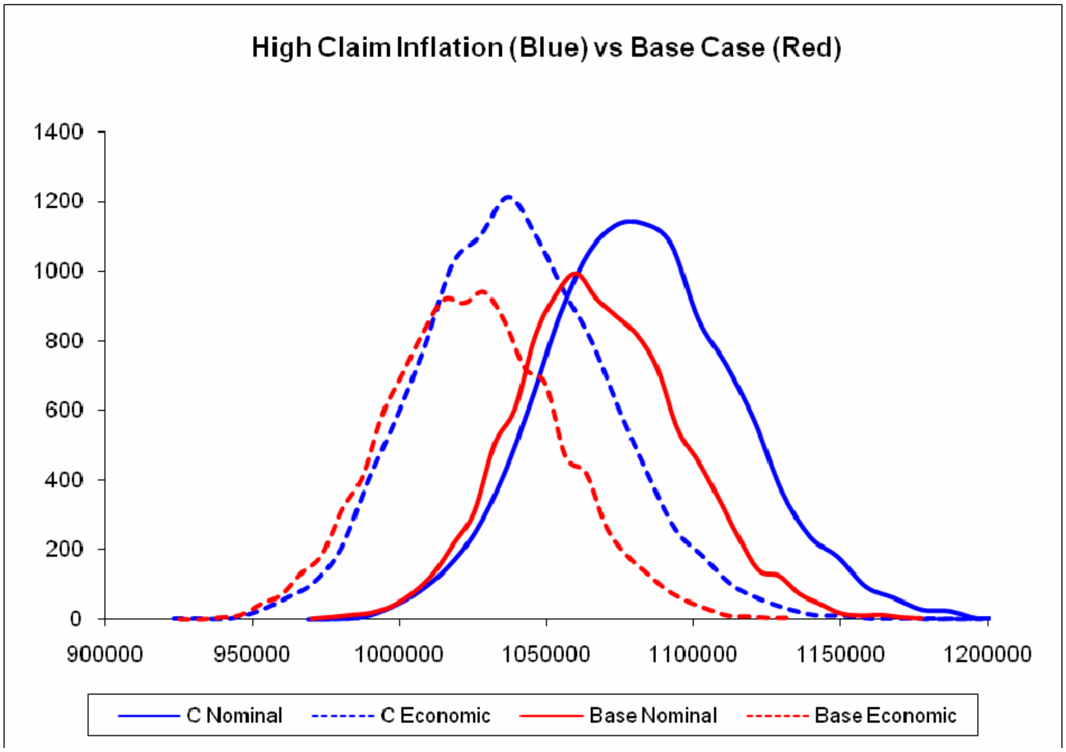


Figure 4

Case	Mean	Standard Deviation	Percentiles 5th	Percentiles 95th	90% Confidence Interval	Confidence Interval Ratio
C - nominal	1078069.12	33220.69	1025879.25	1134626.99	108747.74	96.89%
C - economic	1034635.49	32372.86	983266.87	1088635.84	105368.97	

Table 2

9.3 High Correlation between Inflation with Nominal Rates

Inflation and nominal interest rates moved in tandem during the 1970s, with correlation reaching 65% to 70%. Based on a 12-month inflation rate, the correlation with interest rates over the period 1970-2007 was 68%. High correlation between inflation and nominal interest rates reduces the range of economic loss reserves. For Case D the correlation factor was 70%. Figure 5 shows how this increase in correlation has little impact on the nominal values of loss reserves, but does reduce the distribution of economic values. In this case, the confidence interval ratio drops to 88%.

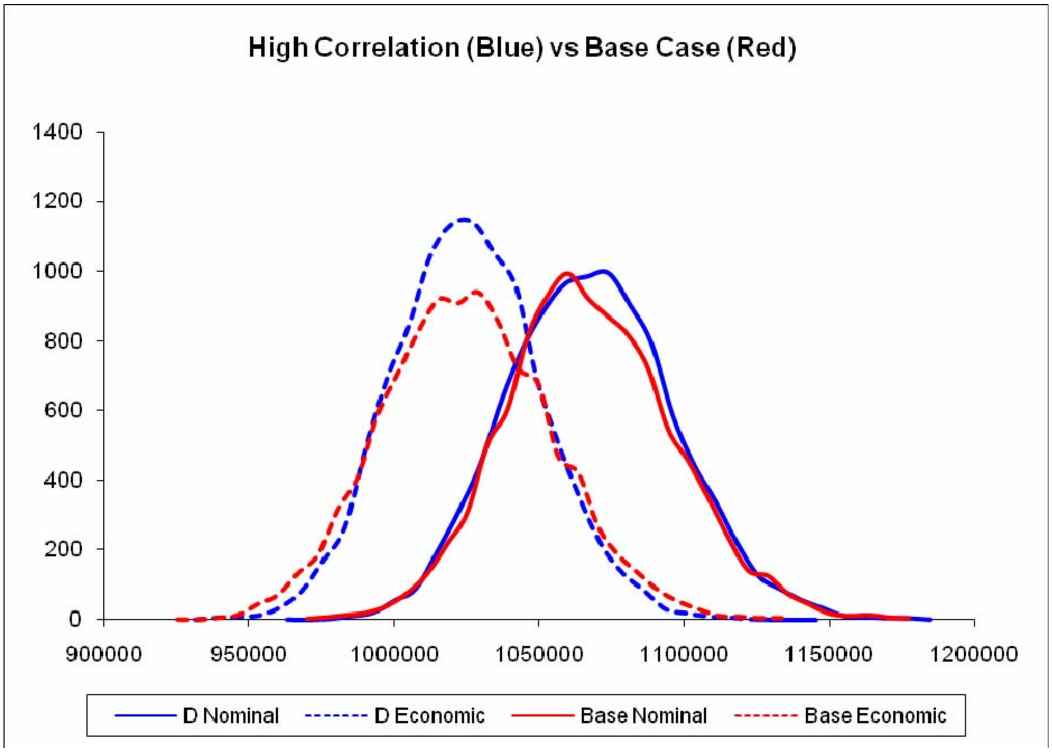


Figure 5

Case	Mean	Standard Deviation	Percentiles 5th	95th	90% Confidence Interval	Confidence Interval Ratio
D - nominal	1064157.75	28783.10	1018181.31	1112651.8	94470.49	88.40%
D - economic	1021531.44	25496.24	980743.24	1064253.59	83510.35	

Table 3

9.4 Periods of High and Volatile Inflation

In the situation of high and volatile inflation, such as in the 1970s, the problem of using nominal loss reserves to determine reserve ranges is exacerbated. For Case E, the current inflation rate is increased to 10% (from the base case 3.54%) and the inflation volatility is increased to 6% (from 1.9%). Figure 6 shows how this change increases the level and range of the distribution compared with the base case. Table 4 provides the key metrics for Case E; the confidence intervals are much wider and the confidence interval ratio is 88%.

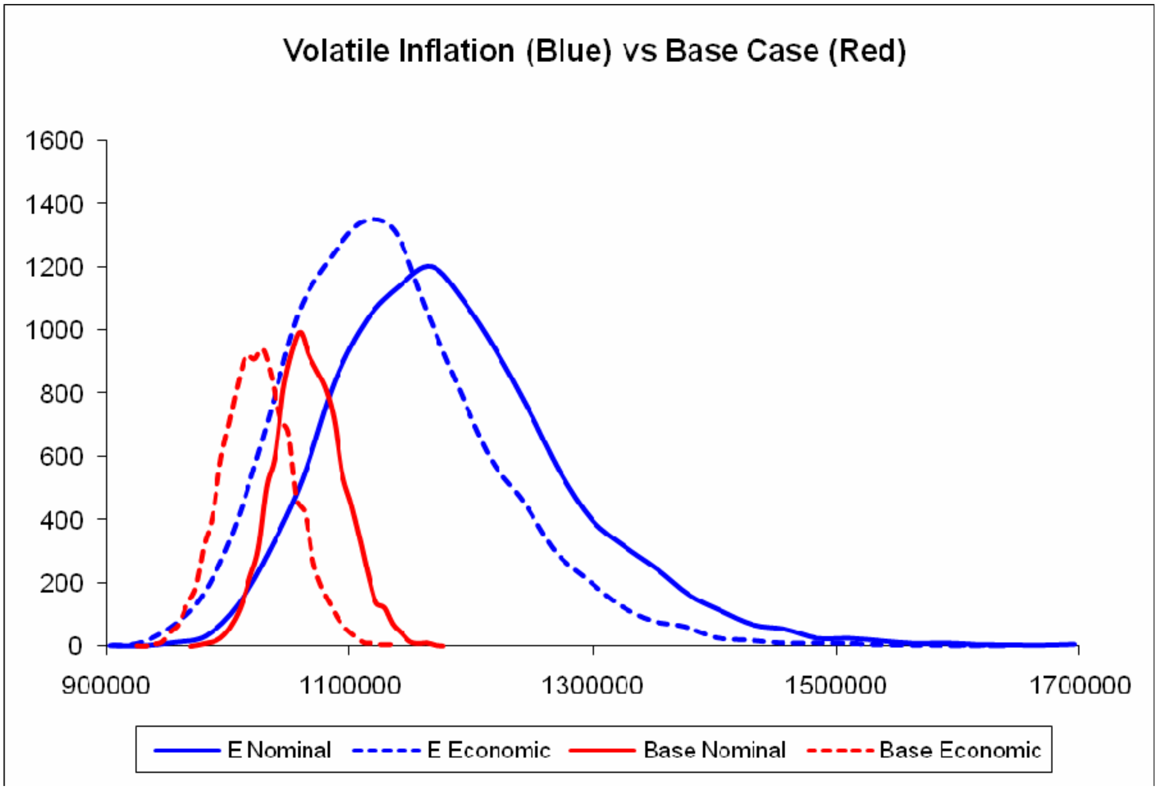


Figure 6

Case	Mean	Standard Deviation	Percentiles 5th	95th	90% Confidence Interval	Confidence Interval Ratio
E – nominal	1173532.64	97382.53	1036359.22	1351072.81	314713.59	87.50%
E – economic	1121600.19	85091.77	1000580.60	1275965.70	275385.10	

Table 4

9.5 Summary of Results

Based on the many simulations run for this research, the economic mean is smaller than the nominal mean. Under most circumstances, the economic value reserve ranges are slightly smaller than the nominal value ranges. This is not always the case under the fixed-claim model. The economic value range will be smaller than the nominal value ranges if claims cost inflation is very high relative to the CPI inflation, if correlation is high between the nominal interest rate and inflation, or if inflation becomes highly volatile.

9.6 Sensitivity Analysis

Sensitivity tests for all the parameters used were run to determine the impact of changes of each parameter. Case B was used as the base case, and each parameter was changed in turn over the ranges shown in Table 5. The results of a series of 5000 simulations of 1000 claims are summarized in the table below. For example, the first line of Table 5 indicates that changing the long run mean value for inflation over the range from 2% to 12% had no significant effect on the confidence interval range; in all cases, the economic value range was approximately 100% of the nominal value range. The next line indicates that changing the speed of mean reversion for the inflation rate over the range 0.1 to 0.3 increased the confidence interval range, in a linear manner, from 96% to 105%. Based on these results, the factors that have the most effect on the relationship between the confidence interval range of economic loss reserves and nominal loss reserves are the inflation volatility, the volatility of the short-term nominal interest rate, the correlation between interest rates and inflation, and the slope of the regression of claim costs against general inflation. These are the values that it is most important to measure accurately. A detailed discussion of the results for each parameter is provided in the appendix.

Model	**Parameter**	**From**	**To**	**Increment**	**Results (ratios)**	**90% CI Range**	**Type**
Inflation	Mean	2%	12%	2%	No effect	98-101%	N/A
Inflation	Speed	0.1	0.3	0.05	Increase	96-105%	Linear
Inflation	Vol	1%	8%	1%	Decrease	115-87%	Concave
Nominal	LT Mean	2%	10%	2%	No effect	99-103%	N/A
Nominal	LT Speed	0.02	0.10	0.02	No effect	100-102%	N/A
Nominal	LT Vol	0.6%	1.2%	0.2%	No effect	100-101%	N/A
Nominal	ST Speed	0.02	0.10	0.02	No effect	100-103%	N/A
Nominal	ST Vol	1%	5%	1%	Increase	96-141%	Linear
Fixed Claim	K	0.1	0.4	0.1	Increase	99-105%	Linear
Fixed Claim	M	0.3	0.8	0.1	Decrease	106-98%	Linear
Fixed Claim	N	0.5	2	0.5	No effect	99-102%	N/A
Loss	SD	200	1000	200	No effect	98-101%	Linear
Decay	Factor	0.2	0.8	0.1	Increase	92-102%	Convex
Correlation	Correlation	0%	100%	20%	Decrease	109-58%	Convex
Claim Cost	Slope	0.4	2.0	0.4	Decrease	127-94%	Linear

Table 5

10. CONCLUSION

Property-liability insurance companies have traditionally valued their loss reserves on a nominal basis due to statutory requirements. These requirements do not reflect the economic value of the future payments and distort insurance company financial statements. Nominal loss reserves overstate the impact of inflation on reserves, though only slightly under the current economic environment, as they ignore the relationship between inflation and nominal interest rates. The economic impact on loss reserves of a change in inflation is commonly offset by a similar shift in the nominal interest rate and by the high claims cost inflation. Loss reserve ranges based on nominal values accentuate this problem. Recent proposals advocate the use of fair value accounting for loss reserves, which would replace nominal values with economic values. In this study a loss reserve model was developed to quantify the uncertainty introduced by stochastic interest rates and inflation rates and to compare reserve ranges based on nominal and economic values. The results demonstrate a variety of scenarios under which the reserve ranges based on economic values can be either smaller or larger than the nominal value ranges. However, use of economic values for loss reserves would better serve the insurance industry and its regulators. The key reason for encouraging the use of economic value ranges is that they properly reflect the true measure of the uncertainty involved in loss reserving. An additional benefit is that the ranges are smaller in many circumstances, and the current economic environment seems to be moving toward those situations. Claim cost inflation and the level and volatility of inflation appear to have an upward trend. Economic value reserves would provide more credible values of the cost and uncertainty of future loss payments, and in the cases mentioned before, would have a smaller confidence interval range.

APPENDIX - SENSITIVITY ANALYSIS

For any stochastic model, the number simulations run in a study is an important determinant of the consistency and accuracy of the results. Running too few simulations can lead to widely varying results and erroneous conclusions. Too many simulations on the other hand waste time and computer resources. For simple models, statistical analysis can be used to determine the appropriate number of simulations to run. However, in this project, which consists of five separate stochastic models, that approach is not feasible. Instead, we ran the model multiple times for selected numbers of simulations and then calculated the variability of the 90% confidence interval ratio, the key variable used in this study. (This value is the ratio of the 90 percent confidence interval based on the economic value of loss reserves to the 90 percent confidence interval based on the nominal value of loss reserves.) For each number of simulations, the model was run eight times and the coefficient of variation of the confidence interval ratios was calculated. The optimal number of simulations was the point where the coefficient of variation did not continue to decline when additional simulations were run. The starting point was 1000 simulations, which generated a coefficient of variation of 2.83%. The number of simulations was increased, first to 2500, then 5,000 and 7,500. The coefficient of variation gradually declined to 1.12%. Running 10,000 simulations did not reduce the variability further, so this combination (10,000 simulations of 1000 claims) was used to run the individual cases (A through E) described in the paper.

Due to limitation in computational power, a smaller number of simulations were used to run the sensitivity analyses. As this required multiple runs for each variable over a range of feasible values, we used 5000 simulations and 1000 claims for this aspect of the project. Although the coefficient of variation of the 50% confidence interval ratios was slightly higher for this combination, at 1.54%, this was still sufficient to show the general effect of changing each parameter over the relevant range.

No. of Simulation	**50% C.I. Range Ratio CV**	**90% C.I. Range Ratio CV**	**90% C.I. Range Ratio**
1000	4.85%	2.83%	99.72% - 108.68%
2500	1.91%	1.90%	98.07% - 103.71%
5000	1.85%	1.61%	98.51% - 103.57%
7500	1.54%	1.12%	100.42% - 103.10%
10000	1.12%	1.15%	98.82% - 101.99%

Table 1 - A

Table 2-A shows the level of the variable changed and the corresponding 90% confidence interval ratio for each sensitivity test. The long-term mean inflation rate was varied from 2% to 12%, but in each case the confidence interval ratio remained approximately 100%. This value exhibited no trend over this range. Varying the speed of mean reversion from 0.1 to 0.3 did impact the confidence interval ratio in a systematic manner, although the effect was not large. The confidence interval ratio increased from 98% when the speed of mean reversion was 0.1, to 105% when the speed was increased to 0.3. As discussed in the body of the paper, the greatest impact occurred when the inflation volatility parameter changed. When this parameter was 1%, the confidence interval ratio was 115%. As the inflation volatility parameter increases, the confidence interval ratio declines at first but then remains at approximately 88% when inflation volatility is 4% or higher.

Changes in the long-term mean of the nominal interest rate (2%-10%), the speed of mean reversion of the long-term mean (0.02-0.10), the volatility of the long-term mean (0.6%-1.2%), and the speed of mean reversion of the short-term mean (0.02-0.10) had no consistent effect on the confidence interval ratio. However, increasing the volatility of the short-term mean interest rate over the range of 1% to 5% had a significant effect, opposite to the effect of increasing the volatility of the inflation rate. The confidence interval ratios increase as volatility increases.

Not much data are yet available to determine the appropriate parameters for the D'Arcy-Gorvett fixed claim model, but the results of the sensitivity tests are as expected. The higher the proportion of a claim that is fixed in value when the claim occurs (k), the higher the confidence interval ratio; the higher the proportion of the claim that is not fixed in value until the claim is settled (m), the lower the confidence interval ratio. The rate of fixing a claim's value (as n increases) had no consistent effect on the confidence interval ratio.

Losses were assumed to be log-normally distributed with a mean of 1000 and a standard deviation of 500. Increasing the standard deviation of each claim, over the range of 200 to 1000, had no consistent effect on the confidence interval ratio. Changing the decay factor representing what portion of unsettled claims were settled each year had a slight impact on the confidence interval range; a higher decay factor led to a higher confidence interval range. Changing the correlation between the inflation rate and the nominal interest rate from 0% to 100% had a significant impact on the confidence interval ratio. The higher the correlation, the lower the confidence interval ratio. Increasing the slope in the claim cost regression formula over the range of 0.4 to 2.0 also decreased the confidence interval ratio.

The purpose of the sensitivity analysis is to indicate which of the many parameters used in this model have the greatest impact on the results and the conclusions of this paper. In almost all cases, the conclusion that the use of economic values to determine loss reserves would lead to smaller reserve ranges is supported. Attention should be focused on measuring the parameters with the greatest impact on determining loss reserves and their ranges under either nominal or economic values. Thus, measures of interest rate and inflation volatility, the correlation between inflation and interest rates, and the relationship between claim costs and general inflation should be studied closely.

Inflation Model—Long-Term Mean

Test Value	2.00%	4.00%	6.00%	8.00%	10.00%	12.00%		
90% C.I. Ratio	101.24%	99.79%	98.62%	97.78%	98.80%	100.36%		

Inflation Model—Mean-Reversion Speed

Test Value	0.1	0.15	0.2	0.25	0.3			
90% C.I. Ratio	97.67%	95.76%	101.77%	101.53%	105.47%			

Inflation Model—Volatility

Test Value	1.00%	2.00%	3.00%	4.00%	5.00%	6.00%	7.00%	8.00%
90% C.I. Ratio	114.82%	100.53%	92.95%	89.74%	88.84%	87.87%	88.57%	87.36%

Nominal Interest Rate Model—Long-Term Mean

Test Value	2.00%	4.00%	6.00%	8.00%	10.00%
90% C.I. Ratio	99.35%	100.78%	103.06%	100.94%	99.63%

Nominal Interest Rate Model—Long-Term Mean-Reversion Speed

Test Value	0.02	0.04	0.06	0.08	0.10
90% C.I. Ratio	102.28%	100.74%	99.93%	100.96%	101.29%

Nominal Interest Rate Model—Long-Term Volatility

Test Value	0.60%	0.80%	1.00%	1.20%
90% C.I. Ratio	99.95%	100.88%	100.76%	100.88%

Nominal Interest Rate Model—Short-Term Mean-Reversion Speed

Test Value	0.02	0.04	0.06	0.08	0.1
90% C.I. Ratio	100.48%	102.70%	101.70%	99.78%	99.82%

Nominal Interest Rate Model—Short-Term Volatility

Test Value	1.00%	2.00%	3.00%	4.00%	5.00%
90% C.I. Ratio	96.45%	105.61%	117.30%	125.39%	140.72%

Fixed Claim Model—Fixed Portion at time 0 (*k*)

Test Value	0.1	0.2	0.3	0.4
90% C.I. Ratio	99.35%	100.45%	101.11%	105.10%

Fixed Claim Model—Portion unknown until settlement (*m*)

Test Value	0.3	0.4	0.5	0.6	0.7	0.8
90% C.I. Ratio	105.52%	106.42%	99.88%	98.00%	97.01%	97.73%

Fixed Claim Model—Speed of fixed settlement (*n*)

Test Value	0.5	1	1.5	2
90% C.I. Ratio	100.66%	101.93%	99.37%	100.81%

Loss Model—Standard Deviation

Test Value	200	400	600	800	1000
90% C.I. Ratio	101.12%	98.96%	98.20%	99.17%	99.07%

Decay Model—Annual Decay Factor

Test Value	0.2	0.3	0.4	0.5	0.6	0.7	0.8
90% C.I. Ratio	92.47%	96.10%	98.73%	99.71%	101.48%	101.64%	101.44%

Correlation—Correlation between Inf and Int. Rate

Test Value	0%	20%	40%	60%	80%	100%
90% C.I. Ratio	108.83%	107.28%	102.77%	95.08%	81.15%	58.15%

Claim Cost Regression—Slope

Test Value	0.40	0.80	1.20	1.60	2.00
90% C.I. Ratio	126.52%	116.09%	109.97%	101.04%	94.42%

Table 2-A

11. REFERENCES

[1] Actuarial Standards Board, *Actuarial Standard of Practice No. 20*, "Discounting of Property and Casualty Loss and Loss Adjustment Expense Reserves," http://www.actuarialstandardsboard.org/pdf/asops/asop020_037.pdf

[2] Ahlgrim, Kevin, Stephen P. D'Arcy, and Richard W. Gorvett, "The Effective Duration and Convexity of Liabilities for Property-Liability Insurers Under Stochastic Interest Rates," *Geneva Papers on Risk and Insurance Theory*, 2004, Vol. 29, No. 1, pp. 75-108.

[3] Ahlgrim, Kevin, Stephen P. D'Arcy, and Richard W. Gorvett, "Modeling Financial Scenarios: A Framework for the Actuarial Profession," *Proceedings of the Casualty Actuarial Society*, Vol. XCII, 2005, pp. 60-98.

[4] A. M. Best, *Best's Aggregates & Averages, Property-Casualty*, selected years.

[5] American Academy of Actuaries, "Fair Valuation of Insurance Liabilities: Principles and Methods," 2002, http://www.actuary.org/pdf/finreport/fairval_sept02.pdf.

[6] Australian Prudential Regulation Authority, *GPS 310*, "Audit and Actuarial Reporting and Valuation," February 2006, http://www.apra.gov.au/General/upload/GPS-310-Audit-and-Actuarial-Valuation.pdf

[7] Barnett, Glen and Ben Zehnwirth, "Best Estimates for Reserves," *Casualty Actuarial Society Forum*, Fall 1998, pp. 1-54.

[8] Berquist, James R. and Richard E. Sherman, "Loss Reserve Adequacy. Testing: A Comprehensive, Systematic Approach," *Proceedings of the Casualty Actuarial Society*, Vol. LXIV, 1977, pp. 123-184.

[9] Butsic, Robert P., "The Effect of Inflation on Losses and Premiums for Property-Liability Insurers," *Casualty Actuarial Society Discussion Paper Program*, 1981, pp. 58-102.

[10] Butsic, Robert P., "Determining the Proper Interest Rate for Loss Reserve Discounting: An Economic Approach," *Casualty Actuarial Society Discussion Paper Program*, Fall 1988, pp. 147-186.

[11] Casualty Actuarial Society, "Fair Value of P&C Liabilities: Practical Implications," 2004, http://www.casact.org/pubs/fairvalue/FairValueBook.pdf.

[12] Casualty Actuarial Society, "Task Force on Actuarial Credibility," 2005, http://www.casact.org/about/reports/tfacrpt.pdf.

[13] CAS Working Party on Quantifying Variability in Reserve Estimates, *Casualty Actuarial Forum*, Fall 2005, pp. 29-146.

[14] CEA Information Papers, "Solvency II: Understanding the process," 2007, http://www.cea.assur.org/cea/download/publ/article257.pdf.

[15] D'Arcy, Stephen P., "Revisions in Loss Reserving Techniques Necessary to Discount Property Liability Loss Reserves" *Proceedings of the Casualty Actuarial Society*, Vol. 74, 1987, pp. 75 100.

[16] D'Arcy, Stephen and Richard W. Gorvett, "Measuring the Interest Rate Sensitivity of Loss Reserves," *Proceedings of the Casualty Actuarial Society*, Vol. 87, 2000, pp. 365-400. http://www.casact.org/pubs/proceed/proceed00/00365.pdf.

[17] Federal Reserve Economic Data, http://research.stlouisfed.org/fred2/.

[18] Fisher, Irving, "The Relation of Interest to Money and Prices," Chapter 19 in *The Theory of Interest* (New York: Macmillan, 1930).

[19] Girard, Luke N., "An Approach to Fair Valuation of Insurance Liabilities Using the Firm's Cost of Capital," *North American Actuarial Journal*, 6:2, 2002, pp. 18–46.

[20] Hettinger, Thomas, "What Reserve Ranges Makes You Comfortable," *Midwestern Actuarial Forum Fall Meeting*, 2006, http://www.casact.org/affiliates/maf/0906/hettinger.pdf.

[21] Hilsenrath, Jon, "When Nerves Get Short, Credit Gets Tight," *Wall Street Journal*, February 19, 2008, p. C 1.

[22] Institute of Actuaries of Australia, *Professional Standard 300*, "Valuations of General Insurance Claims," August 2007, http://www.actuaries.asn.au/NR/rdonlyres/0A2FF56A-1489-404F-BBF3-CB7D916B269E/2908/PS300ApprovedbyCouncil28August2008.pdf

[23] Kirschner, Gerald S., Colin Kerley, and Belinda Isaacs, "Two Approaches to Calculating Correlated Reserve Indications Across Multiple Lines of Business," *Casualty Actuarial Society Forum*, Fall 2002, pp. 211-246.

[24] Mack, Thomas, "Distribution-free Calculation of the Standard Error of Chain Ladder Reserve Estimates," *AST1N Bulletin*, 23:2, 1993, pp. 213-225.

[25] Masterson, Norton E., "Property-Casualty Insurance Inflation Indexes: Communicating with the Public," *Casualty Actuarial Society Discussion Paper Program*, 1981, pp. 344-370.

[26] Masterson, Norton E., "Economic Factors in Property/Casualty Insurance Claims Costs," *Best's Review*, June 1987, pp. 50-52.

[27] McClenahan, Charles L., "Estimation and Application of Ranges of Reasonable Estimates," *Casualty Actuarial Society Forum*, Fall 2003, pp. 213-230.

[28] Miller, Mary Frances, "From the President: Are You Part of the Solution?" *Actuarial Review*, February 2004, vol. 31, no. 1, http://www.casact.org/pubs/actrev/feb04/pres.htm.

[29] Murphy, Daniel M., "Unbiased Loss Development Factors," *Proceedings of the Casualty Actuarial Society LXXXI*, 1994, pp. 154-222.

[30] Narayan, Prakash and Thomas V. Warthen, "A Comparative Study of the Performance of Loss Reserving Methods Through Simulation," *Casualty Actuarial Society Forum*, Summer 1997, pp. 175-196.

[31] Patel, Chandu C., and Alfred Raws III, "Statistical Modeling Techniques for Reserve Ranges: A Simulation Approach," *Casualty Actuarial Society Forum*, Fall 1998, pp. 229-255.

[32] Pecora, Jeremy P., "Setting the Pace," *Best's Review*, January 2005, pp. 87-88.

[33] Richards, William F., "Evaluating the Impact of Inflation on Loss Reserves," *Casualty Actuarial Society Discussion Paper Program*, Fall 1981, pp. 384-400.

[34] Sarte, Pierre-Daniel G., "Fisher's Equation and the Inflation Risk Premium in a Simple Endowment Economy," *Economic Quarterly*, Vol. 84, Fall 1998, pp. 53-72

[35] Shapland, Mark R., "Loss Reserve Estimates: A Statistical Approach for Determining 'Reasonableness,'" *Casualty Actuarial Society Forum*, Fall 2003, pp. 321-360.

[36] Standard & Poor's, "Insurance Actuaries: A Crisis of Credibility," 2003.

[37] Taylor, Greg C., "Separation of Inflation and Other Effects from the Distribution of Non-Life Insurance Claim Delays," *ASTIN Bulletin*, 9:1-2, 1977, pp. 219-230.

[38] Van Ark, William R., "Gap in Claims Cost Trends Continues to Narrow," *Best's Review*, March 1996, pp. 22-23.

[39] Venter, Gary G., "Refining Reserve Runoff Ranges," *ASTIN Colloquium*, Orlando, Florida, 2007.

[40] Wiser, Ronald F., Jo Ellen Cockley, and Andrea Gardner, "Loss Reserving," Chapter 5 in *Foundations of Casualty Actuarial Science* (Arlington, Va.: Casualty Actuarial Society, 2001) pp. 197-285.

[41] Zehnwirth, Ben, "Probabilistic Development Factor Models with Applications to Loss Reserve Variability, Prediction Intervals, and Risk Based Capital," *Casualty Actuarial Society Forum*, Spring 1994, pp. 447-606.

Biographies of the authors

Stephen P. D'Arcy is a Professor of Finance and the John C. Brogan Faculty Scholar in Risk Management and Insurance at the University of Illinois at Urbana-Champaign. He is a Fellow of the Casualty Actuarial Society, a member of the American Academy of Actuaries, Past-President of the American Risk and Insurance Association and Past-President of the Casualty Actuarial Society. He received his B.A. in Applied Mathematics from Harvard College and his Ph.D. in Finance from the University of Illinois. The courses he teaches include an introduction to insurance, property-liability insurance, casualty actuarial mathematics, advanced corporate finance, managing financial risk for insurers, and enterprise risk management. He has taught a seminar on finance and on-line courses on financial risk management and enterprise risk management for the Casualty Actuarial Society.

Steve has won the University of Illinois Campus Award for Excellence in Undergraduate Teaching, the Commerce Council Award for Best Professor in a large class and in a small class, the CBA Alumni Association Award for Excellence in Undergraduate Teaching, twice won the Dorweiler Award for one of the best papers published in the *Proceedings of the Casualty Actuarial Society*, twice shared the *Journal of Risk and Insurance* Award for one of the best papers published in that journal, twice won awards for best papers at Casualty Actuarial Society meetings and won the first American Risk and Insurance Association Innovation in Instruction Award. He recently won the CAS-ARIA Award for best paper published by the American Risk and Insurance Association on casualty actuarial topics. He has been a University Scholar for the University of Illinois and a Faculty Fellow for the National Center of Supercomputing Applications working on predictive modeling in insurance.

Prior to his academic career, he worked as an actuarial student at Aetna Insurance Company and as actuary at CUMIS Insurance Society. He served on the Governor's Task Force on Medical Malpractice in Illinois. His research interests include insurance fraud, financial modeling, dynamic financial analysis, financial pricing models for property-liability insurers, catastrophe insurance futures, public pension funding, and insurance regulation.

Alfred Au is an undergraduate student at the University of Illinois at Urbana-Champaign majoring in Actuarial Science and graduating in May 2008. He is an active candidate for actuarial exams. He has held actuarial internships at Hewitt Associates and HSBC Insurance (Asia Pacific). Alfred is interested in investments and financial modeling.

Liang Zhang is an undergraduate student at the University of Illinois at Urbana-Champaign majoring in Actuarial Science and graduating in May 2008. He is an active candidate for actuarial exams. Liang is interested in investments and financial derivatives.

A Survival Model Approach to Non-Life Run-off Triangle Estimation

Brian A. Fannin, ACAS

Abstract

Motivation. Most standard loss reserving techniques do not explicitly consider the rate at which claims will close, or the expected amount of time that a claim will remain open. Consideration of the time until closure allows one to calculate the amount of time until a block of claims will run-off. Further, it allows one to take explicit assumptions with regard to interest and inflation into account.

Method. By observing the closure rates for claims by age, a survival function is produced. This function can be used to determine the future lifetime of a claim at any age and the number of claims remaining open at any time.

Results. The method applied to a set of sample data generates a complete picture of the future pattern of claim disposal.

Conclusions. The method presented here grounds the projection of future claim run-off in theory common to life actuaries and opens up the life toolset to the analysis of non-life data.

Keywords. Reserving; Survival models.

1. INTRODUCTION

This paper will present a method to estimate the length of time that claims for a book of non-life insurance will remain open. This method is based largely on theory common to life actuaries but rarely used in the non-life field. This technique requires data presented in a manner slightly different than that with which non-life actuaries are accustomed to working. Nevertheless, the concepts are straightforward and intuitive and the data storage and computation requirements are not onerous. Coupled with a standard technique to forecast the emergence of unreported claims, an estimate for the time required for the complete run-off of a portfolio is produced.

1.1 Objective

Loss development techniques have traditionally sought to produce an estimate of the total quantum of losses remaining to be paid, e.g., the total reserve position. More recently, attention has been directed to consideration of the variance around both that estimate and the actual realized value of payments. Timing of payments to be made or a statement about the total amount of time to run-off a reserve is not always considered. When calculating the transfer price of a block of non-life (re)insurance liabilities, or calculating the amount of capital required to support the run-off, this is highly relevant. In those cases where a discounted value of reserves is needed, the standard approach is to take the results of an analysis of a paid loss triangle. Doing so doesn't allow one to directly observe or consider the manner in which a typical claim settles.

1.2 Outline

The paper proceeds as follows: First, the survival model is briefly reviewed. Next, the data required for the technique is described and the estimation of the survival model parameters is outlined. Results for a set of test data are shown and discussed. The method is then compared with several well-known methods. We conclude by addressing several unresolved issues and also by discussing some of the applications of this technique.

2. BACKGROUND AND METHODS

2.1 Review of Survival Model Mathematics

2.1.1 The Survival Function

A survival function, denoted $S(x)$, measures the probability that a value will be greater than or equal to some threshold x. When so stated, the function is easily seen to be equal to 1 minus the cumulative probability function or

$$S(x) = 1 - F(x) \tag{2.1}$$

This function follows the normal rules associated with probability distribution functions, with the additional requirement that x be non-negative. So, $S(0) = 1$ and $S(\infty) = 0$, or "the probability that a life will survive past age 0 is 1 and the probability that a life will survive to age infinity is zero." The terms "life" and "age" need not refer to an actual life, be it human or otherwise, but may refer to anything that has a well-defined temporal start and end point.

Although $S(x)$ can be defined continuously, the function is often given in a discrete form, using integral values for the age x. For convenience, the notation ${}_a p_x$, where x represents a starting age and a represents some future time period is often used. This can be read as "the probability that a life aged x survives for an additional time period a." Note that this probability is conditional on having attained age x. Mathematically, this is stated as follows:

$${}_a p_x = \frac{S(a+x)}{S(x)} = \frac{1-F(a+x)}{1-F(x)} = \frac{\int_{a+x}^{\infty} f(y)dy}{\int_{x}^{\infty} f(y)dy} \tag{2.2}$$

The notation ${}_aq_x$ represents the probability that a life aged x will terminate within time period a. This is the logical complement of the probability implied by ${}_ap_x$ (a life must either survive or terminate within a stated period of time) and is therefore equal to

$$ {}_aq_x = 1 - {}_ap_x. \tag{2.3}$$

When a is omitted a time period of one year is assumed. Given a set of factors for ages x through $x+a$-1, one can calculate the probability of survival for any duration a by multiplying successive factors as follows:

$$ {}_a p_x = p_x p_{x+1} \cdots p_{x+a-1} = \prod_{i=0}^{a-1} p_{x+i} \tag{2.4}$$

We use the random variable $K(x)$ to describe the future lifetime for a life aged x. Its expectation and variance for discrete probabilities are as follows:

$$ E[K(x)] = \sum_{i=0}^{\infty} {}_{i+1} p_x \tag{2.5}$$

$$ E[K^2(x)] = \sum_{i=0}^{\infty} (2i+1)_{i+1} p_x \tag{2.6}$$

$$ Var(K(x)) = E[K^2(x)] - E[K(x)]^2 = \sum_{i=0}^{\infty} (2i+1)_{i+1} p_x - \left(\sum_{i=0}^{\infty} {}_{i+1} p_x\right)^2 \tag{2.7}$$

The derivation of the above formulae can be found in London [8] or Bowers [4].

2.1.2 Estimation of the Survival Function

London describes two different types of studies that may be performed to estimate a survival function. A longitudinal study examines a cohort of lives from age zero until the time of death. A cross-sectional study examines a group of lives of various ages for a fixed period of time.

For reasons that will be made clear below, the focus of this paper is a cross-sectional study. Here, a set of lives are observed between two points in time. For each life, the quantities y_i and z_i represent the age at which the observation period begins and ends, respectively. Note that the life may not survive until age z_i. For an age interval $(x, x+1]$, the quantities $x+r_i$ and $x+s_i$ are defined as the ages at which life i is scheduled to enter and exit that age interval. For example, if one observed a group of lives between 1 January 2004 and 31 December 2005, a life with birth date 16 February 1972 would have the following values:

$y_i = 31.87$
$z_i = 33.87$

For age interval (31, 32)	For age interval (32, 33)	For age interval (33, 34)
$x+r_i = 31.87$	$x+r_i = 32.00$	$x+r_i = 33.00$
$x+s_i = 32.00$	$x+s_i = 33.00$	$x+s_i = 33.87$

For each age interval, the probability of death within one year, q_x is estimated using the following estimator:

$$\hat{q}_x = \frac{d_x}{\sum_{i=1}^{n}(s_i - r_i)} \tag{2.6}$$

where n represents the number of observed lives and d_x represents the number of observed deaths. London shows that this estimator can be derived using the method of moments or maximum likelihood.

In cases where the exact age is not known, s_i and r_i are taken to be 1 and 0, respectively. In this case, the estimate of q_x is simply equal to:

$$\hat{q}_x = {}^{d_x}\!/_{n_x} \tag{2.7}$$

The estimated survival function is constructed by combining equations 2.3 and 2.4 as follows:

$$\hat{S}(x) = \prod_{i=0}^{x-1} \hat{p}_x = \hat{p}_0 \hat{p}_1 \ldots \hat{p}_{x-1} = (1-\hat{q}_0)(1-\hat{q}_1)\ldots(1-\hat{q}_{x-1}) \tag{2.8}$$

London makes two assumptions.

(1) Each $\hat{p}_x$ is binomially distributed with mean p_x and variance $\dfrac{p_x q_x}{n_x}$.

(2) The $\hat{p}_x$ s are independent.

These two assumptions allow us to state the following about the sample estimate of the expected future lifetime:

$$E\left[\hat{K}(x)\right] = E\left[\sum_0^{\infty} {}_{i+1}\hat{p}_x\right] = \sum_0^{\infty} E\left[{}_{i+1}\hat{p}_x\right] = \sum_0^{\infty} {}_{i+1}p_x = E\left[K(x)\right], \tag{2.9}$$

$$E\left[\hat{K}^2(x)\right] = E\left[\sum_0^{\infty}(2i+1)_{i+1}\hat{p}_x\right] = \sum_0^{\infty}(2i+1)E\left[{}_{i+1}\hat{p}_x\right] = \sum_0^{\infty}(2i+1)_{i+1}p_x = E\left[K^2(x)\right], \tag{2.10}$$

$$Var(\hat{K}(x)) = E\left[\hat{K}^2(x)\right] - \left(E\left[\hat{K}(x)\right]\right)^2 = E\left[K^2(x)\right] - \left(E\left[K(x)\right]\right)^2 = Var(K(x)). \tag{2.11}$$

In other words, the sample estimate of expected future lifetime is an unbiased estimate whose variance is independent of sample size.

A number of other techniques exist to estimate the survival function. Their implementation and appropriateness will not be explored in this paper.

2.2 Survival Model Methods as Applied to Non-Life Run-Off

A claim may be regarded as analogous to a life. It begins and ends at a fixed point in time. Its future remaining lifetime at any point is a random variable. A group of homogenous claims will likely exhibit similar survival patterns in the same way that humans with common characteristics will exhibit similar mortality. In the same way that human lifespans change over time, due to any number of factors such as nutrition, environment, changes in lifestyle, or advances in medicine, claim survival patterns may also change over time. A number of factors may influence non-life survival characteristics: claim department practice, legislative changes, behavior of insureds or cedants, to name but a few.

In general, a claim cannot be observed from time zero, the date of accident. There is generally a lag between when a claim occurs and when the claim is reported to a (re)insurer. This lag will vary depending on the characteristics of the claim and the type of coverage. First-party primary claims will be reported more quickly than third-party excess claims. This means that a claim may already be several years old when it can first be observed. For this reason, claim survival functions can only be estimated using the cross-sectional study described above.

2.2.1 The Data

The data was taken from a transactional database, which showed a history of claim payments made in each year, the date the claim occurred, and the status of the claim. Here, we define age as the calendar year minus the accident (or underwriting or reporting) year, plus one.[1] Note that when the age is so defined, the values for r_i and s_i in equation (2.6) are 0 and 1, respectively.

If no payment is made, a record is still kept to indicate that the claim remains open. This allows one to determine whether or not the claim will remain open in the following year.[2] So, for each payment year, for each age, one can calculate the total number of claims open as well as the number of claims that will terminate in the following year. The figures were summed for all payment years. The results are shown in Appendix A.

Note that this data may also be presented in a triangle format. The resultant triangles would be the number of claims open and the incremental number of claims closed during the period. These are shown as Appendices C and D.

2.2.2 The Method

With the data so arranged, the calculations proceed simply. Refer to n_x as the number of claims of age x and d_x as the number of claims of age x which will close. An estimator for the probability of claim closure for each age is given by formula 2.7.

Note that if the data is given as a triangle, summing across payment years is equivalent to taking the sum of the accident year rows.

At this point, a model has been developed for the expected value of the future life of all claims that have been reported. To forecast the emergence of new claims, a standard chain-ladder technique can be used. This will yield projections of the number of claims with respective ages for all future time periods. The same survival function can be applied to this set of IBNR claims.

The future lifetime for the book is equal to the maximum of the future lifetimes for all claims. As will be seen below, this is not necessarily the same as the expected future lifetime for the youngest claim present in the sample. The expected future lifetime is a quantity which depends on attained age. For non-life claims, it is often true that the longer a claim has been open, the longer it can be expected to remain open.

---

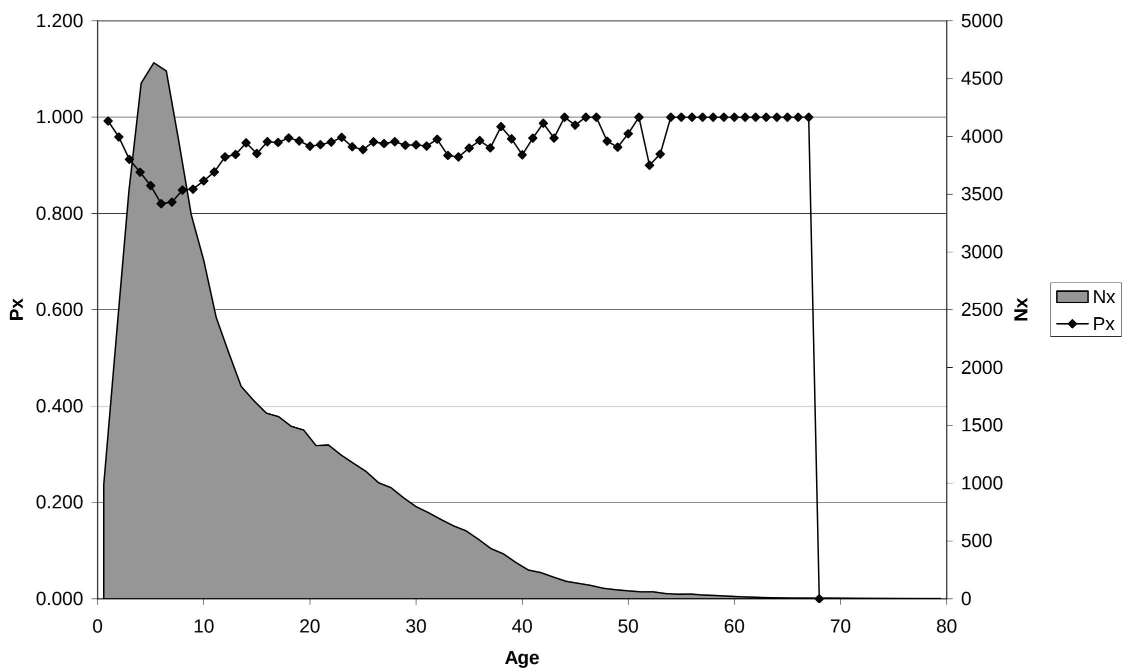
[1] Note that some authors refer to this quantity as "lag."
[2] Note that for the most recent year, it is impossible to determine whether or not a claim will terminate.

3. RESULTS AND DISCUSSION

3.1 Results

The forecast method was applied to a set of data that includes, among other things, excess bodily injury claims. The data has been randomly altered to conceal its identity, however the broad conclusions remain. Claim payment records for over 40 years were available. The earliest payment year includes information on currently open claims, so the oldest potential age can be, and indeed is, greater than 40 years. The oldest age in the sample was 68 years.

The chart below plots p_x against the age of a claim. The shaded surface shows n_x.



The likelihood of a claim persisting for an additional year drops for claims of low age, but then raises to a relatively high and constant survival probability beyond 14 years. Claims older than 14 years are very likely permanent bodily injury claims that will last as long as the claimant remains alive.

The fluctuation in probabilities beginning around age 38 is due to a reduction in sample data. Specifically, the number of observed claim closures drops below 10 at this age and is zero for some ages. This is a worrisome result. In effect, what it means is that one cannot truly know what's happening to claims that have been open for a very long time. That is to say that the likelihood that a very old claim will close in any given period of time is not easily estimable via statistical methods given this particular sample data.

This is a situation with which casualty actuaries are familiar. When reserving, one usually has the problem of how to estimate a tail factor. There are a number of techniques discussed in the non-life literature as to how to go about this. When revising a mortality table, one not only adjusts and extrapolates the estimates for high ages—the "tail" of the table—one smoothes the estimates for all ages. This revision of sample estimates is referred to as "graduation." London [9] gives a useful introduction to several graduation methods used by life actuaries. Contrast this with the typical non-life approach where age-to-age and tail factors are each calculated and judgmentally adjusted individually.

In this case, the sample estimate was adjusted by using the Whitaker-Henderson method of graduation. This technique can be considered ad hoc. The intent is to produce a revised set of estimates that represents a blend between smoothness and reproduction of the sample estimates. To do this, one minimizes the sum of the differences between the estimates and the squared difference between the sample estimates and revised values. A parameter ε controls the relative weight one places on smoothness and reproduction of the sample.

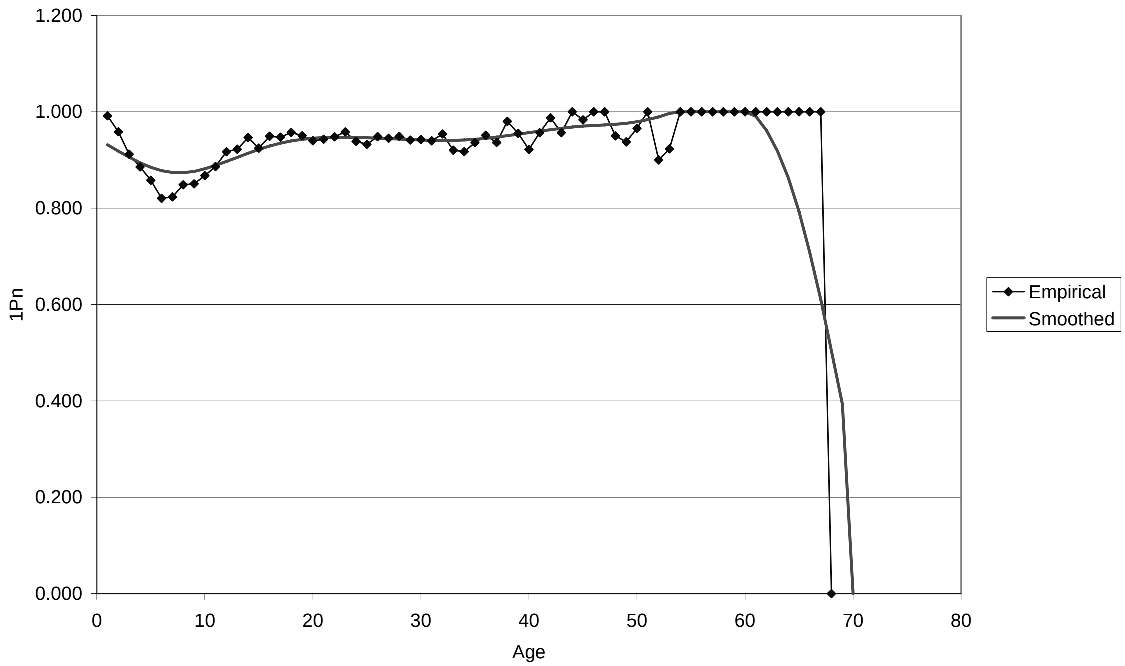
This quantity is given as follows:

$$\sum_{x=0}^{\max-3}(\Delta^3 p'_x)^2 + \varepsilon\sum_{x=0}^{\max}\left(p'_x - \hat{p}_x\right)^2 \tag{3.1}$$

where p' is the revised estimate.

In addition to smoothing the results, the technique can also be used to extrapolate beyond the maximum age in the sample. In this case, 70 was selected as the maximum feasible age of a claim. A claim age of 70 would imply a claimant age of 70 plus the age of the claimant at the time of injury. This is well within a reasonable maximum for a human life for claimants of a very young age, but not claimants who make a claim later. However, it is possible that beneficiaries or a claimant's estate may also receive claim payments. Further, the claimant may be a corporation or some other nonhuman entity. In either case, the age of a claim could be greater than the feasible length of one human life.

The chart below shows the results of the smoothing.



This is but one option. One could also extrapolate the survival model or construct a parametric survival model. An alternative would be to rely on a human mortality table. This could be appropriate for lifetime bodily injury cases. Diss and Sherman [5] use this method to estimate a tail factor for workers compensation business. (Note that there has been some research into differences between the mortality for bodily injury claimants and the general population. In particular, see Barnett [2] and Gillam [7].)

Yet another alternative would be to replace the sample estimates with judgmentally derived survival probabilities, possibly determined in conjunction with the claims department. Note that when speaking with non-actuaries it is likely far easier to pose the question "What is the likelihood that a claim that has been open for 40 years will stay open for another year?" than to ask "Is a tail factor of 1.025 at development year 40 reasonable?"

The smoothed results are given in Appendix B, along with an estimate of the expected future lifetime for each age. For this sample, the expected future lifetime is 20.3; it should take at least another 20 years for the business to completely run-off. However, there is a possibility that it will take quite a while longer. Assuming a normal approximation and using the standard deviation as given in formula 2.7, there is a 5% chance that this book could take 46.1 years to completely close—a difference of over 25 years.

3.2 Comparison with Other Techniques

Observation of the rate at which non-life claims close is not a new idea. Although claim count information is used less often, several well-known papers present techniques for handling this data. Following are comments on how this method compares with others.

Both Adler and Kline [1] and Berquist and Sherman [3] discuss a claims closure ratio. This ratio is defined as the number of claims closed in a given development period divided by the number of ultimate claims. Thus, the claims closure ratio depends on an estimate of the ultimate number of claims having already been made. In both cases, they presume that the future ratio of closed claims to ultimate for any development period will be the same as the most recent calendar year. Adler and Kline presume that the rate of claim closure is a stable figure that depends on the amount of claims remaining to be closed for a particular accident year. Note that this is different than what is presumed here. Here it is assumed that the survival of a particular claim is determined by the characteristics of the claim itself. The implication of both Adler and Kline and Berquist and Sherman is that a claim has an expected lifetime, which is more or less fixed, and that its time of settlement can change only because of the workflow characteristics of the claim department.

Fisher and Lange [6] describe a claims disposal ratio. Here too, they calculate this as the ratio of claims disposed of in any particular year to the total number of claims. Because they are working with report year data, the number of claims is known for each year and need not be estimated.

Teng and Sherman [10] present a reserving technique that utilizes an estimate of claims closure ratios similar to what is presented here. The closure ratio is calculated as the number of claims closed in any particular period to the number of claims reported up to the beginning of that period. Because closed claims will always remain in the population of claims reported to date, this quantity is not the same as the probability that a claim will terminate given that it survives to a particular development period. In fact, what Teng and Sherman are estimating is $1 - S(x)$. Because $S(x)$ depends on the individual p_xs, one could argue that the method presented here may be more appropriate given that it develops a specific estimate of the survival probability at each age.

3.3 Enhancements

There are a number of ways that this technique could be enhanced. At present, no distinction is made as to the way in which a claim is closed. If one were aware of certain effects, such as an active commutation strategy, or the influence of particular cedants or insureds, those claims could be removed from the sample.

The Whitaker-Henderson method is but one option for graduation of an empirical survival model. One could also apply the standard battery of smoothing and trending methods. As noted earlier, the use of a survival model does not obviate the need to select a tail factor. However, unlike some techniques applied to triangle data, a survival model requires the actuary to posit an upper bound for the length of time that a claim will remain open and state the likelihood of attaining that age.

It is commonly accepted that claim closure patterns change over time due to any number of influences. Less common is an objective method to forecast those changes. Life actuaries do attempt to project mortality trends into the future. Adoption of those techniques may help shed light on the dynamics of non-life claim behavior.

In order to convert this method into one for which an estimate of reserves could be calculated, as estimate of the size of prospective payments would have to be incorporated. Among the advantages of constructing the reserving model in this way are that one can integrate estimates for future inflation explicitly. This is effectively what Teng and Sherman have done.

4. CONCLUSIONS

The use of survival models, though understood in principle, is not common to non-life actuaries. The ability to examine data in this way opens up a number of interesting possibilities, including the use of techniques developed in the fields of population growth and demography. In the view of this author, equally important is a philosophical shift away from triangulated data towards a more fundamental consideration of the dynamics of the claim. Every actuary appreciates that the dynamics of claim generation and settlement are complex and change over time, but the methods currently available for the analysis of aggregate claim triangles do not easily lend themselves to taking these forces into account. It is hoped that this approach will serve as a step towards changing that.

Acknowledgment

The author gratefully acknowledges Dave Clark of Munich Re America who provided some very useful comments on an earlier (much different) draft of this paper. The author would also like to thank Dr. Frank Schiller for his clarification of the graduation of mortality tables and comments on mortality models in general. Ana Mata and Susan Macaulay provided very thorough and worthwhile comments. Their assistance is much appreciated. Errors, omissions and anything ridiculous are, sadly, the author's sole responsibility.

Appendix A

Age	*N*	*D*	*qx*	*px*	*S*(*x*)
1	985	8	0.008	0.992	0.992
2	2270	94	0.041	0.959	0.951
3	3522	309	0.088	0.912	0.867
4	4461	510	0.114	0.886	0.768
5	4636	659	0.142	0.858	0.659
6	4566	820	0.180	0.820	0.541
7	3958	699	0.177	0.823	0.445
8	3322	503	0.151	0.849	0.378
9	2927	438	0.150	0.850	0.321
10	2433	322	0.132	0.868	0.279
11	2126	242	0.114	0.886	0.247
12	1838	152	0.083	0.917	0.227
13	1715	133	0.078	0.922	0.209
14	1607	86	0.054	0.946	0.198
15	1575	119	0.076	0.924	0.183
16	1492	76	0.051	0.949	0.174
17	1459	77	0.053	0.947	0.164
18	1325	57	0.043	0.957	0.157
19	1331	66	0.050	0.950	0.150
20	1243	75	0.060	0.940	0.141
21	1172	67	0.057	0.943	0.132
22	1101	57	0.052	0.948	0.126
23	1003	42	0.042	0.958	0.120
24	961	59	0.061	0.939	0.113
25	875	59	0.067	0.933	0.105
26	796	41	0.052	0.948	0.100
27	744	41	0.055	0.945	0.094
28	686	35	0.051	0.949	0.090
29	631	37	0.059	0.941	0.084
30	588	34	0.058	0.942	0.079
31	515	31	0.060	0.940	0.075
32	435	20	0.046	0.954	0.071
33	389	31	0.080	0.920	0.066
34	314	26	0.083	0.917	0.060
35	249	16	0.064	0.936	0.056
36	226	11	0.049	0.951	0.054
37	188	12	0.064	0.936	0.050
38	151	3	0.020	0.980	0.049
39	133	6	0.045	0.955	0.047
40	115	9	0.078	0.922	0.043
41	92	4	0.043	0.957	0.041
42	79	1	0.013	0.987	0.041
43	69	3	0.043	0.957	0.039
44	60	0	0.000	1.000	0.039
45	60	1	0.017	0.983	0.038
46	46	0	0.000	1.000	0.038

Age	*N*	*D*	*qx*	*px*	*S*(*x*)
47	39	0	0.000	1.000	0.038
48	40	2	0.050	0.950	0.036
49	32	2	0.063	0.938	0.034
50	29	1	0.034	0.966	0.033
51	23	0	0.000	1.000	0.033
52	20	2	0.100	0.900	0.030
53	13	1	0.077	0.923	0.027
54	11	0	0.000	1.000	0.027
55	7	0	0.000	1.000	0.027
56	6	0	0.000	1.000	0.027
57	8	0	0.000	1.000	0.027
58	6	0	0.000	1.000	0.027
59	7	0	0.000	1.000	0.027
60	6	0	0.000	1.000	0.027
61	3	0	0.000	1.000	0.027
62	3	0	0.000	1.000	0.027
63	3	0	0.000	1.000	0.027
64	3	0	0.000	1.000	0.027
65	2	0	0.000	1.000	0.027
66	2	0	0.000	1.000	0.027
67	2	0	0.000	1.000	0.027
68	1	1	1.000	0.000	0.000

Appendix B

Age	Empirical p_x	Smoothed p_x	Capped p_x	$S(x)$	$\hat{K}(x)$	Std.Dev.$(\hat{K}(x))$	95th
1	0.992	0.931	0.931	0.931	11.203	13.527	33.454
2	0.959	0.918	0.918	0.855	11.110	13.754	33.734
3	0.912	0.906	0.906	0.775	11.162	14.040	34.256
4	0.886	0.894	0.894	0.693	11.363	14.380	35.015
5	0.858	0.885	0.885	0.613	11.713	14.761	35.993
6	0.820	0.878	0.878	0.538	12.203	15.166	37.150
7	0.823	0.874	0.874	0.470	12.816	15.571	38.428
8	0.849	0.874	0.874	0.411	13.523	15.950	39.759
9	0.850	0.876	0.876	0.360	14.290	16.281	41.069
10	0.868	0.881	0.881	0.318	15.077	16.546	42.292
11	0.886	0.889	0.889	0.282	15.842	16.737	43.371
12	0.917	0.897	0.897	0.253	16.549	16.855	44.273
13	0.922	0.906	0.906	0.229	17.170	16.908	44.981
14	0.946	0.914	0.914	0.209	17.690	16.907	45.500
15	0.924	0.922	0.922	0.193	18.103	16.867	45.846
16	0.949	0.929	0.929	0.179	18.410	16.798	46.040
17	0.947	0.935	0.935	0.168	18.625	16.710	46.110
18	0.957	0.939	0.939	0.158	18.761	16.610	46.082
19	0.950	0.943	0.943	0.149	18.837	16.503	45.982
20	0.940	0.945	0.945	0.140	18.871	16.390	45.830
21	0.943	0.947	0.947	0.133	18.879	16.271	45.642
22	0.948	0.947	0.947	0.126	18.872	16.147	45.431
23	0.958	0.947	0.947	0.119	18.864	16.015	45.206
24	0.939	0.947	0.947	0.113	18.865	15.873	44.974
25	0.933	0.946	0.946	0.107	18.880	15.720	44.737
26	0.948	0.945	0.945	0.101	18.914	15.551	44.493
27	0.945	0.944	0.944	0.095	18.970	15.364	44.242
28	0.949	0.943	0.943	0.090	19.051	15.155	43.979
29	0.941	0.942	0.942	0.085	19.158	14.918	43.696
30	0.942	0.941	0.941	0.080	19.290	14.649	43.385
31	0.940	0.941	0.941	0.075	19.444	14.342	43.034
32	0.954	0.940	0.940	0.071	19.614	13.992	42.629
33	0.920	0.940	0.940	0.066	19.792	13.597	42.156
34	0.917	0.941	0.941	0.063	19.964	13.153	41.598
35	0.936	0.943	0.943	0.059	20.115	12.663	40.943
36	0.951	0.945	0.945	0.056	20.229	12.131	40.182
37	0.936	0.947	0.947	0.053	20.294	11.562	39.313
38	0.980	0.950	0.950	0.050	20.302	10.964	38.336
39	0.955	0.953	0.953	0.048	20.247	10.342	37.259
40	0.922	0.956	0.956	0.046	20.123	9.706	36.088
41	0.957	0.960	0.960	0.044	19.925	9.065	34.836
42	0.987	0.963	0.963	0.042	19.654	8.427	33.515
43	0.957	0.966	0.966	0.041	19.315	7.795	32.137
44	1.000	0.968	0.968	0.040	18.916	7.169	30.707
45	0.983	0.970	0.970	0.038	18.467	6.542	29.227
46	1.000	0.972	0.972	0.037	17.977	5.905	27.690

Age	Empirical p_x	Smoothed p_x	Capped p_x	$S(x)$	$\hat{K}(x)$	Std.Dev.($\hat{K}(x)$)	95th
47	1.000	0.973	0.973	0.036	17.452	5.250	26.087
48	0.950	0.974	0.974	0.035	16.889	4.573	24.411
49	0.938	0.976	0.976	0.034	16.277	3.887	22.670
50	0.966	0.979	0.979	0.034	15.600	3.223	20.901
51	1.000	0.984	0.984	0.033	14.842	2.640	19.184
52	0.900	0.989	0.989	0.033	13.990	2.229	17.656
53	0.923	0.997	0.997	0.033	13.036	2.096	16.484
54	1.000	1.005	1.000	0.033	12.036	2.096	15.484
55	1.000	1.013	1.000	0.033	11.036	2.096	14.484
56	1.000	1.019	1.000	0.033	10.036	2.096	13.484
57	1.000	1.024	1.000	0.033	9.036	2.096	12.484
58	1.000	1.025	1.000	0.033	8.036	2.096	11.484
59	1.000	1.021	1.000	0.033	7.036	2.096	10.484
60	1.000	1.010	1.000	0.033	6.036	2.096	9.484
61	1.000	0.991	0.991	0.032	5.083	2.048	8.451
62	1.000	0.961	0.961	0.031	4.250	1.912	7.395
63	1.000	0.919	0.919	0.029	3.536	1.722	6.369
64	1.000	0.863	0.863	0.025	2.938	1.502	5.408
65	1.000	0.793	0.793	0.020	2.444	1.268	4.529
66	1.000	0.708	0.708	0.014	2.038	1.028	3.730
67	1.000	0.610	0.610	0.008	1.701	0.778	2.981
68	0.000	0.503	0.503	0.004	1.393	0.488	2.197
69	0.000	0.393	0.393	0.002	1.000	0.000	0.000
70	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Appendix C

Number of open claims

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1990	22	39	54	87	79	92	63	48	42	30	27	25	25	21	22	24
1991	13	24	44	57	61	71	40	35	31	27	26	20	18	17	16	
1992	10	30	54	66	63	69	53	49	45	44	39	35	27	22		
1993	9	41	56	69	67	60	63	49	48	36	37	34	31			
1994	14	58	86	107	146	139	118	123	111	93	82	79				
1995	30	99	146	192	208	217	251	202	176	171	102					
1996	40	94	152	189	236	221	230	221	160	130						
1997	40	149	215	300	329	415	297	342	248							
1998	28	78	122	195	198	236	236	197								
1999	29	78	200	391	414	424	375									
2000	36	120	291	391	352	366										
2001	54	143	257	270	309											
2002	27	74	116	130												
2003	12	23	41													
2004	8	15														
2005	4															

Appendix D

Incremental number of closed claims

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1990	0	1	1	8	13	24	15	10	12	5	1	2	4	1	1	4
1991	0	0	0	5	9	20	16	10	8	6	2	6	0	4	6	
1992	0	0	7	8	9	14	9	9	7	3	7	4	6	4		
1993	0	0	4	7	14	8	9	13	7	1	7	9	12			
1994	0	2	5	13	20	31	18	19	22	14	19	24				
1995	0	3	16	18	26	34	48	34	30	59	26					
1996	1	4	16	20	47	33	38	42	48	36						
1997	0	2	5	19	30	88	48	85	67							
1998	0	0	8	14	27	42	69	62								
1999	0	3	9	60	72	106	93									
2000	0	1	22	43	84	125										
2001	0	7	12	44	87											
2002	0	4	16	30												
2003	0	2	12													
2004	0	4														
2005	0															

5. REFERENCES

[1] Adler, M., and C.D. Kline Jr., "Evaluating Bodily Injury Liabilities Using a Claims Closure Model," *Casualty Actuarial Society Discussion Paper Program* 1988, pp. 1-66.
[2] Barnett, J., B. Schill, and G. Venter, "Review of Report of Committee on Mortality for Disabled Lives," *Casualty Actuarial Society Forum,* Winter 1991 (Arlington, Va.: Casualty Actuarial Society, 1991).
[3] Berquist, J., and R. Sherman, "Loss Reserve Adequacy Testing: A Comprehensive Systematic Approach," *Proceedings of the Casualty Actuarial Society* 1977, Vol. LXVII, pp. 123-184.
[4] Bowers, N.L., et al., *Actuarial Mathematics*, 1st ed. (Chicago, Il. Society of Actuaries, 1986).
[5] Diss, G., and R. Sherman, "Estimating the Workers Compensation Tail," *Proceedings of the Casualty Actuarial Society* 2005, Vol XCII, pp.579-678.
[6] Fisher, W.H., and J.T. Lange, "Loss Reserve Testing: A Report Year Approach," *Proceedings of the Casualty Actuarial Society* 1973, Vol. LX, pp. 189-207.
[7] Gillam, W., "Injured Worker Mortality," *Proceedings of the Casualty Actuarial Society* 1993, Vol. LXXX, pp. 34-54.
[8] London, D., *Survival Models and Their Estimation,* 2nd ed. (Winstead, Ct.: ACTEX Publications, 1988).
[9] London, D., *Graduation: The Revision of Estimates* (Winstead, Ct.: ACTEX Publications, 1985).
[10] Teng, Michael, "Projecting Workers Compensation Losses Using Open Claim Count and Average Loss Payment, and Application to Analysis of California Workers Compensation Loss Development," *Casualty Actuarial Society Forum*, Fall 2001, pp. 347-392.

Abbreviations and notations

$S(x)$, survival function

${}_aq_x$, probability that a life aged x will die within the next a years

r_i, difference between the age when observation begins and the most recent integral age

d_x number of deaths observed during a period of observation

${}_ap_x$, probability that a life aged x will survive for a years

$K(X)$, the future lifetime of a life aged x

s_i, difference between the next integral age and the age at which the life exits observation

n_x, number of lives at age x during the observation period

Biography of the Author

Brian Fannin is an actuary at the Munich Re Group, responsible for the pricing and underwriting of retrospective reinsurance contracts for the Customized Portfolio Solutions unit. Prior to his current role, he spent three years in Munich Re's Integrated Risk Management unit where he participated on a variety of ERM projects. Prior work experience includes treaty and facultative reinsurance and primary commercial lines pricing. Brian may be contacted at BFannin@MunichRe.com

Hierarchical Growth Curve Models for Loss Reserving

James Guszcza, FCAS, MAAA

Abstract

Hierarchical or multilevel modeling extends traditional GLM or non-linear models by giving certain of the model parameters their own probability sub-models. Hierarchical modeling can be viewed as an extension of Bayesian credibility theory that allows one to build models for data that are grouped along a dimension containing multiple levels. In particular, hierarchical modeling can be used to analyze longitudinal datasets containing multiple observations for each of several subjects. A contention of this paper is that traditional loss reserving triangles are most naturally regarded as longitudinal datasets. Non-linear hierarchical models – known also as non-linear mixed effects models – therefore provide a natural and flexible framework in which to model loss development across multiple accident years. The use of non-linear growth curves together with multilevel modeling techniques allows one to build models that are at once parsimonious and easy to interpret. Finally, because they incorporate growth curves, such models obviate the need to specify tail factors.

Keywords: Stochastic loss reserving, hierarchical models, multilevel models, nonlinear mixed effects models, growth models, repeated measurements, longitudinal data, Bayesian credibility, shrinkage, R.

1. INTRODUCTION

Loss reserving theory and practice is undergoing a renaissance due to a recent proliferation of stochastic reserving techniques. To cite but a few examples, recent authors have applied regression analysis (Barnett and Zehnwirth [1]), generalized linear models (England and Verrall [2]), loss development growth curves together with maximum likelihood estimation (Clark [3]), and Bayesian methods (Meyers [4]) to model loss development data. Statistical modeling techniques are increasingly supplementing or supplanting spreadsheet-based projection methods for estimating ultimate losses.

This paper will propose yet another statistical framework for modeling loss triangles: *nonlinear hierarchical models*. These models are also commonly known as *nonlinear mixed effects* [NLME] *models*. The contention of this paper is that this class of models provides a highly flexible and natural framework within which the loss development process can be analyzed. The goals of this paper are twofold: to introduce the concept of hierarchical models and to illustrate how hierarchical models can be used in loss reserving.

Section 2 will sketch some of the basic theory of hierarchical models and also provide a hypothetical example illustrating how hierarchical modeling can be used to analyze longitudinal (or "repeated measurements") data. The relationship between hierarchical modeling and Bayesian credibility theory will also be discussed. These topics are not specific to loss reserving in particular, but are discussed in order to set the stage for Section 3.

Section 3, the main section of the paper, will broaden the discussion of hierarchical models to include non-linear model forms. Motivated by the very interesting presentation of Clark [3], our hierarchical models will incorporate the Weibull and Loglogistic "growth curve" functional forms. These models will be applied to the same loss triangle data analyzed by Clark and others. Many such growth curves are possible, but the Weibull and Loglogistic are two natural options. One way of understanding the loss reserving models proposed here is that they add "random effects" to the types of growth curve models introduced by Clark.

No attempt will be made in this paper to estimate reserve variability, which is beyond the scope of this introductory paper. This will be the subject of a future paper.

2. HIERARCHICAL MODELS

Generally speaking, hierarchical models are used when the data at hand are *grouped* in some important way. Examples include:

- The relationship between standardized test scores and prior grades of students from different high schools.
- Performance of a state's high schools, where schools are grouped into school districts.
- Expected workers compensation claims for exposures with various NCCI class codes.
- Expected loss ratio relativities for a personal auto carrier's various state territories.
- The growth of a collection of soybean plants, measured at various times since planting.

The first two examples are typical of the examples discussed in the social science literature (e.g., Gelman and Hill [5]). The third and fourth examples are classic problems of actuarial science, but are similar in form to the first two examples.

The final example is typical of hierarchical modeling applications in such fields as biology (e.g., modeling the growth of plants and animals) and pharmacology (e.g., modeling the effect of a drug over time). Many such examples are given in the book by Pinheiro and Bates [6]. In cases such as these, we have multiple measurements of each subject, performed at different points in time. Such multilevel datasets are commonly referred to as "longitudinal," "panel," or "repeated measurements" datasets. The primary goal of this paper is to convince the reader that loss reserving triangles can reasonably be regarded as longitudinal datasets, to which hierarchical modeling techniques naturally apply.

The central concept of hierarchical models is that certain model parameters are *themselves* modeled. In other words, not all of the parameters in a hierarchical model are directly estimated from the data. Rather, (some of) the model parameters are calculated from estimates of the model's *hyperparameters*, which are in turn estimated from the data. Model parameters that are themselves given models are sometimes referred to as "random effects." They are to be distinguished from "fixed effects," which are not modeled, but are instead estimated directly from the data. "Mixed effects models," therefore, refer to models that contain both modeled and non-modeled parameters.

A note on terminology: this paper generally follows Gelman and Hill in favoring the language of hierarchical models over the "random/fixed/mixed effects" terminology. However, the phrase "random effects" will occasionally be used as shorthand for model parameters that are given sub-models. Many authors, including Pinheiro and Bates [6], speak mainly in terms of "mixed effects models." Note that Pinheiro and Bates wrote the "nlme" R function that was used to fit the hierarchical models described in this paper.

At this point an example might aid the discussion. Consider a hypothetical company that sells personal auto insurance in each of eight roughly equal-sized regions. We have data for the number of policies in force by region as of January 1, 2005, 2006, 2007, and 2008. We thus have 8*4=32 data points in all. We would like to build a model that could be used to forecast the number of policies in force, by region, in the coming years.

Using notation suggested by Gelman and Hill, let i denote the data point number and range from 1 to 32; similarly let j denote the region number. The term $j[i]$ will denote the group to which data point i belongs. For example, $j[5]=2$ because the fifth data point is an observation from Region 2. Two modeling strategies immediately suggest themselves.

Model 1 (complete pooling of data): First, we could simply pool the data from all eight regions and regress PIF (policies in force) on time.

$$PIF = \alpha + \beta t + \varepsilon$$

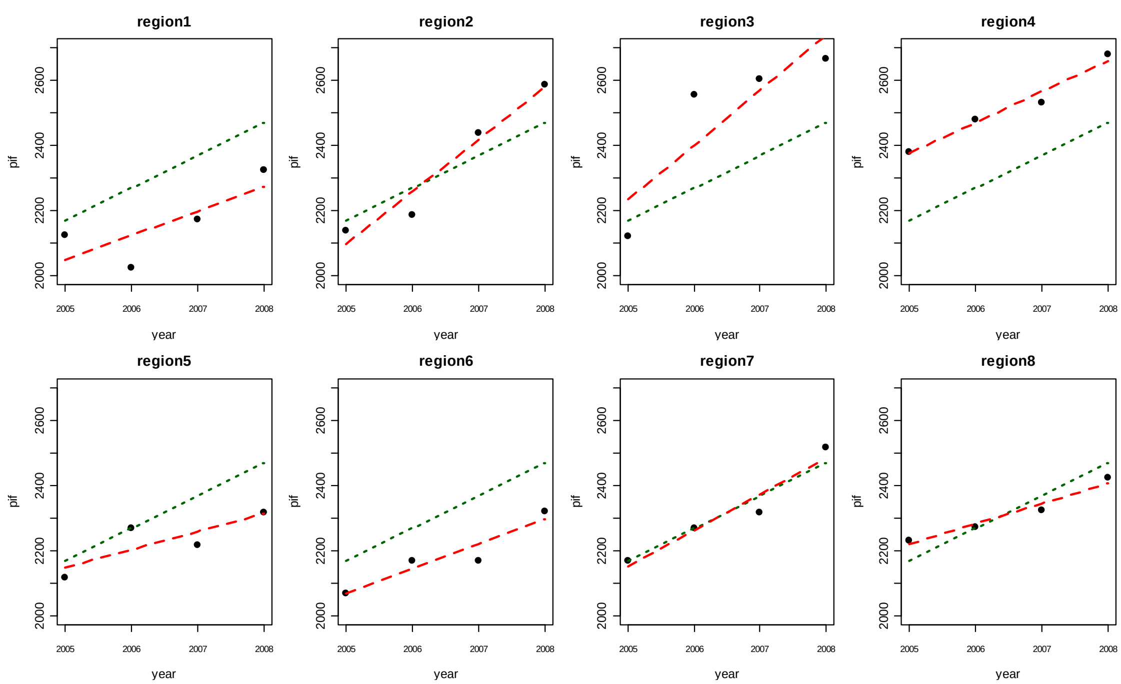
where $\varepsilon \sim N(0,\sigma^2)$. In this case the 32 data points would be used to estimate the three parameters $\{\alpha, \beta, \sigma\}$. Here we are effectively ignoring region.

Model 2 (separate models by region): Second, we could run eight separate regression models, one for each region.

$$PIF = \left\{\alpha^j + \beta^j t + \varepsilon^j\right\}_{j=1,2,\ldots,8}.$$

Note that each of these eight regression models is fit using only four data points.

These models are plotted in the figure below. The dotted lines represent Model 1 and are the same across all regions. The dashed lines represent Model 2 and vary from region to region. This plot illustrates why neither option is entirely satisfactory. At one extreme, the "pooled" model clearly provides poor fits in, for example, regions 1 and 4. At the other extreme, one might doubt that the data is sufficiently credible to support the fitting of eight region-specific models. For example, the first data point in region 3 appears to exert too much leverage on that model's parameters. A slope closer to that of the "pooled" model might be more believable.



Model 3 (include region indicator variables): Of course other conventional strategies are possible. For example, one could fit a no-intercept model that includes a separate indicator variable for each of the eight regions:

$$PIF = \beta_1(region == 1) + \ldots + \beta_8(region == 8) + \beta_9 t + \varepsilon.$$

This is a compromise between models 1 and 2. Like Model 1, it is a single "pooled" model that is fit to all of the data. Like Model 2, it allows us to capture region-specific aspects of the data. This is an improvement, but perhaps is still not ideal. We are still estimating 10 parameters – $\{\beta_1, \ldots, \beta_8, \beta_9, \sigma\}$

– using 32 data points. We face the danger of building an over-parameterized model. (Of course not all of the eight region indicators will necessarily be significant in the model. One or more of the indicators might be dropped.) The need to potentially add region/time interaction terms presents the possibility of further over-parameterization. In the extreme case where we need a separate intercept term and interaction with time for each region, we would need to estimate a model eight different intercepts and eight different slopes. This would essentially return us to Model 2.

Model 4 (random intercepts): Hierarchical modeling offers a different type of compromise. In this simple example, rather than estimate a separate "β" parameter for each region directly from the data, we specify a *Gaussian sub-model* of which eight region-specific intercept parameters $\{\alpha_1, \ldots, \alpha_8\}$ are random draws. Therefore, unlike $\{\beta_1, \ldots, \beta_8\}$ in Model 3, these so-called "random intercepts" $\{\alpha_1, \ldots, \alpha_8\}$ are not "estimated directly from the data." Rather, they are derived from the *hyperparameters* of the Gaussian sub-model.

Explicitly, this "random intercepts" hierarchical model can be written:

$$PIF = \alpha_1 + ... + \alpha_8 + \beta t + \varepsilon \quad where \quad \alpha_j \sim N(\mu_\alpha, \sigma_\alpha^2) \quad and \quad \varepsilon \sim N(0, \sigma^2)\,.$$

Or more compactly:

$$PIF_i \sim N(\alpha_{j[i]} + \beta t_i, \sigma^2) \quad where \quad \alpha_j \sim N(\mu_\alpha, \sigma_\alpha^2)\,.$$

In some circles it is conventional to call such a model a "mixed effects" model. The "slope" parameter β is called a "fixed effect," while the $\{\alpha_1, \ldots, \alpha_8\}$ parameters are called "random effects."

This hierarchical model contains four hyperparameters which can be estimated using maximum likelihood or a related optimization technique:

$$\hat{\mu}_\alpha = 2068.0 \quad \hat{\beta} = 100.06 \quad \hat{\sigma} = 81.13 \quad \hat{\sigma}_\alpha = 123.94\,.$$

Compare this with the 10 parameters estimated from the non-hierarchical regression model with a separate indicator variable for each region.

As noted above, the intercept "random effect" parameters $\{\alpha_1, \ldots, \alpha_8\}$ are derived using the model's estimated hyperparameters. Readers familiar with credibility theory might have anticipated that the formula used to do this is:

$$\hat{\alpha}_j = Z_j \cdot (\bar{y}_j - \beta \bar{t}_j) + (1 - Z_j) \cdot \mu_\alpha \quad where \quad Z_j = \frac{n_j}{n_j + \sigma^2 / \sigma_\alpha^2}\,.$$

In actuarial parlance, each random intercept α_j is a credibility-weighted average of the region-specific intercept and the average (μ_α) of all of the region-specific intercepts. The credibility factor Z_j is determined in a familiar way using the number of observations for each region (n_j), the variance of the region specific intercepts (σ^2_α), and the residual variation σ^2.

Models 1 and 2, illustrated above, are special cases of this hierarchical model in a precise sense. As $\sigma^2_\alpha \rightarrow 0$, $Z_j \rightarrow 0$ and the hierarchical model approaches Model 1. As $\sigma^2_\alpha \rightarrow \infty$, $Z_j \rightarrow 1$ and the hierarchical model approaches Model 2 (Gelman and Hill [5] p. 258).

As an aside, it should be apparent that Bühlmann's credibility model is a specific instance of hierarchical models. If we remove the time covariate t, Model 4 becomes

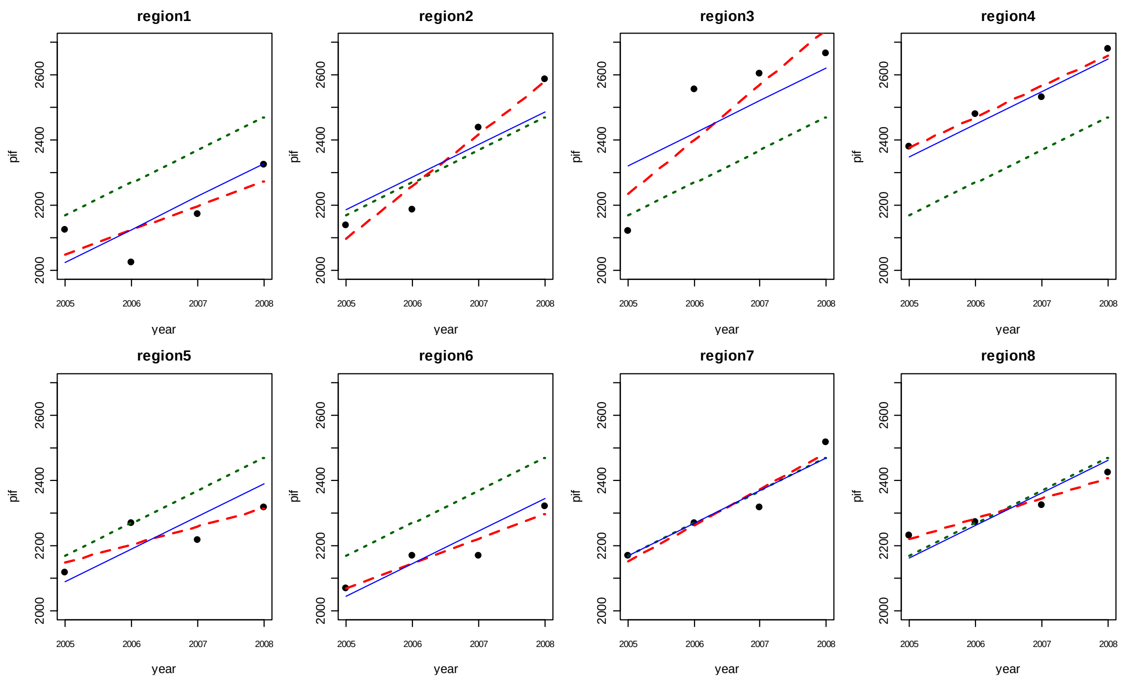
$$PIF_i \sim N(\alpha_{j[i]}, \sigma^2) \quad where \quad \alpha_j \sim N(\mu_\alpha, \sigma^2_\alpha).$$

And the credibility weighting expression becomes:

$$\hat{\alpha}_j \approx Z_j \cdot \bar{y}_j + (1 - Z_j) \cdot \mu \quad where \quad Z_j = \frac{n_j}{n_j + \sigma^2 / \sigma^2_\alpha}.$$

Frees [7 section 4.7] provides a helpful discussion of the ways in which several well-known credibility models are specific types of hierarchical models.

In the figure below, the predicted values of Model 4 (solid line) are added to the predicted values of Models 1 and 2. In certain cases (such as Regions 1 and 3) Model 4 appears to be an improvement over Model 2. This is because the more parsimonious Model 4 is not seriously leveraged by these regions' "2005" data points. For regions 2 and 8, on the other hand, Model 2 seems to fit the data better.



Model 5 (random slopes and intercepts): Because of the seemingly suboptimal fit on the model in Regions 2 and 8, one might consider adding a "slope random effect" to model 4. Explicitly:

$$PIF_i \sim N\left(\alpha_{j[i]} + \beta_{j[i]} \cdot t_i, \sigma^2\right) \quad where \quad \begin{pmatrix} \alpha_j \\ \beta_j \end{pmatrix} \sim N\left(\left[\mu_\alpha, \mu_\beta\right], \Sigma\right) \quad , \quad \Sigma = \begin{bmatrix} \sigma_\alpha^2 & \sigma_{\alpha\beta} \\ \sigma_{\alpha\beta} & \sigma_\beta^2 \end{bmatrix}.$$

Model 5 contains six hyperparameters: $\{\mu_\alpha,\mu_\beta,\sigma_\alpha,\sigma_\beta,\sigma_{\alpha\beta},\sigma\}$, two more than Model 4. Because Models 4 and 5 are nested models, we can compare their expected predictive accuracy by comparing their log-likelihoods and Akaike Information Criterion [AIC] statistics.

	LL	*d.f.*	*AIC*
Model 4	-186.20	4	380.40
Model 5	-184.32	6	380.64

Recall that AIC = -2*LL + 2*d.f., as can be confirmed from the above table. In a phrase, AIC is log-likelihood penalized for the number of hyperparameters in the model. The model with the lower AIC statistic is thought to make a better trade-off between complexity and goodness-of-fit, and is therefore expected to make more accurate predictions of future data.

Adding the further "random effect" to vary the slopes (in addition to varying the intercepts) results in an improved log-likelihood; but a slightly worse AIC. This comparison suggests that it would be wise to favor the more parsimonious Model 4 above the slightly better fitting Model 5. The AIC comparison suggests that Model 5 might over-fit the data.

General Observations: Before turning to loss reserving, it is worth making a few general observations about the implications of hierarchical modeling for actuarial work. First, the hierarchical/multilevel modeling framework is a generalization of current actuarial modeling practice in two important ways.

- Actuaries often face a dilemma when faced with multilevel modeling situations. For example, should one pool one's data and build a single countrywide predictive model to be used in all states? Or should one build separate models by state? These options are analogous to Models 1 and 2 above. In the light of the above discussion, it should be clear that these two options are extreme cases (as the variance of a hierarchical model's random effects approach 0 and ∞, respectively) of a suitably specified hierarchical model.
- Bayesian credibility models are specific types of hierarchical models. Just as generalized linear models (GLMs) have provided a unifying framework for traditional minimum bias calculations, hierarchical modeling theory provides a unifying framework for Bayesian credibility modeling. This is helpful both pedagogically and practically. Pedagogically, it is helpful to understand the connection between Bayesian credibility and linear modeling. Practically, multilevel modeling packages can be used to perform Bayesian credibility calculations. In the same way that GLM modeling is less cumbersome than performing minimum bias calculations, hierarchical modeling packages allow one to perform Bayesian credibility calculations with a minimum of ad hoc programming. Furthermore, multilevel modeling packages make it easy to employ rigorous statistical methodology – such as graphical diagnostics and comparison of goodness-of-fit statistics – in one's work.

A second observation is that multilevel modeling potentially allows one to achieve a much better fit at the expense of adding only a few additional hyperparameters to a conventional GLM model. In the above example, Model 4 contains only one more hyperparameter than Model 1, but it provides a much better fit to the data. This is because the "scoring equation" for Model 4 contains nine parameters $\{\alpha_1, \ldots, \alpha_8, \beta\}$ as opposed to Model 1's insufficient $\{\alpha, \beta\}$ parameters. In short, actuaries can consider specifying hierarchical GLM models (HGLMs) as an alternative to purely "fixed effects" GLM models.

A related point is that the hierarchical modeling framework works well even if one's data contains a very large number of levels. The above example could easily be modified to involve four years of PIF data in each of 1000 counties. Model 4, with its four hyperparameters, or Model 5, with its six hyperparameters, would be no less applicable to this data. By comparison a traditional, non-hierarchical model would potentially need hundreds of indicator variables. In short, the hierarchical modeling framework provides a natural way to handle "massively categorical" variables in one's modeling work. This is because hierarchical modeling implicitly allows one to perform Bayesian credibility weighting within a GLM model building context.

These observations are not specific to loss reserving, but they set the stage for the hierarchical growth curve approach to loss reserving to be outlined in the next section.

3. HIERARCHICAL MODELS FOR LOSS RESERVING

The preceding section might have seemed like a long detour away from the topic of loss reserving. But it reviewed some of the concepts needed to build a hierarchical model of the loss development process. Consider a garden variety 10-by-10 loss triangle. Each of the 55 non-missing cells contains cumulative losses (CL), indexed by accident year AY and development period dev. We will treat this loss triangle as a multilevel dataset, in which each of the 10 accident years is a separate level. This will allow us to build a hierarchical model in which we "regress" cumulative losses CL on development period dev. The major disanalogy with the illustrative example in the previous section is that we must replace the linear regression with a non-linear model.

Pinheiro and Bates discuss three advantages of nonlinear hierarchical models, each of which apply in the context of loss reserving:

- **Interpretability.** The modeling approach to be outlined here requires that one explicitly model the loss development process in a specific functional form. Judgment as well as background empirical or theoretical knowledge can be used to guide the choice of nonlinear functional form.
- **Parsimony.** A well-chosen nonlinear function can model a non-linear process with fewer parameters than a linear model with multiple polynomial terms. In addition, as illustrated in the previous section, the hierarchical modeling approach potentially allows one to replace a potentially large number of subject-specific indicator variables and interaction terms with a small number of hyperparameters.

- **Validity beyond the observed range of the data.** Of course it is always dangerous to use a model to extrapolate beyond the data. However, the approach to be outlined here at least offers a framework within which one can harness one's background knowledge when specifying a model. Such an approach is less likely to lead one astray than a less parsimonious or more atheoretical "curve-fitting" approach.

Sample Dataset: To illustrate, we will work with the sample loss reserving dataset analyzed by Clark [3]. For ease of viewing, the cumulative loss numbers in the table below numbers have been divided by 1,000. These numbers are rounded only for the purpose of display; no rounding was done in performing the calculations.

Cumulative Losses in 1000's

AY	*12*	*24*	*36*	*48*	*60*	*72*	*84*	*96*	*108*	*120*	*reported*	*est ult*	*reserve*
1991	358	1,125	1,735	2,183	2,746	3,320	3,466	3,606	3,834	3,901	3,901	3,901	0
1992	352	1,236	2,170	3,353	3,799	4,120	4,648	4,914	5,339		5,339	5,434	95
1993	291	1,292	2,219	3,235	3,986	4,133	4,629	4,909			4,909	5,379	470
1994	311	1,419	2,195	3,757	4,030	4,382	4,588				4,588	5,298	710
1995	443	1,136	2,128	2,898	3,403	3,873					3,873	4,858	985
1996	396	1,333	2,181	2,986	3,692						3,692	5,111	1,419
1997	441	1,288	2,420	3,483							3,483	5,672	2,189
1998	359	1,421	2,864								2,864	6,787	3,922
1999	377	1,363									1,363	5,644	4,281
2000	344										344	4,971	4,627
chain link	3.491	1.747	1.455	1.176	1.104	1.086	1.054	1.077	1.018	1.000	34,358	53,055	18,697
chain ldf	14.451	4.140	2.369	1.628	1.384	1.254	1.155	1.096	1.018	1.000			
growth curve	6.9%	24.2%	42.2%	61.4%	72.2%	79.7%	86.6%	91.3%	98.3%	100.0%			

To provide a baseline for comparison, the results of a simple chain ladder calculation are displayed along with the raw data. All data was used to calculate each of the link ratios; and the 120→ultimate "tail factor" is assumed to be 1.0. According to this calculation, the expected total outstanding losses are approximately $18.7M. The implied "growth curve" is simply the reciprocal of the sequence of loss development factors.

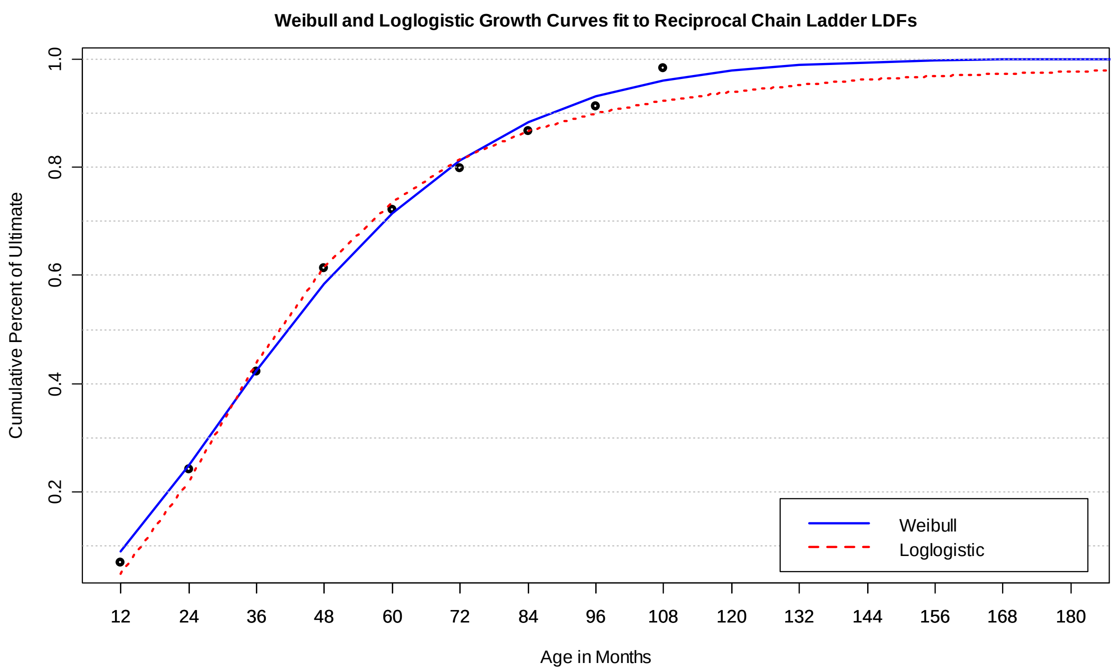
Clark's Models: The nine "growth curve" numbers resulting from the simple chain ladder exercise can be viewed as a piecewise linear approximation to a continuous growth curve. Clark considers two such growth curves, the Weilbull and Loglogistic, and integrates each of them into two models of the loss triangle data. The Weilbull growth curve has the form:

$$G(x \mid \omega, \theta) = 1 - \exp\left(-(x/\theta)^{\omega}\right).$$

The Loglogistic curve has the form:

$$G(x \mid \omega, \theta) = \frac{x^{\omega}}{x^{\omega} + \theta^{\omega}} .$$

Purely for illustration, we can fit each of these curves to the reciprocal of the chain ladder loss development factors (LDFs) displayed above. The resulting curves are displayed below, together with the reciprocal of the nine chain ladder LDFs.



This plot confirms that both the Weibull and Loglogistic growth curves are plausible candidates for modeling the loss development process. Each of the curves fits the reciprocal LDF pattern reasonably well. Note that the Loglogistic growth curve has a "heavier tail" than the Weibull, implying a longer loss development process and higher estimated ultimate losses. Note also that neither of the curves fits empirical development pattern perfectly. The Loglogistic curve fits the earlier data points better; whereas the Weibull curve is a bit closer to the final data point. In practice, one's background knowledge of the likely length of the loss development process would be used to decide between these, or other, growth curves. Following Clark, we fit sample models incorporating each of these growth curves and compare the results.

Clark proposes two models of the loss data. The first is called the "Loss Development Factor" (LDF) model, and can be expressed:

$$CL_{AY,dev} = ULT_{AY}\left[1 - G(dev \mid \omega, \theta)\right].$$

The function G can be the Weibull, the Loglogistic, or any other suitable growth function. The LDF model contains 12 parameters: $\{ULT_{1991},\ldots, ULT_{2000}, \omega, \theta\}$.

(Note that Clark in fact models incremental rather than cumulative losses, and therefore specifies a formula that differs accordingly. Specifically, Clark's formula is

$$IL_{AY;x,y} = ULT_{AY}\left[(y \mid \omega, \theta) - G(x \mid \omega, \theta)\right]$$

where $IL_{AY;x,y}$ denotes the incremental losses in accident year AY between ages x and y. This is advantageous in that random noise at age x will not be propagated through ages $x+1$, $x+2$, and so on. For readability and ease of exposition, the models discussed in this paper are cast in terms of cumulative, rather than incremental, losses. However, it is a simple exercise to recast these models, as done above, in terms of incremental losses.)

Clark's calls his second model a "Cape Cod" model. Here the unknown parameters ULT_{AY} are replaced with $PREM_{AY} \cdot ELR$:

$$CL_{AY,dev} = PREM_{AY} \cdot ELR\left[1 - G(dev \mid \omega, \theta)\right].$$

$PREM_{AY}$ denotes on-leveled premium for accident AY (a known quantity). This model incorporates the Cape Cod assumption of a constant expected loss ratio (ELR) across all accident years. As a result, this model contains only three unknown parameters, $\{ELR,\omega,\theta\}$, as opposed to the LDF model's 12. The Cape Cod model is therefore less prone to overfitting the available data (in this illustration, 55 data points) than the LDF model. Clark points out that the less parsimonious LDF model results in more parameter variance, in turn resulting in more variance around the estimated reserves.

Baseline Hierarchical Model: It is possible to build hierarchical counterparts to each of Clark's models. Let us begin with Clark's LDF model. Rather than estimate the 10 parameters $\{ULT_{1991},\ldots, ULT_{2000}\}$ directly from the data, we can model them in hierarchical fashion. Explicitly:

$$\begin{gathered} CL_{AY,dev} = ULT_{AY}\left[1 - \exp\left(-(x/\theta)^{\omega}\right)\right] + \varepsilon_{AY,dev} \\ ULT_{AY} \sim N(\mu_{ULT}, \sigma_{ULT}^2) \\ Var(\varepsilon_{AY,dev}) = \sigma^2 \hat{CL}_{AY,dev} \end{gathered}$$

This will be our baseline model. All of the alternate models to be discussed subsequently will be modifications of this baseline. The baseline model contains five unknown hyperparameters that must be estimated from the data: $\{\mu_{ULT}, \omega, \theta, \sigma_{ULT}, \sigma\}$. Specifying a sub-model of ULT_{AY} in the above fashion is analogous to replacing the region-specific indicator variables in the previous section's PIF example with the "random intercepts" α_j.

Note that rather than assuming constant variance for each loss amount, we are assuming that the within-variance is proportional to the fitted value, where σ^2 is the proportionality constant. This corresponds to the over-dispersed Poisson assumption found in both England and Verrall [2] and Clark. We will relax this assumption shortly.

This model can easily be fit using the "nlme" ("non-linear mixed effects") function in R. (Please refer to the note at the end of this paper for information on how to obtain R and the nlme function.) The R code needed to do this is quite straightforward:

```
start.vals <- c(ult=5000, omega=1.4, theta=45)
w1 <- nlme(cum ~ ult*(1 - exp(-(dev/theta)^omega))
            , fixed = list(ult~1, omega~1, theta ~ 1)
            , random = ult ~ 1 | AY
            , weights = varPower(fixed=.5)
            , data=dat, start = start.vals)
```

Note most stochastic reserving techniques, this one included, require that one organize one's data in matrix rather than triangular form. The appendix to this paper displays the data in the form that it is read in prior to submitting the above R code.

We must supply starting values in order to estimate the parameters of a non-linear hierarchical model (starting values are not needed for linear hierarchical models). Choosing the appropriate starting values is something of an art. Still, in this particular case the model converges to the correct solution for a wide range of starting values. For example, replacing the above starting values with {10000, 2.0, 100} does not change the resulting model. However further changing the starting value of "ult" to 15000 causes the model not to converge. Changing the starting value for "omega" to 3.0, on the other hand, causes the model to converge to an incorrect solution. (A quick glance at a residual plot makes it clear that the solution is incorrect.) In most cases it should be possible to select a workable set of starting values using the estimated ultimate losses and implied growth curve from a simple chain ladder analysis.

Submitting the above R code yields the following estimates of the model's five hyperparameters. The model runs in seconds.

$$\mu_{ULT} = 5306.6 \quad \omega = 1.306 \quad \theta = 46.64 \quad \sigma_{ULT} = 543.03 \quad \sigma = 2.955$$

This model's AIC statistic is 725.76. Also, the *p*-values associated with μ, ω, and θ are all less than

.0001. The parameter error associated with this model is therefore fairly low.

We note in passing that the ω and θ parameter estimates for Clark's Weilbull LDF model are 1.297 and 48.885, respectively. These are reasonably consistent with our results.

The parameters and estimated ultimate losses and loss reserves resulting from the baseline model are displayed in the table below. The model's parameters (not hyperparameters) are listed in the omega, theta, and ULT columns. Because they were not given "random effects," omega and theta are the same for each accident year. We will shortly investigate the effect of adding random effects to the ω and θ parameters.

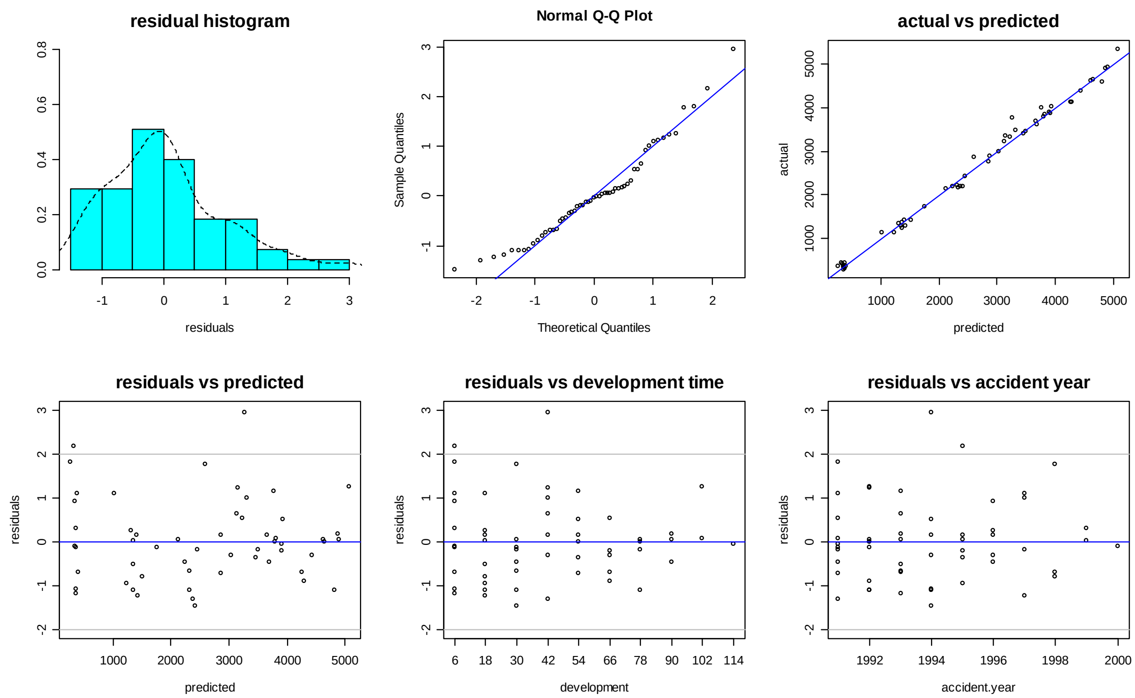
The key difference between this model and Clark's LDF model, is that here the estimated ultimate losses in the *ULT* column are *not* estimated directly from the data. Rather, they are derived from the estimates of the model's hyperparameters. Note that the average value of the *ULT* column is 5306.6, which is the same as the estimate of μ_{ULT}.

Parameters and Estimated Reserves - Baseline Model

AY	*dev*	*omega*	*theta*	*growth*	*reported*	*eval120*	*eval240*	*ULT*	*reserves*
1991	114	1.306	46.638	96.0%	3,901	3,943	4,073	4,074	172
1992	102	1.306	46.638	93.8%	5,339	5,239	5,412	5,413	74
1993	90	1.306	46.638	90.6%	4,909	5,207	5,379	5,380	470
1994	78	1.306	46.638	85.9%	4,588	5,423	5,602	5,603	1,015
1995	66	1.306	46.638	79.3%	3,873	4,777	4,935	4,936	1,062
1996	54	1.306	46.638	70.2%	3,692	5,052	5,219	5,220	1,528
1997	42	1.306	46.638	58.2%	3,483	5,512	5,694	5,695	2,212
1998	30	1.306	46.638	43.0%	2,864	5,850	6,043	6,044	3,180
1999	18	1.306	46.638	25.0%	1,363	5,255	5,429	5,430	4,067
2000	6	1.306	46.638	6.6%	344	5,101	5,270	5,271	4,927
total								53,066	**18,708**

The baseline model's estimate of the total outstanding losses is roughly \$18.7M. This is virtually identical to the chain ladder's outstanding loss estimate displayed above. However, this similarity is a coincidence. The two models' reserve estimates differ considerably by accident year. For example, the chain ladder model's estimate of accident year 1998's outstanding losses is \$3.92M, in contrast with the baseline hierarchical model's estimate of \$3.18M.

Next we can inspect the standardized residuals and fitted values:



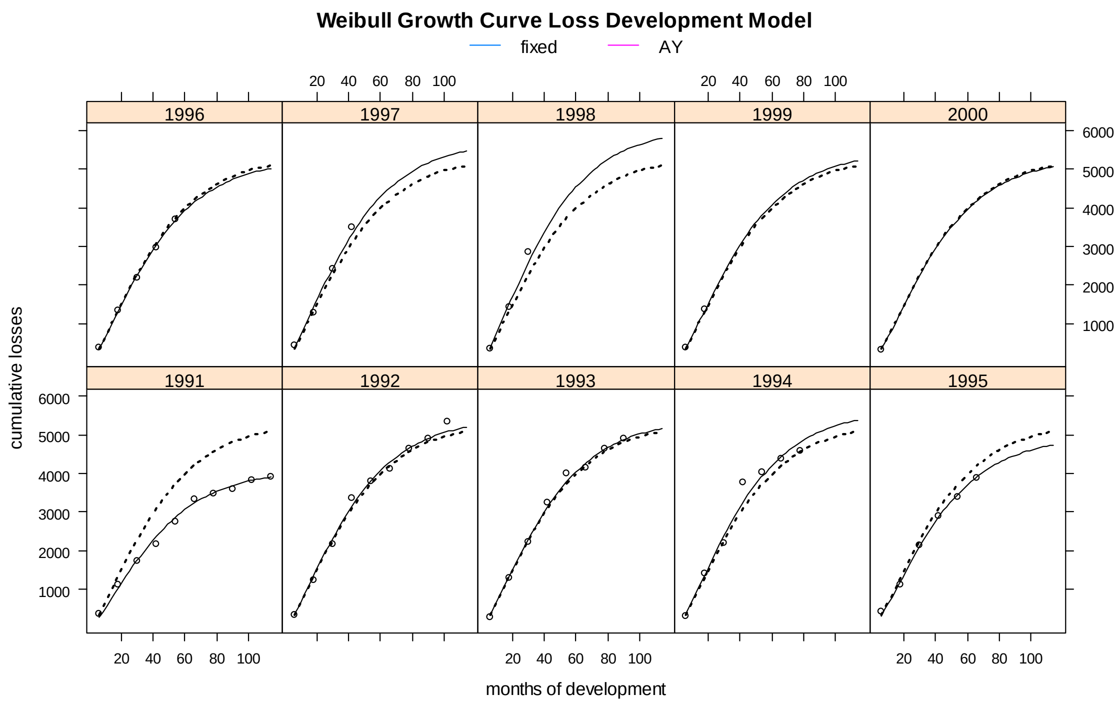
These diagnostic plots together indicate that the model fits the data reasonably well. However, the model is not perfect. The upper left two plots indicate that the standardized residuals are not quite normally distributed. Still, the deviation from normality is perhaps within the realm of acceptability. The "actual vs predicted" plot indicates a good fit. Consistent with this, the "residuals vs predicted" plot indicates that most of the standardized residuals are less than 2.0 in absolute value. A close inspection of this plot reveals an undulating pattern in the residuals: the model has a slight but systematic tendency for the model to under-estimate cumulative losses in the range of \$1M-\$3M and over-estimate cumulative losses in the \$3M-\$4M range. This suggests that the Weibull curve does not perfectly characterize the development of cumulative losses.

The general conclusion while the model could perhaps be improved upon, the overall fit is good. Four points are worth emphasizing:

- The model fits the data well despite the fact that it contains only five hyperparameters. In contrast, Clark's non-hierarchical LDF model contains 12 parameters; and the chain ladder analysis requires us to estimate nine link ratios (not including the arbitrary tail factor that must be added).
- Unlike Cape Cod-type models (to be described below), it is not necessary to bring in premium data or assume a constant expected loss ratio across accident years.

- This five-parameter model can be used to project losses to their ultimate values (or any intermediate value) without the need for a tail factor.
- The model's parsimony is made possible both by the hierarchical modeling methodology as well as the use of a non-linear growth function *G*.

Relating to this last point, another way to evaluate the model's fit is to superimpose each accident year's estimated growth curve on top of the cumulative loss observations. In the plot below, the (identical) dotted curves represent the "fixed" Weibull curve implied by the hyperparameters $\{\omega, \theta, \mu_{ULT}\}$. The solid curves are the accident year-specific Weibull curves implied by ω and θ as well as the derived parameters $\{ULT_{1991},\ldots,ULT_{2000}\}$.



These plots further support the conclusion that our baseline hierarchical growth model fits the data well. In addition, they illustrate the basic intuition motivating the approach. Following Clark, we are modeling loss development as a *growth process*, in much the same way that a biostatistician would model the growth of a group of trees or soybean plants. In the latter cases, each "subject" is an individual tree or soybean plant and each observation is a measurement of size at various ages. In loss reserving, each "subject" is the aggregate claims from an accident year and each observation is the aggregate cumulative losses at various development ages.

Before continuing, it is worth commenting on the growth curve plots for accident years 1991 and

1998. Note that the 1991 growth curve is different from the other year's growth curves. This is reflected the ULT_{1991} parameter of 4.074M, which is more than 20% lower than the average μ_{ULT}=5.3066M. However, we have 10 AY 1991 observations, all of which fall squarely on the 1991-specific growth curve. This suggests that the low value of ULT_{1991} is justified.

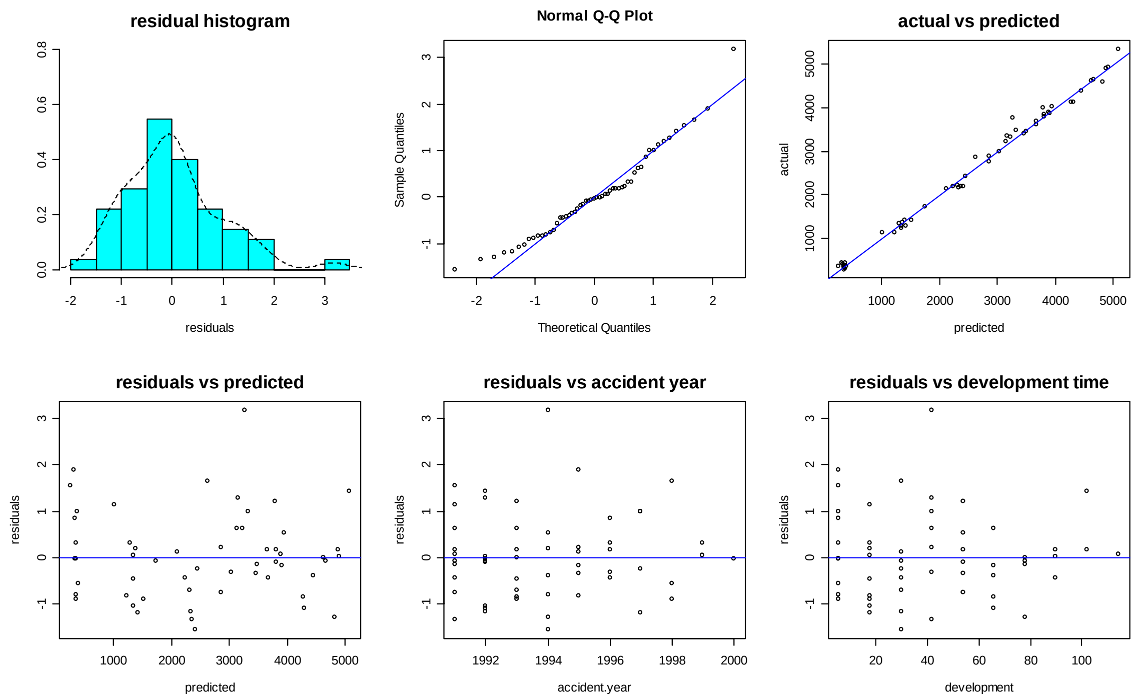
In contrast, the estimated ultimate losses for 1998 are approximately 6.044M, 14% higher than average. This is of course driven by only three data points, which have greater 12→24 and 24→36 developments than their counterparts in other accident years. The chain ladder method produces an even higher estimate of 1998 ultimate losses: 6.787M. The hierarchical model's estimate therefore falls between the global average μ_{ULT} and the chain ladder estimate. This is illustrative of the way in which the hierarchical model implicitly uses a type of "credibility weighting" to "shrink" the accident-year specific estimates towards the global mean. The amount of "shrinkage" is more pronounced for more recent accident years. The most extreme amount of shrinkage occurs for accident year 2000: the estimated ultimate losses for this year are $5.271M, only a fraction of a percent lower than the global mean of $5.3066M. Little credibility is given to the single data point for accident year 2000.

Relaxing the Process Variance Assumption: Recall that the baseline model contains the assumption that the within-variance is proportional to the fitted value. We can replace this with the weaker assumption that:

$$Var(\varepsilon_{AY,dev}) = \sigma^2 (C\hat{L}_{AY,dev})^{2\varsigma}.$$

In other words, rather than pre-specify that ζ=0.5, we can introduce ζ as a further model hyperparameter to be estimated. This means that our model will contain six, rather than five, hyperparameters. (In R, this is achieved by simply removing the "fixed=0.5" from inside the "varPower" expression.)

The resulting estimate is $\zeta \approx 0.37$. Although not displayed here, the estimated loss reserves of this model are, in aggregate, only $100,000 (or 0.5%) less than that of the baseline model. The residual plot indicated an improved residual histogram, but otherwise little difference in the goodness of fit. For simplicity we will therefore continue with the baseline model.



Random Shape Effect: The baseline model incorporates the assumption that the different accident years' ultimate losses vary randomly about a mean value: $ULT_{AY} \sim N(\mu_{ULT}, \sigma^2_{ULT})$. It also incorporates the assumption that the shape (ω) and scale (θ) characterizing the loss development process do *not* vary by accident year. Just as we were able to vary slope – in addition to intercept – by region in the previous section's PIF example, here we have the option of allowing ω and/or θ to vary by accident year.

To illustrate, we expand our model to include varying shape parameters $\{\omega_{1991},\ldots, \omega_{2000}\}$ by accident year. Specifically:

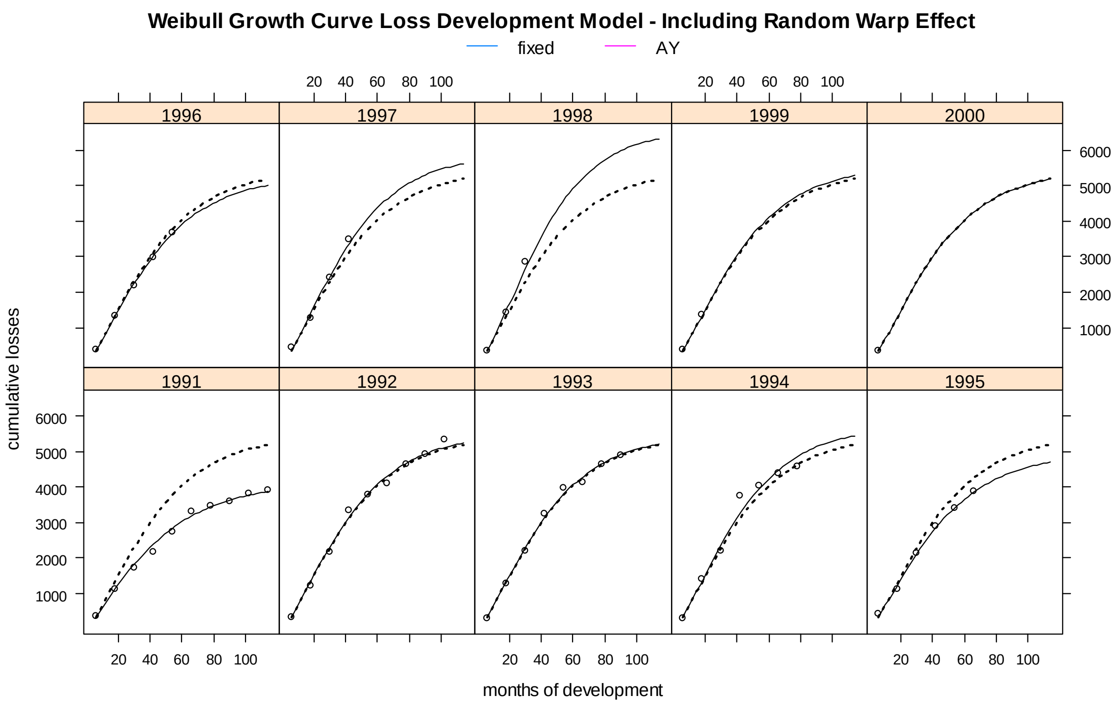
$$CL_{AY,dev} = ULT_{AY}\left[1-\exp\left(-(x/\theta)^{\omega}\right)\right]+\varepsilon_{AY,dev}$$
$$\begin{pmatrix} ULT_{AY} \\ \omega_{AY} \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_{ULT} \\ \mu_{\omega} \end{pmatrix}, \Sigma\right), \quad \Sigma = \begin{pmatrix} \sigma^2_{ULT} & \sigma_{ULT,\omega} \\ \sigma_{ULT,\omega} & \sigma^2_{\omega} \end{pmatrix}.$$
$$Var(\varepsilon_{AY,dev}) = \sigma^2 \hat{CL}_{AY,dev}$$

This model contains the two new hyperparameters σ_{ω} and $\sigma_{\omega,ULT}$ in addition to the baseline model's five hyperparameters. The resulting model parameters and associated loss reserve estimates are displayed below:

Model Parameters and Estimated Reserves

AY	*dev*	*omega*	*theta*	*growth*	*reported*	*eval120*	*eval240*	*ULT*	*reserves*
1991	114	1.189	47.202	95.8%	3,901	3,907	4,101	4,105	203
1992	102	1.313	47.202	93.5%	5,339	5,281	5,462	5,463	124
1993	90	1.311	47.202	90.2%	4,909	5,259	5,440	5,441	532
1994	78	1.332	47.202	85.5%	4,588	5,491	5,667	5,668	1,080
1995	66	1.265	47.202	78.8%	3,873	4,744	4,933	4,935	1,061
1996	54	1.292	47.202	69.7%	3,692	5,052	5,236	5,238	1,546
1997	42	1.347	47.202	57.6%	3,483	5,662	5,835	5,835	2,352
1998	30	1.410	47.202	42.5%	2,864	6,368	6,525	6,525	3,661
1999	18	1.317	47.202	24.7%	1,363	5,325	5,504	5,505	4,142
2000	6	1.308	47.202	6.5%	344	5,229	5,410	5,411	5,067
total								54,126	**19,768**

Note that the parameters in the "omega" column now vary by accident year. The expected ultimate reserves are approximately \$1M (5%) higher than those of the more parsimonious baseline model. It is interesting to note that nearly half of this increase comes from the increase in AY 1998 reserves from \$3.18M in the baseline model to \$3.661M here. At the same time the ω shape parameter for AY 1998 is 1.410 – the highest of all accident years. Allowing the shape parameter to vary by accident year therefore results in an accident year 1998 reserve estimate that is nearly as high as that of the chain ladder model's estimate.



The AY 1998 component of the above plot suggests that adding the random shape effect gives the most recent observation from AY 1998 more leverage over that accident year's growth curve. This might be appropriate – perhaps accident year 1998's claims are expected to be of higher ultimate severity. We can also note that the random shape model's AIC is 720.79, down from the baseline model's 725.76. This suggests that the random shape model offers a better tradeoff between complexity and goodness of fit.

Of course, it is equally possible that the most recent 1998 observation is an outlier, in which case we would want to mitigate its leverage on the ultimate loss estimate. Assuming the latter is correct, we would favor the baseline model over the random shape alternative model. For simplicity, we will continue to work with the baseline model.

Random Scale Effect: We can similarly allow the scale parameter θ to vary by accident year. Doing so causes the AIC measure to deteriorate from 725.76 to 729.76. Therefore on this dataset, allowing θ to vary by accident year does not offer a sufficient improvement in fit to justify the additional complexity. It is interesting to note that the (θ, σ_0) hyperparameters of this model are 46.6375 and 0.0000094, respectively. In other words the estimate of θ is nearly identical in the baseline and random scale models; and the estimated size of the random scale effect is negligible.

To summarize, where Clark's LDF model requires a separate ultimate loss parameter for each accident year, we allow ultimate loss (*ULT*) to randomly vary by accident year using a Gaussian sub-model. In addition, there is perhaps some justification for allowing the shape parameter (ω) to similarly vary by accident year. But doing so heightens the danger of overfitting the data. In the absence of compelling prior knowledge in support of including a random shape effect, one might be inclined to exclude it. Finally, the data indicates that the scale parameter (θ) does not vary by accident year. There is therefore no justification for including a random scale effect.

Loglogistic Growth Curves: Next, we can test the effect of replacing the Weibull growth curve with a Loglogistic growth curve: $G(x|\omega,\theta)=x^{\omega}/(x^{\omega}+\theta^{\omega})$. This is achieved by changing a single line of our R code:

```
start.vals <- c(ult=5000, omega=1.4, theta=45)
l1 <- nlme(cum ~ ult*(dev^omega)/((dev^omega) + (theta^omega))
            , fixed = list(ult~1, omega~1, theta ~ 1)
            , random = ult ~ 1 | AY
            , weights = varPower(fixed=.5)
            , data=dat, start = start.vals)
```

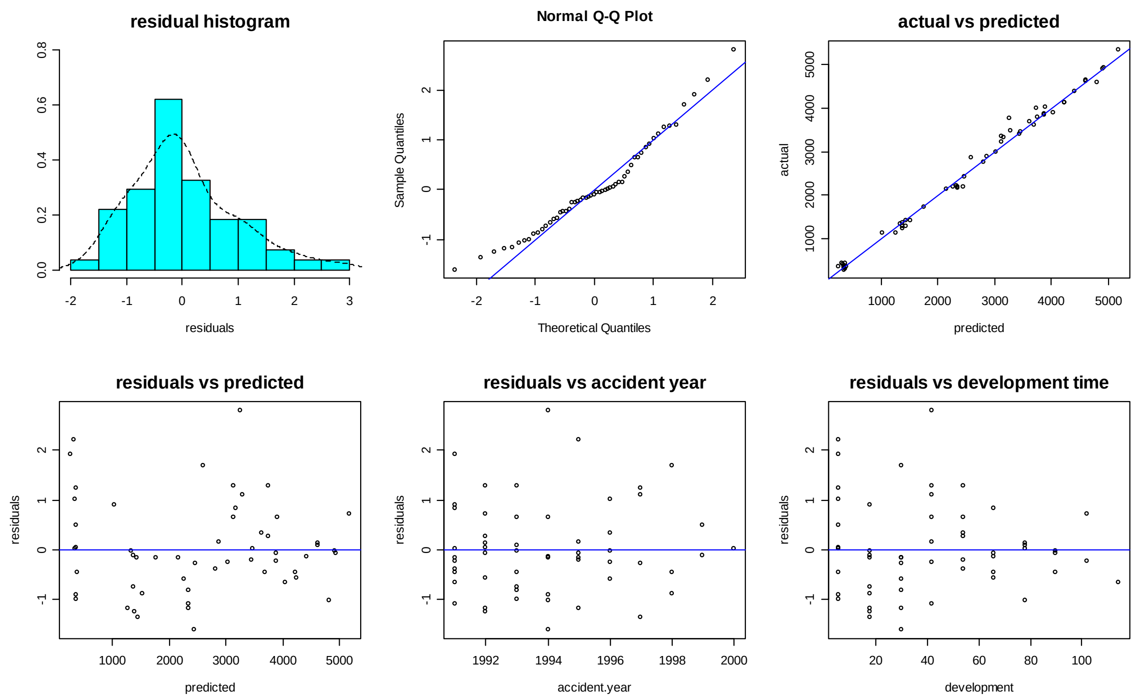
As with the baseline Weibull model, we allow only the *ULT* parameter to vary by accident year – no random shape or scale effects are included. The resulting hyperparameter estimates are:

$$\mu_{ULT} = 6898.3 \quad \omega = 1.403 \quad \theta = 49.14 \quad \sigma_{ULT} = 702.8 \quad \sigma = 3.109\,.$$

Note in passing that Clark reports ω and θ parameter estimates of 1.434 and 48.63, respectively for his Loglogistic LDF model.

Immediately we can see that the Loglogistic model will result in considerably higher loss reserve estimates than the Weibull model: the μ_{ULT} hyperparameter was 5306.6 for the Weibull model, compared with 6898.3 for the Loglogistic model.

The residual plots suggest that the Loglogistic model also fits the data fairly well. It is not clear from these plots that the Loglogistic model fits the data substantially better or worse than the baseline Weibull model.



The model parameters and expected loss reserves are displayed below:

Model Parameters and Estimated Reserves

AY	*dev*	*omega*	*theta*	*growth*	*reported*	*eval120*	*eval240*	*ULT*	*reserves*
1991	114	1.404	49.135	76.5%	3,901	4,099	4,756	5,269	1,368
1992	102	1.404	49.135	73.6%	5,339	5,471	6,348	7,034	1,694
1993	90	1.404	49.135	70.0%	4,909	5,458	6,333	7,017	2,107
1994	78	1.404	49.135	65.7%	4,588	5,696	6,609	7,322	2,734
1995	66	1.404	49.135	60.2%	3,873	5,020	5,825	6,454	2,580
1996	54	1.404	49.135	53.3%	3,692	5,294	6,142	6,805	3,113
1997	42	1.404	49.135	44.5%	3,483	5,742	6,662	7,381	3,898
1998	30	1.404	49.135	33.3%	2,864	6,055	7,026	7,784	4,920
1999	18	1.404	49.135	19.6%	1,363	5,454	6,329	7,012	5,648
2000	6	1.404	49.135	5.0%	344	5,372	6,234	6,906	6,562
total								68,984	**34,626**

Again, these results are broadly consistent with those reported by Clark. As anticipated, the estimated reserve amount – \$34.6M – is quite a bit higher than the \$18.7 estimated by the baseline Weibull model. But as Clark points out, one should be careful using a heavy-tailed model such as the Loglogistic to extrapolate too many years beyond the data. If, following Clark, we compute the reserves using losses projected to 240 months (the "eval240" column in the table above), the resulting reserve estimate is \$27.9M. Again, this is broadly consistent with Clark's result (\$28.9M). This is more realistic than using the Loglogistic model to extrapolate the results "to infinity." However, the result is still somewhat disconcerting: the reserve estimate after arbitrarily truncating the Loglogistic growth curve at 240 months is still nearly 50% higher than the corresponding Weibull models' reserve estimate.

The moral is that much hinges on the form of the growth curve one chooses for one's model. The advantage discussed by Pinheiro and Bates – validity of the model beyond the observed range of the data – is meaningful only to the extent that the model has been chosen wisely. In practice the considerations one would use to choose a growth curve are similar to considerations that are used in choosing a tail factor. The above display shows that, according to the Loglogistic model, the losses are only 76.5% developed as of 120 months. In contrast, the baseline Weibull model implies that the losses are 96% developed as of 120 months. One's general knowledge of how rapidly the types of claims being modeled develop should be considered when deciding which is the more appropriate growth curve, or whether additional growth curves should be investigated.

"Cape Cod" Models: If we have access to exposure information in addition to loss development data, it is easy to recast our hierarchical growth model into what might be called "Cape Cod" form. In the Cape Cod method, one assumes that expected ultimate loss ratio is constant across accident years and either estimates it from the data or simply introduces it as a model assumption. In the hierarchical modeling framework, we can dispense with the assumption that loss ratio is common across accident years. Rather, we can provide a sub-model for the various accident years' loss ratios, just as we provided a sub-model for the various accident years' ultimate losses in the baseline model. Still, we are acting in the original spirit of the Cape Cod method in the sense that we include the average loss ratio across all accident years as a model hyperparameter.

We will modify our original baseline Weibull model:

$$\begin{gathered} CL_{AY,dev} = ULT_{AY}\left[1-\exp\left(-(x/\theta)^{\omega}\right)\right] + \varepsilon_{AY,dev} \\ ULT_{AY} \sim N(\mu_{ULT}, \sigma^2_{ULT}) \\ Var(\varepsilon_{AY,dev}) = \sigma^2 \hat{CL}_{AY,dev} \end{gathered} .$$

The "Cape Cod" counterpart is:

$$\begin{gathered} CL_{AY,dev} = prem_{AY} LR_{AY}\left[1-\exp\left(-(x/\theta)^{\omega}\right)\right] + \varepsilon_{AY,dev} \\ LR_{AY} \sim N(\mu_{LR}, \sigma^2_{LR}) \\ Var(\varepsilon_{AY,dev}) = \sigma^2 \hat{CL}_{AY,dev} \end{gathered} .$$

In other words, we replace the hyperparameters $\{\mu_{ULT}, \sigma_{ULT}\}$ with $\{\mu_{LR}, \sigma_{LR}\}$.

The corresponding modification of our R code is equally minor:

```
prem <- seq(from=0, length=10, by=400) + 10000
prem <- rep(prem, 10:1)
start.vals <- c(lr=.5, omega=1.4, theta=45)
cc.w1 <- nlme(cum ~ prem*lr*(1 - exp(-(dev/theta)^omega))
            , fixed = list(lr~1, omega~1, theta ~ 1)
            , random = lr ~ 1 | AY
            , weights = varPower(fixed=.5)
            , data=dat, start = start.vals)
```

(Note that the loss triangle analyzed by Clark and others was originally not accompanied by premium information. Clark therefore assumed that the premium was \$10M in 1991 and increased by \$400,000 in each subsequent year. This is done in the first two lines of code above.)

Recall that the parameter estimates for the baseline Weibull model are:

$$\mu_{ULT} = 5306.6 \quad \omega = 1.306 \quad \theta = 46.64 \quad \sigma_{ULT} = 543.03 \quad \sigma = 2.955 .$$

In contrast the parameter estimates for the "Cape Cod" Weibull model are:

$$\mu_{LR} = 0.4634 \quad \omega = 1.317 \quad \theta = 49.91 \quad \sigma_{LR} = 0.0383 \quad \sigma = 2.977 .$$

It is comforting to note that the estimates of both process error (σ) and of the parameters determining the average shape of the loss development curve ({ω,θ}) are fairly consistent across both of these models.

Although it will not be reproduced here, the residual plot for the "Cape Cod" model is virtually identical to that of the baseline model. The various parameters and resulting loss reserve estimates for the Cape Cod Weibull model are displayed below:

Model Parameters and Estimated Reserves -- Cape Cod Weibull Model

AY	*dev*	*prem*	*omega*	*theta*	*lr*	*growth*	*reported*	*eval120*	*lr*prem*	*reserves*
1991	114	10,000	1.317	46.910	0.408	96.0%	3,901	3,952	4,082	181
1992	102	10,400	1.317	46.910	0.519	93.8%	5,339	5,229	5,401	62
1993	90	10,800	1.317	46.910	0.498	90.5%	4,909	5,208	5,380	470
1994	78	11,200	1.317	46.910	0.501	85.8%	4,588	5,433	5,611	1,023
1995	66	11,600	1.317	46.910	0.429	79.1%	3,873	4,818	4,977	1,103
1996	54	12,000	1.317	46.910	0.440	70.0%	3,692	5,114	5,283	1,591
1997	42	12,400	1.317	46.910	0.467	57.9%	3,483	5,608	5,792	2,309
1998	30	12,800	1.317	46.910	0.486	42.6%	2,864	6,016	6,215	3,350
1999	18	13,200	1.317	46.910	0.439	24.7%	1,363	5,613	5,798	4,435
2000	6	13,600	1.317	46.910	0.446	6.4%	344	5,871	6,064	5,720
total									54,604	**20,245**

The total reserves estimate by this model is $20.2M: about 8% higher than the baseline Weibull result. Most of the additional $1.5M of estimated reserves come the increased reserve estimates for accident years 1998-2000. This is an expected and sensible result. The ultimate loss estimates for the earlier accident years, where more loss development information is available, are less affected by the premium information. Conversely, the more recent the accident year, the less loss development data is available. Therefore, the ultimate loss estimates depend more heavily on the model's *LR* hyperparameter (the "Cape Cod" loss ratio estimate) together with the premium information.

Recall that Clark's Cape Cod model contains only three parameters (ω, θ, *ELR*) in contrast with his LDF model's 11 parameters. Because we are building hierarchical models there is not such a dramatic difference between our baseline model and its "Cape Cod" counterpart. Each of these models contains five hyperparameters.

Each of these models – the baseline and the Cape Cod variant – offers an advantage over its non-hierarchical counterpart:

- The hierarchical baseline model is less prone to overparameterization because it does require not a separate ultimate loss parameter for each accident year. The parameters $\{ULT_{1991}, \ldots, ULT_{2000}\}$ are replaced with the $\{\mu_{ULT},\sigma_{ULT}\}$ hyperparameters.

- The "Cape Cod" hierarchical model does not require one to assume a constant loss ratio across all accident years. This hierarchical model approaches the Clark Cape Cod model as the hyperparameter $\sigma_{LR} \rightarrow 0$.

We are not arguing that the hierarchical Cape Cod model is not an improvement on its baseline counterpart. On the contrary, the Cape Cod hierarchical model is preferable because adding exposure information will typically yield improved ultimate loss estimates, especially for more the recent, data-sparse, accident years. This, not greater parsimony, is the benefit it offers over the baseline hierarchical model.

Reserve Variability: Estimating the variability around a hierarchical growth model's loss reserve estimates (reserve variability) will be the topic of a future paper. For now a few brief comments must suffice. The problem of estimating reserve variability is twofold: we must estimate the variability resulting from the stochastic nature of the loss development process (process variance); and we must also estimate the variability resulting from the uncertainty around our models' hyperparameters (parameter variance). The future paper will outline a simulation-based approach to estimate the variability arising from both of these sources. In particular, Markov Chain Monte Carlo (MCMC) simulation, a technique widely used in contemporary Bayesian statistics, will be used to estimate parameter variance.

Of course, model risk – illustrated above by the dramatic effect that the choice of growth functions has on one's ultimate loss estimate – will remain a serious issue even after process and parameter variance have been accounted for.

4. CONCLUSION

Hierarchical modeling in actuarial science is an idea whose time has come. Hierarchical models encompass Bayesian credibility theory and therefore allow actuaries to perform credibility calculations within a statistical modeling framework. Moreover, hierarchical models allow one to easily integrate credibility concepts into one's GLM or non-linear modeling activities. By incorporating sub-models of various model parameters, hierarchical models allow one to strengthen an estimate for a sparsely populated segment of one's data by appropriately weighting it with the overall average estimate for the population as a whole. This integrates the fundamental insight of Bayesian credibility into a statistical modeling framework. For classification ratemaking and predictive modeling applications, actuaries can consider adding hierarchical structure to their generalized linear models in order to account for the variation along such "massively categorical" dimensions as territory or class code.

Turning to loss reserving, hierarchical modeling is useful in that it provides a natural way to analyze longitudinal (or "repeated measures") datasets. The point of view of this paper is that traditional loss reserving triangles can be viewed as longitudinal datasets in which each accident year is a "subject" and the cumulative or incremental losses at various development times constitute a series of repeated observations.

Unlike ratemaking and other general insurance predictive modeling applications, loss reserving is best approached using non-linear models. Following Clark, we have explored the use of the Weibull and Loglogistic growth curves for modeling the loss development process. We have done this in a non-linear hierarchical modeling (or "non-linear mixed effects models" – NLME) context. Hierarchical modeling allows us to specify sub-models for one or more of the parameters that determine the loss development process. The result is a natural and flexible framework in which to build parsimonious loss reserving models. Furthermore, the use of growth curves eliminates the need to specify arbitrary tail factors.

A Note Regarding Software

All models discussed in this paper were fit using the freely available R statistical computing package. R is available at http://www.r-project.org. Once the base R package has been installed, the multilevel modeling packages "lmer" and "lme" can easily be added.

Acknowledgments

I would like to thank Bill Stergiou for several enjoyable conversations about hierarchical modeling and also for introducing me to the Gelman-Hill textbook. I also thank Dave Clark, Keith Curley, Ravi Kumar, Glenn Meyers, Bill Stergiou, and Steve Talley for their many helpful comments on an earlier draft of this paper. When I have not taken their advice, it has been at my peril. Any errors are entirely my own.

Appendix: Raw Loss Triangle Data, as Imported into R

AY	dev	cum
1991	6	357.848
1991	18	1124.788
1991	30	1735.33
1991	42	2182.708
1991	54	2745.596
1991	66	3319.994
1991	78	3466.336
1991	90	3606.286
1991	102	3833.515
1991	114	3901.463
1992	6	352.118
1992	18	1236.139
1992	30	2170.033
1992	42	3353.322
1992	54	3799.067
1992	66	4120.063
1992	78	4647.867
1992	90	4914.039
1992	102	5339.085
1993	6	290.507
1993	18	1292.306
1993	30	2218.525
1993	42	3235.179
1993	54	3985.995
1993	66	4132.918
1993	78	4628.91
1993	90	4909.315
1994	6	310.608
1994	18	1418.858
1994	30	2195.047
1994	42	3757.447
1994	54	4029.929
1994	66	4381.982
1994	78	4588.268
1995	6	443.16
1995	18	1136.35
1995	30	2128.333
1995	42	2897.821
1995	54	3402.672
1995	66	3873.311
1996	6	396.132
1996	18	1333.217
1996	30	2180.715
1996	42	2985.752
1996	54	3691.712
1997	6	440.832
1997	18	1288.463
1997	30	2419.861
1997	42	3483.13
1998	6	359.48
1998	18	1421.128
1998	30	2864.498
1999	6	376.686
1999	18	1363.294
2000	6	344.014

5. REFERENCES

[1] Barnett, G., and B. Zehnwirth, "Best Estimates for Reserves," *PCAS* 2000, Vol. 87, 245-321.
[2] England, P. D., and R. J. Verrall, "Stochastic Claims Reserving in General Insurance," *British Actuarial Journal* 2002, Vol. 8, 443-544.
[3] Clark, D. R., "LDF Curve Fitting and Stochastic Loss Reserving: A Maximum Likelihood Approach," Casualty Actuarial Society *Forum*, Fall 2003, 41-91.
[4] Meyers, G., "Estimating Predictive Distributions for Loss Reserve Models" *Variance*, Vol 1, No. 2, Fall 2007, 248-272.
[5] Gelman, A. and J. Hill (2007). Data Analysis Using Regression and Multilevel/Hierarchical Models, Cambridge University Press, New York.
[6] Pinheiro, J. C. and D. M. Bates (2000). Mixed-Effects Models in S and S-Plus, Springer-Verlag, New York.
[7] Frees, E. (2006). Longitudinal and Panel Data Analysis and Applications in the Social Sciences, Cambridge University Press, New York.

A Stochastic Framework for Incremental Average Reserve Models

Roger M. Hayne, Ph.D., FCAS, MAAA

Abstract

Motivation. Chain ladder forecasts are notoriously volatile for immature exposure periods. The Bornhuetter-Ferguson method is one commonly used alternative but needs a priori estimates of ultimate losses. Berquist and Sherman presented another alternative that used claim counts as an exposure base and used trended incremental severities to "square the triangle." A significant advantage of the Berquist and Sherman method is the simultaneous estimate of underlying inflation. Though not the first to do so, this paper looks to extend the incremental severity method to a stochastic environment. Rather than using logarithmic transforms or (generalized) linear models, used in many other approaches, we use maximum likelihood estimators, bringing to bear the strength of that approach avoiding limiting assumptions necessitated when taking logarithms.

Method. Given that incremental severities can be looked at as averages over a number of claims, the law of large numbers would suggest those averages follow an approximately normal distribution. We then assume the variance of the incremental payments in a cell are proportional to a power of the mean (with the constant of proportionality and power constant over the triangle). We then use maximum likelihood estimators (MLEs) to estimate the incremental severities, along with the inherent claims inflation to "square the triangle." We also use properties of MLEs to estimate the variance-covariance matrix of the parameters, giving not only estimates of process but also of parameter uncertainty for this method. Not only do we consider the model described by Berquist and Sherman, but we also set the presentation in a more general framework that can be applied to a wide range of potential underlying models.

Results. A reasonably common and powerful method now presented in a stochastic framework allowing for assessment of variability in the forecasts of the method.

Availability. The R script for these estimates appear on the CAS Web Site.

Keywords. Stochastic reserving, maximum likelihood, normal-*p*, incremental severity method, PPCI

1. INTRODUCTION

The chain ladder method has long been recognized as leading to potentially volatile forecasts for immature exposure periods. As a result, other methods that depended on information in addition to the amounts to date were soon used to augment the chain ladder method for less mature ages. These methods include the Bornhuetter-Ferguson method [1], incremental severity methods shown in Berquist and Sherman [2], and the operational time models from Wright [3], among others. In effect, these approaches replace the multiplicative model inherent in the chain ladder with additive increments. The Bornhuetter-Ferguson method looks to historical development and an a priori estimate of ultimate losses to derive these additive increments, while the incremental severity method considers incremental average costs per ultimate claim (or other unit of exposure) and a measure of inflationary trend to derive these increments. In the discussion by Berquist and Sherman, the trend itself is estimated from the data.

Thus by adding a single parameter trend to be estimated from the data, Berquist and Sherman avoided assumptions about the relative adequacy of pricing or the need of deriving a priori ultimate loss estimates by exposure year. Of course, they do require a measure of relative exposure, usually claim counts.

There has been much published about stochastic generalizations of the chain ladder method. Verall and England [4] presents a very nice summary. We will not touch on those here, but rather attempt to re-cast the incremental severity method in a stochastic light.

In the present paper we first consider the incremental severity method in a stochastic framework. We note that the incremental severities are themselves averages over a number of observations and, as a result of the law of large numbers, would likely have a distribution that is asymptotically normal. This is a very significant observation and was made by Stelljes [5] and provides a bit of support to at least one answer to the question of what statistical model to use. Stelljes assumes that the development pattern follows a mixed exponential over time and does not measure the trend inherent in the data.

We however, start with the classic incremental severity model (allowing for different averages at each age) but measure the inflation inherent in the loss experience. Not only does this allow for a broader range of runoff curves, it also allows for systematic negative incremental amounts, making it possible to model not only paid amounts (net of recoverable) but also incurred amounts. In addition, rather than making somewhat restrictive assumptions about the underlying variance structure as present in Stelljes that allows the use of non-linear regression, we will take a somewhat more general approach of maximum likelihood estimators allowing more flexible assumptions regarding the underlying variance structure.

In this paper we not only derive parameter estimates for our model, including inherent trend, but also estimates of the standard deviation of those parameter estimates, often called the standard error of the parameters. The standard error can be used to measure the significance of the parameter as well as the parameter uncertainty inherent in the forecasts of this model. We also derive estimates of the distribution of outcomes for this model, not to be confused with the distribution of potential outcomes for the liabilities under review.

1.1 Research Context

In the context of reserves for a book of liabilities at a point in time, there is a wide range of possible outcomes, some of which may be more likely than others. We call this entire range of outcomes along with their likelihoods the "distribution of outcomes" for the liabilities under consideration. This observation seems to have pervaded the analysis of reserves for decades. Traditional reserving approaches, although relying on deterministic methods, usually had the actuary applying a variety of those methods with the unstated goal of providing at least a subjective view of the distribution of outcomes, or at least the portion of that distribution that contained "reasonable estimates."

More recently, though, questions of just how "good" the "reasonable estimates" were led to consideration of stochastic methods to rigorously quantify that uncertainty. Statements such as "My selection for unpaid liabilities is $a million. In my view it is just as likely that the ultimate unpaid liabilities will be between $x million and $y million as outside that range and in addition, it is very unlikely that the ultimate unpaid liabilities will be below $w million or above $z million" provide much more useful information to a principal than "My best estimate is $a million and I believe a range of reasonable estimates is between $b million and $c million." Because of this there has been increased focus on models that will assist the actuary in estimating the distribution of outcomes.

Just as no traditional reserve method completely captures all the complexities possible for all lines of business, it is not likely that any current stochastic model can capture all those complexities. Because of this, results presented here should <u>not</u> be interpreted as estimates of the distribution of outcomes, but rather the distribution of possibilities <u>under the specific assumptions of the single model we present</u>.

1.2 Objective

The incremental average cost method has long been a very powerful alternative to the chain ladder method that can be quite volatile for more immature exposure periods. The Cape Cod and Bornhuetter-Ferguson methods are often used as alternatives that try to overcome this problem. There has been research setting all of these methods in stochastic frameworks. Our objective is to take another powerful alternative to the chain ladder method, the incremental average loss method presented by Berquist and Sherman [2], and set it into a stochastic framework.

One substantial contribution of the Berquist and Sherman approach is the estimation of trend in the averages from the averages themselves. This is in contrast to the necessary external trend usually necessary in stochastic versions of both the Bornhuetter-Ferguson and Cape Cod methods.

Another weakness of many stochastic generalizations of traditional methods is the necessity of assumptions about the form of the distributions used. Because of the central limit theorem, averages of independent samples from a distribution are asymptotically normal, thus suggesting a form for the distributions in the stochastic model.

Another inherent limitation of most stochastic generalizations is the necessity of assuming all incremental amounts are positive. This limits the generalization of those methods in the case of incurred losses, or in the case of consistent downward paid development. The use of the normal distribution allows more flexibility in handling consistent negative incremental averages.

The goal of this paper is to set the traditional incremental average method in a stochastic framework taking advantage of the ease of computation afforded by the normal distribution and ability to handle negative values. In addition to moving the average cost method into a stochastic framework, this paper also shows the relative ease of moving to a completely non-linear environment, thereby avoiding the constraints inherent in linear or generalized linear methods, echoing the comments of Venter in several venues, including [7].

1.3 Outline

In Section 2 we set out our stochastic generalization of the incremental average method presented in Berquist and Sherman [2]. Section 3 discusses the results of applying these methods to the adjusted paid automobile bodily injury liability data in that paper. We present our conclusions in Section 4 with Appendix A showing the derivatives used in the estimation along with the R script that we used in the calculations.

2. BACKGROUND AND METHODS

Klugman, Panjer, and Willmot [6] present a very clear and concise discussion of maximum likelihood estimates (MLEs). We will make use of that approach in this paper.

For this paper C_{ij} denotes payments made or the change in incurred losses (defined as payments plus case reserve estimates) for exposure (policy, accident, underwriting, etc.) period (year, quarter, month, etc.) i during development period j. For convenience here we will assume the same frequency for both i and j, and hence the resulting development triangle will have the same number of rows as columns, denoted as n here. Without loss of generality, we will talk in terms of accident and development years.

For each accident year we have some measure of loss exposure, either an exposure count or an estimate of ultimate claim counts. Exposure count, such as earned car years for automobile coverages, generally does not require estimation. The same cannot be said for claim counts that must be estimated and hence should be treated as random variables. We will not make that generalization here but rather leave it as a future project.

We do note that, just as there are a number of models that can be used to estimate ultimate loss amounts, there are a number of approaches that can be used to estimate the ultimate number of claims. If the number of reported counts is deemed to be a reliable and stable base, that is, if there has been no change in the definition or nature of reported claims during the experience period under consideration, they often provide a measure of exposure that matures more quickly than losses and hence those estimates will likely have less inherent uncertainty, i.e., lower standard error, than losses. It might well be that consideration of both chain ladder estimates and those of an incremental average frequency method, such as presented here applied to claim counts, using earned exposures as an exposure base, could provide reasonable estimate of ultimate reported counts for use here.

In any event, we will denote this measure of relative exposure as E_i for accident year i. We will thus focus on the incremental averages A_{ij} defined by equation (2.1).

$$A_{ij} = \frac{C_{ij}}{E_i}. \tag{2.1}$$

The traditional incremental severity method then "squares the triangle" with trended averages as in equation (2.2).

$$A_{ij} = \alpha_j \tau^i,\ i = 2, 3, \ldots, n;\ j = n - i + 2, \ldots, n. \tag{2.2}$$

We will effectively take this same approach to frame a stochastic model based on this method. It is not unusual, see for example Venter [6], to assume that the variance of the incremental amounts is a power of their expected value. We will take this same approach. However, since we will allow the expected values to be negative we will, without loss of generality, we take the variance to be a power of the square of the mean. Also we are taking the constant of proportionality among the variances as an exponential, thereby allowing the parameter to take on any value. However, we note that the variance of the average of n items is inversely proportional to the number of items so we further adjust our assumed variances to reflect the potential for a different number of exposures or claims in the various accident years. For this we let e denote the number of exposures or claims for the year. Following the notation in [6] we will assume the relationships in (2.3), suppressing subscripts for the moment.

$$\begin{aligned}&\mathrm{E}(A)=\mu.\\&\mathrm{Var}(A)=e^{\kappa-e}\mu^{2p}.\end{aligned}\tag{2.3}$$

Now, since the A_{ij} are averages, the law of large numbers implies that they are asymptotically normal with parameters given in (2.4), again suppressing subscripts for the moment.

$$A\sim \mathrm{N}\left(\mu,e^{\kappa-e}\mu^{2p}\right).\tag{2.4}$$

Since we are concerned with maximum likelihood estimates, the negative log likelihood for this distribution will be key to our analysis. Since we have a normal distribution the likelihood function is relatively simple and given by (2.5).

$$\mathrm{f}\left(x;\mu,\kappa,p\right)=\frac{1}{\sqrt{2\pi e^{\kappa-e}\mu^{2p}}}e^{-\frac{(x-\mu)^2}{2e^{\kappa-e}\mu^{2p}}}.\tag{2.5}$$

This gives a negative log likelihood for a single variable given in (2.6).

$$\begin{aligned}\mathrm{l}\left(x;\mu,\kappa,p\right)&=-\ln\left(\mathrm{f}\left(x;\mu,\kappa,p\right)\right)\\&=-\ln\left(\frac{1}{\sqrt{2\pi e^{\kappa-e}\mu^{2p}}}e^{-\frac{(x-\mu)^2}{2e^{\kappa-e}\mu^{2p}}}\right)\\&=\frac{1}{2}\left(\kappa-e+\ln\left(2\pi\mu^{2p}\right)\right)+\frac{\left(x-\mu\right)^2}{2e^{\kappa-e}\mu^{2p}}.\end{aligned}\tag{2.6}$$

We note the incremental amounts A_{ij} under consideration are averages of a number of observations. If we assume the observations are themselves independent, then the central limit theorem would imply that they have asymptotically normal distributions. For this reason we will assume that the A_{ij} variables are all independent and have normal distributions. We generalize the incremental severity model with the parametric model shown in (2.7).

$$A_{ij}\sim \mathrm{N}\left(\alpha_j\tau^i,e^{\kappa-e_i}\left(\alpha_j\tau^i\right)^{2p}\right).\tag{2.7}$$

With observations in a typical loss triangle we get the negative log likelihood function given in (2.8).

$$\mathrm{l}\left(A_{11},A_{12},\ldots,A_{n1};\alpha_1,\alpha_2,\ldots,\alpha_n,\tau,\kappa,p\right)=\sum_{(i,j)\in S}\frac{\kappa-e_i+\ln\left(2\pi\left(\alpha_j\tau^i\right)^{2p}\right)}{2}+\frac{\left(A_{ij}-\alpha_j\tau^i\right)^2}{2e^{\kappa-e_i}\left(\alpha_j\tau^i\right)^{2p}}.\tag{2.8}$$

The set S in (2.8) denotes the set of all index pairs for which data are available. If the data were available in a full triangle, with n rows and n columns then S would follow the form given in (2.9).

$$S = \{(i, j) | i = 1, 2, \dots, n, \ j = 1, 2, \dots, n - i + 1\}. \tag{2.9}$$

However, we will not restrict ourselves to this regular case. We also note in formula (2.8) the e_i values are known constants (the natural logs of the number of exposures for accident year i, not parameters to be estimated.

Once parameters that minimize the negative log likelihood function are determined, then it is straight-forward to obtain estimates of the distribution of outcomes under the assumption that this model and the resulting parameters completely describe the loss emergence phenomenon. Let us denote these estimates by $\hat{\alpha}_k$, $\hat{\kappa}$, $\hat{\tau}$, and $\hat{p}$. Under our assumptions we can now conclude that the distribution of average future payments for each year is given by (2.10).

$$R_i \sim \mathrm{N}\left(\sum_{j=n-i+2}^{n} \hat{\alpha}_j \hat{\tau}^i, \sum_{j=n-i+2}^{n} e^{\hat{\kappa} - e_i} \left(\hat{\alpha}_j \hat{\tau}^i \right)^{2\hat{p}} \right). \tag{2.10}$$

This then gives the effect of process uncertainty on the total forecast incremental severity by accident year. This does not, however, address the issue of parameter uncertainty. Just as the standard error provides insight into parameter uncertainty in usual regression applications, the information matrix can be helpful in estimating the variance-covariance matrix of the parameters. For this, we first define the Fisher Information Matrix as the matrix of expected values of the Hessian of the negative log likelihood function. That is, the matrix whose element in i^{th} row and j^{th} column is the second derivative of the negative log likelihood function, once with respect to the i^{th} variable and once with respect to the j^{th}. We show these expectations, along with both the elements of the gradient and Hessian of the negative log likelihood function in the appendix to this paper. The inverse of the information matrix is then an approximation for the variance-covariance matrix for the parameters.

Since the mean and variance for individual incremental averages are functions of the parameters, we elected to estimate the distribution of future amounts both by exposure period and in total using simulation. For this we first selected the parameters from a multivariate normal distribution with expected values equal to the MLE estimates and variance-covariance matrix equal to the inverse of the information matrix. Given those parameters, we then randomly selected future incremental averages in each cell using the relationship in (2.7). We added up the indications by exposure year and multiplied by the denominator (claim count or exposure count) to obtain a single observation for an exposure year and then added all those simulations together to get a single observation of the total future amount.

At this juncture if we wished to assume that claim counts, instead of being deterministic, were themselves stochastic, but independent of the incremental severities, we could simulate the ultimate number of claims by exposure year at this juncture to add a provision for uncertainty in those estimates in the final forecast.

3. RESULTS AND DISCUSSION

As an example of this model, the top portion of Exhibit 1 shows the incremental averages based on automobile bodily injury liability data from Berquist and Sherman [2]. The last column is the forecast ultimate claim counts from Exhibit J of that paper. The incremental severities are based on adjusted paid losses in Exhibit N divided by these claim count estimates.

The bottom portion of Exhibit 1 shows the parameter estimates derived by minimizing the negative log likelihood function shown in (2.8). Shown in the "standard error" row is the square root of the diagonal of the approximate parameter variance-covariance matrix.

Exhibit 2 shows scatter plots of the standardized residuals from the fitted model, calculated as the ratio of the difference between the historical average minus the expected average from the model, divided by the estimated standard deviation by cell. The first three charts show the residuals first by calendar year, then by accident year, and finally by development lag. The last histogram shows the simulated range of forecasts from 25,000 simulations. The line on that histogram presents the distribution assuming independence and the mean and variance by cell implied by the parameter estimates.

Exhibit 3 shows the expected averages and related variances by cell indicated by the estimated parameters and the model shown in (2.7). Exhibit 4 shows the indicated mean forecast and standard deviation by accident year and for all years combined. Exhibit 4 also shows the forecasts for the next calendar year, both with and without parameter uncertainty. These estimates can be used to assess how well emerging experience fits with what is forecast by the model, a critical test for the on-going application of just about any model.

Since the model in (2.7) assumes the incremental averages are independent, the future average forecast is simply the sum of the future indications by accident year, as is the variance for the future forecast, assuming process uncertainty only. The resulting means and standard deviations, after multiplication by the number of claims are shown under the "Process Only" columns.

The remaining columns summarize the results of the simulation. We first randomly simulated a selection of parameters given the parameter estimates and the approximate variance-covariance matrix, using a multivariate normal distribution. Given those parameters, we then randomly simulated individual incremental averages by cell using a normal distribution with the mean and standard deviation shown in (2.7). We then totaled the results for one simulation to derive both the simulated future average estimates by accident year and then, after multiplying by claim counts, the total indicated future amounts. The averages and standard deviations in the right portion of that exhibit represent the mean and standard deviation of the simulated amounts as are the fifth percentile and 95th percentile (the 90% probability interval) for the simulations. These last columns thus present an estimate of the distribution of possible forecasts from this model, given the loss data in the top of Exhibit 1.

As can be seen, parameter uncertainty clearly contributes substantially to the uncertainty in the forecasts for this model. The standard deviation including parameter uncertainty is nearly three times that for process uncertainty only. In addition, as one would expect there are correlations in the forecasts among accident years, particularly since the forecast for an accident year depends not only on the losses for that year but also on the losses and forecasts for previous years. If the accident years were independent, then the standard deviation for the total would equal the square root of the sum of the squares of the standard deviations for the various years. That calculation yields approximately 1.1 million, compared with the final 1.5 million shown in Exhibit 4.

Although we do not show the results of the calculations, the model and estimation process reacts as one should expect with negative values. A simple test would simply replace the incrementals in a column with their negatives. When doing this all values of the parameters and variance-covariance matrix remain unchanged, except with a sign change in the parameter estimates and covariances related to the affected column.

The R script used to derive these estimates are also shown in Appendix A. Generally the approach is quite straight forward. Key to deriving the estimates is the function R nlminb. As with many optimization routines, this function requires a starting value. In this case, we first selected a starting value for τ as the trend in the averages for the first development period (unless that trend generates an error, in which case we selected 1.03). We then estimated the initial α_j values as the averages of the averages, discounted at the initial τ estimate, and selected the initial values for κ and p as the natural logarithm of the largest exposure number and 1.5, respectively (somewhat arbitrarily).

This R function also allows for different iteration increments for the various variables to be optimized. Users should consult the documentation that accompanies R for this function. We selected relative scaling among variables inversely proportional to the initial averages for the α_j variables and five for the remaining three.

4. CONCLUSIONS

Although we focused on a very simple model of incremental averages, nothing in what we have done relies on the specific structure of the underlying model. This is in contrast to many stochastic approaches that require non-negative incrementals, and the necessity of making additional assumptions about the distributions of the incremental amounts. The framework we chose, along with the central limit theorem, suggests the normal distribution for the incremental averages.

As shown in (2.8), this distribution leads to a rather convenient form for the negative log likelihood function. Together with the ability to differentiate the assumed model for the average and resulting standard deviation makes this approach easily expandable to other models for the incremental averages. Coupled with powerful, reasonably easy-to-use, and affordable statistical software such as the language R, actuaries now have quite flexible tools to use to expand the models used in estimating future losses, even beyond the simple model presented here.

Supplementary Material

The R script used for these calculations is stored electronically on the CAS Web Site.

Appendix A

In order to derive estimates of parameter uncertainty we need the matrix of second derivatives of the negative log likelihood function. In this appendix we list those derivatives.

Recall from (2.8) the negative log likelihood function is given by

$$1\left(A_{11}, A_{12}, \ldots, A_{n1}; \alpha_1, \alpha_2, \ldots, \alpha_n, \tau, \kappa, p\right) = \sum_{(i,j)\in S} \frac{\kappa - e_i + \ln\left(2\pi\left(\alpha_j \tau^i\right)^{2p}\right)}{2} + \frac{\left(A_{ij} - \alpha_j \tau^i\right)^2}{2e^{\kappa - e_i}\left(\alpha_j \tau^i\right)^{2p}}.$$

Suppressing arguments and parameters we thus have the following first partial derivatives:

$$\frac{\partial 1}{\partial \alpha_k} = \sum_{i\in\{i|(i,k)\in S\}} \frac{p}{\alpha_k} - \frac{\alpha_k\tau^i\left(A_{ik}-\alpha_k\tau^i\right)+p\left(A_{ik}-\alpha_k\tau^i\right)^2}{e^{\kappa-e_i}\alpha_k\left(\alpha_k\tau^i\right)^{2p}}.$$

$$\frac{\partial 1}{\partial \kappa} = \sum_{(i,j)\in S} \frac{1}{2}\left(1-\frac{\left(A_{ij}-\alpha_j\tau^i\right)^2}{e^{\kappa-e_i}\left(\alpha_j\tau^i\right)^{2p}}\right).$$

$$\frac{\partial 1}{\partial \tau} = \sum_{(i,j)\in S} \frac{pi}{\tau} - \frac{i\alpha_j\tau^i\left(A_{ij}-\alpha_j\tau^i\right)+pi\left(A_{ij}-\alpha_j\tau^i\right)^2}{\tau e^{\kappa-e_i}\left(\alpha_j\tau^i\right)^{2p}}.$$

$$\frac{\partial 1}{\partial p} = \sum_{(i,j)\in S} \ln\left(\alpha_j\tau^i\right)\left(1-\frac{\left(A_{ij}-\alpha_j\tau^i\right)^2}{e^{\kappa-e_i}\left(\alpha_j\tau^i\right)^{2p}}\right).$$

These then give the following second derivatives:

$$\frac{\partial^2 1}{\partial \alpha_k^2} = \sum_{i\in\{i|(i,k)\in S\}} \left(\frac{\alpha_k^2\tau^{2i}+4p\alpha_k\tau^i\left(A_{ik}-\alpha_k\tau^i\right)+p(2p+1)\left(A_{ik}-\alpha_k\tau^i\right)^2}{\alpha_k^2 e^{\kappa-e_i}\left(\alpha_k\tau^i\right)^{2p}} - \frac{p}{\alpha_k^2}\right).$$

$$\frac{\partial^2 1}{\partial \alpha_k \partial \alpha_m} = \frac{\partial^2 1}{\partial \alpha_m \partial \alpha_k} = 0, m \neq k.$$

$$\frac{\partial^2 1}{\partial \kappa \partial \alpha_k} = \frac{\partial^2 1}{\partial \alpha_k \partial \kappa} = \sum_{i\in\{i|(i,k)\in S\}} \frac{\alpha_k\tau^i\left(A_{ik}-\alpha_k\tau^i\right)+p\left(A_{ik}-\alpha_k\tau^i\right)^2}{e^{\kappa-e_i}\alpha_k\left(\alpha_k\tau^i\right)^{2p}}.$$

$$\frac{\partial^2 1}{\partial \tau \partial \alpha_k} = \frac{\partial^2 1}{\partial \alpha_k \partial \tau} = \sum_{i\in\{i|(i,k)\in S\}} \frac{i\alpha_k^2\tau^{2i}+(4p-1)i\alpha_k\tau^i\left(A_{ik}-\alpha_k\tau^i\right)+2p^2 i\left(A_{ik}-\alpha_k\tau^i\right)^2}{e^{\kappa-e_i}\tau\alpha_k\left(\alpha_k\tau^i\right)^{2p}}.$$

$$\frac{\partial^2 1}{\partial p \partial \alpha_k} = \frac{\partial^2 1}{\partial \alpha_k \partial p} = \sum_{i\in\{i|(i,k)\in S\}} \frac{1}{\alpha_k} + \frac{2\ln\left(\alpha_k\tau^i\right)\alpha_k\tau^i\left(A_{ik}-\alpha_k\tau^i\right)+\left(2p\ln\left(\alpha_k\tau^i\right)-1\right)\left(A_{ik}-\alpha_k\tau^i\right)^2}{e^{\kappa-e_i}\alpha_k\left(\alpha_k\tau^i\right)^{2p}}.$$

$$\frac{\partial^2 1}{\partial \kappa^2} = \sum_{(i,j)\in S} \frac{\left(A_{ij}-\alpha_j\tau^i\right)^2}{e^{\kappa-e_i}\left(\alpha_j\tau^i\right)^{2p}}.$$

$$\frac{\partial^2 1}{\partial \tau \partial \kappa} = \frac{\partial^2 1}{\partial \kappa \partial \tau} = \sum_{(i,j)\in S} \frac{i\alpha_j\tau^i\left(A_{ij}-\alpha_j\tau^i\right)+pi\left(A_{ij}-\alpha_j\tau^i\right)^2}{\tau e^{\kappa-e_i}\left(\alpha_j\tau^i\right)^{2p}}.$$

$$\frac{\partial^2 l}{\partial p \partial \kappa} = \frac{\partial^2 l}{\partial \kappa \partial p} = \sum_{(i,j)\in S} \frac{\ln\left(\alpha_j \tau^i\right)\left(A_{ij} - \alpha_j \tau^i\right)^2}{e^{\kappa - e_i}\left(\alpha_j \tau^i\right)^{2p}}.$$

$$\frac{\partial^2 l}{\partial \tau^2} = \sum_{(i,j)\in S} \frac{i^2 \alpha_j^2 \tau^{2i} + \left(4ip - i + 1\right) i \alpha_j \tau^i \left(A_{ij} - \alpha_j \tau^i\right) + (2ip+1) pi \left(A_{ij} - \alpha_j \tau^i\right)^2}{\tau^2 e^{\kappa - e_i}\left(\alpha_j \tau^i\right)^{2p}} - \frac{pi}{\tau^2}.$$

$$\frac{\partial^2 l}{\partial \tau \partial p} = \frac{\partial^2 l}{\partial p \partial \tau} = \sum_{(i,j)\in S} \frac{i}{\tau} + \frac{2i\alpha_j \tau^i \ln\left(\alpha_j \tau^i\right)\left(A_{ij} - \alpha_j \tau^i\right) + \left(2pi \ln\left(\alpha_j \tau^i\right) - 1\right)\left(A_{ij} - \alpha_j \tau^i\right)^2}{\tau e^{\kappa - e_i}\left(\alpha_j \tau^i\right)^{2p}}.$$

$$\frac{\partial^2 l}{\partial p^2} = \sum_{(i,j)\in S} \frac{2\ln^2\left(\alpha_j \tau^i\right)\left(A_{ij} - \alpha_j \tau^i\right)^2}{e^{\kappa - e_i}\left(\alpha_j \tau^i\right)^{2p}}.$$

The information matrix then requires the expected values of these derivatives. To this end recall that because of (2.7) we have the following relationships:

$$\mathrm{E}\left(A_{ij}\right) = \alpha_j \tau^i, \text{and}$$
$$\mathrm{Var}\left(A_{ij}\right) = \mathrm{E}\left(\left(A_{ij} - \alpha_j \tau^i\right)^2\right) = e^{\kappa - e_i}\left(\alpha_j \tau^i\right)^{2p}.$$

We can then derive the entries of the information matrix as follows:

$$\mathrm{E}\left(\frac{\partial^2 l}{\partial \alpha_k^2}\right) = \frac{2p^2}{\alpha_k^2} \sum_{i\in\{i|(i,k)\in S\}} 1 + \sum_{i\in\{i|(i,k)\in S\}} \frac{1}{e^{\kappa - e_i}\left(\alpha_k^2\right)^p \tau^{2i(p-1)}}.$$

$$\mathrm{E}\left(\frac{\partial^2 l}{\partial \alpha_k \partial \alpha_m}\right) = 0, m \neq k.$$

$$\mathrm{E}\left(\frac{\partial^2 l}{\partial \kappa \partial \alpha_k}\right) = \mathrm{E}\left(\frac{\partial^2 l}{\partial \alpha_k \partial \kappa}\right) = \frac{p}{\alpha_k} \sum_{i\in\{i|(i,k)\in S\}} 1.$$

$$\mathrm{E}\left(\frac{\partial^2 l}{\partial \tau \partial \alpha_k}\right) = \mathrm{E}\left(\frac{\partial^2 l}{\partial \alpha_k \partial \tau}\right) = \frac{2p^2}{\tau \alpha_k} \sum_{i\in\{i|(i,k)\in S\}} i + \frac{1}{\tau \alpha_k} \sum_{i\in\{i|(i,k)\in S\}} \frac{i}{e^{\kappa - e_i}\left(\alpha_k^2\right)^{p-1} \tau^{2i(p-1)}}.$$

$$\mathrm{E}\left(\frac{\partial^2 l}{\partial p \partial \alpha_k}\right) = \mathrm{E}\left(\frac{\partial^2 l}{\partial \alpha_k \partial p}\right) = \frac{p\ln\left(\alpha_k^2\right)}{\alpha_k} \sum_{i\in\{i|(i,k)\in S\}} 1 + \frac{2p\ln\left(\tau\right)}{\alpha_k} \sum_{i\in\{i|(i,k)\in S\}} i.$$

$$\mathrm{E}\left(\frac{\partial^2 l}{\partial \kappa^2}\right) = \sum_{(i,j)\in S} 1.$$

$$\mathrm{E}\left(\frac{\partial^2 l}{\partial \tau \partial \kappa}\right) = \mathrm{E}\left(\frac{\partial^2 l}{\partial \kappa \partial \tau}\right) = \frac{p}{\tau} \sum_{(i,j)\in S} i.$$

$$\mathrm{E}\left(\frac{\partial^2 l}{\partial p \partial \kappa}\right) = \mathrm{E}\left(\frac{\partial^2 l}{\partial \kappa \partial p}\right) = \sum_{(i,j)\in S} \frac{\ln\left(\alpha_j^2 \tau^{2i}\right)}{2}.$$

$$\mathrm{E}\left(\frac{\partial^2 l}{\partial \tau^2}\right) = \frac{2p^2}{\tau^2} \sum_{(i,j)\in S} i^2 + \sum_{(i,j)\in S} \frac{i^2}{e^{\kappa - e_i} \tau^{2+2i(p-1)} \left(\alpha_i^2\right)^{p-1}}.$$

$$\mathrm{E}\left(\frac{\partial^2 l}{\partial \tau \partial p}\right) = \mathrm{E}\left(\frac{\partial^2 l}{\partial p \partial \tau}\right) = \sum_{(i,j)\in S} \frac{i-1}{\tau} + \frac{p}{\tau}\left(\sum_{(i,j)\in S} i \ln\left(\alpha_j^2\right) + \sum_{(i,j)\in S} i \ln\left(\tau^{2i}\right)\right).$$

$$\mathrm{E}\left(\frac{\partial^2 l}{\partial p^2}\right) = \sum_{(i,j)\in S} \frac{\ln^2\left(\alpha_j^2 \tau^{2i}\right)}{2}.$$

The calculations in this paper made use of the following R script:

```
library(mvtnorm)
library(MASS)

A0=matrix(c(178.73,361.03,283.69,264.00,137.94,61.49,15.47,8.82,
  196.56,393.24,314.62,266.89,132.46,49.57,33.66,NA,
  194.77,425.13,342.91,269.45,131.66,66.73,NA,NA,
  226.11,509.39,403.20,289.89,158.93,NA,NA,NA,
  263.09,559.85,422.42,347.76,NA,NA,NA,NA,
  286.81,633.67,586.68,NA,NA,NA,NA,NA,
  329.96,804.75,NA,NA,NA,NA,NA,NA,
  368.84,NA,NA,NA,NA,NA,NA,NA),8,8,byrow=TRUE)
dnom=c(7822,8674,9950,9690,9590,7810,8092,7594)

# Input (A0) is a development array of incremental averages with a the
# exposures (claims) used in the denominator appended as the last column.
# Assumption is for the same development increments as exposure
# increments and that all development lags with no development have #
# been removed.  Data elements that are not available are indicated as
# such.  This should work (but not tested for) just about any subset of
# an upper triangular data matrix.  Another requirement of this code is
# that the matrix contain no columns that are all zero.

# Matrix shape, m rows, n columns
m=(nrow(A0))[1]
n=(ncol(A0))[1]

# Generate a matrix to reflect exposure count in the variance structure
logd=log(matrix(dnom,m,n))

# Set up matrix of rows and columns, makes later calculations simpler
```

```
r=row(A0)
c=col(A0)


# msk is a mask matrix of allowable data, upper triangular assuming same
# development increments as exposure increments, msn picks off the first
# forecast diagonal
msk=(m-r)>=c-1
msn=(m-r)==c-2

# Negative loglikelihood function, to be minimized
l.obj=function(a,A) {
    e=outer(a[n+2]^(1:m),a[1:n])
    v=exp(a[n+1]-logd)*(e^2)^a[n+3]
    t1=log(2*pi*v)/2
    t2=(A-e)^2/(2*v)
  sum(t1+t2,na.rm=TRUE)}

# Gradient of the objective function
l.grad=function(a,A) {
    e=outer(a[n+2]^(1:m),a[1:n])
    v=exp(a[n+1]-logd)*(e^2)^a[n+3]
    da=colSums(a[n+3]-(e*(A-e)+a[n+3]*(A-e)^2)/
      v,na.rm=TRUE)/a[1:n]
    yy=1-(A-e)^2/v
    dk=sum(yy/2,na.rm=TRUE)
    dp=sum(yy*log(e^2)/2,na.rm=TRUE)
    du=sum((a[n+3]*r/a[n+2])-
      (r*e*(A-e)+a[n+3]*r*(A-e)^2)/(a[n+2]*v),na.rm=TRUE)
  c(da,dk,du,dp)}

 # Hessian of the objective function
l.hess=function(a,A) {
    e=outer(a[n+2]^(1:m),a[1:n])
    v=exp(a[n+1]-logd)*(e^2)^a[n+3]
    daa=diag(
          colSums((e^2+4*a[n+3]*e*(A-e)+
            a[n+3]*(2*a[n+3]+1)*(A-e)^2)/v-a[n+3],
          na.rm=TRUE)/a[1:n]^2)
    dak=colSums((e*(A-e)+a[n+3]*(A-e)^2)/v,na.rm=TRUE)/a[1:n]
    dat=colSums((r*e^2+(4*a[n+3]-1)*r*e*(A-e)+
          2*a[n+3]^2*r*(A-e)^2)/v,
          na.rm=TRUE)/(a[1:n]*a[n+2])
    dap=colSums(msk+(log(e^2)*e*(A-e)+
          (a[n+3]*log(e^2)-1)*(A-e)^2)/v,na.rm=TRUE)/a[1:n]
    dkk=sum((A-e)^2/v,na.rm=TRUE)
    dkt=sum((r*e*(A-e)+a[n+3]*r*(A-e)^2)/(a[n+2]*v),na.rm=TRUE)
    dkp=sum(log(e^2)*(A-e)^2/(2*v),na.rm=TRUE)
    dtt=sum((r^2*e^2+(4*r*a[n+3]-r+1)*r*e*(A-e)+
            (2*r*a[n+3]+1)*a[n+3]*r*(A-e)^2)/v-a[n+3]*r,
            na.rm=TRUE)/a[n+2]^2
    dtp=sum(r+(r*e*log(e^2)*(A-e)+
            (a[n+3]*r*log(e^2)-1)*(A-e)^2)/v,na.rm=TRUE)/a[n+2]
    dpp=sum(log(e^2)^2*(A-e)^2/(2*v),na.rm=TRUE)
    dm1=matrix(c(dak,dat,dap),n,3)
    dm2=matrix(c(dkk,dkt,dkp,dkt,dtt,dtp,dkp,dtp,dpp),3,3)
```

```
    rbind(cbind(daa,dm1),cbind(t(dm1),dm2))}

  # Set up starting values, take trend from first column, unless it errors
  # out (because of 0 or negatives) in which case take 3% as a default
  tmp=na.omit(data.frame(x=1:m,y=log(A0[,1])))
  trd=1.03
  trd=exp(coef(lm(tmp$y~tmp$x))[2])
  a0=c(colSums(A0/(trd^c),na.rm=TRUE)/colSums(msk+0*A0,na.rm=TRUE),log(max(dnom))
,trd,1.5)

  max=list(10000,10000)
  names(max)=c("iter.max","eval.max")

  # Actual minimization
  mle= nlminb(a0,l.obj,gradient=l.grad,hessian=l.hess,
    scale=c(abs(1/a0[1:n]),rep(5,3)),A=A0,control=max)

  # mean and var are model fitted values, stres standardized residuals
  mean=outer(mle$par[n+2]^(1:m),mle$par[1:n])
  var=exp(mle$par[n+1]-logd)*(mean^2)^mle$par[n+3]
  stres=(A0-mean)/sqrt(var)

  # Calculate the information matrix using second derivatives of the
  # log likelihood function

  # Second with respect to alpha parameters
  aa=diag(
    (2*mle$par[n+3]^2*
      colSums(msk+0*A0,na.rm=TRUE)/
        mle$par[1:n]^2)+
      colSums((msk+0*A0)/
        outer(exp(mle$par[n+1]-log(dnom))*mle$par[n+2]^(2*(1:m)*(mle$par[n+3]-
1)),
          (mle$par[1:n]^2)^mle$par[n+3])
        ,na.rm=TRUE)
      )

  # Second with respect to alpha and kappa
  ak=(mle$par[n+3]/mle$par[1:n])*
    colSums(msk+0*A0,na.rm=TRUE)

  # Second with respect to alpha and tau
  at=(2*mle$par[n+3]^2/(mle$par[n+2]*mle$par[1:n]))*
    colSums((msk+0*A0)*r,na.rm=TRUE)+
      colSums((msk+0*A0)*outer((1:m)/(exp(mle$par[n+1]-log(dnom))*
        mle$par[n+2]^(2*(1:m)*(mle$par[n+3]-1))),
          1/(mle$par[1:n]^2)^(mle$par[n+3]-1)),
      na.rm=TRUE)/(mle$par[n+2]*mle$par[1:n])

  # Second with respect to alpha and p
  ap=(mle$par[n+3]*log(mle$par[1:n]^2)/mle$par[1:n])*
      colSums((msk+0*A0),na.rm=TRUE)+
    (mle$par[n+3]*log(mle$par[n+2]^2)/mle$par[1:n])*
      colSums((msk+0*A0)*r,na.rm=TRUE)

  # Second with respect to kappa
```

```
kk=sum((msk+0*A0),na.rm=TRUE)

# Second with respect to kappa and tau
kt=mle$par[n+3]*sum((msk+0*A0)*r,na.rm=TRUE)/mle$par[n+2]

# Second with respect to kappa and p
kp=sum((msk+0*A0)*log(outer(mle$par[n+2]^(2*(1:m)),
  mle$par[1:n]^2)),na.rm=TRUE)/2

# Second with respect to tau
tt=2*mle$par[n+3]^2*sum((msk+0*A0)*r^2,na.rm=TRUE)/mle$par[n+2]^2+
  sum((msk+0*A0)*
    outer((1:m)^2/(exp(mle$par[n+1]-log(dnom))*mle$par[n+2]^
        (2+2*((1:m)*(mle$par[n+3]-1)))),
      1/(mle$par[1:n]^2)^(mle$par[n+3]-1)),
    na.rm=TRUE)

# Second with respect to tau and p
tp=sum((msk+0*A0)*(r-1),na.rm=TRUE)/mle$par[n+2]+mle$par[n+3]*(
  sum((msk+0*A0)*outer(1:m,
    log(mle$par[1:n]^2)),
    na.rm=TRUE)+
  sum((msk+0*A0)*r*log(mle$par[n+2]^(2*r)),na.rm=TRUE))/
  mle$par[n+2]

# Second with respect to p
pp=sum((msk+0*A0)*log(outer(mle$par[n+2]^(2*(1:m)),
  mle$par[1:n]^2))^2,na.rm=TRUE)/2

# Create information matrix in blocks
m1=matrix(c(ak,at,ap),n,3)
m2=matrix(c(kk,kt,kp,kt,tt,tp,kp,tp,pp),3,3)
inf=rbind(cbind(aa,m1),cbind(t(m1),m2))

# Variance-covariance matrix for parameters, inverse of information
# matrix
vcov=solve(inf)

# Initialize simulation array to keep simulation results
sim=matrix(0,0,m+1)
smn=matrix(0,0,m+1)

# Simulation for distribution of future amounts
# Want 10,000 simulations, but exceeds R capacity, so do
# in batches of 5,000
nsim=5000
smsk=aperm(array(c(msk),c(m,n,nsim)),c(3,1,2))
smsn=aperm(array(c(msn),c(m,n,nsim)),c(3,1,2))

for (i in 1:5) {

# Randomly generate parameters from multivariate normal
spar=rmvnorm(nsim,mle$par,vcov)

# Arrays to calculate simulated means
ttoi=array(c(outer(spar[,n+2],1:m,"^")),c(nsim,m,n))
```

```
   alph=aperm(array(c(spar[,1:n]),c(nsim,n,m)),c(1,3,2))
   esim=alph*ttoi

   # Arrays to calculate simulated variances
   ksim=array(exp(outer(spar[,n+1],log(dnom),"-")),c(nsim,m,n))
   psim=array(spar[,n+3],c(nsim,m,n))
   vsim=ksim*(esim^2)^psim

   # Randomly simulate future averages
   temp=array(rnorm(nsim*m*n,c(esim),sqrt(c(vsim))),c(nsim,m,n))

   # Combine to total by exposure period and in aggregate
   # notice separate array with name ending in "n" to capture
   # forecast for next accounting period
   sdnm=t(matrix(dnom,m,nsim))
   fore=sdnm*rowSums(temp*!smsk,dims=2)
   forn=sdnm*rowSums(temp*smsn,dims=2)

   # Cumulate and return for another 5,000
   sim=rbind(sim,cbind(fore,rowSums(fore)))
   smn=rbind(smn,cbind(forn,rowSums(forn)))
   }

   summary(sim)
   summary(smn)

   # Scatter plots of residuals & Distribution of Forecasts
   windows()
   par(mfrow=c(2,2))
   plot(na.omit(cbind(c(r+c-1),c(stres))),
     main="Standardized Residuals by CY",xlab="CY",
     ylab="Standardized Residual",pch=18)
   plot(na.omit(cbind(c(r),c(stres))),
     main="Standardized Residuals by AY",xlab="AY",
     ylab="Standardized Residual",pch=18)
   plot(na.omit(cbind(c(c),c(stres))),
     main="Standardized Residuals by Lag",xlab="Lag",
     ylab="Standardized Residual",pch=18)
   proc=list(x=(density(sim[,m+1]))$x,
       y=dnorm((density(sim[,m+1]))$x,
         sum(matrix(c(dnom),m,n)*mean*!msk),
         sqrt(sum(matrix(c(dnom),m,n)^2*var*!msk))))
   truehist(sim[,m+1],ymax=max(proc$y),
     main="All Years Combined Future Amounts",xlab="Aggregate")
   lines(proc)

   # Summary of mean, standard deviation, and 90% confidence interval from
   # simulation, similar for one-period forecast
   sumr=matrix(0,0,4)
   sumn=matrix(0,0,4)

   for (i in 1:(m+1)) {
sumr=rbind(sumr,c(mean(sim[,i]),sd(sim[,i]),quantile(sim[,i],c(.05,.95))))
sumn=rbind(sumn,c(mean(smn[,i]),sd(smn[,i]),quantile(smn[,i],c(.05,.95))))
  }
```

5. REFERENCES

[1] Bornhuetter, R. and R. Ferguson, "The Actuary and IBNR," *Proceedings of the Casualty Actuarial Society* 1972, Vol. LIX, 181-195.

[2] Berquist, J. and R. Sherman, "Loss Reserve Adequacy Testing: A Comprehensive Systematic Approach," *PCAS* 1977, Vol. LXIV, 123-184.

[3] Wright, T.S., "Stochastic Claims Reserving When Past Claims Numbers are Known," *Proceedings of the Casualty Actuarial Society* 1992, Vol. LXXIX, 255-361.

[4] England, P. and R. Verrall, "Stochastic Claims Reserving in General Insurance," *British Actuarial Journal* 2000, Vol. 8, No. 3, 443-518.

[5] Stelljes, S., "A Nonlinear Regression Model for Incurred But Not Reported Losses," *Casualty Actuarial Society Forum*, Fall 2006, 353-377

[6] Klugman, S.A., H.H. Panjer, and G.E. Willmot, *Loss Models: From Data to Decisions,* 1st ed.(New York: John Wiley and Sons, , 1998), 55-67.

[7] G. Venter, "Generalized Linear Models beyond the Exponential Family with Loss Reserve Applications," *CAS E-Forum*, Summer 2007, http://www.casact.org/pubs/forum/07sforum/07s-venter1.pdf.

Abbreviations and notations
MLE, maximum likelihood estimator

Biography of the Author

Roger Hayne is a Consulting Actuary in the Pasadena, CA office of Milliman, Inc. He is a Fellow of the Casualty Actuarial Society, a Member of the American Academy of Actuaries and holds a Ph.D. in mathematics from the University of California. Roger is an active volunteer in the CAS, serving on several CAS committees and task forces as chair of several, and also as Vice President – Research and Development for the CAS. He has published numerous papers in the *CAS Forum*, the *Proceedings of the Casualty Actuarial Society (PCAS)*, and *Variance*. One of his *PCAS* papers was awarded the 1995 Dorweiller Prize.

Exhibit 1

Incremental Adjusted Average Paid Losses Per Ultimate Claim
Berquist & Sherman Automobile Liability Data

Accident	Months of Development								Forecast
Year	12	24	36	48	60	72	84	96	Counts
1969	178.73	361.03	283.69	264.00	137.94	61.49	15.47	8.82	7,822
1970	196.56	393.24	314.62	266.89	132.46	49.57	33.66		8,674
1971	194.77	425.13	342.91	269.45	131.66	66.73			9,950
1972	226.11	509.39	403.20	289.89	158.93				9,690
1973	263.09	559.85	422.42	347.76					9,590
1974	286.81	633.67	586.68						7,810
1975	329.96	804.75							8,092
1976	368.84								7,594

Estimates

	α_1	α_2	α_3	α_4	α_5	α_6	α_7	α_8
Parameter	143.78	316.77	251.78	197.68	102.53	46.23	21.36	7.36
Std. Error	6.20	11.54	9.16	7.62	5.25	3.75	3.07	2.41

	z	t	p
Parameter	8.5871	1.1265	0.5782
Std. Error	0.2321	0.0077	0.0303

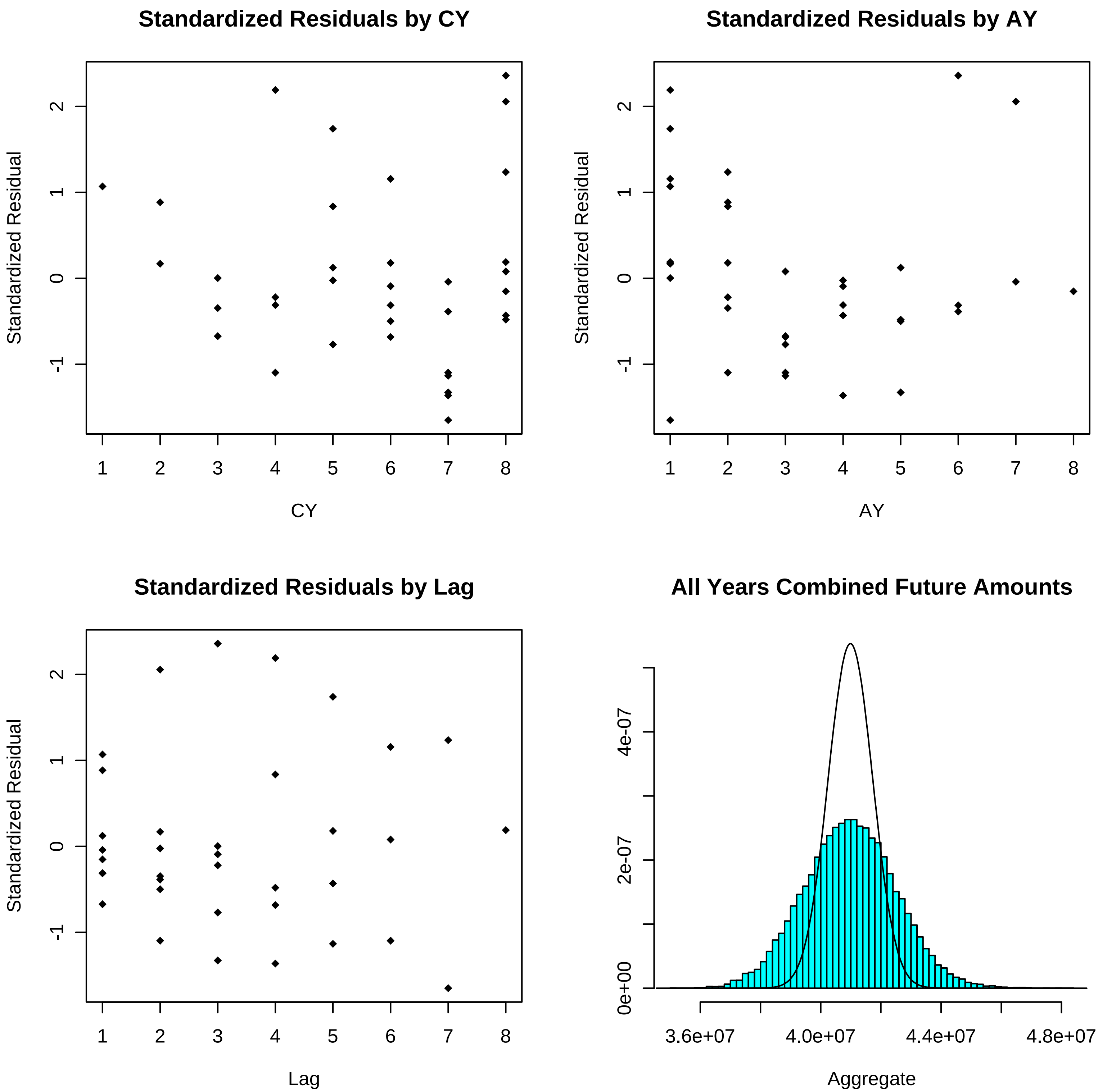

Standardized Residuals by CY
Standardized Residual
CY
Standardized Residuals by AY
Standardized Residual
AY
Standardized Residuals by Lag
Standardized Residual
Lag
All Years Combined Future Amounts
0e+00
2e-07
4e-07
3.6e+07
4.0e+07
4.4e+07
4.8e+07
Aggregate

Exhibit 3

Incremental Adjusted Average Paid Losses Per Ultimate Claim
Berquist & Sherman Automobile Liability Data

Forecast Expected

Accident Year	24	36	48	60	72	84	96	Total
1969								
1970							9.34	9.34
1971						30.54	10.52	41.06
1972					74.43	34.40	11.85	120.68
1973				185.96	83.84	38.75	13.34	321.90
1974			403.89	209.48	94.45	43.65	15.03	766.50
1975		579.48	454.96	235.97	106.39	49.17	16.93	1,442.91
1976	821.26	652.77	512.50	265.81	119.84	55.39	19.07	2,446.64

Forecast Variance

Accident Year	24	36	48	60	72	84	96	Total
1969								
1970							8.19	8.19
1971						28.10	8.19	36.29
1972					80.84	33.11	9.65	123.60
1973				235.51	93.74	38.40	11.19	378.84
1974			709.12	331.88	132.10	54.11	15.77	1,242.97
1975		1,039.02	785.45	367.61	146.32	59.93	17.47	2,415.80
1976	1,657.07	1,270.62	960.54	449.55	178.93	73.29	21.36	4,611.37

Exhibit 4

Incremental Adjusted Average Paid Losses Per Ultimate Claim
Berquist & Sherman Automobile Liability Data

Estimates of Accident Year Future Loss Forecasts

Accident	Process Only		Including Parameter Uncertainty			
		Standard		Standard	Percentile	
Year	Mean	Deviation	Mean	Deviation	5%	95%
1969	0	0	0	0	0	0
1970	80,981	26,503	80,551	36,442	24,148	144,035
1971	408,500	63,754	407,019	82,070	274,928	545,616
1972	1,169,365	106,448	1,169,765	137,850	945,662	1,399,015
1973	3,087,023	172,060	3,086,394	233,709	2,702,457	3,476,160
1974	5,986,335	216,225	5,984,922	344,212	5,425,005	6,551,203
1975	11,676,044	307,380	11,671,230	549,685	10,783,705	12,583,860
1976	18,579,788	375,626	18,581,701	808,465	17,258,898	19,916,569
Total	40,988,036	572,742	40,981,581	1,513,557	38,528,696	43,485,373

Forecasts for Next Calendar Year

Accident	Process Only		Including Parameter Uncertainty			
		Standard		Standard	Percentile	
Year	Mean	Deviation	Mean	Deviation	5%	95%
1969	0	0	0	0	0	0
1970	80,981	24,817	80,551	36,442	24,148	144,035
1971	303,859	52,742	302,553	68,934	192,431	418,164
1972	721,230	87,122	721,793	105,826	551,032	898,662
1973	1,783,372	147,171	1,783,236	172,967	1,502,286	2,075,631
1974	3,154,365	207,974	3,154,597	240,834	2,764,684	3,559,245
1975	4,689,180	260,836	4,686,348	309,909	4,179,644	5,204,351
1976	6,236,615	309,130	6,236,267	372,667	5,629,261	6,854,599
Total	16,969,602	489,384	16,965,345	652,968	15,893,889	18,045,385

Estimating the Ultimate Liability for a Non-Insurance Company's Revised Warranty Product

Orin M. Linden, Ph.D., FCAS, MAAA, ARM
James B. Kahn, FCAS, MAAA
Brian Ko

Abstract

Motivation. Non-insurance companies are offering ever greater enhancements to their warranty programs, many times as a competitive tool to strengthen market position. Yet, oftentimes very little analysis is performed to understand the cost of these changes. This paper discusses how warranties are accrued for on a manufacturer's balance sheet and offers examples of methods to estimate these costs.
Method. Most of the paper's discussion centers around projecting actual payments over time using an approach similar to an incremental loss development triangle approach, properly adjusted for exposure and inflation changes. Other methods discussed include Bornhuetter-Ferguson, Average Age of Warranty Claim Times Annual Spend, Active Life, and Calendar Year Payments to Revenue Approaches.
Results. The most appropriate projection method may depend on factors such as available data or the nature of the company's product.
Conclusions. Actuarial projections of warranty costs rooted in common actuarial reserving and pricing techniques are appropriate for estimating the future liabilities for the warranty liabilities.
Keywords. Warranty/Service Contracts; Parts and Labor Cost; Reserving; Pricing.

1. INTRODUCTION

Quite simply put, if a company has a product, it will likely be offering warranties associated with it. While warranty coverage for automobile mechanical breakdown has become more commonplace over the years, the latest technological explosions have led to more and more customers being offered a bewildering array of warranties for products ranging from most personal and home appliances to highly specific products such as jet engines.

For the most part, history and the various state laws that deal with the appropriateness of items for their intended use have dictated that a manufacturer's basic warranty would be in place for no additional charge to the customer. However, many companies offer the consumer an opportunity to purchase an "extended" warranty or service agreement, which would provide for some combination of (1) additional years of coverage, or (2) coverage of additional costs, or both (1) and (2). In addition, a manufacturer may offer a maintenance agreement that provides for maintaining a product according to a recommended maintenance schedule. Some manufacturers will guarantee coverage over the entire "lifetime" of a given product under the terms of their "lifetime" warranties.

2. WHY WORRY?

Warranties come in all types and shapes. They may be implied (i.e., an expectation of fitness for the use intended) or expressed (actually stated orally or in writing). They may cover a short period of time such as 30 days or the lifetime of a product, however defined. They may cover all or part of a product (e.g., a five-year bumper-to-bumper vs. a 10-year drivetrain auto warranty). Some will cover parts only; others, labor as well. Specific warranties may cover related costs such as a roofing tile company covering the cost of removal and dumping of the defective product.

Some companies offer a basic warranty and nothing else. The basic warranty may be simple (e.g., one year parts and labor) or quite complicated (e.g., auto warranties that have different life spans and coverages for different parts and may require servicing from other companies for purchased parts such as tires or radios). Indeed, many companies offer a dazzling assortment of both products and warranty options. Customers may receive a basic warranty automatically but the terms may vary by product. When a manufacturer offers greater coverage through an extended warranty or service contract, these additional warranties are sold for a variety of purposes: to generate profit, to differentiate a product, or to underscore the inherent quality of a product. The terms of an extending warranty can be just as puzzling as the standard or basic warranty. They also may cover replacement parts, labor, and peripheral costs for differing time periods.

To make matters more interesting, depending on the product and economic situation, warranties can often be used as competitive tools. In such cases manufacturers may decide to extend the coverage offered by their basic or extended warranties for several purposes:

- Showing a company's belief in the inherent quality of their product;
- Adding a differentiator where consumers view products as inherently similar;
- Changing warranties to be similar to changes adopted by competitors;

- Encouraging product buyers to also purchase maintenance services from the manufacturer;
- Encouraging sales of an existing product or a new product where the manufacturer may have had quality issues in the past; and
- Encouraging sales from a new manufacturer or a new product without a history of product quality.

The costs of providing warranties vary by manufacturer but are large overall. One source we are familiar with put the industry reserve level at $39 billion for the third quarter of 2006.[1] Others sources estimate the total cost of warranty reserves as greater than 2% of revenue for manufacturers. Such costs rival those paid by manufacturers for insuring their fortuitous risk exposure. Because most manufacturer warranties are short term, this likely approximates manufacturers' annual spend for this exposure. By contrast, the Insurance Information Institute puts the United States annual spend for workers compensation premium at approximately $42 billion, a similar amount.

Funding for warranty accruals is required to be disclosed according to FASB Interpretation 45[2] which states that, among other things, a manufacturer is required to disclose the approximate terms of warranties, the accounting policy and methodology for funding, and the carried reserves for product warranties. By their nature, warranty claims tend to have a very high frequency/low severity exposure and the law of large numbers generally works well. These costs are generally able to be estimated by actuarial methods.

The principal obstacles in performing such analyses tend to be data issues. Simply put, many companies do not seem to code warranty claims in the level of detail that insurers code losses. As a result, in performing the analysis of warranty claim costs, the greatest amount of time is usually spent scurrying around looking for data sources that could, with some preparation, be coded and thus be much more readily available to an analyst. While such improvements in coding would be beneficial to manufacturers for accrual purposes, the

---

[1] "Warranty Reserve Levels." *Warranty Week*, 30 January 2007. http://www.warrantyweek.com/archive/ww20070130.html. Accessed 30 January 2008.

[2] Financial Accounting Standards Board Interpretation No. 45 – Guarantor's Accounting and Disclosure Requirements for Guarantees, Including Indirect Guarantees of Indebtedness of Others. http://www.fasb.org/pdf/fin%2045.pdf

authors believe that far greater benefits of better coding would appear in both the control of costs and in costing changes in the basic or extended warranties that companies consider routinely for competitive and cost control purposes. Like insurance claims, a single calendar year brings in claims from sales in all prior revenue years in which warranties are still live. However, unlike insurance claims, the amounts are generally relatively small, paid quickly, and do not take a great deal of time to develop. Thus, changes in coding will begin providing useful data almost immediately for all lengths of warranties. In the cases of relatively inexpensive items sold in great bulk, such as tools, useful information may be available in as little as a week.

While some warranties are sold by third parties who sell them as a business model, we will be focusing on the ultimate warranty cost associated with product manufacturers' warranty programs. Although warranty costs are not technically insurance costs, many of the same characteristics can apply to the warranty programs in explaining yearly cost emergence. The CAS literature has a number of articles dealing with the insurance aspect of mechanical breakdown insurance. There are much fewer publications dealing with the non-insurance sector and general product warranty accruals.

Generally speaking, warranty programs among industries and companies are rarely identical. As such, it is important to be able to determine what general steps to take in modifying an existing warranty program. Equally important, a manufacturer should consider carefully how to code their internal claims and information systems to be able to handle any changes for the warranty program in the future.

3. WARRANTY BACKGROUND

3.1 Warranty Characteristics

As mentioned previously, specifics of company warranties are rarely identical to each other. Some characteristics of warranty programs observed and worth noting are listed below:

- Mechanical breakdown coverage has historically been the most prevalent type of warranty with actuarial involvement. Several companies have offered extended warranty coverage for automobiles for several decades.
- Automobile mechanical breakdown losses are generally "back-end loaded" as manufacturer warranties often inure to the benefit of extended warranty coverages. Normal wear and tear of a vehicle usually results in mechanical breakdown after a period of several years.
- Because of the propensity for mechanical breakdown losses to increase over the course of time, many companies earn their premium accordingly, often using methodologies like the reverse rule of 78s or something similar to approximate the true payment patterns over time and match the premium earnings accordingly. The NAIC's Statement of Actuarial Opinion requirement requires opining actuaries to mention the Unearned Premium Reserve component associated with "Long Duration Contracts." Many of these long duration contracts are warranty contracts.[3] Companies should be aware of profitability considerations to determine if any premium deficiency issues would need to be separately handled.
- Many manufacturers have drags on earnings as a result of their warranty liabilities. Many of these difficulties have arisen from misestimation of ultimate liabilities involving the long tails of "back-end loaded" types of products.
- Some types of warranties are offered for products that rarely have warranty events, whereas others are offered for products that companies expect to pay losses routinely. In the latter case, an extended warranty is often bundled with a service contract to maintain the product. Estimation of costs would obviously be quite different amongst the various products and types of warranties.
- Companies offering "lifetime" warranties generally have a long tail of claims emergence. It is also possible for products with lifetime warranties that historical cost emergence patterns may not be indicative of costs seen in the future. There are

many reasons why this may be so. For instance, product specifications may have been changed over time. Secondly, distribution channels or locations may have been changing over the course of time, and may be distorting companies' historical data. Additionally, a company's data may only include a limited number of years. It is certainly possible that expected lifetimes may be longer than the historical timeframe, but well within the bounds of time as specified in revised extended warranty periods. In such cases, it is important to try to identify external sources of data or internal company experts to help assess cost emergence for periods outside of the experience period.

- Generally speaking, ALAE costs for warranty programs are either minimal or non-existent. Companies may want to consider ULAE costs involved with an internal claims department that handles their warranty claims.
- There is usually a very short lag period between the reporting of a loss and the actual payment of loss by the company issuing the warranty. As such, reserving for known claims is typically not a major issue. Case reserves are not generally set up and funding may be part of the overall warranty accrual process.
- It is not uncommon for companies issuing warranties to pay for losses that are outside the scope of coverage (either for cause of loss or time) as a "goodwill" measure especially in the first 12 months after sale. Such payments may be made often enough to be routine and may be considerably larger than the true costs covered by the warranty. Other times, they may be rare. An understanding of how often an entity pays goodwill losses is necessary to fully understand the cost impact involved in the expansion of a company's warranty program. However, it is our experience that many companies do not keep track of these payments. Nor do they have guidelines as to when they should be paid. As a result, large costs may be incurred without the benefits anticipated.

---

[3] National Association of Insurance Commissioners, "Instructions to the Annual Statement for Property/Casualty," www.naic.org.

3.2 Competitive Environment of Warranties

Most companies consider what their competitors offer in the marketplace in establishing their warranties. Many companies that issue extended warranties routinely revise their existing programs based upon what their competitors have been offering. Many times, salespeople are adamant that not providing warranties with as much coverage as the competitors (both in terms of covered perils as well as years of coverage) places the company at a competitive disadvantage, The perception in the marketplace is that the company in question does not "stand behind the quality of its product." The result is that companies, for fear of losing market share, are often proactive in either making revisions to their current program to be more in line with their competition, or offering their own broader warranty coverage before their competitors do. The result of the competitive environment is that even in the cases where a company may offer a broader coverage than their competition in an attempt to gain market share, there is often a limited time period before their competitors begin to provide their own similar product essentially "leveling the playing field' despite the company's best efforts to differentiate themselves.

Often revisions to existing warranty programs are made without adequate studies as to the immediate impact such revisions could have on a company's financial statement. Yet analysis can often be done to better understand the financial costs of such changes. The techniques discussed here in this paper are designed to estimate the additional costs associated with the broadening of coverage.

3.3 External Regulation and Distribution of Warranties

A recent example involving a warranty company (not a manufacturer) illustrates what can potentially go wrong given the current regulatory environment. A warranty company, the majority of whose business pertained to used-car extended warranty business, recently ran into such high financial problem, that its reinsurer also runs the risk of becoming insolvent. By the time the state insurance department became involved, the company's insurer, backing a portion of the warranties, was itself on the brink of insolvency. According to one source, in the overall long list of creditors in such a case, it is fair to assume that the warranty policyholders themselves would be very low on the list of those who can make claims on the insurance company's assets. As it turned out, even though some state departments of commerce had denied the warranty company a license in their states, and had even issued a "cease and desist" order against selling these warranties in their states, thousands of warranty policies continued to be issued, by auto dealers whose commissions exceeded as much as 50% of the overall premium.[4]

Thus, even in situations where state commerce departments have banned companies from doing business in a state, the many third-party agents selling warranty policies have made it difficult to effectively carry out such a ban. It is easy to speculate that this phenomenon may become more and more prevalent especially as an increasing number of people begin to purchase personal electronics from some of the larger appliance stores. In many of these cases warranty may not be issued by the manufacturer or the store, but rather by a third-party company whose performances may not be tied to the results of the warranty company itself. These instances could lead to similar situations as the warranty companies may have exorbitant potential liabilities on the books without fully grasping the magnitude of their aggregate exposures to liability.

4. DATA ISSUES

As mentioned earlier, a company may choose to address expansion of their warranty in either a proactive or reactive manner in an attempt to differentiate themselves from their competition and increase sales. It is possible, if not likely, that extensions of years of coverage may be made without immediately performing a formal actuarial cost analysis. However, some companies have had several extended warranty type products in the marketplace through many different attempts at broadening the options available for the consumer. As such, with proper data coding of historical claims, there would be an opportunity to use this actual historical company cost emergence information to formally project the proposed cost changes to an existing warranty.

4.1 Loss Coding

As with any pricing or reserving project, costing and reserving for warranty exposures generally starts with historical information. The following list of issues should be considered in establishing a warranty database or refining an existing database. The items below should be considered, though the list should not be limited to those noted:

4 Cummins, H.J., "Worthless Warranty," *Minneapolis Star Tribune*, December 12, 2007, pages D1, D3.

- ***Type of Product*** – As historical loss information may be different by type of product, and as terms and conditions of warranties may have varied historically by type of product, this information should be properly coded within a company's historical claims database. To the extent that products are sold internationally, the warranty costs should be segregated both by geographic area and by currency so that differing geographical usage patterns and currency exchange rate fluctuations can be addressed. Similar considerations will apply to revenue coding as well.
- ***Date of Purchase, Shipment, or Installation*** – The date for when a product is purchased, shipped, or installed (whichever is the key date to start the beginning of the warranty) should be captured in the company's claims database.
- ***Date of Claim Occurrence*** – The date of claims occurrence, as defined by the company for the purposes of providing coverage, should be captured in the company's claims database.
- ***Date of Claim Report*** **–** The date a claim is first reported to the company or recorded on the company's records, as defined by the company, should be captured in the company's claims database. As mentioned previously, the lag period for warranty coverage is usually not significantly long from the date of occurrence to the date of claim report.
- ***Date of Claim Payment*** **–** The date or dates a claim is paid by the company should be recorded in the claims database. Warranty costs are usually paid quickly after the claim is reported to the company.
- ***Repair Type and Cost*** – The type of repair made, as defined by the company, should be captured in the company's claims database. In its simplest form, this may be a split between materials and labor costs. This may be further sub-divided, sometimes significantly, at the company's discretion. The amount paid for each type of claim category associated with a

given claim should also be captured in the company's claims database. If payment is made for costs that, in theory, should not have been paid for, these figures should also be tracked. Companies should consider the level of detail in the coding of repair types and costs very carefully. Generally, more detailed coding of these items should be encouraged as future changes in product design and production or changes to warranty programs may consider the explicit addition or deletion of various costs.

- ***Location of Claim*** – The location of incident or repair for each type of claim should be captured in the company's claims database. This category could be important for several reasons. For instance, it may be necessary to understand and segregate payments made as a result of catastrophes that led to claims being submitted under warranties. For, example, a building product warranty may be tapped to cover damage in a hurricane. Additionally, frequency of claims, costs of repairs, or legal environment could vary significantly by location due to local weather and economic conditions. Potentially, a company may ultimately decide for exclusions, pricing differences, or distributions of products in certain states depending upon the results of the observed history of claims emergence.
- ***Basic vs. Extended Warranty*** – For losses that are coded by the company in their claims database, a distinction should be made as to which losses are associated with a historical basic warranty type, and which losses are associated with an extended warranty. Warranties change from time to time. The exact warranty that was purchased should be coded along with the date of warranty purchase so that the exact terms of warranties can be analyzed.
- ***Length of Warranty or Service Agreement*** – The term length of the warranty associated with a given claim should be captured by the company.
- ***Cost of Labor by Geographical Area*** – To the extent that labor is a covered cost, and the associated rates vary by geographical area, the historical costs of labor and the associated repairs should be tracked. If this

information is not immediately available, a reasonable approximation might be calculated for some procedures by using outside sources including the Internet.

- ***Historical Pro Rata Percentages*** – Some warranties contain a provision where, for a certain number of initial years, warranty costs are covered at full cost. Following this period, many cover some costs at a prorated percentage of the warranty policy. Furthermore, some may cover some types of losses at a prorated percentage, but entirely exclude other types of coverage after this initial period. To the extent that losses are coded, the company should also code the detail in their claims database as to the pro rata percentage that was historically paid for a given claim amount. This would be valuable information to have, should the "full coverage" time allowance be amended in the future.

4.2 Exposure and Revenue Coding

Additionally, the following list of items should be monitored, but not necessarily in the same database that contains the claims information. There should, however, be an appropriate way to group these figures to those captured in the claims database:

- ***Number of Units Sold*** – The company should consider the number of units sold historically including an appropriate way to group these to track historical sales. They should further track this information, if possible, as to what types of warranties were sold (or provided at sale) for each unit. If possible, the company should consider coding the type of warranty purchased or provided by serial number so as to make matching warranty type to sales possible.
- ***Revenue*** – The company should track the historical sales volume including an appropriate way to group this for appropriate tracking of exposure volume. They should further track this information, if possible, as to what types of warranties were sold (or not sold) for each unit. Revenue should

clearly delineate cost of product and cost of extended warranty by type of warranty and product.

- ***Cost of Warranty Type Sold*** – The company should track the prices that all their different types of warranties have cost the consumer historically.
- ***Commissions*** – The company should track the commission paid and who sold the warranties historically.
- ***Rejection Database*** – The company should track claims that have been historically rejected and the reasons why they were rejected. This could perhaps be tracked in conjunction with those that were paid "outside the scope" of the warranty policy. A database such as this is a good way to track claims information should the company one day expand the terms of its warranty program to include types of loss that had not been covered historically. It should be noted that, in such a situation, the company should be aware that the frequency of these types of losses may be understated from what they should expect to see with a broadening of coverage as the consumer may understand that these costs would not be covered under their warranty and as such not report the incident to the company.

Unfortunately, the above mentioned items may not come from the same source or department from within a company. In other cases, different users of data may carry it in varying levels of detail, causing reconciliation to be difficult. Companies may even have third-parties handle certain data items. As a result, the task of matching revenue, product and warranty is not always a straightforward exercise. Having a designated person from the company that can help in coordinating and providing the necessary data in a usable format in these situations can be vital.

Database systems may vary widely from company to company but the actual data can usually be exported in a universally recognized format and imported into other platforms. Examples include commonly used database formats such as comma-delimited files (.csv) or text files (.txt). However, conversions may cause a loss of information regarding data field types. For example, fields that were originally formatted as a "date" value may be imported as generic "text" by default, unless the user manually specifies otherwise. Without such care, any subsequent operation that relies on the field being a "date" value would lead to difficulties.

Data types are also important because of space considerations. Designating a field that could suffice with an "integer" value (2 bytes) to be a "double-precision" value (8 bytes) results in using four times more space than is necessary. Over multiple fields and millions of records, the volume of misallocated space can become quite large. The size of the data file is an important factor in determining how smoothly the project proceeds. While many data manipulation tasks may seem simple enough in theory, working with large files can require lengthy computer run-times, in addition to causing issues with software constraints. For example, if the user does not have access to specialized data analysis software such as SAS, and elects instead to use a database management program with say a 2GB limit, the file can quickly reach the limit. This, in turn, can cause the file to become corrupt and unusable or cause important information to be lost in conversion. Taking the extra effort to correctly define data fields at the onset of a large undertaking is highly recommended as it will generally save great amounts of time throughout the rest of the project.

In addition to the above, companies should consider whether other sources of payment may be available to the purchaser of a product and consider whether the warranty acquired or purchased is primary or secondary. For example, a building owner may have property insurance that covers the failure of a building materials product such as siding in a windstorm. The company needs to consider whether the warranty coverage will be primary or secondary to the other sources and state this explicitly.

5. TECHNIQUES OF COST PROJECTION

5.1 General Considerations

Warranty cost projection can be viewed in a manner similar to pricing if it is prospective and similar to loss reserving for accrual purposes, with bodies of data grouped appropriately by year and with historical "warranty year" costs being observed and projected to ultimate. Although many exposure bases can be considered, such as number of units sold, methodologies can often be most easily applied by reviewing historical loss cost emergence as a percentage of sales figures where it can preliminarily be assumed that, with both costs and revenue being inflation-sensitive, the overall ratio should not need to be significantly adjusted for changes in inflation over a reasonable time period. Of course, it is highly recommended that yearly ratios be observed to see if an increasing annual trend is observed, signifying that overall costs may be increasing at a greater rate than revenue. Adjustments should be considered in such a situation.

As previously mentioned, report and payment lags for warranty policies are usually relatively short. As such, future cost emergence of existing basic and extended warranties in-force can be generally viewed as occurrences that have not yet happened but will ultimately give rise to warranty claims, as pipeline claims are usually small in total costs compared to the total warranty exposures. Whoever performs a warranty analysis may consider performing further breakdowns of cost emergence between labor, parts, or any other type of cost payments, depending on the purpose of the analysis and the available data.

Many warranty policies specify that no losses should be paid for any of the periods beyond the terms of the warranty agreement. Any warranty costs paid beyond these periods should be quantified and properly denoted as either "goodwill" payments or some other appropriate notation. A company should understand how often these types of payments are made, how much they cost, and adjust for this in any pricing or accrual exercises.

For the purposes of a company establishing a warranty accrual on their balance sheet for future costs, a company may want to factor in the time value of money by establishing a discounting procedure, depending on their accounting requirements. One assumption to consider would be an after-tax, risk-free discount rate based on something appropriate such as the latest year's average Treasury bill rates. A similar consideration exists for costing and pricing new warranties or making changes to existing programs.

Depending on data availability and scope of analysis, it is sometimes better to perform an analysis on all the data of a particular product group, rather than on an individual type of product or model basis or any other classification deemed appropriate by the manufacturer or warranty issuer. Such an analysis increases the volume of the data. In such cases, if there is to be any further analysis by individual sub-line of products, loss costs should ultimately balance back to the overall total loss cost figures based on the latest year's exposure distributions. A normalization procedure should be incorporated not unlike off-balance factor calculations that are utilized in territorial rate filings for insurance companies.

Some companies allow warranties to be extended at any time, not just at product purchase. In such cases, the analyst needs to consider whether there is an adverse selection process going on such that these buyers can be expected to have higher costs. If so, the pricing of the warranty needs to consider this. Similarly, some companies offer service contracts designed to maintain products in addition to extended or basic warranties. In such cases, the analyst should attempt to understand whether the warranty costs are likely to be lower as the products are well maintained and/or that part failure may be prevented by replacement under the maintenance contract. Such analyses are only possible if the company has implemented coding procedures that allow such issues to be tracked.

Finally, we note that many times warranty coverage allows some sort of continuation of the warranty coverage as transfer of ownership or title takes place for a covered product. While warranties differ in terms of how they handle the transfer of ownership (either disallowing transfer entirely, allowing transfer in all cases, or allowing transfer for a limited number of times), any changes to the existing structure going forward should be reflected in the future analyses.

5.2 A Costing Example: Expansion of Number of Years

5.2.1 Yearly Emergence

Perhaps the most common type of warranty expansion is allowing for coverage for an additional number of years. In our example, we will assume that a company will be expanding the current basic program from three years of coverage to five years of coverage. The new basic warranty will cover the costs of parts and labor in the fourth and fifth years. In this case, we will consider that the company has offered extended warranties historically, and thus has their own claims experience for the fourth and fifth year after installation for the customers who have purchased extended warranties historically. (We note that without actual data, one could consider other techniques such as curve fits or loss experience for similar products). To complicate matters a little bit, we will also assume that historically the extended warranties upon which we are basing our cost estimates had labor costs covered only through year 3. Additionally, the old program had covered parts/material costs at 100% through year 3 and at a prorated percentage thereafter. The new proposal will cover parts/material at 100% for years 4 and 5. As such, historical data would show the cost in years 4 and 5 without the impact of labor and 100% part replacement in these time periods.

Calculations would thus need to quantify the amounts of labor cost and non-prorated part cost for these years in the determination of the new projections.

To determine an estimate of additional cost emergence for all the changes, one can review the company's database and group the cost emergence in subsequent 12-month intervals for each year of product sale. This type of grouping and corresponding emergence will look like a policy-year loss development triangle but will consider at each evaluation warranty claims that occurred within successive 12-month periods from the date of sale of each item. Each corresponding policy year (year of sale) will not be fully "developed" for the first 12 months of emergence until the policy (sale) year is 24-months-old (as policies sold on the last day of the 12th month of a calendar year will be 12-months-old when the policy/sale year is 24-months-old). There will also be some (usually) short time period in which the newest warranty claims are evaluated and settled.

The data will now need to be adjusted to consider the impact of extending the non-prorated period. This will consist of two adjustments: adjusting the claim value to 100% for calendar years 4 and 5 (after the end of the historical non-prorated periods), and putting in a provision for additional labor costs for years 4 and 5. Adjustments can be made fairly easily by reading the company's historical warranty policy to estimate the prorated percentage to apply to years 4 and 5. They can be adjusted to 100% simply by dividing historical payments by these percentages.

To estimate the provision for additional labor costs, we multiply the number of historical units repaired or replaced under warranty claims by the average cost of labor per unit. If coding is available, such costs can be directly calculated based on historical payments. Otherwise, an estimate of these labor costs can be estimated by appropriate company personnel by geographical location, or an overall national basis. It is possible, of course, that the overall national labor average may vary going forward if the company's mix of geographical location changes over the course of time (or may have historical distortions if the mix has changed over time). We note that in performing these calculations the average cost of labor utilized will generally vary by year due to inflation.

The impact of extending the non-prorated period from three to five years would be estimated by subtracting the unadjusted calendar-year payments (as a percentage of revenue) for years 4 and 5 from the newly adjusted calendar-year payments (as a percentage of revenue) for years 4 and 5. It is possible that prorated percentages after year 5 may change as well. Adjustments for such changes in percentage will need to be contemplated in such a situation. The calculations will be similar to the adjustments already considered (excluding the cost of labor).

If no pro rata percentage of losses existed in the old extended warranty program, and no additional charges for new coverages would be applied, this calculation would be nothing more than a simple addition of the estimated costs for years 4 and 5 (as a percentage of revenue).

If projections are based on ratios of units sold, not as percentages of revenue, appropriate adjustments for inflation should be made as unit sales do not reflect inflation.

As a final note, it may be necessary to consider the impact of yearly cost projections for periods where no data exists. For example, if a company is interested in expanding the warranty coverage period to 40 years but had no data beyond 10 years, one could estimate an annual decay percentage through year 10, and then project the yearly decline in coverage to year 40. Alternatively, other curve fits should be considered. In such cases, the analyst should identify company engineers or product experts and discuss the results to determine whether such a pattern would truly make sense so many years into the future, or if the expected product lifetime is likely to alter yearly emergence after a given point in time. Any additional information as to the characteristics of the product should be reflected appropriately.

5.2.2 Percentage of First Year Costs

Many companies cannot segregate their product sales revenue by type of warranty (basic or extended). However, most companies can expect that basic warranty costs can be related to annual revenue as extended warranties almost surely contain terms at least as good as the basic warranty. Thus, if we are estimating the cost of adding more years to a basic warranty, another way to project future cost emergence involves making an overall determination of the ratio of the costs of the basic warranty in the first 12 months to total annual sales. Estimations can then be made as to relationships of future years' emergence to emergence in the first year warranty costs for the basic warranty only by using the historical ratio from actual data.

5.3 Expansion of Covered Costs and Introduction of New Covered Perils A Costing Example: Expansion of Number of Years

A company may look to expand the types of losses being covered by an extended warranty rather than solely the number of years. Many different variations of this expansion may be considered: (1) a company may consider paying for certain types of warranty costs during the same time that the basic warranty is in place, effectively providing broader coverage, (2) a company may consider providing for a coverage, which is currently being provided at full or pro rated value for some years, but not during a proposed expanded time period (for example, some companies include things like labor cost initially, but then only costs for materials after a certain point in time), and (3) a company may be considering adding coverage for a type of warranty cost that is not being provided in any current form.

Depending on the type of data available for a company, one may wish to calculate expansion of covered costs either directly (by multiplying the expected number of incidents of coverage by an appropriate expected cost) or by taking differences of costs based on existing programs.

In the first situation, one may be able to properly estimate frequency of incidence that would require the new type of coverage (or determining estimation during the appropriate future time period). Such a frequency might be approximated in many ways. Historical frequency of similar claims might be useful. Alternatively, sampling or surveying the users of the product or internal personnel might be useful. Finally, the rejection database might also give some indication of future frequency.

The severity to apply to the expected incidents could be based on the average value of a claim observed historically, or based on some external proxy such as the average cost of repair. With any inflation sensitive component such as labor rates, it is important to factor yearly cost of living adjustments into the appropriate calculations either by trending historical costs or by detrending the most recent information available. Adding broader coverage during the basic warranty timeframe could consider using this approach as well.

In the second situation involving differences in cost estimations, a company may have existing information for a program that includes the new coverage (even for a longer period of time than what is being proposed with the expansion). Historical yearly cost emergence as a percentage of revenue including the new covered peril as well as the historical yearly cost emergence excluding the new peril might be tracked for each selected time period. In such a case, a simple subtraction of the two data sets can be performed, with the difference being the new additional cost. Any differences in the overall emergence between the two data segments should be accounted for and adjusted if necessary.

6. METHODS FOR RESERVING

Accounting guidance for warranties that are a part of a product purchase is provided by FASB Interpretation No. 45[5], which clarifies that a liability for expected costs of a warranty must be recognized at the inception of the warranty. Accounting guidance for extended warranties appears in FASB Technical Bulletin No. 90-1, Accounting for Separately Priced Extended Warranty and Product Maintenance Contracts[6]. The situation is a bit different as it deals with revenue recognition and not expected costs. The guidance notes:

> 3. Revenue from separately priced extended warranty and product maintenance contracts should be deferred and recognized in income on a straight-line basis over the contract period except in those circumstances in which sufficient historical evidence indicates that the costs of performing services under the contract are incurred on other than a straight-line basis. In those circumstances, revenue should be recognized over the contract period in proportion to the costs expected to be incurred in performing services under the contract.
>
> 4. Costs that are directly related to the acquisition of a contract and that would have not been incurred but for the acquisition of that contract (incremental direct acquisition costs) should be deferred and charged to expense in proportion to the revenue recognized. All other costs, such as costs of services performed under the contract, general and administrative expenses, advertising expenses, and costs associated with the negotiation of a contract that is not consummated, should be charged to expense as incurred.
>
> 5. A loss should be recognized on extended warranty or product maintenance contracts if the sum of expected costs of providing services under the contracts and unamortized acquisition costs exceeds related unearned revenue. Extended warranty or product maintenance contracts should be grouped in a consistent

---

5 Financial Accounting Standards Board Interpretation No. 45 – Guarantor's Accounting and Disclosure Requirements for Guarantees, Including Indirect Guarantees of Indebtedness of Others. http://www.fasb.org/pdf/fin%2045.pdf.
6 Financial Accounting Standards Board Technical Bulletin No. 90-1 – Accounting for Separately Priced Extended Warranty and Product Maintenance Contracts. http://www.fasb.org/pdf/ftb%2090-1.pdf.

> manner to determine if a loss exists. A loss should be recognized first by charging any unamortized acquisition costs to expense. If the loss is greater than the unamortized acquisition costs, a liability should be recognized for the excess. [7]

We note that the reader should check with their accountants or other accounting experts for the exact treatment of warranty accruals.

The methodologies above are mainly utilized for estimating the cost of warranty covers. Following the accounting guidance above, the methodology for reserving may be based on these calculations or a simpler method can often be utilized so long as the revenue recognition for extended warranties follows expected costs. In the following we will discuss the recognition of the loss cost portion itself. This relates directly to establishing the reserve and runoff pattern of basic warranties included with the sale at no extra cost. Some additional work may be required for extended warranties to insure that the revenue recognition follows these expected loss costs and that expenses and any loss in excess of revenues are appropriately recognized. In the particular case of extended warranties the analyst should also consider whether the costs of claims in the settlement process are significant enough to establish a separate reserve for pipeline claims in-transit.

6.1 Bornhuetter-Ferguson Test A Priori

Monitoring results, especially for the first time, is an important measure to be taken with a warranty program. The methodologies listed previously contemplate expected emergence in a given calendar year. Year-to-year fluctuations among actual results can be handled through a Bornhuetter-Ferguson methodology by using this expected cost emergence in a given year as an a priori together with the expected unpaid percentage of cost emergence. With this method, the calculations made to cost or price the existing coverages can be unwound as to year of expected costs as a percentage of revenue. These percentages, for the unexpired portion of the warranties in place can then be applied (often on a discounted basis) to revenue for the year of sale to create a reserve. Revisions to yearly cost estimates can be made or, at the very least, differences between actual and expected cost emergence can be observed and studied. A series of refinements to selected yearly costs can be made

---

7 Accounting guidance for extended warranties appears in FASB Technical Bulletin No. 90-1, Accounting for Separately Priced Extended Warranty and Product Maintenance Contracts, Copyright © 1990, Financial Accounting Standards Board, Page 4.

appropriately going forward at least annually, if not quarterly. [8]

6.2 Average Age of Warranty Claim Times Annual Spend

This is likely the simplest of all warranty accrual methods. In this method one computes the average age of warranty claims in a given period of time. This figure is then applied to the average annual expenditure on warranties (appropriately adjusted for inflation) to determine the increase in warranty reserve. For example, if the average age of warranty claim is five years past the date of sale or installation and the average spend is $1 million per year, then for a period of five years of emergence, a reserve of $5 million (plus any adjustments for inflation) would be established.

This methodology appears to work best where the volume and mix of product is similar year to year, the annual average cost of each specific warranty incident is not very variable and the average annual outlay for warranty claims is fairly constant. Products such as tools and small electronics may often utilize this method successfully.

6.3 Active Life Approach

This methodology utilizes methods similar to those discussed in costing to come up with probabilities of a claim by report year under each type of warranty issued or sold. These probabilities are multiplied by the number of products sold by warranty type, in a given time period, to arrive at an expected frequency by report year. These are in turn multiplied by an expected cost (inflation adjusted) to arrive at a warranty reserve.

This method is well suited for warranties of particularly long duration where probability of defects is fairly constant.

6.4 Calendar-Year Payments to Revenue Approach

This method is similar to the active life approach but much simpler to apply in practice. Actual warranty costs by year of product sale are accumulated for at least one and usually two or more years. These costs are divided by either revenue or number of products sold in each year of sale. Ratios of claim costs by age as a percentage of revenue (or average cost per product sold) are selected and the unexpired warranty year fundings are computed as the

---

[8] Bornhuetter, Ronald L. and Ronald E. Ferguson., "The Actuary and IBNR", *PCAS* LIX, 1972, pp. 181-195.

ratios by age times the revenue sold by year. This method works well where warranty costs are fairly static as a percentage of revenue but the product line volume varies significantly by year.

7. BUSINESS CONSIDERATIONS

7.1 Warranty Language

Many of the basics of standard insurance contracts are often overlooked in producing a warranty. Unlike insurance, there is no industry-wide organization writing standardized warranty wording for companies to adopt or rewrite. Warranties often vary greatly by type of product being serviced, oftentimes being in completely different types of industries. Moreover, warranty language may vary company to company for similar products and may even vary across similar warranties in the same company. Because of this, it is often difficult to have universal, standard types of warranties in place. As a result, companies could conceivably find themselves paying more claim dollars than they otherwise would with universal, standard contracts.

For instance, warranty language often does not include wording that would make their policy secondary to other sources of recovery such as homeowners insurance or any other sources of recovery. As a result, the warranty writer may in effect pay for claims either instead of other recoveries, or even worse, in addition to other sources of recovery. In contrast, it is common for other types of guarantees such as credit card companies to cover the collision damage to a rental vehicle. In such cases, card issuers routinely make this coverage secondary to the renter's own insurance policy.

Finally, warranty policies many times omit language that would limit payment to some specified amount such as the value of a completely new replacement item that may very well be less than replacing individual parts for outdated products. Some sort of limitation language would potentially be advantageous to both the warranty companies (that may pay fewer dollars of claims) as well as the customer (who may choose to get a cheaper, but more state-of-the-art model of product should they choose).

7.2 Internal Operations

Major companies who issue warranties often do not have staff fully dedicated to monitoring the results of the warranty program. It is likely that in the overall costs of such companies, warranty costs are a small (but meaningful) fraction of the overall cost measures of a company. As such, devoting full-time staff for such tasks as fully estimating the cost of warranty policy revisions or estimating the quarterly accruals is often neglected (even though warranty policies involve many different segments such as marketing, claims, accounting, and other financial departments, to name a few).

An interesting example mentioned previously concerns the potential payment of many claims as "goodwill" measures even though the warranty language may not allow for such coverage. With poor internal communication as to how often goodwill claims are paid, and poor monitoring as to the actual dollars of payments by year that are effectively true goodwill claims, it will be difficult, if not impossible, to understand the true warranty costs and ultimately, the liability for balance sheet accruals. As such, goodwill payments tend to be made without the benefit of a strategic plan and it is not always clear that these non-required payments have a benefit equal to their costs. Further, companies do not always track the amount or payment detail of such payments and cannot always analyze whether these payments are actually undermining the company's attempt to sell for-profit extended warranties.

While many companies may want to focus their efforts solely on the external marketing of their extended warranty product, it is well worth the effort to monitor the amount of goodwill claims that have been routinely paid in the marketplace on their basic warranties. Those with a historically large preponderance of goodwill type claims that might be covered under extended warranties offered by the company may find it difficult to find purchasers of the extended warranty product. Such loose standards may effectively dictate that consumers have been receiving extended warranty coverage for basic warranty cost. While this is not necessarily a terrible result, a company should be thoughtful in its use of such payments and integrate them into its strategic marketing plan for both basic and extended warranties.

7.3 External Sales Force

As mentioned earlier, products with corresponding warranties, many times have a third-party agency force whose results are not dependent on the overall results of the warranty program. Often, the third-party sales force finds themselves with commissions that may very well be in excess of half the warranty price itself.

In such circumstances, it is not difficult to imagine where conflicts of interest could exist between sales forces and the results of the warranty company. As an example, let us look at the case where a salesperson may be allowed, if not encouraged, to sell a one-year service agreement several years after the one-year basic manufacturer's warranty has expired. Should the consumer, based on their own observed experience, see that each of their last few years of experience has cost them more in repairs than the upcoming cost of the warranty, they would not be hesitant to purchase the policy. This adverse selection practice may go on for several years if a company issuing the warranty does not adequately monitor the activity and subsequently change their contract wording, their pricing scheme, their agency force, or some combination of all.

8. CONCLUSIONS

Although there seems to be a growing market for extended warranties as competitive pressures continue to exist and new and complex products are brought to the market, it is fair to say that companies oftentimes do not devote enough resources to studying the financial impact brought on by either the introduction of a new warranty or the modification of an existing warranty.

Interestingly, many companies have complex data systems already set up for monitoring historical warranty claims, and may need relatively few changes to internal coding to better monitor their perpetually changing programs. Such changes are very likely to pay for themselves many times over in terms of providing data to make cost-effective decisions. In today's competitive marketplace, information is crucial and those companies who can access their data in detail have a clear competitive edge.

Historically, many companies have found that their warranty accruals have been understated, sometimes resulting in large balance sheet corrections that catch the eye of financial analysts. It is becoming increasingly important for companies to monitor their warranty results going forward both with revision of internal data coding as well as with actuarial type analyses involving cost projections. Once projections have been contemplated, it is important that information continue to be monitored appropriately given all the potential sources of adverse selection ("goodwill" claims, third-party agency force, etc.).

Even the simplest of actuarial loss projection methods can prove invaluable in a company's understanding of historical cost emergence, or more significantly, their future cost emergence. In the competitive marketplace, the ability to quantify the potential impact competitors' warranty changes on a company's own book will allow for more informed business decisions and most importantly, more profitable financial results.

Appendix A

- Cost of Adding Widget Repair from 6 to 25 Years – EXHIBIT 1
- Cost of Extending Full Value Warranty Period from 4 to 6 Years – EXHIBIT 2
- Warranty Cost for First Five Years Based on Percentage of Year 1 (Basic Warranty) Costs – EXHIBIT 3

9. REFERENCES

[1] Bornhuetter, Ronald L. and Ronald E. Ferguson., "The Actuary and IBNR," *Proceedings of the Casualty Actuarial Society* LIX, 1972, pp. 181-195.

[2] Cummins, H.J., "Worthless Warranty," *Minneapolis Star Tribune*, December 12, 2007, pages D1, D3.

[3] Financial Accounting Standards Board Interpretation No. 45 – Guarantor's Accounting and Disclosure Requirements for Guarantees, Including Indirect Guarantees of Indebtedness of Others. http://www.fasb.org/pdf/fin%2045.pdf.

[4] Financial Accounting Standards Board Technical Bulletin No. 90-1 – Accounting for Separately Priced Extended Warranty and Product Maintenance Contracts. http://www.fasb.org/pdf/ftb%2090-1.pdf

[5] National Association of Insurance Commissioners – Instructions to the Annual Statement for Property/Casualty, www.naic.org.

[6] Spraker, Greg, "Warranty Reserve Levels," *Warranty Week*, 30 January 2007. http://www.warrantyweek.com/archive/ww20070130.html. Accessed 30 January 2008.

Abbreviations and notations

ALAE, allocated loss adjustment expense
FASB, Financial Accounting Standards Board
NAIC, National Association of Insurance Commissioners
ULAE, unallocated loss adjustment expense

Biographies of the Authors

Orin M. Linden is the Practice Leader for Insurance & Financial Services for Watson Wyatt in New York, NY. He is a Fellow of the Casualty Actuarial Society, a Member of the American Academy of Actuaries and holds an Associate in Risk Management. He received his B.A. degree in Mathematics (with Honors) from Queens College of the City University of New York, and holds Masters and Ph.D. degrees from the Belfer Graduate School of Science at Yeshiva University. Mr. Linden has over 30 years of actuarial experience, having held positions at two insurance companies, and has been a partner at two of the Big 4 accounting firms prior to joining Watson Wyatt. He has made numerous speeches at CAS and other functions and published numerous papers and reviews in the *Proceedings of the Casualty Actuarial Society* and other publications. His paper, "Estimating Satellite Liabilities," co-written with Alan Gould, was awarded the 2000 Ronald Bornhuetter Loss Reserve Prize.

James B. Kahn is a Senior Consultant with Watson Wyatt in New York, NY. He is a Fellow of the Casualty Actuarial Society and a Member of the American Academy of Actuaries. He received his B.A. degree in Mathematics (with an application in Probability and Statistics) from the Johns Hopkins University in Baltimore, MD. Mr. Kahn has over 18 years of actuarial experience, having held various positions at a primary insurance company, a reinsurance company, and consulting firms prior to coming to Watson Wyatt.

Brian Ko is a Consultant with Watson Wyatt in New York, NY. He received his B.A. degree in Mathematics and International Studies from the University of Pennsylvania's School of Arts and Sciences in Philadelphia, PA, and his B.S. degree in Finance from the University of Pennsylvania's Wharton School. Mr. Ko has four years of actuarial experience, having held various positions at the Insurance Services Office prior to coming to Watson Wyatt.

The Prediction Error of Bornhuetter-Ferguson

Thomas Mack

Abstract: Together with the Chain Ladder (CL) method, the Bornhuetter-Ferguson (BF) method is one of the most popular claims reserving methods. Whereas a formula for the prediction error of the CL method has been published already in 1993, there is still nothing equivalent available for the BF method. On the basis of the BF reserve formula, this paper develops a stochastic model for the BF method. From this model, a formula for the prediction error of the BF reserve estimate is derived.

Moreover, the model gives important advice on how to estimate the parameters for the BF reserve formula. For example, it turns out that the appropriate BF development pattern is different from the CL pattern. This is a nice add-on as it makes BF a standalone reserving method that is fully independent from CL. The other parameter required for the BF reserve is the well-known initial estimate for the ultimate claims amount. Here the stochastic model clearly shows what has to be meant with "initial."

In order to apply the formula for the prediction error, the actuary must assess his uncertainty about both sets of parameters, about the development pattern and about the initial ultimate claims estimates. But for both, much guidance can be drawn from the estimates themselves and from the run-off data given. Finally, a numerical example shows how the resulting prediction error compares to the one of the CL method.

Keywords: Loss reserving, Bornhuetter-Ferguson, Stochastic model, Prediction error.

1. INTRODUCTION

For most insurance companies and their auditors, the use of the Chain Ladder method (CL) and of the Bornhuetter-Ferguson method (BF) has become a certain standard or benchmark in claims reserving. This means that these methods are applied in almost every case, and only if they seem to fail, one looks for other methods. Originally, these methods gave only a point estimate for the claims reserve. But this was not satisfactory because then one could not decide whether the estimates differ significantly or not. Moreover, for the calculation of risk-based capital and of premium loadings one needs to assess the prediction error of the estimate (i.e., the standard deviation of the true claims reserve from the point estimate).

In 1993, a formula for the prediction error of the CL reserve estimate was published (Mack (1993) or the more comprehensive version Mack (1994)), which in the mean time is widely used. This formula gives an answer to the question of significant differences to other methods and measures the variability of the true reserves for business segments where CL is acceptable. But for BF, such a formula is still missing. This may seem strange because BF is even simpler than CL. But this simplicity is just the problem. The prediction error consists of two components, the process error and the estimation (or parameter) error. Whereas the estimation error basically always can be calculated via the laws of error propagation, for the process error a stochastic model of the claims process is required. The latter was feasible in the CL case because the way in which the CL age-to-age factors are estimated contains implicit information on the underlying stochastics. In the BF case, no clear procedure on how to estimate the parameters has been established. In such a situation, many models may seem admissible.

The stochastic model for BF introduced in this paper is very similar in its structure to the CL model of Mack (1993) but adequately reflects the two fundamental differences between CL and BF. The first difference is the fact that the CL reserve is directly proportional to the claims amount known so far whereas the BF reserve does not depend at all on the known claims amount. This is reflected in an additional independence assumption of the BF model. The second difference is the fact that the BF reserve estimate includes the full tail of the claims development whereas the standard CL reserve (i.e., without additional tail factor) only considers the development until a given last development year. The latter fact implies that the parameter estimation for the BF model has also to consider the tail of the development where there is no data and some judgment is required. Therefore, we do not give a unique estimation formula for the tail parameters but discuss two alternative ways to cope with this problem. In any case, the development pattern suggested by the BF model turns out to be different from the well-known CL pattern. This makes BF to a really standalone reserving method. But still, the actuary may make his own selections regarding the development pattern, especially for the tail.

In addition to the development pattern, the BF reserve formula requires another element, an initial estimate for the ultimate claims amount. Of course, the uncertainty of this estimate must have a high impact on the prediction error. As this estimate usually comes from outside (e.g., from pricing) or is simply set by the actuary on the basis of his knowledge of the business, its uncertainty must be assessed from outside of the run-off triangle, too. And an actuary who is able to set (or accept) a point estimate should also be able to quantify (or ask for quantification of) the uncertainty of this estimate. Moreover, from the stochastic model important advice can be derived for the assessment of these estimates and their uncertainty. Altogether, this means that the prediction error of the BF reserve estimate depends largely on the (more or less subjective) assessment of the actuary as it is already the case with the BF reserve estimate itself.

Section 2 gives a short review of the BF method and of its connections and differences to the CL method. Section 3 describes the appropriate stochastic BF model. Section 4 shows two ways to estimate or select the model parameters. The estimation of the standard error of the parameters is discussed in Section 5 where also the formula for the prediction error and its components is derived. Section 6 gives a numerical example and Section 7 concludes.

2. THE BF METHOD

Let $C_{i,k}$ denote the cumulative claims amount (either paid or incurred) of accident year i after k years of development, $1 \le i, k \le n$, and v_i be the premium volume of accident year i where n denotes the most recent accident year. Then $C_{i,n+1-i}$ denotes the currently known claims amount of accident year i. Let further $S_{i,k} = C_{i,k} - C_{i,k-1}$ denote the incremental claims amount (with $C_{i,0} = 0$) and U_i the (unknown) ultimate claims amount of accident year i. Then $R_i = U_i - C_{i,n+1-i}$ is the (unknown true) claims reserve for accident year i. Let finally $S_{i,n+1} = U_i - C_{i,n}$ be the incremental claims amount after development year n (tail development).

Bornhuetter-Ferguson (1972) introduced their method to estimate R_i in order to cope with a major weakness of the CL method. Therefore we first consider this weakness. CL uses link ratios (age-to-age factors) $\hat{f}_k$ and a tail factor $\hat{f}_\infty$ in order to project the current claims amount $C_{i,n+1-i}$ to ultimate, i.e., it estimates $\hat{U}_i^{CL} = C_{i,n+1-i} \cdot \hat{f}_{n+2-i} \cdot \ldots \cdot \hat{f}_n \cdot \hat{f}_\infty$, and therefore the CL reserve is

$$\hat{R}_i^{CL} = \hat{U}_i^{CL} - C_{i,n+1-i} = C_{i,n+1-i}\left(\hat{f}_{n+2-i} \cdot \ldots \cdot \hat{f}_\infty - 1\right).$$

This means that the reserve strongly depends on the current amount $C_{i,n+1-i}$, which can, for example, lead to a nonsense reserve $\hat{R}_i^{CL} = 0$ for accident years where currently no claims are paid or reported, which is not unusual in excess-of-loss reinsurance for the most recent accident year(s).

The BF reserve estimate avoids this dependency from the current claims amount $C_{i,n+1-i}$. It is

$$\hat{R}_i^{BF} = \hat{U}_i\left(1 - \hat{z}_{n+1-i}\right)$$

where

$\hat{U}_i = v_i \hat{q}_i$ with a prior estimate $\hat{q}_i$ for the ultimate claims ratio $q_i = U_i/v_i$ of accident year i,

$\hat{z}_k \in [0, 1]$ is the estimated percentage of the ultimate claims amount that is expected to be known after development year k.

The term $\hat{q}_i$ is called 'prior" (or "initial") as opposed to the posterior estimate $(C_{i,n+1-i} + \hat{R}_i^{BF})/v_i$ for the ultimate claims ratio, which is based on the prior $\hat{q}_i$ and is different iff $C_{i,n+1-i} \neq \hat{z}_{n+1-i} v_i \hat{q}_i$, i.e., if the current claims amount deviates from its estimated expectation. The percentages z_1, z_2, ... constitute the expected cumulative development pattern and $1 - \hat{z}_{n+1-i}$ is therefore an estimate for the percentage of the expected outstanding claims of accident year i.

Having already an estimate $\hat{U}_i$, the question may arise why BF does not simply use $\hat{R}_i = \hat{U}_i - C_{i,n+1-i}$ as reserve estimate. In that case, the reserve estimate would become the higher, the smaller the current amount $C_{i,n+1-i}$ is and would again strongly depend on $C_{i,n+1-i}$. With CL, the reserve estimate behaves just in the opposite way, i.e., is the smaller, the smaller $C_{i,n+1-i}$ is. Here BF takes a neutral position: It does not care about the size of $C_{i,n+1-i}$ at all, i.e., it considers the deviation between the observed amount $C_{i,n+1-i}$ and the expected amount $\hat{z}_{n+1-i}\hat{U}_i$ as purely random and by no means indicative for the future development. Altogether, the essential feature of the BF method is to avoid any dependency between $C_{i,n+1-i}$ and $\hat{R}_i^{BF}$.

In order to apply the BF method, the actuary has to estimate the parameters q_i and z_k for all i and k. In practice, the ultimate claims ratios q_i are estimated in various ways, mainly based on additional pricing and market information in such a way that any expected differences between the accident years are reasonably reflected. The z_k are usually derived from the (selected) CL link ratios $\hat{f}_2,\ldots,\hat{f}_n$ together with a selected tail factor $\hat{f}_\infty$ in the following way:

$$\hat{z}_n = \hat{f}_\infty^{-1},\ \hat{z}_{n-1} = \left(\hat{f}_n \cdot \hat{f}_\infty\right)^{-1},\ \ldots,\ \hat{z}_1 = \left(\hat{f}_2 \cdot \ldots \cdot \hat{f}_n \cdot \hat{f}_\infty\right)^{-1}.$$

The systematic use of the CL link ratios assumes that the outstanding claims part is a direct multiple of the already known part at each point of the development. This contradicts the basic BF idea of the independence between $C_{i,n+1-i}$ and $\hat{R}_i^{BF}$, i.e., between past and future claims, which was fundamental for the origin of the BF method. At least, with the use of the CL pattern, the BF method cannot really claim to be a standalone reserving method. Moreover, in the following we will see that the stochastic BF model suggests a different way to estimate the BF development pattern.

3. A STOCHASTIC MODEL UNDERLYING THE BF METHOD

From the BF reserve formula it is clear that the appropriate model for BF has to be cross-classified of the type

$$E(C_{i,k}) = x_i z_k \text{ or equivalently } E(S_{i,k}) = x_i y_k \text{ for } 1 \le i \le n \text{ and } 1 \le k \le n+1.$$

Because of $x_i y_k = (x_i a)(y_k/a)$ for any $a > 0$, x_i and y_k are only unique up to a constant factor. Thus we can—without loss of generality—impose the restriction $y_1+\ldots+y_n+y_{n+1} = 1$. This yields $E(U_i) = E(S_{i,1} + \ldots + S_{i,n+1}) = x_i$ and shows that x_i can be considered to be a measure of volume for accident year i. We therefore will assume in addition that $Var(U_i)$ is proportional to x_i or $Var(U_i/x_i)$ proportional to $1/x_i$. This is the usual assumption for the influence of the volume on the variance. Furthermore, the fundamental BF property of independence between past and future claims suggests to assume that all increments $S_{i,k}$ of the same accident year are independent – the independence of the accident years themselves being a standard assumption anyway. Note that the independence within the accident years does not hold in the CL model of Mack (1993).

Thus we work with the following model for the increments $S_{i,k}$, $1 \le i \le n$, $1 \le k \le n+1$:

(BF1) All increments $S_{i,k}$ are independent.

(BF2) There are unknown parameters x_i, y_k with $E(S_{i,k}) = x_i y_k$ and $y_1+\ldots+y_{n+1} = 1$.

(BF3) There are unknown proportionality constants s_k^2 with $Var(S_{i,k}) = x_i s_k^2$.

From these assumptions, we deduce

$$E(R_i) = x_i(y_{n+2-i} + \ldots + y_{n+1}) = x_i(1 - z_{n+1-i}) \quad \text{with} \quad z_k := y_1 + \ldots + y_k,$$

which shows that the expected claims reserve has the same form as the BF reserve estimate. Furthermore, we have

$$Var(U_i) = Var(S_{i1} + \ldots + S_{i,n+1}) = x_i\left(s_1^2 + \ldots + s_{n+1}^2\right),$$

which shows that $Var(U_i)$ is proportional to x_i as intended.

This model is thought to be the most general model fitting to the philosophy of the BF method. Like with the CL model and as suggested by having an own parameter y_k for the expectation in each column, it here, too, makes sense to assume that the variability constant s_k^2 is the same for all $S_{i,k}$ within each column k but differs from column to column. The simpler assumption $Var(S_{i,k}) = c x_i y_k$ for all i, k seems to contradict to reality as has already been mentioned by Taylor (2002) because then "the coefficient of variation of the claim size is inversely related to the mean claim size," which is "opposite of what one observes." Moreover, this last variance assumption is just a special case of (BF3) and thus less general. Finally, this variance assumption would imply that all y_k be > 0, which is not the case with (BF3), and which would prevent using the model for incurred claims amounts where negative incremental claims are not uncommon.

Like with the CL model of Mack (1993), this model is heavily parametrized, especially for the late development years. But, of course, the actuary may—depending on the data—apply additional regression assumptions in order to reduce the number of parameters and to stabilize the estimates. This is shown in the numerical example below.

From the above model, we deduce further

$$Var(R_i) = x_i\left(s_{n+2-i}^2 + \ldots + s_{n+1}^2\right).$$

As background for the next section, we note that with $x_1, \ldots, x_n$ known,

$$\hat{y}_k = \sum_{i=1}^{n+1-k} S_{i,k} \Big/ \sum_{i=1}^{n+1-k} x_i \,, \tag{1}$$

is a linear minimum variance unbiased estimate of y_k, $1 \le k \le n$, and

$$\hat{s}_k^2 = \frac{1}{n-k} \sum_{i=1}^{n+1-k} \left(S_{i,k} - x_i \hat{y}_k\right)^2 \Big/ x_i \tag{2}$$

is an unbiased estimate of s_k^2, $1 \le k \le n$-1.

4. PARAMETER ESTIMATION FOR THE BF MODEL

From the model above we clearly see what is meant with calling $\hat{U}_i$ a "prior" or "initial" estimate: It has to be an estimate $\hat{x}_i$ for the unconditional (= prior, initial) expectation $x_i = E(U_i)$ and not for the "posterior" expectation $E(U_i \mid C_{i,n+1-i})$, given $C_{i,n+1-i}$. This shows that the claims amount $C_{i,n+1-i} = S_{i,1} + \ldots + S_{i,n+1-i}$ known so far should not be the main basis for the estimate $\hat{x}_i$. For example, it would be wrong to use for $\hat{x}_i$ the posterior estimate $C_{i,n-i} + \hat{R}_i^{BF(n-1)}$ of last year's reserving because this is an estimate for $E(U_i \mid C_{i,n-i})$ and not for $E(U_i)$. Even a very large random claim that happened in accident year i and is already known must not change the estimate $\hat{x}_i$ as long as it fits the randomness assumed in the pricing model. As an extreme example, we might have an accident year where $\hat{x}_i < C_{i,n+1-i}$. Thus, the estimate $\hat{U}_i$ should be prior to making the known claims experience $C_{i,k}$ of accident year i a decisive basis of the estimate. But this does not mean that the prior estimate $\hat{x}_i$ cannot change during the claims development.

To fix ideas, let us assume that $\hat{x}_i$ originally stems from pricing (which has taken place before the end of development year 1). Usually, the pricing is based on the (trended) claims experience of the preceding accident years (i.e., on the years i-1, i-2, …) and on assumptions on the future claims cost inflation. This basic information develops from year to year because the claims experience of the preceding years develops as well as the relevant inflation index. Thus, we can reprice the business of accident year i every later year and thus arrive at updated estimates for $x_i = E(U_i)$. We may even include the claims experience of the accident years i, i+1, … into this repricing of accident year i as long as it can be translated to the portfolio of accident year i. In any case, the own claims experience $C_{i,n+1-i}$ should only have a marginal influence on $\hat{x}_i$ otherwise we would rather estimate $E(U_i \mid C_{i,n+1-i})$. Thus, the estimate $\hat{x}_i$ may change over the years but normally not to a large extent, at least if the first estimate for x_i came from a sound pricing.

When the actuary does not have the result of a complete repricing available, he has at least the data $\{v_i, C_{ik}\}$ of the run-off triangle. On basis of this data and some rather general information on rate level changes, he may follow the procedure outlined in Mack (2006) which is not a full repricing but brings all accident years on about the same claims ratio level as basis for the calculation of the initial ultimate claims ratio $\hat{q}_i$.

After these clarifying remarks, we assume that the initial estimate $\hat{U}_i$ of Section 2 fulfills the requirements for being an estimate of $x_i = E(U_i)$. Thus we write $\hat{U}_i$ instead of $\hat{x}_i$ in the following. Having now an estimate $\hat{U}_i$ for $E(U_i)$, we are only left with the task to estimate y_k and s_k^2. The main problem here is the fact that we have only very few observations for the late development years. As we do not have any observations beyond development year n, we cannot estimate the tail ratio y_{n+1} without further assumptions. An outside estimate may be gained from similar portfolios with more accident years where the claims experience of later development years than year n is available. Without such information, the actuary may arrive at an estimate $\hat{y}_{n+1}$ by extrapolation from $\hat{y}_1,...,\hat{y}_n$ (which are not available yet). Similarly, an estimate for s_n^2 cannot be obtained from the only available observation of column n alone but may be obtained by extrapolation, too. Therefore, in order to fix ideas for an iterative procedure, we first consider the situation where we have already reasonable estimates $\hat{y}_{n+1}$, $\hat{s}_1^2,..., \hat{s}_n^2$. Then we can get a weighted least squares estimate (i.e., with the weights inversely proportional to the variances) for $y_1, \ldots, y_n$ by minimizing

$$Q = \sum_{i=1}^{n} \sum_{k=1}^{n+1-i} \frac{\left(S_{i,k} - \hat{U}_i \hat{y}_k\right)^2}{\hat{U}_i \hat{s}_k^2}$$

under the constraint $\hat{y}_1 + ... + \hat{y}_n = 1 - \hat{y}_{n+1}$. As starting values for the minimization we can use

$$\hat{\hat{y}}_k = \sum\nolimits_{i=1}^{n+1-k} S_{i,k} \Big/ \sum\nolimits_{i=1}^{n+1-k} \hat{U}_i \,, \tag{3}$$

(see (1)) but these will usually not fulfill the constraint.

In most cases the data will not be so stable that the resulting least squares estimates $\hat{y}_1,...,\hat{y}_n$ seem reliable enough to leave them as they are (especially for k large). Therefore, the actuary will apply a smoothing procedure to select his own final $\hat{y}_1^*,...,\hat{y}_n^*,\hat{y}_{n+1}^*$ (i.e., including a possible revision of the tail ratio in view of the other $\hat{y}_k^*$) with $\hat{y}_1^* + ... + \hat{y}_n^* + \hat{y}_{n+1}^* = 1$.

On the basis of the fact that the actuary will in any case make some own selections due to the few data, he can dispense with the above exact minimization and just proceed as follows: He starts with the raw estimates $\hat{\hat{y}}_k, 1 \le k \le n,$ as given in (3) and applies some manual smoothing and extrapolating in order to arrive at his final selection for $\hat{y}_1^*,...,\hat{y}_n^*,\hat{y}_{n+1}^*$ fulfilling $\hat{y}_1^* + ... + \hat{y}_n^* + \hat{y}_{n+1}^* = 1$. In view of (2), he then estimates s_k^2 by

$$\hat{\hat{s}}_k^2 = \frac{1}{n-k}\sum_{i=1}^{n+1-k}\left(S_{i,k} - \hat{U}_i\hat{y}_k^*\right)^2 \Big/ \hat{U}_i\,, \quad 1 \le k \le n\text{-}1, \tag{4}$$

and again applies some smoothing in order to select his final $\hat{s}_1^{2*},\ldots,\hat{s}_{n-1}^{2*}$ and an extrapolation to obtain $\hat{s}_n^{2*}$. Note that $\hat{s}_{n+1}^{2*}$ cannot be obtained in this way because it usually has to cover several development years as is the case for $\hat{y}_{n+1}$, too. Therefore, $\hat{s}_{n+1}^{2*}$ may be arrived at by interpolating a regression of $\hat{s}_k^{2*}$ against $\left|\hat{y}_k^*\right|$ at the point $\left|\hat{y}_{n+1}^*\right|$. (Note that some $\hat{y}_k$ may be negative.) The whole estimation procedure is shown in the numerical example.

A more formal way to estimate the parameters y_k, s_k^2 (in case of rather stable data) would be as follows: On the basis of $\hat{\hat{y}}_k, 1 \le k \le n,$ according to (3), we decide on the formula for a smoothing regression, e.g., $ln(\hat{y}_k) = \alpha - \beta \cdot k$ for k above some $k_1 < n$ (assuming $y_k > 0$ there), which then is extrapolated until some final development year $k_2 > n$. Then we calculate $\hat{\hat{s}}_k^2$ (according to (4) but using the smoothened $\hat{y}_k$ for $k > k_1$). The resulting values $\hat{\hat{s}}_1^2,\ldots,\hat{\hat{s}}_{n-1}^2$ are now kept fixed and used in the above constrained minimization of Q to obtain better values for $\hat{y}_1,\ldots,\hat{y}_{k_1},\alpha,\beta$ under the constraint

$$\hat{y}_1 + \ldots + \hat{y}_{k_1} + exp\left(\alpha - \beta(k_1+1)\right) + \ldots + exp\left(\alpha - \beta k_2\right) = 1.$$

Note that in Q we have to leave out the term for $(i, k) = (1, n)$ because now we do not yet have a value for $\hat{s}_n$. This minimization yields our selections for all $\hat{y}_k^*$: The values for $k = 1, \ldots, k_1$ are obtained directly, those for $k = k_1+1, \ldots, n$ are taken from the smoothing regression and $\hat{y}_{n+1}^*$ is obtained by adding up the extrapolated values of the regression up to development year k_2. Using these $\hat{y}_k^*$, we calculate new values $\hat{\hat{s}}_k^2$ according to (4) and plot $ln\left(\hat{\hat{s}}_k^2\right)$ for $k > k_1$ against $\left|\hat{y}_k^*\right|$ or $ln\left(\left|\hat{y}_k^*\right|\right)$ in order to select appropriate values for $\hat{s}_k^{2*}$, especially for $k = n$ (over $\left|\hat{y}_n^*\right|$) and $k = n+1$ (over $\left|\hat{y}_{n+1}^*\right|$). Of course, we could now apply another constraint minimization with these new values of $\hat{s}_k^{2*}$, but usually this will not change much. Note that the values of $\hat{s}_k^{2*}$ for $k > k_1$ will be overestimated a little as we did not change the degrees of freedom in formula (4) for $\hat{\hat{s}}_k^2$ which would have been possible as the regression employs fewer parameters.

As the result of each of these two estimation procedures we have selected $\hat{y}_1^*,\ldots,\hat{y}_n^*,\hat{y}_{n+1}^*$ and $\hat{s}_1^{2*},\ldots,\hat{s}_n^{2*},\hat{s}_{n+1}^{2*}$ from which we estimate the BF claims reserve by

$$\hat{R}_i^{BF} = \hat{U}_i\left(\hat{y}_{n+2-i}^* + \ldots + \hat{y}_{n+1}^*\right) = \hat{U}_i\left(1 - \hat{z}_{n+1-i}^*\right) \text{ with } \hat{z}_k^* = \hat{y}_1^* + \ldots + \hat{y}_k^*.$$

$\hat{s}_1^{2*},\ldots,\hat{s}_n^{2*},\hat{s}_{n+1}^{2*}$ will be needed for the prediction error.

The properties of the above estimators can be sketched as follows:

(a) $\hat{y}_1^*,\ldots,\hat{y}_n^*,\hat{y}_{n+1}^*$ are pairwise (slightly) negatively correlated as they have to add up to unity.

(b) $\hat{y}_1^*,\ldots,\hat{y}_n^*,\hat{y}_{n+1}^*$ and therefore also $\hat{z}_1^*,\ldots,\hat{z}_{n+1}^*$ are practically independent from $\hat{U}_1,\ldots,\hat{U}_n$ as the latter do not really influence the size of any $\hat{y}_k^*$ because these have to add up to unity in any case and because of selections and regressions used.

(c) $\hat{R}_i^{BF}$ and R_i are independent (due to BF1).

(d) $E(\hat{U}_i) = E(U_i) = x_i$, $1 \le i \le n$.

(e) $E(\hat{y}_k^*) = y_k$, $1 \le k \le n+1$, and therefore $E(\hat{z}_k^*) = z_k$, $1 \le k \le n+1$.

(f) $E(\hat{s}_k^{2*}) = s_k^2$, $1 \le k \le n+1$.

In (d) – (f) we have simply assumed that the actuary's selections are unbiased.

The unbiasedness of the reserve estimate $\hat{R}_i^{BF}$ follows directly from these properties:

$$E(\hat{R}_i^{BF}) = E(\hat{U}_i)E(1-\hat{z}_{n+1-i}^*) = x_i(1-z_{n+1-i}) = E(R_i).$$

Note that the raw estimates $\hat{\hat{y}}_k$ according to (3) are identical to the estimates $\hat{\beta}_k$ in Mack (2006) which were shown there as being suggested directly by the BF reserve formula itself. In any case and even without any smoothing of $\hat{\hat{y}}_k$, the resulting development pattern will turn out to be different from the CL pattern (see also the numerical example below).

Now we are prepared to derive the formula for the prediction error.

5. THE PREDICTION ERROR OF THE BF METHOD

As one is interested in the future variability only, given the data observed so far, the mean squared error of prediction of any reserve estimate $\hat{R}_i$ is defined to be

$$msep(\hat{R}_i) = E\left(\left(\hat{R}_i - R_i\right)^2 \middle| S_{i,1},\ldots,S_{i,n+1-i}\right).$$

According to (BF1), $R_i = S_{i,n+2-i}+\ldots+S_{i,n+1}$ is independent from $S_{i,1}, \ldots, S_{i,n+1-i}$. Also, the BF reserve estimate $\hat{R}_i^{BF}$ can be taken as being independent from $S_{i,1}, \ldots, S_{i,n+1-i}$ (as these play at most a marginal role when selecting $\hat{U}_i$ and $\hat{y}_{n+2-i}^*,\ldots,\hat{y}_{n+1}^*$), more precisely, R_i and $\hat{R}_i^{BF}$ are taken to be commonly independent from $S_{i,1}, \ldots, S_{i,n+1-i}$. Thus we have

$$msep(\hat{R}_i^{BF}) = E\left(\left(\hat{R}_i^{BF} - R_i\right)^2\right)$$

$$= Var\left(\hat{R}_i^{BF} - R_i\right) + \left(E\left(\hat{R}_i^{BF}\right) - E\left(R_i\right)\right)^2$$

$$= Var\left(\hat{R}_i^{BF}\right) + Var\left(R_i\right),$$

i.e., the mean squared error of prediction is the sum of the (squared) estimation error $Var\left(\hat{R}_i^{BF}\right)$ and of the (squared) process error $Var\left(R_i\right)$.

For the process error we simply have

$$Var\left(R_i\right) = Var\left(S_{i,n+2-i}\right) + \ldots + Var\left(S_{i,n+1}\right) = x_i\left(s_{n+2-i}^2 + \ldots + s_{n+1}^2\right),$$

which will be estimated by

$$\hat{V}ar\left(R_i\right) = \hat{U}_i\left(\hat{s}_{n+2-i}^{2*} + \ldots + \hat{s}_{n+1}^{2*}\right).$$

For the estimation error of $\hat{R}_i^{BF} = \hat{U}_i\left(1 - \hat{z}_{n+1-i}^*\right)$, we use the general formula

$$Var\left(XY\right) = \left(E\left(X\right)\right)^2 Var\left(Y\right) + Var\left(X\right)Var\left(Y\right) + Var\left(X\right)\left(E\left(Y\right)\right)^2$$

for independent random variables X and Y and obtain

$$Var\left(\hat{R}_i^{BF}\right) = \left(E\left(\hat{U}_i\right)\right)^2 Var\left(\hat{z}_{n+1-i}^*\right) + Var\left(\hat{U}_i\right)Var\left(\hat{z}_{n+1-i}^*\right) + Var\left(\hat{U}_i\right)\left(1 - E\left(\hat{z}_{n+1-i}^*\right)\right)^2$$

$$= \left(x_i^2 + Var\left(\hat{U}_i\right)\right)Var\left(\hat{z}_{n+1-i}^*\right) + Var\left(\hat{U}_i\right)\left(1 - z_{n+1-i}\right)^2.$$

Whereas we have already estimators $\hat{U}_i$ for x_i and $\hat{z}_{n+1-i}^*$ for z_{n+1-i}, we still need estimates for $Var\left(\hat{U}_i\right)$ and $Var(\hat{z}_{n+1-i}^*)$, i.e., we have to quantify the precision of $\hat{U}_i$ and $\hat{z}_{n+1-i}^*$.

The standard error $s.e.\left(\hat{U}_i\right)$, i.e., an estimate for $\sqrt{Var\left(\hat{U}_i\right)}$, cannot be obtained from the estimation error $s.e.\left(\hat{R}_i^{BF(n-1)}\right)$ of last year's reserving because this would ignore the variability of $C_{i,n-i}$, which has to be included into $s.e.\left(\hat{U}_i\right)$. Like $\hat{U}_i$ itself, $s.e.\left(\hat{U}_i\right)$ is best be obtained from a repricing of the business. But one has to be cautious there. For example, the variability of the posterior claims ratio estimates $\hat{U}_1^{post}/v_1, \ldots, \hat{U}_n^{post}/v_n$ would underestimate $s.e.\left(\hat{U}_i/v_i\right)$ because these estimates are positively correlated via the common estimates $\hat{z}_k^*$. Similarly, also the initial estimates $\hat{U}_1, \ldots, \hat{U}_n$ will usually be positively correlated. Thus the formula

$$\left(s.e.\left(\hat{U}_i\right)\right)^2 = \frac{v_i}{n-1}\sum_{j=1}^{n} v_j\left(\frac{\hat{U}_j}{v_j} - \hat{q}\right)^2 \quad \text{with} \quad \hat{q} = \sum_{j=1}^{n}\hat{U}_j \Big/ \sum_{j=1}^{n} v_j \qquad (5)$$

(which is analogous to (1), (2) for BF3) is applicable only if the initial estimates $\hat{U}_j$ can be assumed to be uncorrelated. But even then, using the real premiums v_j would include the market cycle of premium adequacy into $s.e.(\hat{U}_i)$, which would overestimate $s.e.(\hat{U}_i)$ in those situations where we can predict the market cycle rather well. Thus, we should remove the influence of the market cycle from (5) by using on-level premiums $\tilde{v}_j$. In addition, we should correct for any positive correlation between the $\hat{U}_i$s by replacing the term $n - 1$ of (5) with for example, $n - \sqrt{n}$ for a constant correlation coefficient $\hat{\rho}_{ij}^U = 1/\sqrt{n}$ between $\hat{U}_i$ and $\hat{U}_j$ or with (approximately) $n - \sqrt{2n}$ for a decreasing correlation coefficient $\hat{\rho}_{ij}^U = 1/(1+|i-j|)$; the precise formula being $n - \sum_{i,j} \rho_{ij}^U \sqrt{\frac{v_i}{v_+}\frac{v_j}{v_+}}$ with $v_+ = \sum_{i=1}^{n} v_i$.

Usually, these standard errors $s.e.(\hat{U}_i)$ will not change much over the years. Of course, we will have slight changes as long as the $\hat{U}_i$ change. But even at the end of the development, we will not know $E(U_i)$ much more precisely than at the beginning. The actuary should examine the plausibility of the resulting values of $s.e.(\hat{U}_i)$, for instance in the following way: If we assume a normal distribution, then the interval $(\hat{U}_i - 2 \cdot s.e.(\hat{U}_i), \hat{U}_i + 2 \cdot s.e.(\hat{U}_i))$ will contain the true $E(U_i)$ with 95% probability. Thus, if the size of the interval is plausible, then $s.e.(\hat{U}_i)$ is plausible, too.

Next, we have to decide on how to estimate

$$Var(1-\hat{z}_{n+1-i}^*) = Var(\hat{z}_{n+1-i}^*) = Var(\hat{y}_1^* + \ldots + \hat{y}_{n+1-i}^*) = Var(\hat{y}_{n+2-i}^* + \ldots + \hat{y}_{n+1}^*).$$

From property (a) we see that we will be on the safe side when we replace $Var(\hat{y}_1^* + \ldots + \hat{y}_{n+1-i}^*)$ with $Var(\hat{y}_1^*) + \ldots + Var(\hat{y}_{n+1-i}^*)$. But whereas the latter sum increases with each additional term, this is not the case with $Var(\hat{y}_1^* + \ldots + \hat{y}_{n+1-i}^*)$ as finally $Var(\hat{y}_1^* + \ldots + \hat{y}_{n+1}^*) = Var(1) = 0$. Therefore we replace $Var(\hat{z}_k^*) = Var(1-\hat{z}_k^*)$ for small k with $Var(\hat{y}_1^*) + \ldots + Var(\hat{y}_k^*)$ and for large k with $Var(\hat{y}_{k+1}^*) + \ldots + Var(\hat{y}_{n+1}^*)$. More precisely, we replace—still being on the safe side—

$$Var(\hat{z}_k^*) \text{ with } min(Var(\hat{y}_1^*) + \ldots + Var(\hat{y}_k^*),\ Var(\hat{y}_{k+1}^*) + \ldots + Var(\hat{y}_{n+1}^*)).$$

Due to $\hat{y}_k^* \approx \hat{\hat{y}}_k \approx \sum_{j=1}^{n+1-k} S_{j,k} \Big/ \sum_{j=1}^{n+1-k} x_j$, we can assume that

$$Var(\hat{y}_k^*) \approx Var\left(\sum_{j=1}^{n+1-k} S_{j,k} \Big/ \sum_{j=1}^{n+1-k} x_j \right) = \frac{s_k^2}{\sum_{j=1}^{n+1-k} x_j}, \quad 1 \le k \le n.$$

Therefore we estimate $Var\left(\hat{y}_k^*\right)$ by

$$\left(s.e.\left(\hat{y}_k^*\right)\right)^2 = \frac{\hat{s}_k^{2*}}{\sum_{j=1}^{n+1-k} \hat{U}_j}, \quad 1 \leq k \leq n. \tag{6}$$

But the value of $s.e.\left(\hat{y}_{n+1}^*\right)$ must come from outside. Without this, a plausible choice is often $s.e.\left(\hat{y}_{n+1}^*\right) = 0.5\hat{y}_{n+1}^*$, i.e., a coefficient of variation $c.v.\left(\hat{y}_{n+1}^*\right)$ = 50%, assuming a normal distribution with 95% probability within the interval (0; 2 $\hat{y}_{n+1}^*$).

Altogether, our estimate $\left(s.e.\left(\hat{z}_k^*\right)\right)^2$ for $Var\left(\hat{z}_k^*\right)$ is

$$\left(s.e.\left(\hat{z}_k^*\right)\right)^2 = min\left(\left(s.e.\left(\hat{y}_1^*\right)\right)^2 + \ldots + \left(s.e.\left(\hat{y}_k^*\right)\right)^2, \left(s.e.\left(\hat{y}_{k+1}^*\right)\right)^2 + \ldots + \left(s.e.\left(\hat{y}_{n+1}^*\right)\right)^2\right). \tag{7}$$

In any case, we have $s.e.\left(\hat{z}_{n+1}^*\right) = s.e.(1) = 0$. Of course, the actuary will check the plausibility of $s.e.\left(\hat{z}_k^*\right)$ similarly as $s.e.(\hat{U}_i)$ and, if necessary, manually adjust some of the resulting values.

Thus we finally obtain the following estimator for the mean squared error of prediction:

$$\hat{msep}\left(\hat{R}_i^{BF}\right) = \hat{U}_i\left(\hat{s}_{n+2-i}^{2*} + \ldots + \hat{s}_{n+1}^{2*}\right) + \left(\hat{U}_i^2 + \left(s.e.\left(\hat{U}_i\right)\right)^2\right)\left(s.e.\left(\hat{z}_{n+1-i}^*\right)\right)^2 + \left(s.e.\left(\hat{U}_i\right)\right)^2\left(1 - \hat{z}_{n+1-i}^*\right)^2.$$

This is the formula one needs for risk-based capital and premium loading calculations as well as for the construction of a confidence interval for R_i. In order to check the significance of differences between alternative reserve estimates or to construct a confidence interval for $E(U_i)$ one only needs the pure estimation error

$$\left(s.e.\left(\hat{R}_i^{BF}\right)\right)^2 = \left(\hat{U}_i^2 + \left(s.e.\left(\hat{U}_i\right)\right)^2\right)\left(s.e.\left(\hat{z}_{n+1-i}^*\right)\right)^2 + \left(s.e.\left(\hat{U}_i\right)\right)^2\left(1 - \hat{z}_{n+1-i}^*\right)^2.$$

A closer analysis of this formula shows that

$$s.e.\left(\hat{R}_i^{BF}\right)/\hat{U}_i \approx s.e.\left(\hat{z}_{n+1-i}^*\right) \qquad \text{for } \hat{z}_{n+1-i}^* \text{ close to 1,}$$

$$s.e.\left(\hat{R}_i^{BF}\right)/\hat{U}_i \approx s.e.\left(\hat{U}_i\right)/\hat{U}_i \qquad \text{for } \hat{z}_{n+1-i}^* \text{ close to 0,}$$

i.e., for the very green accident years, the uncertainty of the initial ultimate claims estimate is directly transferred to the reserve estimate.

For the overall reserve $R = R_1 + \ldots + R_n$, we have the unbiased estimate $\hat{R}^{BF} = \hat{R}_1^{BF} + \ldots + \hat{R}_n^{BF}$. Its mean squared error of prediction is $msep(\hat{R}^{BF}) = Var(\hat{R}^{BF}) + Var(R)$. For the process error we have

$Var(R) = Var(R_1) + \ldots + Var(R_n)$ due to the independence of the accident years (BF1) and thus get the estimate

$$\hat{V}ar(R) = \sum_{i=1}^{n} \hat{U}_i \left(\hat{s}^{2*}_{n+2-i} + \ldots + \hat{s}^{2*}_{n+1} \right).$$

The estimation error $Var(\hat{R}^{BF})$ is more involved because $\hat{R}_1^{BF},\ldots,\hat{R}_n^{BF}$ are positively correlated via the common parameter estimates $\hat{y}_k^*$ (and in addition via the $\hat{U}_i$s). We have

$$Var\left(\hat{R}^{BF}\right) = \sum_{i=1}^{n} Var\left(\hat{R}_i^{BF}\right) + 2\sum_{i<j} Cov\left(\hat{R}_i^{BF}, \hat{R}_j^{BF}\right).$$

For $Cov\left(\hat{R}_i^{BF}, \hat{R}_j^{BF}\right) = Cov\left(\hat{U}_i(1-\hat{z}^*_{n+1-i}), \hat{U}_j(1-\hat{z}^*_{n+1-j})\right)$ we use the general formula

$$Cov(XY, WZ) = Cov(X, W)\, E(Y)\, E(Z) + Cov(X, W)\, Cov(Y, Z) + E(X)\, E(W)\, Cov(Y, Z)$$

for random variables X, Y, W, Z where the sets $\{X, W\}$ and $\{Y, Z\}$ are independent. We omit the term in the middle, which is of lower order, and obtain

$$Cov\left(\hat{U}_i(1-\hat{z}^*_{n+1-i}), \hat{U}_j(1-\hat{z}^*_{n+1-j})\right) =$$
$$= \rho^U_{ij} \sqrt{Var(\hat{U}_i) Var(\hat{U}_j)}\, E\left(1-\hat{z}^*_{n+1-i}\right) E\left(1-\hat{z}^*_{n+1-j}\right) + \rho^z_{ij} \sqrt{Var(\hat{z}^*_{n+1-i}) Var(\hat{z}^*_{n+1-j})}\, E(\hat{U}_i) E(\hat{U}_j)$$

with the correlation coefficients

$$\rho^U_{ij} = Cov\left(\hat{U}_i, \hat{U}_j\right) \Big/ \sqrt{Var(\hat{U}_i) Var(\hat{U}_j)}\,,$$

$$\rho^z_{ij} = Cov\left(1-\hat{z}^*_{n+1-i}, 1-\hat{z}^*_{n+1-j}\right) \Big/ \sqrt{Var(\hat{z}^*_{n+1-i}) Var(\hat{z}^*_{n+1-j})}\,.$$

Thus, we only have to estimate these correlation coefficients as we have estimates for all the other terms. If the actuary does not has the possibility to obtain data-based estimates for ρ^U_{ij} (e.g., from repricing) and ρ^z_{ij}, he may simply use one of the two estimates $\hat{\rho}^U_{ij}$ as given above (after (5)) and

$$\hat{\rho}^z_{ij} = \frac{\hat{z}^*_{n+1-j}\left(1-\hat{z}^*_{n+1-i}\right)}{\hat{z}^*_{n+1-i}\left(1-\hat{z}^*_{n+1-j}\right)} \quad \text{for } i<j \text{ and } \hat{z}^*_1 \le \ldots \le \hat{z}^*_{n+1}.$$

The latter estimate stems from assuming a Dirichlet distribution (which is a generalization of the Beta distribution) for $\hat{y}^*_1,\ldots,\hat{y}^*_n$. Thus we finally get

$$\left(s.e.(\hat{R}^{BF})\right)^2 = \sum_{i=1}^{n} \left(s.e.(\hat{R}_i^{BF})\right)^2 + 2\sum_{i<j} \hat{C}ov\left(\hat{R}_i^{BF}, \hat{R}_j^{BF}\right)$$

with

$$\hat{C}ov\left(\hat{R}_i^{BF}, \hat{R}_j^{BF}\right) = \hat{\rho}_{ij}^{U}\, s.e.(\hat{U}_i)\, s.e.(\hat{U}_j)\left(1-\hat{z}_{n+1-i}^*\right)\left(1-\hat{z}_{n+1-j}^*\right) + \hat{\rho}_{ij}^{z}\, s.e.(\hat{z}_{n+1-i}^*)\, s.e.(\hat{z}_{n+1-j}^*)\hat{U}_i\hat{U}_j\,.$$

6. NUMERICAL EXAMPLE

The paid triangle of Exhibit A of Mack (2006), see also Table 0 below, with n = 13 is used as example and we keep the initial ultimate claims estimates $\hat{U}_i$ from there (Exhibit C, column (I)), see Table 2 below, second column. In a first approach, we also keep the development pattern $\hat{z}_k^*$ (= b_k of Exhibit C, row (9), of Mack (2006)), see the row "selected z" in the first block of Table 1 below. This pattern can also be obtained—except for rounding differences—from the raw estimates $\hat{\tilde{y}}_k$ according to (3) by manually smoothing with the selections $\hat{y}_8^*$ = 8%, $\hat{y}_9^*$ = 5%, $\hat{y}_{10}^*$ =3.7%, $\hat{y}_{11}^*$ =2.1%, $\hat{y}_{12}^*$ =1.5%, $\hat{y}_{13}^*$ =1.4% and a tail ratio $\hat{y}_{14}^*$ =3.5%, see the second and third row of Table 1 below. In Mack (2006), this tail ratio was based on the calculation for the incurred data. From the pattern and the initial $\hat{U}_i$ the reserve estimates $\hat{R}_i^{BF1} = \hat{U}_i\left(\hat{y}_{n+2-i}^* + \ldots + \hat{y}_{n+1}^*\right) = \hat{U}_i\left(1-\hat{z}_{n+1-i}^*\right)$ are calculated. These reserves, see the fourth column of Table 2, are thus the same as in Mack (2006) except for rounding differences.

For the prediction error, we first select $\hat{s}_k^{2*}$. For this purpose, we calculate the raw $\hat{\tilde{s}}_k^2$ according to (4) and plot $ln\left(\hat{\tilde{s}}_k^2\right)$ against $\left|\hat{y}_k^*\right|$ for the decreasing part $k \geq 4$. We see that the plot looks reasonably smooth. Crucial cases are always $\hat{\tilde{s}}_{n-1}^2$ and $\hat{\tilde{s}}_{n-2}^2$, which rely on very few data. Here (n=13), according to the plot, $\hat{\tilde{s}}_{n-2}^2$ = 21.8 and $\hat{\tilde{s}}_{n-1}^2$ = 19.5 seem to be rather small. Thus, we adjust these to $\hat{s}_{n-2}^{2*} = \hat{s}_{11}^{2*}$ = 30, $\hat{s}_{n-1}^{2*} = \hat{s}_{12}^{2*}$ = 25, leave $\hat{\tilde{s}}_k^2$, $1 \leq k \leq 10$, as they are, i.e., $\hat{s}_k^{2*} = \hat{\tilde{s}}_k^2$, and manually select from the plot the missing values $\hat{s}_{13}^{2*}$ = 20 (over $\hat{y}_{13}^*$ =1.4%) and $\hat{s}_{14}^{2*}$ = 35 (over $\hat{y}_{14}^*$ =3.5%). With these selections for $\hat{s}_k^{2*}$, we calculate $s.e.\left(\hat{y}_k^*\right)$ for $1 \leq k \leq n$ = 13 according to (6) and find the resulting values and their coefficients of variation plausible. Then, we have to quantify our uncertainty on $\hat{y}_{14}^*$ = 3.5% and select it to be $s.e.\left(\hat{y}_{14}^*\right)$ = 1.5% assuming a 95%-range from 0.5% up to 6.5%. This fits well to the *s.e.* of $\hat{y}_{10}^*$, which is close to $\hat{y}_{14}^*$. Now we calculate $s.e.\left(\hat{z}_k^*\right)$ according to (7). All estimates and selections are shown in the first block of Table 1, where a bold number indicates a pure selection or a change from the raw estimate.

Finally, we have to select $s.e.(\hat{U}_i)$. In this example, we have an extreme premium cycle: The ultimate claims ratios $\hat{U}_i/v_i$ first decrease to 63%, then increase to 277%, then decrease again to 69% (see Mack (2006)). Thus, an application of equation (5) does not make sense. In Mack (2006), on-level premium factors r_i^* were estimated which bring all accident years on about the same claims ratio level. Then, the prior $\hat{U}_i$ were chosen to be

$$\hat{U}_i = v_i r_i^* \hat{m}^* \left(\hat{\hat{y}}_1 + \ldots + \hat{\hat{y}}_{n+1}\right)$$

with $\hat{\hat{y}}_k$ according to (3) and a certain constant factor $\hat{m}^*$. We can assume that the variability of $r_i^* \hat{m}^*$ is small compared to the one of $\hat{\hat{y}}_1 + \ldots + \hat{\hat{y}}_{n+1}$. Then we have

$$Var\left(\hat{U}_i\right) \approx \left(v_i r_i^* \hat{m}^*\right)^2 Var\left(\hat{\hat{y}}_1 + \ldots + \hat{\hat{y}}_{n+1}\right) = \left(v_i r_i^* \hat{m}^*\right)^2 \left(Var(\hat{\hat{y}}_1) + \ldots + Var(\hat{\hat{y}}_{n+1})\right)$$

because the $\hat{\hat{y}}_k$ s are fully independent due to BF1 as they do not have to add up to unity. As in the derivation of (6), we have

$$Var\left(\hat{\hat{y}}_k\right) \approx s_k^2 \Big/ \sum_{j=1}^{n+1-k} \hat{U}_j \text{ , i.e., we take } \left(s.e.(\hat{\hat{y}}_k)\right)^2 = \left(s.e.(\hat{y}_k^*)\right)^2 = \hat{s}_k^2 \Big/ \sum_{j=1}^{n+1-k} \hat{U}_j \text{ .}$$

Finally, in order to get rid of the factor $v_i r_i^* \hat{m}^*$, we consider the coefficient of variation and obtain

$$c.v.\left(\hat{U}_i\right) = \frac{s.e.\left(\hat{U}_i\right)}{\hat{U}_i} \approx \frac{\sqrt{\left(s.e.(\hat{\hat{y}}_1)\right)^2 + \ldots + \left(s.e.(\hat{\hat{y}}_{n+1})\right)^2}}{\hat{\hat{y}}_1 + \ldots + \hat{\hat{y}}_{n+1}} = 6.7\% \text{ .}$$

As we have ignored the variability of $r_i^* \hat{m}^*$ and have eliminated the full premium cycle (which probably would not have been achieved a priori), we deliberately increase this *c.v.* to $c.v.\left(\hat{U}_i\right)$ = 10% for all accident years i. This is considered to be a rather high uncertainty for an estimate of $E(U_i)$ for classical insurance business because, e.g., for $\hat{U}_i/v_i$ = 90%, this corresponds to a wide 95% confidence range of (72%; 108%)—note that this is the range for $E(U_i)$ and not for U_i!

Note further that this approach only works for prior estimates $\hat{U}_i$ that were obtained in this specific way. It cannot be applied to estimates $\hat{U}_i$ obtained differently, e.g., via repricing, because each approach to $\hat{U}_i$ has its own uncertainties. Normally, $c.v.(\hat{U}_i)$ will not be the same for all accident years but will be lower for years with higher volume. In our example, we leave $c.v.(\hat{U}_i)$ = 10% constant (see the third column of Table 2) assuming the varying volume has essentially been caused by writing varying shares of the same treaties. With these selections, we obtain the error estimates shown in the block "Bornh/Ferg 1" of Table 2.

We also may apply the alternative estimation procedure described in Section 4: Then, we do not use the pattern of Mack (2006) but start with the original raw $\hat{\hat{y}}_k$ according to (3) (see second row of Table 1) and select as last payment year k_2 = 20. Looking at the plot of $ln\left(\left|\hat{\hat{y}}_k\right|\right)$ against k, we select k_1 = 3 and take an initial smoothing regression $ln(\hat{y}_k) = \alpha - \beta k$ with $\alpha = -0.03874$ and $\beta = 0.3632$

for $k > k_1$. With the resulting initial values for $\hat{y}_k$, initial values for $\hat{\hat{s}}_1^2, \ldots, \hat{\hat{s}}_{n-1}^2$ are calculated according to (4), which then are kept fixed during the following minimization of Q (without the term for i=1 and k=n=13). The minimum 79.98 is obtained at $\hat{y}_1^*$ = 0.65%, $\hat{y}_2^*$= 4.7%, $\hat{y}_3^*$=13.0%, α = -0.4003, and β = 0.2920, which leads to $\hat{y}_{14}^*$ = 3.9% by adding up the extrapolated values for $\hat{y}_k$ from k=14 to k=20. For the other $\hat{y}_k^*$ (from the new regression) and the resulting $\hat{z}_k^*$ see the block "Alternative Estimates" of Table 1. Then the corresponding new $\hat{\hat{s}}_k^2$ are calculated according to (4) and the resulting values $ln\left(\hat{\hat{s}}_k^2\right)$ are plotted against $\left|\hat{y}_k^*\right|$ for $k > k_1$. In view of this plot, we change $\hat{\hat{s}}_{12}^2$= 18.7 to $\hat{s}_{12}^{2*}$ = 25 and select $\hat{s}_{13}^{2*}$ = 23 and $\hat{s}_{14}^{2*}$ = 36. Finally, we calculate $s.e.\left(\hat{y}_k^*\right)$ according to (6) and select $c.v.\left(\hat{y}_{14}^*\right)$ = 50% which gives $s.e.\left(\hat{y}_{14}^*\right)$= 1.93. The resulting reserves $\hat{R}_i^{BF2}$, see Table 2, block "Bornh/Ferg 2," are slightly higher than $\hat{R}_i^{BF1}$ for the old years and slightly lower for the new ones. The amounts (not the percentages) of the prediction error (using $c.v.(\hat{U}_i)$ = 10% as before) are all a little bit higher. Using $\hat{\rho}_{ij}^U = 1/\left(1+|i-j|\right)$, the overall reserve is $\hat{R}^{BF2}$ = 875,497 with a prediction error of 72,940 consisting of an estimation error of 62,770 and a process error of 37,152.

As comparison we apply the Chain Ladder method, too. All parameters used are given in the last block of Table 1. We have replaced the last four raw age-to-age factors with 1.04, 1.03, 1.02, 1.015, and selected a tail factor of 1.04. The latter is in accordance with the tail ratio of 3.5% - 3.9% used above. From the age-to-age factors we can derive the corresponding cumulative development pattern $\hat{z}_k$ as described in Section 2. The resulting values shown in Table 1 are close to the z-estimates of the two BF approaches but not identical. The implementation of the tail factor into the formulae for the prediction error has been done according to Mack (1999). The raw sigma-parameters (see Mack (1993) or Mack (1999)) have been kept and were supplemented with $\hat{\sigma}_n^2$ = 18 and $\hat{\sigma}_{n+1}^2$ = 40 on basis of a plot of $ln\left(\hat{\sigma}_k^2\right)$ against $ln\left(\left|\hat{f}_k - 1\right|\right)$. Finally, for the tail factor, $s.e.\left(\hat{f}_{n+1}\right)$ = 0.02 was assumed, i.e., a 95%-range from 1.00 to 1.08. This yields the results shown in the last block of Table 2. The CL reserves are close to the ones of BF except for the most recent years 2003 and 2004: In 2003, the CL reserve is about half of the BF reserve, whereas in 2004 the CL reserve is more than twice the BF reserve. This higher volatility is reflected in the markedly higher prediction errors for $i \geq 1999$, caused by a much higher process error. The CL and BF reserve estimates for 1992–2002 are not significantly different (i.e., not different by more than $2 \cdot s.e.\left(\hat{R}_i\right)$). But the reserves for 2003 are judged as being different by either method; the 2004 reserves are only different from the BF viewpoint whereas the CL estimation error is so large that the BF reserve is not judged to be different although it is less than 50% of the CL reserve. This is a good example for the fact that CL often cannot be reasonably applied in the standard way for new accident years in Excess business where almost nothing is paid in the first development year(s).

CONCLUSION

On the basis of the BF reserve formula, this paper has developed a stochastic model for the BF method that incorporates the fundamental BF property of the independence between past and future claims amounts (see model assumption BF1). Model assumption BF2 is a direct consequence of the BF reserve formula too. Only assumption BF3 is not forced by the method itself but this assumption is rather general and is only needed to derive the formula for the prediction error. Already from assumptions BF1 and BF2 important consequences for a sound application of the method can be drawn. One is the fact that the appropriate BF development pattern should not be derived from the CL age-to-age factors but be calculated independently on basis of formula (3). This makes BF a fully standalone reserving method. Moreover, the stochastic model gives important advice on how to arrive and how not to arrive at the initial estimate for the ultimate claims amount. For example, it shows that a procedure that is often used in automated reserving systems is rather questionable: It is the use of last year's posterior estimate as initial estimate for this year's reserving. On the other hand, the model shows that the initial estimate for an individual accident year may change over time as the information that has led to the estimate develops.

The independence assumption BF1 may seem more restrictive than the corresponding assumption of the CL model of Mack (1993). The required independence between the incremental amounts within every accident year may be violated, e.g., by changes in the reserving process or in the reporting behavior. In the CL model, this independence is not required, but a similar requirement can be deduced from the CL model: It is the fact that the individual development factors $C_{i,k+1}/C_{ik}$ must be uncorrelated within every accident year. This needs not be fulfilled in the BF model but can be violated by the same changes as mentioned before. As a consequence, we obtain a way of how to decide which model better suits the data by checking these independence/uncorrelatedness properties. Here we see the main advantage of having a model: It gives some guidance on how to estimate the parameters and allows various procedures (e.g., tests, plots) to see which model better suits the data. And, last but not least, it gives the possibility to quantify the reserve variability.

Especially for the BF model, the guidance mentioned leaves enough room for the actuary to bring in his specific knowledge of the business as it was always the case with the BF method. He has to select the parameters (as before) and, in addition, must assess his uncertainty about his selections. The guidance given by the model makes this crucial task feasible. And as a reward, the actuary usually will obtain less volatile reserve results than with CL, especially for the most recent accident years (see the example above). This is a big advantage regarding risk modeling and premium loading calculations. Altogether, this paper gives BF a stochastic foundation equivalent to the one already available for CL.

Acknowledgement

I am grateful to Dave Clark, Alois Gisler, and Gerhard Quarg for helpful discussions in the conception phase of this paper and to the CAS reviewers John Gibson and Mark Wenger for their comments that helped to improve the text.

REFERENCES

[1] Bornhuetter, R.L., and R.E. Ferguson, "The Actuary and IBNR," *Proceedings of the Casualty Actuarial Society* 1972, 59(112):181-195, http://www.casact.org/pubs/proceed/proceed72/72181.pdf.

[2] Mack, T., "Distribution-free Calculation of the Standard Error of Chain Ladder Reserve Estimates," *ASTIN Bulletin* 1993, 23(2):213-225, http://www.casact.org/library/astin/vol23no2/213.pdf.

[3] Mack, T., "Measuring the Variability of Chain Ladder Reserve Estimates," *Casualty Actuarial Society Forum,* Spring 1994:101-182, http://www.casact.org/pubs/forum/94spforum/94spf101.pdf.

[4] Mack, T., "The Standard Error of Chain Ladder Reserve Estimates: Recursive Calculation and Inclusion of a Tail Factor," *ASTIN Bulletin* 1999, 29(2):361-366, http://www.casact.org/library/astin/vol29no2/361.pdf.

[5] Mack, T., Parameter Estimation for Bornhuetter-Ferguson. *Casualty Actuarial Society Forum,* Fall 2006:141-157, http://www.casact.org/pubs/forum/06fforum/145.pdf.

[6] Taylor, G.C., Written discussion of "Stochastic Claims Reserving in General Insurance" by P.D. England and R.J. Verrall, *British Actuarial Journal* 2002, 8(3):540-542.

Stochastic Loss Reserving with the Collective Risk Model

Glenn Meyers, FCAS, MAAA, Ph.D.

Abstract
This paper presents a Bayesian stochastic loss reserve model with the following features.

1. The model for expected loss payments depends upon unknown parameters that determine the expected loss ratio for the given accident years and the expected payment for each settlement lag.
2. The distribution of outcomes is given by the collective risk model in which the expected claim severity increases with the settlement lag. The claim count distribution is given by a Poisson distribution with its mean determined by dividing the expected loss by the expected claim severity.
3. The parameter sets that describe the posterior distribution of the parameters in (1) above are calculated with the Gibbs sampler.
4. For each parameter set generated by the Gibbs sampler in (3), the predicted distribution of outcomes is calculated using a Fast Fourier Transform (FFT). The Bayesian predictive distribution of outcomes is a mixture of the distributions of outcomes over all the parameter sets produced by the Gibbs sampler.

This paper concludes by applying this model to the problem of calculating risk margins for loss reserves using a cost of capital formula.

Keywords
Reserving Methods, Reserve Variability, Uncertainty and Ranges, Collective Risk Model, Fourier Methods, Bayesian Estimation

1. Introduction

Over the years, there has been an increasing recognition that consideration of the random nature of the insurance loss process leads to better predictions of ultimate losses. Some of the papers that led to this recognition include Stanard [11] and Barnett and Zehnwirth [1]. Another thread in the loss reserve literature has been to recognize outside information in the formulas that predict ultimate losses. Bornhuetter and Ferguson [2] represents one of the early papers exemplifying this approach.

More recently, papers by Meyers [7] and Verrall [12] have combined these two approaches with a Bayesian methodology. This paper continues the development of the approach started by Meyers and draws from the methodology described by Verrall.

As the actuarial profession improves its ability to describe the variability of its ultimate loss projections, there arises the question on how one should take this variability into account when setting loss reserves. One proposal originated by the International Association of Insurance Supervisors (IAIS) calls for a risk margin to be added to the actuarial present value of the future loss payments. This paper applies its loss reserve model to the calculation of risk margins.

A significant accomplishment of the Meyers paper cited above was that it made predictions of the distribution of future losses of real insurers, and successfully validated these predictions on subsequent reported losses. To do this, it was necessary to draw upon data that, while generally available, comes at a price. While this made a good case that the underlying model is realistic, it tended to inhibit future research on this methodology. This paper uses simulated data so that readers can verify all calculations. In addition, this paper includes the code that produced all results and, with minor modifications, it should be possible to use this code for other loss reserving applications.

2. The Collective Risk Model

This paper analyzes a 10 x 10 triangle of incremental paid losses organized by rows for accident years 1, 2, …, 10 and by columns for development lags 1, 2, …, 10. We also have the premium associated with each accident year. Table 1 gives the triangle that underlies the examples in this paper.

Table 1 (000)

AY	Premium	*Lag* 1	*Lag* 2	*Lag* 3	*Lag* 4	*Lag* 5	*Lag* 6	*Lag* 7	*Lag* 8	*Lag* 9	*Lag* 10
1	50,000	7,168	11,190	12,432	7,856	3,502	1,286	334	216	190	0
2	50,000	4,770	8,726	9,150	5,728	2,459	2,864	715	219	0	
3	50,000	5,821	9,467	7,741	3,736	1,402	972	720	50		
4	50,000	5,228	7,050	6,577	2,890	1,600	2,156	592			
5	50,000	4,185	6,573	5,196	2,869	3,609	1,283				
6	50,000	4,930	8,034	5,315	5,549	1,891					
7	50,000	4,936	7,357	5,817	5,278						
8	50,000	4,762	8,383	6,568							
9	50,000	5,025	8,898								
10	50,000	4,824									

Our job is to predict the distribution of losses in the empty cells ($AY + Lag > 11$) and to predict the distribution of the sum of losses in the empty cells.

Let us start by considering two models for the expected loss.

Model 1 – The Cape Cod Model

$$\mathrm{E}\left[Loss_{AY,Lag}\right] = Premium_{AY} \cdot ELR_{AY} \cdot Dev_{Lag} \qquad (1)$$

The unknown parameters in this model are ELR_{AY} (AY = 1, 2, …, 10) and Dev_{Lag} (Lag = 1, 2, …, 10). The structure of the parameters is similar to the "Cape Cod" method discussed in Stanard [11] but, as we shall see, this paper's method of parameterizing the model is different.

Model 2 – The Beta Model

In the Cape Cod model, set

$$Dev_{Lag} = \beta\left(Lag/10 \mid a,b\right) - \beta\left((Lag-1)/10 \mid a,b\right) \qquad (2)$$

where $\beta\left(x \mid a,b\right)$ is the cumulative probability of a beta distribution with unknown parameters *a* and *b* as parameterized in Appendix A of Klugman, Panjer and Willmot [5].

The Beta model replaces the ten unknown Dev_{Lag} parameters in the Cape Cod model with the two unknown parameters *a* and *b*. I chose these models as representatives of a multitude of possible models that can be used in this approach. Other examples in this multitude include the models in Meyers [7], who uses a Cape Cod model with constraints on the Dev_{Lag} parameters, and Clark [3], who uses the Loglogistic and Weibull distributions to project Dev_{Lag} parameters into the future.

Let $X_{AY,Lag}$ be a random variable for the loss in the cell (AY,Lag). We describe the distribution of $X_{AY,Lag}$ by the collective risk model, which can be described by the following simulation algorithm.

Simulation Algorithm 1

1. Select a random claim count, $N_{AY,Lag}$ from a Poisson distribution with mean $\lambda_{AY,Lag}$.
2. For $i = 1, 2, \ldots, N_{AY,Lag}$, select a random claim amount, $Z_{Lag,i}$.
3. Set $X_{AY,Lag} = \sum_{i=1}^{N_{AY,Lag}} Z_{Lag,i}$, or if $N_{AY,Lag} = 0$, then $X_{AY,Lag} = 0$.

This paper assumes that the claim severity distributions of Z_{Lag} are given. In our example, we use the Pareto distribution with the cumulative distribution function:

$$F(z) = 1 - \left(\frac{\theta}{z+\theta}\right)^{\alpha}. \tag{3}$$

We set $\alpha = 2$ for all settlement lags. θ will vary by settlement lag as noted in the following table.

Table 2

Lag	1	2	3	4	5	6	7-10
θ (000)	10	25	50	75	100	125	150

Note that the average severity increases with the settlement lag, which is consistent with the common observation that larger claims tend to take longer to settle.

To summarize, we have two models (the Cape Cod and the Beta) that give E[$X_{AY,Lag}$] in terms of the unknown parameters $\{ELR_{AY}\}$ and $\{Dev_{Lag}\}$. We also assume that the claim severity distributions of Z_{Lag} are known. Then for any selected $\{ELR_{AY}\}$ and $\{Dev_{Lag}\}$, we can describe the distribution of $X_{AY,Lag}$ by the following steps.

1. Calculate $\lambda_{AY,Lag} = \dfrac{\mathrm{E}\left[X_{AY,Lag}\right]}{\mathrm{E}\left[Z_{Lag}\right]} = \dfrac{Premium_{AY} \cdot ELR_{AY} \cdot Dev_{Lag}}{\mathrm{E}\left[Z_{Lag}\right]}$.
2. Generate the distribution of $X_{AY,Lag}$ using Simulation Algorithm 1 above.

3. The Posterior Distribution of Model Parameters

Let **X** denote the data in Table 1. Let $\ell\left(\mathbf{X}\,|\left\{ELR_{AY}\right\},\left\{Dev_{Lag}\right\}\right)$ be the likelihood (or probability) of **X** given the parameters $\{ELR_{AY}\}$ and $\{Dev_{Lag}\}$. Note that defining a distribution in terms of a simulation algorithm does not lend itself to calculating the likelihood. To do this we must resort to some math that is described in detail in Appendix B. At this point, the reader should know that we are approximating the likelihood with something called the overdispersed negative binomial distribution.

The maximum likelihood estimator has been historically important and, as we shall see, will also be important in this paper. Over the past decade or so, a number of popular software packages began to include flexible function-maximizing tools that will search over a space that includes a fairly large number of parameters. Excel™ Solver is one such tool. With such a tool, the software programs[1] that accompany this paper calculate the maximum likelihood estimates for the Cape Cod and the Beta models.

The Cape Cod program calculates the maximum likelihood estimate by searching over the space of $\{ELR_{AY}\}$ and $\{Dev_{Lag}\}$, subject to a constraint that $\sum_{Lag=1}^{10} Dev_{Lag} = 1$. The Beta program feeds the results of Equation 2 into the likelihood function used in the Cape Cod program as it searches over the space of $\{ELR_{AY}\}$, *a* and *b*. Table 3 gives the maximum likelihood estimates for each model.

[1] The programs are written in R, a freely downloadable statistical package. See Meyers [6] for a review of this package.

Table 3

	Cape Cod		Beta	
	ELR	*Dev*	*ELR*	*Dev*
AY/Lag				
1	0.89090	0.16948	0.89205	0.15991
2	0.65285	0.26864	0.65670	0.27295
3	0.64448	0.23763	0.69949	0.24156
4	0.55233	0.15539	0.51727	0.16661
5	0.48569	0.07865	0.51696	0.09488
6	0.57259	0.05524	0.53697	0.04410
7	0.56411	0.01771	0.60935	0.01576
8	0.58207	0.00581	0.53487	0.00378
9	0.61922	0.00654	0.68940	0.00044
10	0.52190	0.00491	0.63902	0.00001
			a =	1.90742
			b =	5.78613

Let us now develop the framework for a Bayesian analysis. The likelihood function $\ell\left(\mathbf{X}\,|\,\{ELR_{AY}\},\{Dev_{Lag}\}\right)$ is the probability of **X,** given the parameters $\{ELR_{AY}\}$ and $\{Dev_{Lag}\}$. Using Bayes' Theorem, one can calculate the probability of the parameters $\{ELR_{AY}\}$ and $\{Dev_{Lag}\}$ given the data, **X**.

$$\Pr\left\{\{ELR_{AY}\},\{Dev_{Lag}\}\middle|\mathbf{X}\right\} \propto \ell\left(\mathbf{X}\middle|\{ELR_{AY}\},\{Dev_{Lag}\}\right)\cdot\Pr\left\{\{ELR_{AY}\},\{Dev_{Lag}\}\right\}. \quad (4)$$

A discussion of selecting the prior distribution $\Pr\left\{\{ELR_{AY}\},\{Dev_{Lag}\}\right\}$ is in order. This paper has the advantage that it is working with simulated (i.e., made up) "data" so it is editorially possible to select anything as a prior distribution. However, I would like to spend some time to illustrate one way to approach the problem of selecting the prior distribution when working with real data.

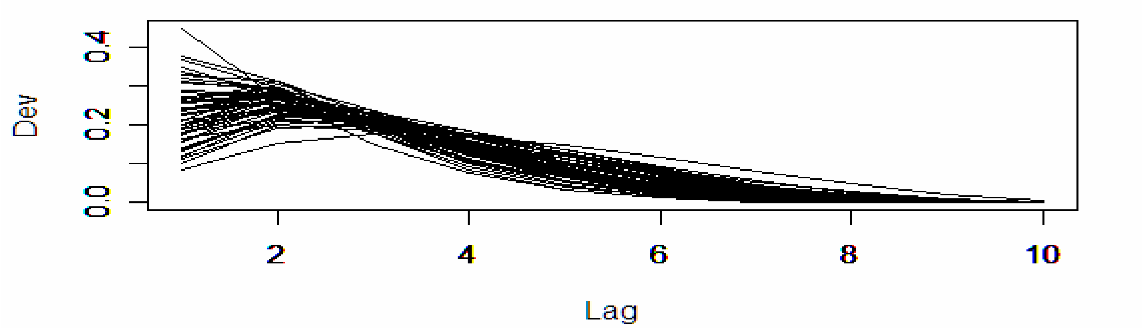
Actuaries always stress the importance of judgment in setting reserves. Actuarial consultants will stress the experience that they have gained by examining the losses of other insurers. Meyers [7] formalizes this by examining the maximum likelihood estimates of the $\{Dev_{Lag}\}$ parameters from the data of 40 large insurers. In an effort to keep the examples in this paper as realistic as possible, I looked at the same data and selected the prior distribution as follows.

Beta Model[2]: $$a \sim \Gamma(\alpha,\theta) \text{ with } \alpha = 75 \text{ and } \theta = 0.02 \tag{5}$$

$$b \sim \Gamma(\alpha,\theta) \text{ with } \alpha = 25 \text{ and } \theta = 0.20 \tag{6}$$

Figure 1 shows the Dev_{Lag} paths generated from a sample of fifty (*a*,*b*) pairs sampled from the prior distribution.

Figure 1



For the Cape Cod model, I calculated the mean and variance of the Dev_{Lag}s simulated from a large sample of (*a,b*) pairs and selected the following parameters for the gamma distribution for each Dev_{Lag}.

Table 4

$\Gamma\backslash Lag$	1	2	3	4	5	6	7	8	9	10
α	11.1010	64.6654	190.1538	34.9314	10.7284	4.4957	2.1298	1.0295	0.4574	0.1556
θ	0.0206	0.0041	0.0011	0.0040	0.0079	0.0101	0.0097	0.0073	0.0039	0.0009

[2] We will use the gamma (Γ) distribution as parameterized in Appendix A of Klugman, Panjer, and Willmot [5].

Before discussing the prior distribution of each ELR_{AY}, let us take a short side trip and look at the compound negative multinomial model[3], described by the following simulation algorithm.

Simulation Algorithm 2

1. For each accident year, select χ_{AY} at random from a gamma distribution with mean 1 and variance c.
2. For each accident year and settlement lag, select a claim count, $N_{AY,Lag}$, at random from a Poisson with mean $\chi_{AY}\cdot\lambda_{AY,Lag}$. (See the end of Section 2 for a description on how to determine the $\lambda_{AY,Lag}$s.)
3. For $i = 1, 2, \ldots, N_{AY,Lag}$, select a random claim amount, $Z_{Lag,i}$.
4. For each accident year and settlement lag, set $X_{AY,Lag} = \sum_{i=1}^{N_{AY,Lag}} Z_{Lag,i}$.

Note that for a given accident year, the $X_{AY,Lag}$s are correlated because of the common χ_{AY} that is in each *Lag*'s expected claim count.

This paper uses the compound negative multinomial model for the losses $X_{AY,Lag}$. At first glance, it might seem that this is different from the collective risk model described in Simulation Algorithm 1. But note that both the Cape Cod and the Beta models treat the ELR_{AY}s as unknown parameters. So by assigning a prior distribution to each ELR_{AY} so that its coefficient of variation squared is equal to the c in the negative multinomial model, we are explicitly modeling a random accident-year effect. With this in mind I selected each

$$ELR_{AY} \sim \Gamma(\alpha,\theta) \text{ with } \alpha = 100 \text{ and } \theta = 0.007. \tag{7}$$

Note that the expected value of each $ELR_{AY} = \alpha\cdot\theta = 0.70$ and the coefficient of variation of each $ELR_{AY} = \sqrt{1/\alpha} = 0.1$.

As we observe data points $x_{AY,Lag}$ in **X**, we gain information about the χ_{AY} in each accident year. As we shall see, treating each ELR_{AY} as an unknown parameter allows us to use this information in predicting the outcomes of future lags.

This paper uses the Gibbs sampler to generate random samples of the $\{ELR_{AY}\}$ and $\{Dev_{Lag}\}$ parameters that represent the posterior distribution. Scollnik [10] introduced the Gibbs sampler to the CAS literature. Verrall [12] gives an application of it to a loss reserving problem.

For the Cape Cod model, this paper implements the Gibbs sampler as follows.

[3] The compound negative multinomial distribution was introduced to the CAS literature by Mildenhall [9].

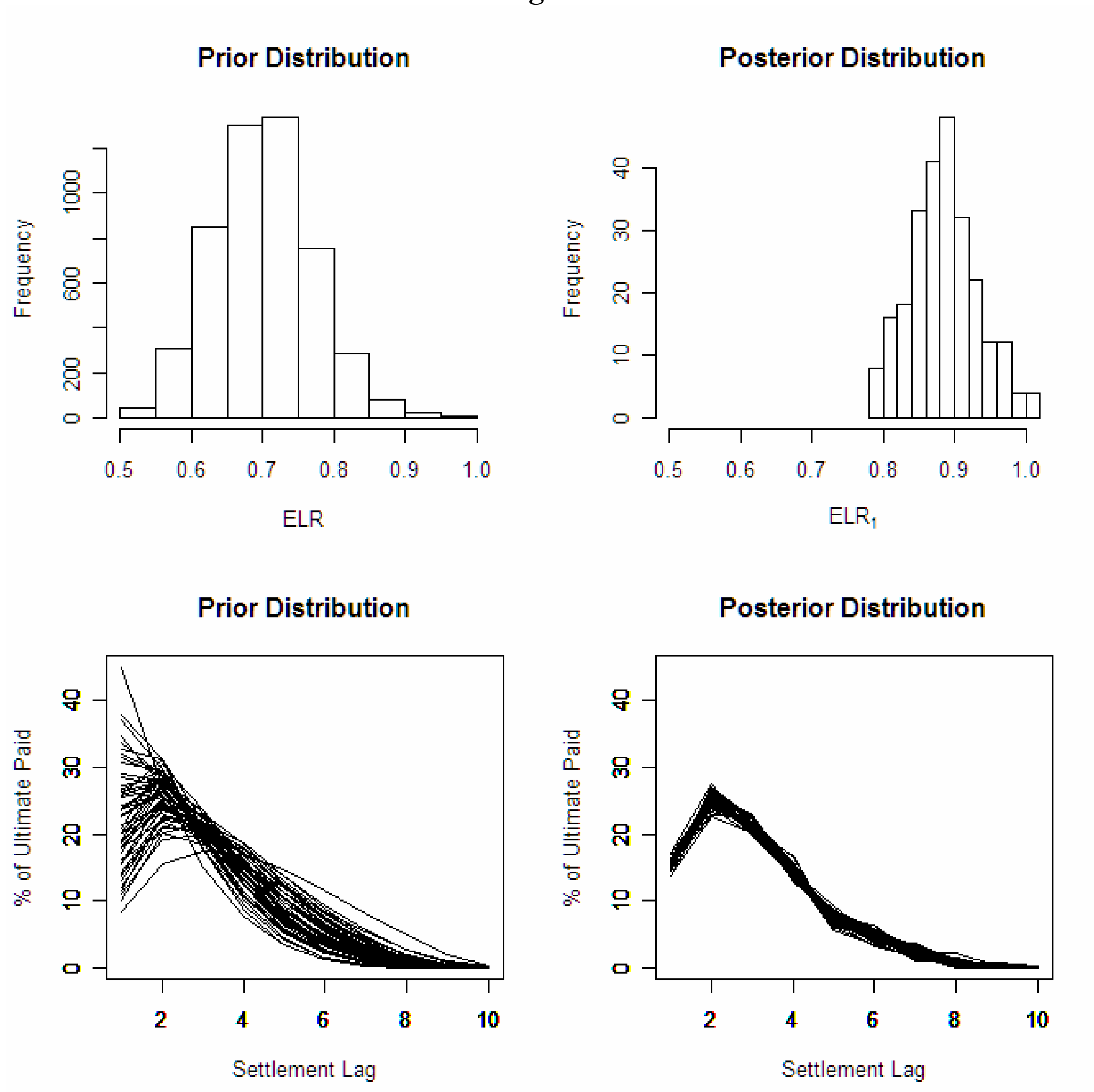
Simulation Algorithm 3

1. Given the data triangle **X**, calculate the maximum likelihood estimates $Dev_{1,Lag}$ for Lag = 1, …,10 and $ELR_{1,AY}$ for AY = 1,…,10. Keep the maximum likelihood, ML, for future reference. Set i = 1.
2. Replace i by i+1, set $Dev_{i,Lag} = Dev_{i\text{-}1,Lag}$ and set $ELR_{i,AY} = ELR_{i\text{-}1,AY}$.
3. For Lag = 1 to 10:
   a. Replace $Dev_{i,Lag}$ with a random number taken from the prior distribution of Dev_{Lag} and calculate its likelihood L.
   b. Select a random number, u, from a uniform (0,1) distribution.
   c. If $L/ML < u$, then return to Step 3a, otherwise continue to the next step.
4. For AY = 1 to 10:
   a. Replace $ELR_{i,AY}$ with a random number taken from the prior distribution of ELR_{AY} and calculate its likelihood L.
   b. Select a random number, u, from a uniform (0,1) distribution.
   c. If $L/ML < u$, then return to Step 4a, otherwise continue to the next step.
5. Return to Step 2 until i is greater than a selected n.

The intuition behind this algorithm is that a parameter "applies" to be included in the Gibbs sample in proportion to its prior probability. Each applicant is "accepted" into the sample in proportion to its likelihood. So the probability of a parameter being included in the sample is the product of the probability of applying times its likelihood, which in turn is equal to its posterior probability. See Equation 4 above.

Figure 2 provides a graphic comparison between the prior distribution and the posterior distribution, as represented by the output of Simulation Algorithm 3. The upper histograms are random samples of ELR_1, taken from its prior and posterior distributions. The lower graphs represent the paths taken from the prior and posterior $\{Dev_{Lag}\}$ distributions.

Figure 2



For the Beta model, we implement the Gibbs sampler as follows.

Simulation Algorithm 4

1. Given the data triangle **X**, calculate the maximum likelihood estimates a_1, b_1 and $ELR_{1,AY}$ for $AY = 1,\ldots,10$. Keep the maximum likelihood, ML, for future reference. Set $i = 1$.
2. Replace i by $i+1$, set $a_i = a_{i-1}$, set $b_i = b_{i-1}$ and set $ELR_{i,AY} = ELR_{i-1,AY}$.
3. For $p_i = a_i$ and then b_i:
   a. Replace p_i with a random number taken from the prior distribution of p, calculate the associated $Dev_{i,Lag}$s using Equation 2 and calculate the likelihood L.
   b. Select a random number, u, from a uniform (0,1) distribution.
   c. If $L/ML < u$, then return to Step 3a, otherwise continue to the next step.
4. For $AY = 1$ to 10:
   d. Replace $ELR_{i,AY}$ with a random number taken from the prior distribution of ELR_{AY} and calculate the likelihood L.
   e. Select a random number, u, from a uniform (0,1) distribution.
   f. If $L/ML < u$, then return to Step 4a, otherwise continue to the next step.
5. Return to Step 2 until i is greater than a selected n.

Each iteration is a step in a Markov chain of random transformations in the parameter space $\{ELR_{AY}\}$ and $\{Dev_{Lag}\}$. It is well know that Markov chains will converge to a limiting distribution and that, when executed as described in these simulation algorithms, the limiting distribution will be the posterior distribution.

The random parameters generated by the first several iterations of the Gibbs sampler may not be distributed as the limiting distribution. So it is a general practice to discard parameters that are generated early in the process. By examining successive blocks of parameters in the examples in this paper, I concluded that using parameters generated after 250 iterations[4] of Simulation Algorithms 3 and 4 was sufficiently accurate for our purposes. Table 5 shows some illustrative results that came out of Simulation Algorithm 4 being applied to the data in Table 1.

Table 5

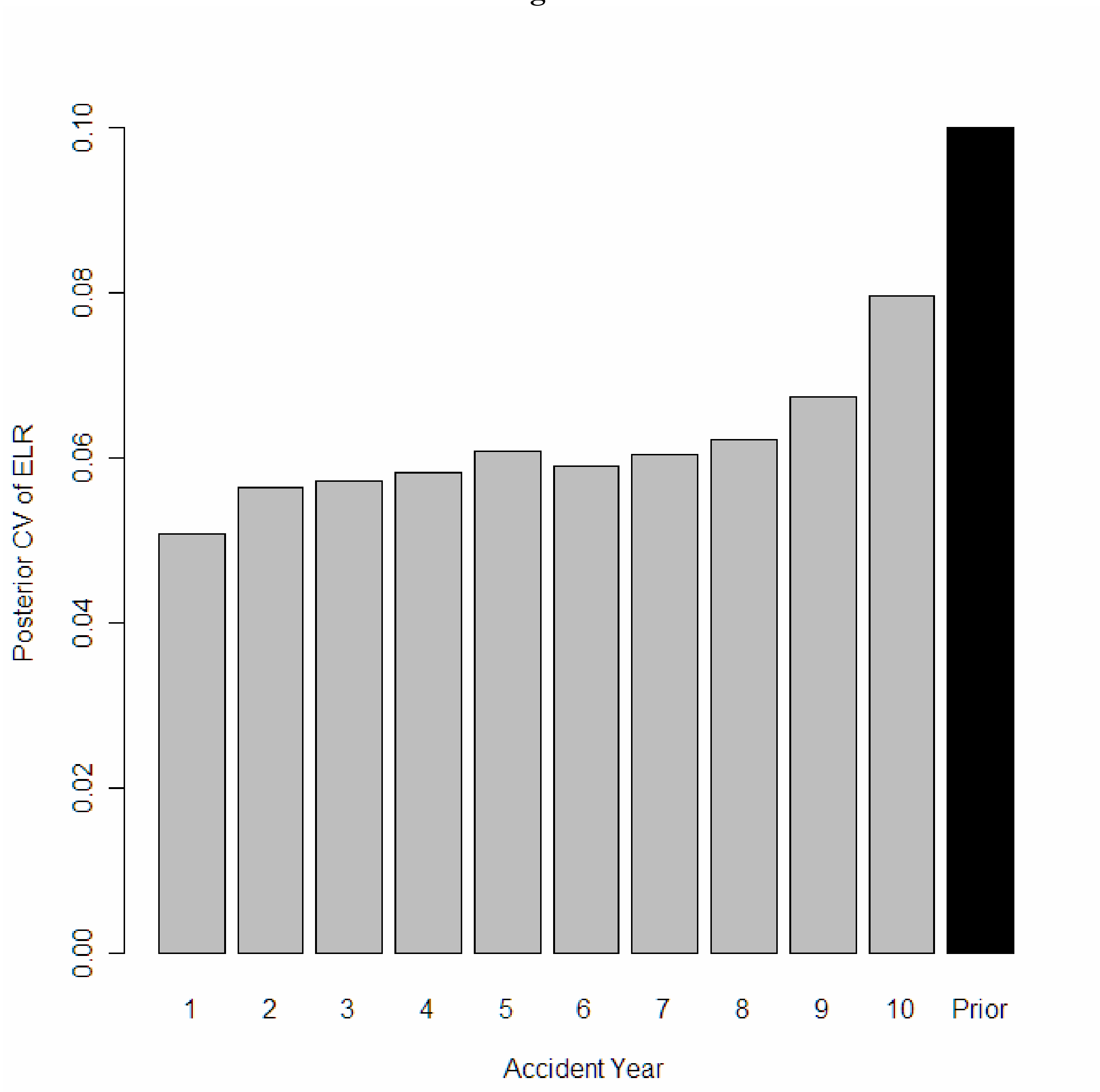
[4] One may find other sources that recommend thousands of iterations. But these sources generally count one draw of a parameter from its prior distribution as one iteration. When counting that way, 250 iterations of Simulation Algorithms 3 and 4 represent 5,000 and 3,000 iterations respectively.

Iteration	ELR_1	ELR_2	ELR_3	ELR_4	ELR_5	ELR_6	ELR_7	ELR_8	ELR_9	ELR_{10}
251	0.75403	0.69942	0.62441	0.56447	0.51833	0.60362	0.61284	0.60487	0.66776	0.63434
252	0.84815	0.73598	0.64300	0.59662	0.50991	0.66961	0.63046	0.65861	0.78867	0.62436
253	0.82959	0.65535	0.62372	0.58285	0.54446	0.65939	0.62067	0.66337	0.77458	0.68514
254	0.82214	0.67833	0.72494	0.58108	0.59692	0.64186	0.64096	0.69660	0.63273	0.71884
255	0.85885	0.70338	0.65320	0.60643	0.57145	0.65768	0.74067	0.64207	0.61538	0.56273
256	0.82655	0.68402	0.71207	0.57685	0.50899	0.62291	0.68097	0.60459	0.73825	0.62867
257	0.86339	0.71486	0.62554	0.55949	0.54898	0.57494	0.63603	0.66952	0.68241	0.61616
258	0.81831	0.64761	0.73752	0.61186	0.63983	0.62646	0.61374	0.67133	0.64861	0.62245
259	0.80801	0.66089	0.70570	0.61823	0.57213	0.62688	0.58704	0.69212	0.62392	0.67231
260	0.81955	0.65917	0.61623	0.64292	0.56440	0.61969	0.61458	0.67270	0.74439	0.59132

Iteration	Dev_1	Dev_2	Dev_3	Dev_4	Dev_5	Dev_6	Dev_7	Dev_8	Dev_9	Dev_{10}
251	0.17353	0.26609	0.23075	0.16171	0.09592	0.04754	0.01863	0.00509	0.00072	0.00002
252	0.17373	0.26219	0.22815	0.16179	0.09773	0.04965	0.02012	0.00576	0.00087	0.00003
253	0.15662	0.25141	0.22857	0.16863	0.10601	0.05625	0.02396	0.00730	0.00119	0.00004
254	0.15514	0.24770	0.22656	0.16906	0.10796	0.05847	0.02559	0.00808	0.00139	0.00005
255	0.16275	0.25121	0.22557	0.16608	0.10487	0.05622	0.02435	0.00760	0.00130	0.00005
256	0.16274	0.24870	0.22378	0.16595	0.10596	0.05768	0.02550	0.00819	0.00145	0.00006
257	0.16549	0.25142	0.22449	0.16497	0.10422	0.05600	0.02436	0.00766	0.00132	0.00005
258	0.15983	0.24720	0.22401	0.16705	0.10721	0.05865	0.02607	0.00842	0.00151	0.00006
259	0.17049	0.25879	0.22734	0.16312	0.09993	0.05165	0.02138	0.00629	0.00099	0.00003
260	0.16584	0.26100	0.23092	0.16494	0.09979	0.05056	0.02034	0.00574	0.00085	0.00003

One can often find interesting information about the uncertainty in the parameter estimates by examining tables of parameters generated by the Gibbs sampler. Figure 3 below show the coefficients of variation (CV) of the loss ratio estimates taken from 2,500 additional iterations of the sample in Table 5. This illustrates how we gain information about the ultimate loss ratio as we get more data from each accident year. Wacek [13] gives another approach to estimating loss ratios as we gain information over time.

Figure 3



4. The Predictive Distribution of Outcomes

Now that we have the posterior distribution estimated from the data of Table 1, we now turn to the problem of predicting future outcomes, $X_{AY,Lag}$, when AY + Lag > 11.

Table 1 (000) (Repeated)

AY	Premium	*Lag* 1	*Lag* 2	*Lag* 3	*Lag* 4	*Lag* 5	*Lag* 6	*Lag* 7	*Lag* 8	*Lag* 9	*Lag* 10
1	50,000	7,168	11,190	12,432	7,856	3,502	1,286	334	216	190	0
2	50,000	4,770	8,726	9,150	5,728	2,459	2,864	715	219	0	$X_{2,10}$
3	50,000	5,821	9,467	7,741	3,736	1,402	972	720	50	$X_{3,9}$	$X_{3.10}$
4	50,000	5,228	7,050	6,577	2,890	1,600	2,156	592	$X_{4,8}$	$X_{4,9}$	$X_{4,10}$
5	50,000	4,185	6,573	5,196	2,869	3,609	1,283	$X_{5,7}$	$X_{5,8}$	$X_{5,9}$	$X_{5,10}$
6	50,000	4,930	8,034	5,315	5,549	1,891	$X_{6,6}$	$X_{6,7}$	$X_{6,8}$	$X_{6,9}$	$X_{6,10}$
7	50,000	4,936	7,357	5,817	5,278	$X_{7,5}$	$X_{7,6}$	$X_{7,7}$	$X_{7,8}$	$X_{7,9}$	$X_{7,10}$
8	50,000	4,762	8,383	6,568	$X_{8,4}$	$X_{8,5}$	$X_{8,6}$	$X_{8,7}$	$X_{8,8}$	$X_{8,9}$	$X_{8,10}$
9	50,000	5,025	8,898	$X_{9,3}$	$X_{9,4}$	$X_{9,5}$	$X_{9,6}$	$X_{9,7}$	$X_{9,8}$	$X_{9,9}$	$X_{9,10}$
10	50,000	4,824	$X_{10,2}$	$X_{10,3}$	$X_{10,4}$	$X_{10,5}$	$X_{10,6}$	$X_{10,7}$	$X_{10,8}$	$X_{10,9}$	$X_{10,10}$

While there are many statistics of interest that one could examine, I chose to examine the predictive distribution of the total reserve:

$$R = \sum_{AY=2}^{10} \sum_{Lag=12-AY}^{10} X_{AY,Lag} . \tag{8}$$

Suppose we have a set of parameters $\{ELR_{AY}\}$ and $\{Dev_{Lag}\}$ calculated from several iterations of the Gibbs sampler. Conceptually, the easiest way to calculate the distribution of outcomes is by repeated use of the following simulation algorithm.

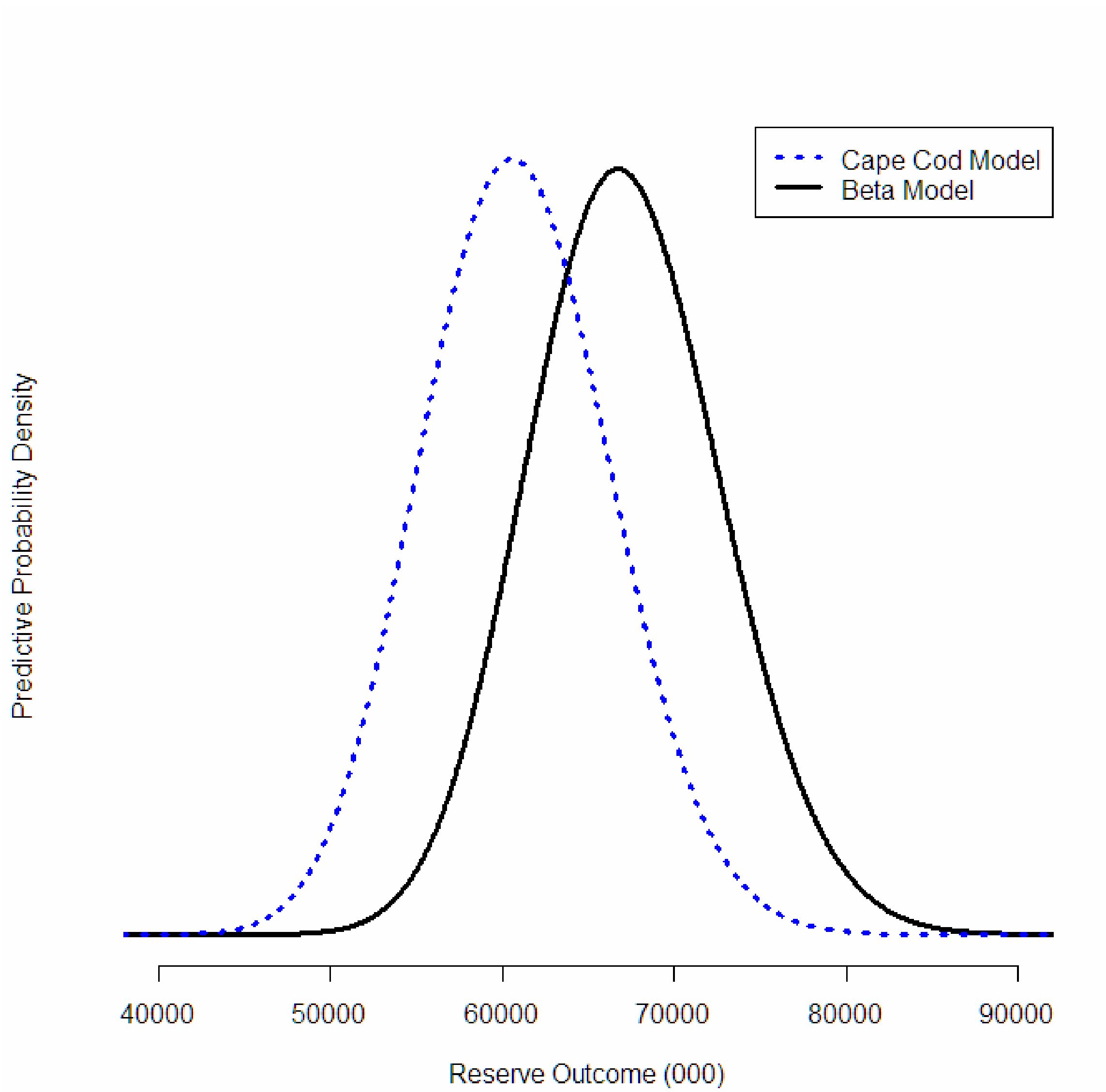
Simulation Algorithm 5

1. Select the parameters $\{ELR_{AY}\}$ and $\{Dev_{Lag}\}$ from a randomly selected iteration.
2. For $AY = 2, \ldots, 10$, do:
    a. For $Lag = 12 - AY$ to 10, do:
        i. Set $\lambda_{AY,Lag} = \dfrac{Premium_{AY} \cdot ELR_{AY} \cdot Dev_{Lag}}{\mathrm{E}\left[Z_{Lag}\right]}$
        ii. Select N at random from a Poisson distribution with mean $\lambda_{AY,Lag}$.
        iii. If $N > 0$, for $i = 1, \ldots, N$ select claim amounts, $Z_{i,Lag}$, at random from the claim severity distribution for the Lag.
        iv. If $N > 0$, set $X_{AY,Lag} = \sum_{i=1}^{N} Z_{i,Lag}$, otherwise set $X_{AY,Lag} = 0$.
3. Set $R = \sum_{AY=2}^{10} \sum_{Lag=12-AY}^{10} X_{AY,Lag}$.

I expect that many actuaries will be satisfied with using this simulation algorithm to calculate the predictive distribution. However, this paper uses a Fast Fourier Transform (FFT) to calculate the predictive distribution. While it is very technical and harder to implement, it is faster and it produces more accurate results (relative to the model assumptions). Appendix A describes how to implement the FFT for this paper's application.

Figure 4 plots the density functions for the predictive distributions derived from the data in Table 1. For each model, I ran 500 iterations of the Gibbs sampler and discarded the first 250 because they are less likely to represent the posterior distributions.

Figure 4



The predictive means and standard deviations are:

- 60,871,000 and 5,487,000 for the Cape Cod model; and
- 67,183,000 and 5,605,000 for the Beta model.

The difference in the predictive means for the two models is 5,982,000, illustrating the fact that we do face "model risk." If one wants to reflect model risk, one could modify Simulation Algorithm

5 by randomly selecting parameters from the $\{ELR_{AY}\}$ and $\{Dev_{Lag}\}$ lists provided by the Gibbs samples for each model.

5. Risk Margins in Loss Reserves

Now that we have demonstrated a method that quantifies the uncertainty in the estimates of future loss payments, we now turn to exploring how this information might be used to post a loss reserve on a financial statement. The art of accounting has always had difficulty in dealing with uncertainty. A common practice, when possible, is to value a liability at its market value. When there is no active market, as is typically the case for loss reserves, the fallback position is to use a model to calculate the cost that an insurance market would "theoretically" charge to transfer the risky reserve.

As this paper is being written, there is still active debate on whether and how to do this. Meyers [8] provides some background and references on this subject. This section only addresses the "how."

I should add that in preparing this section I immeasurably benefited from the discussions that led to the paper jointly written by Kaufman, Broughton, Buchanan, and Meyers [4]. That paper discusses a variety of methods to calculate risk margins for loss reserves, whereas this paper illustrates only one of those methods.

The formula discussed here is called the Capital Cash Flow (CCF) risk margin. In words, this formula assumes that investors in a reinsurer would need to put up (or allocate) capital to take on the loss reserve risk by a ceding insurer. As claims are settled, the reinsurer expects to be able to release the capital over time. The CCF risk margin is the profit that the reinsurer would need to be persuaded to take on this risky venture.

We will now discuss the details. Let:

- i = Risk-free rate of return on investments.
- r = Total rate of return demanded by the reinsurer for taking additional insurance risk.
- C_t = Amount of capital required to (or allocated to) support an insurance portfolio at time t.

First look at the cash flow of the insurance transaction.

- At the beginning of the first year, at time $t = 0$, investors contribute a sum of C_0 to the reinsurer, which earns a risk-free rate of return, i, over the next year.
- At time $t = 0$, the reinsurer collects M_{CCF} from the ceding insurer and immediately transfers it to its investors. Equivalently, one could say that the investor contributes $C_0 - M_{CCF}$ to the reinsurer.
- At time $t = 1$, the investors expect to keep C_1 invested in the reinsurer, and they expect to receive a cash flow $C_0(1+i) - C_1$ at the end of year 1. Since the loss the reinsurer is required to pay and C_1 are uncertain, they discount the value of the amount returned at the risky rate of return $r > i$.
- Continuing on to time t, the investors expect to keep C_t invested in the reinsurer, and they expect a cash flow of $C_{t-1}(1+i) - C_t$ at the end of year t.

Since the cash flows are uncertain, it is appropriate to discount the cash flow at the risky rate of return, r. This leads to the following expression.

$$C_0 = M_{CCF} + \sum_{t=1}^{\infty} \frac{C_{t-1}(1+i) - C_t}{(1+r)^t}. \tag{9}$$

This equation implies

$$\begin{aligned} M_{CCF} &= C_0 - \sum_{t=1}^{\infty} \frac{C_{t-1}(1+i) - C_t}{(1+r)^t} \\ &= \frac{C_0(1+r-1-i)}{1+r} + \frac{C_1(1+r-1-i)}{(1+r)^2} + \frac{C_2(1+r-1-i)}{(1+r)^3} + \ldots \\ &= (r-i)\sum_{t=0}^{\infty} \frac{C_t}{(1+r)^{t+1}}. \end{aligned} \tag{10}$$

The following table shows how to calculate C_t for the example in this paper fit with the Beta model.

Table 6 (000)

(1) t	(2) L_t^{Nom}	(3) ΔL_t^{Nom}	(4) L_t^{Disc}	(5) TVaR_t^{Nom}	(6) $\Delta\text{TVaR}_t^{Nom}$	(7) TVaR_t^{Disc}	(8) C_t
0	67,183	27,103	61,224	80,617	28,086	72,373	11,149
1	40,080	18,847	36,993	52,531	21,984	47,799	10,805
2	21,233	11,391	19,809	30,547	14,167	28,033	8,224
3	9,843	5,978	9,270	16,380	8,224	15,129	5,859
4	3,864	2,653	3,671	8,156	4,315	7,570	3,899
5	1,211	940	1,160	3,841	2,075	3,581	2,422
6	271	237	261	1,766	856	1,659	1,398
7	34	33	33	909	803	877	845
8	1	1	1	106	106	103	102

(1) The time, *t*, after the liability is set.

(2) The nominal expected value of future payments, $L_t^{Nom} = \sum_{AY=2+t}^{10} \sum_{Lag=AY}^{10} \mathrm{E}\left[X_{AY,Lag}\right]$.

(3) $\Delta L_t^{Nom} = L_t^{Nom} - L_{t+1}^{Nom}$.

(4) The discounted liability, $L_t^{Disc} = \sum_{k=t}^{8} \frac{\Delta L_k^{Nom}}{(1+i)^{k-t+0.5}}$, where i = 6%.

(5) The nominal Tail-Value-at-Risk, i.e., the conditional expected value of the nominal random losses, $\sum_{AY=2+t}^{10} \sum_{Lag=AY}^{10} X_{AY,Lag}$, given that they exceed their 99th percentile. The density functions for the nominal losses are plotted on Figure 5 for each *t*.

(6) $\Delta\text{TVaR}_t^{Nom} = \text{TVaR}_t^{Nom} - \text{TVaR}_{t+1}^{Nom}$.

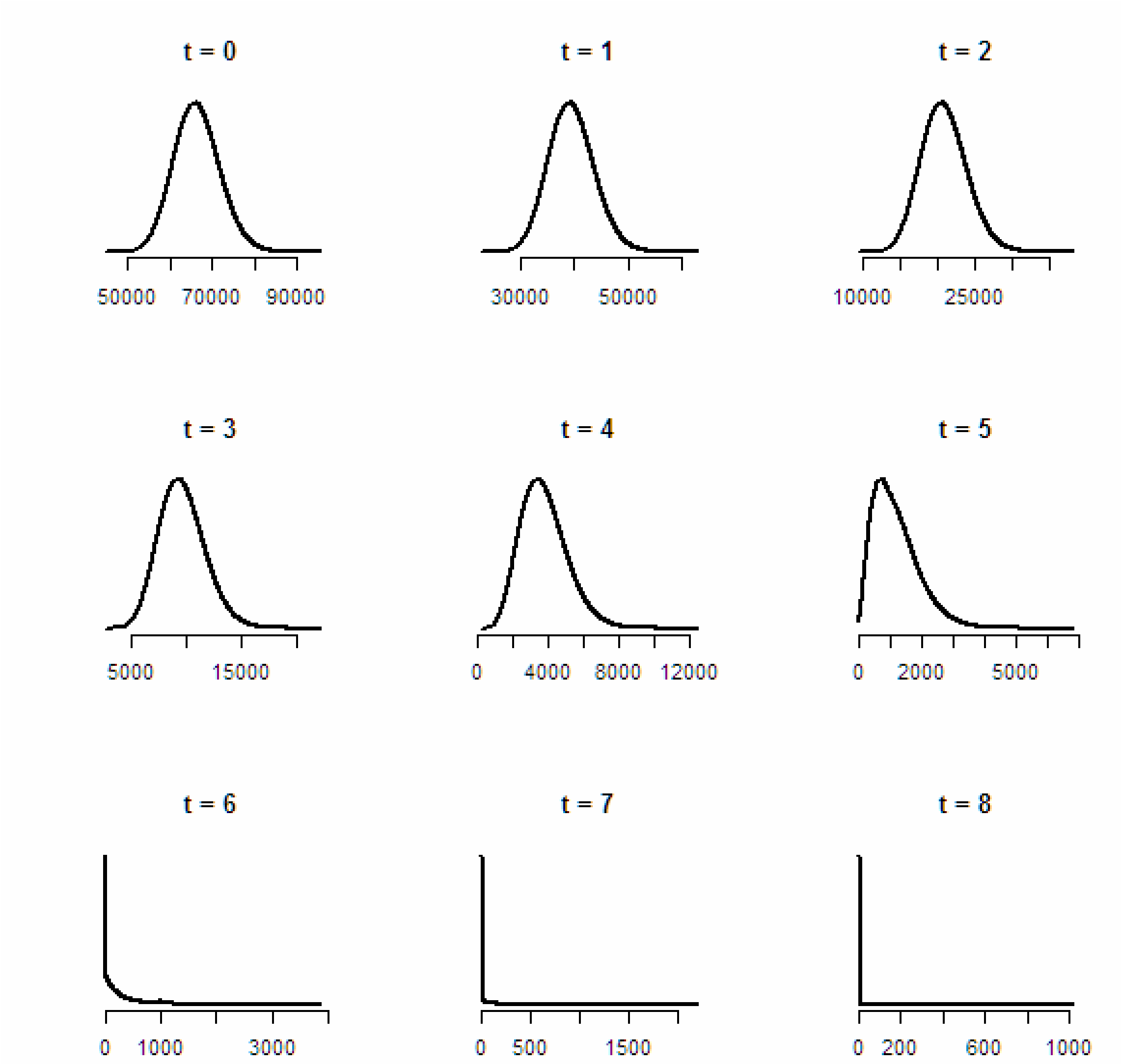
(7) The discounted $\text{TVaR}_t^{Disc} = \sum_{k=t}^{8} \frac{\Delta\text{TVaR}_k^{Nom}}{(1+i)^{k-t+0.5}}$.

(8) The needed capital at time *t* is expected to be $C_t = \text{TVaR}_t^{Disc} - L_t^{Disc}$.

Now that we have the C_ts, we can then use Equation 10, with r = 10%, to calculate M_{CFF} = 1,368,000, which is 2.2% of the discounted liability, 61,224,000.

Figure 5 (000)

Density Functions for the Nominal Losses as They Run Off



- In the latter stages of the runoff, there are a small number of potentially large claims (limited to 1,000,000) that occasionally are paid. Thus, you see the spikes at zero. The density function was plotted for those loss amounts for which the cumulative distribution function was less than 0.999999.

The risk margin calculation above was based on the nominal TVaR for the insurer's own losses. This is tantamount to assuming that the reinsurer has no other business to diversify the losses. If the liability is ever transferred, it will almost surely be transferred to a sizeable reinsurer with a diverse portfolio of losses. Rather than specify the characteristics of the reinsurer, a good approximation to the reinsurer's cost of capital would be to base the calculation of the distribution of the insurer's uncertainty in the expected values, as generated by the $\{ELR\}$ and the $\{Dev\}$ parameters in the Gibbs sampler. Table 7 calculates the risk margin under this assumption.

Table 7 (000)

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
t	L_t^{Nom}	ΔL_t^{Nom}	L_t^{Disc}	TVaR_t^{Nom}	$\Delta\text{TVaR}_t^{Nom}$	TVaR_t^{Disc}	C_t
0	67,183	27,103	61,224	76,583	29,581	69,488	8,264
1	40,080	18,847	36,993	47,002	21,079	43,202	6,208
2	21,233	11,391	19,809	25,923	13,294	24,092	4,283
3	9,843	5,978	9,270	12,629	7,270	11,850	2,580
4	3,864	2,653	3,671	5,359	3,514	5,076	1,405
5	1,211	940	1,160	1,845	1,381	1,763	603
6	271	237	261	464	397	447	186
7	34	33	33	67	65	65	33
8	1	1	1	3	3	3	2

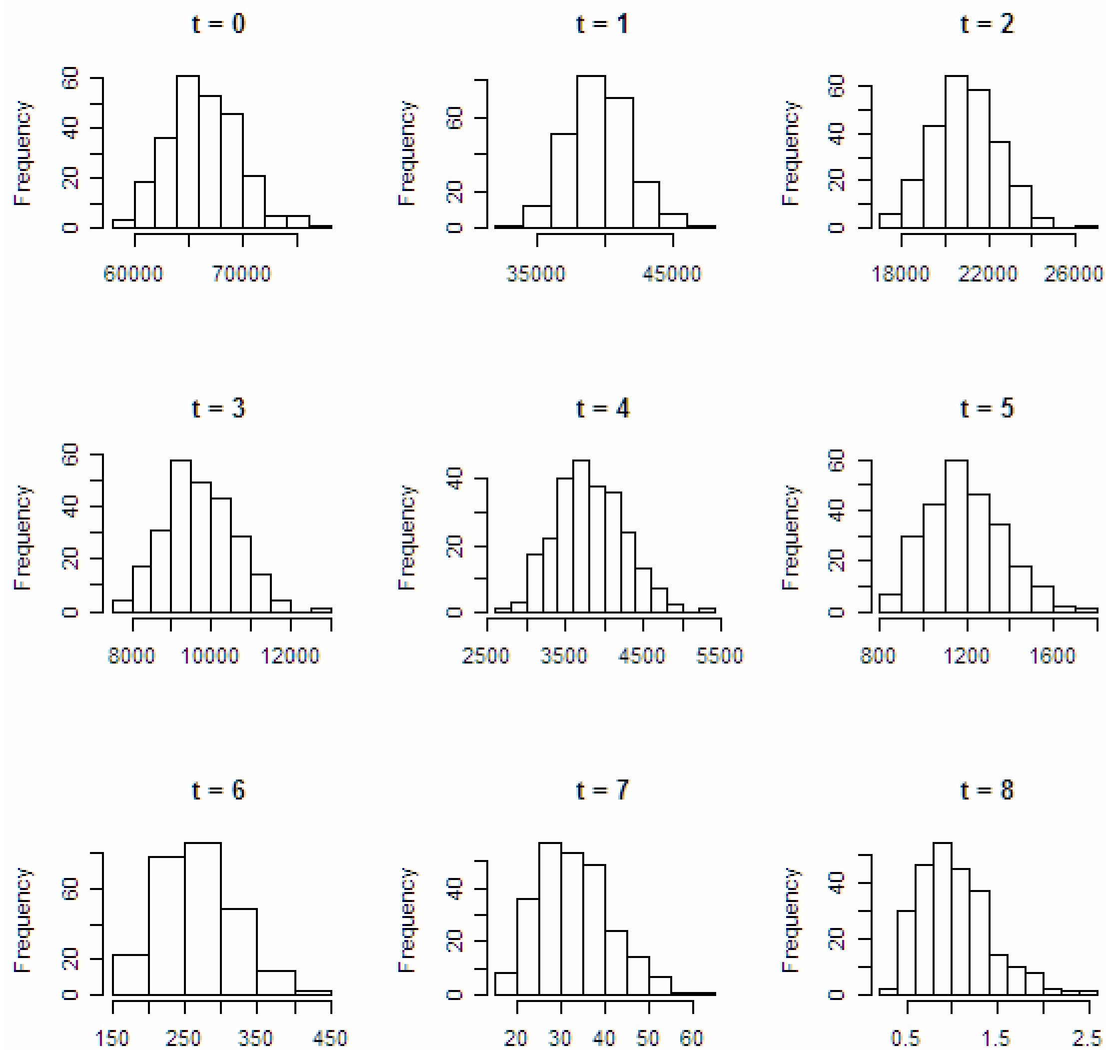
The explanation of the columns is the same as for Table 6 except for Column 5.

(5) The nominal Tail-Value-at-Risk at the 99% level, where the random element is the expected value of the Gibbs sample, $\sum_{AY=2+t}^{10} \sum_{Lag=AY}^{10} Premium_{AY} \cdot ELR_{AY} \cdot Dev_{Lag}$. The histograms of the sums calculated from the Gibbs sample are plotted on Figure 6 for each t.

Now that we have the C_ts, we can then use Equation 10, with r = 10%, to calculate M_{CFF} = 758,000 which is 1.2% of the discounted liability, 61,224,000.

Figure 6

Histograms of the Expected Runoff Scenarios Taken from the Gibbs Sample



With the exception of workers compensation insurance, it is standard statutory accounting practice in the USA to post loss reserves at nominal, not discounted values. A common justification for this practice is that it provides a cushion for the risk in the posted reserve. In the above examples, the difference between the nominal and discounted expected values of the liability is 67,183,000 – 61,224,000 = 5,959,000. This difference is noticeably larger than the 1,368,000 and 758,000 risk margins calculated in the examples above.

Note that the CCF risk margin is sensitive to three factors that many consider when accessing risk:

1. The volatility of the future payouts as quantified by C_t. If desired, one can consider only parameter risk.
2. How long the insurer is exposed to the risk, as quantified by how C_t decreases over time.
3. The premium the market places on risk, as quantified by $r - i$.

Note that proposals for risk margins based solely on statistics taken from a predictive distribution, such as percentiles, do not address (2) and (3) above. The American practice of posting reserves at their nominal value does not address (1) and (3) above.

References

1. Barnett, Glen and Benjamin Zehnwirth, "Best Estimates for Reserves," *Proceedings of the Casualty Actuarial Society* LXXXVII, 2000, pp.245-321. http://www.casact.org/pubs/proceed/proceed00/00245.pdf

2. Bornhuetter, Ronald L. and Ronald E. Ferguson, "The Actuary and IBNR," *Proceedings of the Casualty Actuarial Society* LIX, 1972, pp.181-195. http://www.casact.org/pubs/proceed/proceed72/72181.pdf

3. Clark, David R., "LDF Curve Fitting and Stochastic Loss Reserving: A Maximum Likelihood Approach," *Casualty Actuarial Society Forum*, Fall 2003, pp. 41-92 http://www.casact.org/pubs/forum/03fforum/03ff041.pdf

4. Kaufman, Allan, Jonathan Broughton, Bob Buchanan, and Glenn Meyers, "Cost of Capital Method of Determining Risk Margins—Dealing with Neglected Issues: Observations from Three Empirical Analyses," 2007 GIRO Convention Proceedings.

5. Klugman, Stuart A., Harry A. Panjer, and Gordon E. Willmot, *Loss Models: From Data to Decisions*, Second Edition, John Wiley and Sons, 2004.

6. Meyers, Glenn G., "The R Programming Language," *Actuarial Review*, vol. 33, no. 4, 2006. http://www.casact.org/newsletter/index.cfm?fa=viewart&id=5311

7. Meyers, Glenn G., "Estimating Predictive Distributions for Loss Reserve Models," *Variance* Fall 2007, 1(2):248-272, http://www.variancejournal.org/issues/01-02/248.pdf.

8. Meyers, Glenn G., "Should Reserves Include Risk Margins?—International Developments," *Actuarial Review*, vol. 34, no. 4, 2007b. http://www.casact.org/newsletter/index.cfm?fa=viewart&id=5499

9. Mildenhall, Stephen J., "Correlations and Aggregate Loss Distributions With an Emphasis on the Iman-Conover Method." *Casualty Actuarial Society Forum,* Winter 2006, pp. 103-204. http://www.casact.org/pubs/forum/06wforum/06w107.pdf

10. Scollnik, David P. M., "Introduction to Markov Chain Monte Carlo Methods and Their Actuarial Applications," *Proceedings of the Casualty Actuarial Society* LXXXIII, 1996, pp. 114-165 http://www.casact.org/pubs/proceed/proceed96/96114.pdf

11. Stanard, James N., "A Simulation Test of Prediction Errors of Loss Reserve Estimation Techniques," *Proceedings of the Casualty Actuarial Society* LXXII, 1985, pp.124-148. http://www.casact.org/pubs/proceed/proceed85/85124.pdf

12. Verrall, Richard J., "Obtaining Predictive Distributions for Reserves Which Incorporate Expert Opinion," *Variance,* Spring 2007, 1(1):53-80. http://www.variancejournal.org/issues/01-01/053.pdf

13. Wacek, Michael G., "The Path of the Ultimate Loss Ratio Estimate," *Variance,* Fall 2007, 1(2):173-192. http://www.variancejournal.org/issues/01-02/173.pdf

Appendix A. Collective Risk Model Mathematics with Fast Fourier Transforms

This paper describes the collective risk model in terms of a simulation algorithm. Given the speed of today's personal computers, it is practical to actually do the simulations in a reasonable amount of time. This appendix describes how to do many of the calculations to a higher degree of accuracy in a significantly shorter time using FFTs.

The advantage to using FFTs is that the time-consuming task of calculating the distribution of the sum of random variables is transformed into the much faster task of multiplying the FFTs of the distributions. Simulation Algorithms 1, 2, and 5 show that the collective risk model requires the calculation of the distribution of the sum of random claim amounts. Furthermore, Simulation Algorithm 5 requires the calculation of the distribution of the sum of losses over different accident years and settlement lags.

This appendix has three sections. Since the FFTs work on discrete random variables, the first section shows how to discretize the claim severity distribution in such a way that the limited average severities of the continuous severity distribution are preserved. The second section will show how to calculate the probabilities associated with the collective risk model. The third section will show how to calculate the predictive distribution for the outstanding losses.

A.1 Discretizing the Claim Severity Distributions

The first step is to determine the discretization interval length h. Variable h, which depended on the size of the insurer, was chosen so the 2^{14} (16,384) values spanned the probable range of annual losses for the insurer. Specifically, let h_1 be the sum of the insurer's ten-year premium divided by 2^{14}. The h was set equal to 1,000 times the smallest number from the set
$\{5, 10, 20, 25, 40, 50, 100, 125, 200, 250, 500, 1000\}$ that was greater than $h_1/1000$. This last step guarantees that a multiple, m, of h would be equal to the policy limit of 1,000,000.

The next step is to use the mean-preserving method (described in Klugman [5, p. 656] to discretize the claim severity distribution for each settlement lag. Let $p_{i,Lag}$ represent the probability of a claim with severity $h \cdot i$ for each settlement lag. Using the limited average severity (LAS_{Lag}) function determined from claim severity distributions, the method proceeds in the following steps.

1. $p_{0,Lag} = 1 - \text{LAS}_{Lag}(h)/h$.
2. $p_{i,Lag} = (2 \cdot \text{LAS}_{Lag}(h \cdot i) - \text{LAS}_{Lag}(h \cdot (i-1)) - \text{LAS}_{Lag}(h \cdot (i+1)))/h$ for i = 1, 2, ..., m-1.

3. $p_{m,Lag} = 1 - \sum_{i=0}^{m-1} p_{i,Lag}$.
4. $p_{ik} = 0$ for $i = m + 1, ..., 2^{14} - 1$.

A.2 Calculating Probabilities for the Compound Poisson Distribution

The purpose of this section is to show how to calculate the probabilities of losses defined by the collective risk model as defined in Simulation Algorithm 1. The math described in this section is derived in Klugman [5, Section 6.91]. The calculation proceeds in the following steps.

1. Set $\vec{\mathbf{p}}_{Lag} = \left\{ p_{0,Lag}, \ldots p_{2^{14}-1,Lag} \right\}$.
2. Calculate the expected claim count, $\lambda_{AY,Lag}$, for each accident year and settlement lag using Equation 2, $\lambda_{AY,Lag} \equiv \mathrm{E}\left[Paid\ Loss_{AY,Lag} \right] / \mathrm{E}\left[Z_{Lag} \right]$.
3. Calculate the Fast Fourier Transform (FFT) of $\vec{\mathbf{p}}_{Lag}$, $\Phi\left(\vec{\mathbf{p}}_{Lag}\right)$.
4. Calculate the FFT of each aggregate loss random variable, $X_{AY,Lag}$, using the formula

$$\Phi\left(\vec{\mathbf{q}}_{AY,Lag}\right) = e^{\left(\Phi\left(\vec{\mathbf{p}}_{Lag}\right)-1\right)}.$$

   This formula is derived in Klugman[5, Section 6.91].

5. Calculate $\vec{\mathbf{q}}_{AY,Lag} = \Phi^{-1}\left(\Phi\left(\vec{\mathbf{q}}_{AY,Lag}\right)\right)$, the inverse FFT of the expression in Step 4 above.

The vector ,$\vec{\mathbf{q}}_{AY,Lag}$, contains the probabilities of the discretized compound Poisson distribution defined by Simulation Algorithm 1.

A.3 Calculating Probabilities for the Predictive Distribution

To calculate the predictive distribution of the reserve outcomes by the methods in this paper, one needs the $\{ELR_{AY}, Dev_{Lag}\}$ parameter set that was simulated by the Gibbs sampler as described in Section 3 above.

1. For each parameter set, denoted by i, and $AY+Lag > 11$, do the following.

   a. Calculate the expected loss, $Premium_{AY,i} \cdot ELR_{AY,i} \cdot Dev_{Lag,i}$.

   b. Calculate the FFT of the aggregate loss $X_{AY,Lag,i}$, $\Phi\left(\vec{\mathbf{q}}_{AY,Lag,i}\right)$ as described in Step 4 in section A.2 above.

2. For each parameter set, i, calculate the product $\Phi\left(\vec{\mathbf{q}}_i\right) \equiv \prod_{AY=2}^{10} \prod_{Lag=12-AY}^{10} \Phi\left(\vec{\mathbf{q}}_{AY,Lag,i}\right)$.

3. Calculate the FFT of the mixture over all i, $\Phi\left(\vec{\mathbf{q}}\right) = \dfrac{\sum_i \Phi\left(\vec{\mathbf{q}}_i\right)}{n}$, where n is the number of Gibbs samples.

4. Invert the FFT, $\Phi\left(\vec{\mathbf{q}}\right)$, to obtain the vector, $\vec{\mathbf{q}}$, which describes the distribution of the of the reserve outcomes.

Here are the formulas to calculate the mean and standard deviation of the reserve outcomes:

- Expected Value $= h \cdot \sum_{j=0}^{2^{14}-1} j \cdot \vec{\mathbf{q}}_j$.
- Second Moment $= h^2 \cdot \sum_{j=0}^{2^{14}-1} j^2 \cdot \vec{\mathbf{q}}_j$.
- Standard Deviation $= \sqrt{\text{Second Moment} - \left(\text{First Moment}\right)^2}$.

Figure 4 has plots of the $\vec{\mathbf{q}}$'s for the Cape Cod and the Beta models.

Appendix B. An Approximate Likelihood Calculation for the Collective Risk Model

The goal of this appendix is to show how to calculate approximate likelihoods $\ell\left(\mathbf{X}\,|\,\{ELR_{AY}\},\{Dev_{Lag}\}\right)$ for the Cape Cod model and $\ell\left(\mathbf{X}\,|\,\{ELR_{AY}\},a,b\right)$ for the Beta Model, where the distribution of each $X_{AY,Lag}$ is defined by Simulation Algorithm 1 above.

This paper does not follow Meyers [7], which uses FFTs, as described in Appendix A to calculate the likelihood. The reason for this is the speed of calculation. While today's computers can calculate a likelihood with the FFT in a fraction of a second, the use of the Gibbs sampler can require the calculation of millions of likelihoods. My experience is that the approximate likelihood calculation described below cuts the computing time by a factor of 60.

The general strategy for calculating the likelihood is to start by calculating the first two moments of the aggregate loss for each accident year and settlement lag in terms of the expected loss and the first two moments of the claim severity distribution. The next step is to find an overdispersed negative binomial (ODNB) distribution that has the same first two moments. We then approximate the probability of the observed loss with its probability indicated by the ODNB distribution.

The log-likelihood for a given triangle of data is then given by:

$$\sum_{AY=1}^{10}\sum_{Lag=1}^{11-AY}\log\left(\mathrm{ODNB}\left(x_{AY,Lag}\right)\right).$$

Here are the steps for calculating each $\log\left(\mathrm{ODNB}\left(x_{AY,Lag}\right)\right)$:

Step 1 – Calculate the first two moments of $X_{AY,Lag}$.

Let μ_{Lag} and σ^2_{lag} be the mean and variance of claim severity distribution for the given settlement lag. Formulas for these moments are in Klugman [5]. Next calculate the expected claim count, $\lambda_{AY,Lag} = \dfrac{Premium_{AY} \cdot ELR_{AY} \cdot Dev_{Lag}}{\mathrm{E}\left[Z_{Lag}\right]}$.

Then the variance of the compound Poisson distribution for $X_{AY,Lag}$ is given by

$$Var\left[X_{AY,Lag}\right] = \lambda_{AY,Lag} \cdot \left(\mu^2_{Lag} + \sigma^2_{Lag}\right).$$

Step 2 – Find an ODNB distribution with the same moments as that of $X_{AY,Lag}$.

We parameterize the negative binomial distribution so that the variance is equal to:

$$\lambda_{AY,Lag} + \frac{\lambda^2_{AY,Lag}}{\kappa_{AY,Lag}}.$$

If each claim has a constant size of $\mu_{AY,Lag}$, its variance is then equal to:

$$\mu^2_{AY,Lag}\left(\lambda_{AY,Lag} + \frac{\lambda^2_{AY,Lag}}{\kappa_{AY,Lag}}\right).$$

Equating the variance from Step 1 with the above variance and solving for κ yields:

$$\kappa_{AY,Lag} = \frac{\lambda_{AY,Lag} \cdot \mu^2_{Lag}}{\sigma^2_{Lag}}.$$

Given the parameters ELR_{AY} and Dev_{Lag}, we approximate the log-likelihood of an observation $x_{AY,Lag}$ follows.

1. Set $n_{AY,Lag} = x_{AY,Lag}/\mu_{Lag}$ rounded to the nearest integer.

2. Set $\log\left(\mathrm{ODNB}\left(x_{AY,Lag}\right)\right) = \log\left(\Pr\left(N = n_{AY,Lag} \mid \lambda_{AY,Lag}, \kappa_{AY,Lag}\right)\right)$.

Appendix C. Computer Code for the Algorithms.

This appendix describes the code that implements the algorithms in this paper. The code is written in R, a computer language that can be downloaded for free at www.R-Project.org[5]. The code itself will be posted in a zip folder that accompanies this paper on the CAS Web Site.

There is one feature of the code that is not described above. Occasionally the Gibbs sampler admits a set of parameters with low likelihood. The presence of such parameters causes subsequent parameters to have a high rejection rate with the result that the algorithm is "trapped." When this happens, the algorithm returns to a randomly selected parameter set that had been accepted earlier.

Here is a description of the files in the zip folder.

1. The Rectangle.csv – This is the triangle in Table 1 expressed in rectangular form so it fits into an R data frame.

2. CRM CCod Posterior.r – This code reads The Rectangle.csv and implements the Gibbs sampler to produce an output file containing sampled $\{ELR_{AY}\}$ and $\{Dev_{Lag}\}$ parameters from the Cape Cod model.

3. CRM CCod Posterior.csv – The output from a run of CRM CCod Posterior.r

4. CRM Beta Posterior.r – This code reads The Rectangle.csv and implements the Gibbs sampler to produce an output file containing sampled $\{ELR_{AY}\}$ and $\{Dev_{Lag}\}$ parameters from the Beta model.

5. CRM Beta Posterior.csv – The output from a run of CRM Beta Posterior.r. Some of the records in this dataset are in Table 5.

6. Predict Outcomes.r – This code takes the output from Files 3 and 5 above and calculates the predictive distribution. It creates graphs like those in Figure 4.

7. Risk Margin.r – This code takes File 5 and calculates the expected losses and TVaRs needed for the risk margin calculation.

---

[5] Meyers [6] provides more information about the R programming language.

8. Risk Margin.xls – This spreadsheet takes the output of File 7 and produces Tables 6 and 7.

Biography of the Author

Glenn Meyers is Vice President and Chief Actuary for ISO Innovative Analytics, a division of ISO devoted to predictive modeling. He holds a bachelor's degree in mathematics and physics from Alma College, a master's degree in mathematics from Oakland University, and a Ph.D. in mathematics from the State University of New York at Albany. Glenn is a Fellow of the Casualty Actuarial Society and a member of the American Academy of Actuaries. Before joining ISO in 1988, Glenn worked at CNA Insurance Companies and the University of Iowa.

Glenn's current responsibilities at ISO include the development of scoring products. Prior responsibilities have included working on ISO Capital Management products, increased limits and catastrophe ratemaking, ISO's, and Property Size-of-Loss Database (PSOLD), ISO's model for commercial property size-of-loss distributions.

Glenn's work has been published in *Variance* and the *Proceedings of the Casualty Actuarial Society*. He is a three-time winner of the Woodward-Fondiller Prize, a two-time winner of the Dorweiller Prize, and a winner of the Dynamic Financial Analysis Prize. He is a frequent speaker at CAS meetings and seminars.

His service to the CAS has included membership on various education and research committees. He currently serves on the International Actuarial Association Solvency Committee and the CAS Board of Directors.

Combined Analysis of Paid and Incurred Losses

B. Posthuma, E.A. Cator, W. Veerkamp, and E.W. van Zwet

Abstract

Motivation. The new solvency regimes now emerging, insist that capital requirements align with the underlying (insurance) risks. This paper explains how a stochastic model built on basic assumptions is used to monitor insurance risk in order to get a clear insight in the aligned economic capital including prudence margins for loss reserves.

Method. The incurred loss of an insurer consists of payments on claims and reserves for claims that have been reported. As all claims are settled eventually, the cumulative paid and incurred losses for a given loss period become equal. Therefore, a joint model for the paid and incurred loss arrays is constructed, following a multivariate normal distribution, conditioned on equality of the total paid and incurred losses for a given loss period. A new class of functions is designed specifically to model development curves.

Results. A simulation experiment proved that a joint model for both paid and incurred loss arrays as described under *Method*, leads to a more accurate prediction of loss reserves. While the standard way of estimating percentiles for the reserve is biased, the alternative method of bootstrapping will lead to more accurate outcomes.

Conclusions. Modeling paid and incurred losses jointly leads to a considerable improvement in loss reserving in terms of accuracy of predictions, as well as specification of percentiles.

Availability. This method is incorporated in software available from the authors.

Keywords: Solvency II, loss reserves, joint model for paid and incurred loss arrays.

1. INTRODUCTION

The new risk based solvency regimes now emerging, such as the Solvency II rules to be implemented in Europe in 2009, insist that capital requirements align with the underlying (insurance) risks. This makes a stochastic loss reserving model a necessity. Such a model needs straightforward assumptions that will allow that:

- risk for expired insurance contracts is integrated together with risk for future contracts, in order to get a complete insight into the risk of the insurance portfolio as a whole, and that
- incomplete data – such as imperfect loss triangles due to varying period lengths or even incidental missing values – is still constructive to the model.

Regression as a descriptive technique with basic probability assumptions often offers the possibility to efficiently create an appropriate stochastic framework.

In short, an insurer will have to examine previous payments to make predictions about all future financial obligations. However, the company needs to know more than just how much money it should expect to pay. The model's stochastic ranges generate economic capital and prudence margins for reserves. Therefore, an adequate assessment of percentile ranges is crucial.

Typically, an insurer will arrange his payments by loss period and development period in a rectangular loss array, which is also sometimes called a run-off table. Since some of the payments lie in the future, this array is not fully observed. The observed part is often referred to as a run-off triangle. We regard the unobserved part of the loss array as a collection of random variables and the goal is to determine their probability distributions as well as possible on the basis of the available data.

Naturally, an extensive literature exists on this important problem. Perhaps the most widely used approach is the chain ladder. Renshaw and Verral (1998) identify the underlying assumptions and Mack (1993) and England and Verral (1999) present ways of estimating the standard error of the prediction. There are countless alternatives to the chain ladder and Schmidt (2007) has compiled a 35-page bibliography on the subject of loss reserving!

Much of the existing literature, however, concerns only a single array of payments—an exception is the Munich Chain Ladder introduced by Quarg and Mack (2004). Indeed, in most cases we have two arrays: an array of payments on settled claims and an array of reserves for claims that have been reported, but not yet settled. We refer to the sum of payments and reserves as "incurred loss."

In this paper, we aim to analyze the paid and incurred loss arrays jointly. As all claims are settled eventually, the reserves vanish and the cumulative paid and incurred loss for a given loss period become equal. On the basis of this observation, we construct a joint model. In our description, each array follows a multivariate normal distribution, conditioned on equality of the total paid and incurred losses for a given loss period.

This paper is organized as follows. In the next section we present an overview of our multivariate normal model for the two arrays. We then proceed to give a more detailed description, defining a particular family of functions that is very useful for modeling development curves. In most cases, we observe only various aggregates of the arrays, but we show that this poses no difficulties. We discuss prediction and parameter estimation. To examine the advantage of our joint model we conducted a simulation experiment. We compared the results of the joint model to those obtained from using only a single array. We find that the joint model shows better results in terms of the mean squared prediction error. We report on these results in the final section.

2. MULTIVARIATE NORMAL MODEL

Let $Y_{lk}^{(1)}$ and $Y_{lk}^{(2)}$ denote the incremental paid and incurred losses for loss period $l = 1,2,...,L$ in development period $k = 1,2...,K$. Suppose that they are all independent normally distributed with means

$$EY_{lk}^{(1)} = \mu_l \Pi_k^{(1)} \quad \text{and} \quad EY_{lk}^{(2)} = \mu_l \Pi_k^{(2)}$$

and variances

$$\mathrm{var}(Y_{lk}^{(1)}) = \tilde{\Pi}_k^{(1)} \text{ and } \mathrm{var}(Y_{lk}^{(2)}) = \tilde{\Pi}_k^{(2)}.$$

We assume

$$\sum_k \Pi_k^{(1)} = \sum_k \Pi_k^{(2)} = 1.$$

It is of course sensible to assume a parametric form for the parameter vectors $\boldsymbol{\mu}$, $\boldsymbol{\Pi}^{(1)}$, $\boldsymbol{\Pi}^{(2)}$, $\tilde{\boldsymbol{\Pi}}^{(1)}$ and $\tilde{\boldsymbol{\Pi}}^{(2)}$. For ease of presentation, we defer this issue to the next section.

The assumed normal distribution of the entries of the loss arrays is often not appropriate. Occasional large claims result in distributions that are skewed to the right. To account for this skewness the entries are sometimes assumed to have the lognormal distribution. A disadvantage of such a model is the incompatibility of the log normal distribution with the negative values that do occur in practice in most arrays, and the incompatibility of the distribution when aggregating data (the sum of two lognormal random variables is not log normally distributed). Also, it will not be feasible to do what we are about to propose — that is, condition on the equality of the row sums of the loss arrays.

We should point out that as a result of the Central Limit Theorem, aggregates of the data are more normally distributed than the individual entries. We feel that the advantages of the multivariate normal model outweigh those of the multivariate lognormal model.

Let $\mathbf{Y^{(1)}1}$ and $\mathbf{Y^{(2)}1}$ denote the row sums of the matrices $\mathbf{Y^{(1)}}$ and $\mathbf{Y^{(2)}}$. Also, we can stretch out $\mathbf{Y^{(1)}}$ and $\mathbf{Y^{(2)}}$ as length KL vectors $\mathbf{y^{(1)}}$ and $\mathbf{y^{(2)}}$, respectively.

Given the event $\{\mathbf{Y^{(1)}1} = \mathbf{Y^{(2)}1}\}$, the vectors $\mathbf{y^{(1)}}$ and $\mathbf{y^{(2)}}$ have multivariate normal distributions. It is not difficult to determine the conditional mean and conditional covariance matrix. Refer to the Appendix for a general formulation.

Because $E\mathbf{Y^{(1)}1} = E\mathbf{Y^{(2)}1}$, the conditional mean of the vectors $\mathbf{y^{(1)}}$ and $\mathbf{y^{(2)}}$ is the same as the unconditional mean. However they are of course no longer independent!

Let $\mathbf{\Sigma_{11}}$ denote the unconditional covariance matrix of the length $2KL$ vector $\mathbf{y} = (\mathbf{y^{(1)}}, -\mathbf{y^{(2)}})$.

$$\mathbf{\Sigma_{11}} = \begin{pmatrix} Cov(\mathbf{y^{(1)}}) & 0 \\ 0 & Cov(\mathbf{y^{(2)}}) \end{pmatrix}, \tag{2.1}$$

where $Cov(\mathbf{y^{(1)}})$ and $Cov(\mathbf{y^{(2)}})$ are the diagonal covariance matrices of $\mathbf{y^{(1)}}$ and $\mathbf{y^{(2)}}$. We use $-\mathbf{y^{(2)}}$ for convenience, since in that case the row sums add up to zero.

Let $\mathbf{\Sigma_{22}}$ denote the covariance matrix of $(\mathbf{Y^{(1)}} - \mathbf{Y^{(2)}})\mathbf{1}$. Then

$$\mathbf{\Sigma_{22}} = \left(\sum_k \tilde{\Pi}_{(k)}^{(1)} + \sum_k \tilde{\Pi}_k^{(2)} \right) \mathbf{I},$$

where $\mathbf{I}$ is the $L \times L$ identity matrix.

Let $\mathbf{\Sigma_{12}} = \mathbf{\Sigma'_{21}}$ denote the covariance between $\mathbf{y}$ and $(\mathbf{Y^{(1)}} - \mathbf{Y^{(2)}})\mathbf{1}$.

The conditional covariance matrix of $\mathbf{y}$ given the event $\{(\mathbf{Y^{(1)}} - \mathbf{Y^{(2)}})\mathbf{1} = \mathbf{0}\}$ is

$$\mathbf{\Sigma} = \mathbf{\Sigma_{11}} - \mathbf{\Sigma_{12}} \mathbf{\Sigma_{22}^{-1}} \mathbf{\Sigma_{21}} \tag{2.2}$$

This completes the global specification of our model. In the next section we give a more detailed description.

3. DETAILED SPECIFICATION OF THE MODEL

In the previous section, we introduced vectors $\boldsymbol{\mu} \in \Re^{\mathrm{L}}$ and $\boldsymbol{\Pi}^{(1)}$, $\boldsymbol{\Pi}^{(2)} \in \Re^{\mathrm{K}}$, to describe the expectations. Define

$$\boldsymbol{\Pi} = \begin{pmatrix} \boldsymbol{\Pi}^{(1)} \\ -\boldsymbol{\Pi}^{(2)} \end{pmatrix}.$$

Mostly, we have a vector of "exposures" $\mathbf{W} \in \Re^{\mathrm{L}}$ representing a volume measure for each loss period, such as the total number of insurance policies. We choose an $L \times p$ matrix $\mathbf{X}$ and a parameter vector $\boldsymbol{\beta} \in \Re^{\mathrm{p}}$, and we model the expected total loss for loss period l

$$\mu_l = W_l e^{(\mathbf{X}\boldsymbol{\beta})_l} \tag{3.1}$$

We have

$$E(Y_{lk}^{(1)}) = W_l e^{(\mathbf{X}\boldsymbol{\beta})_l} \Pi_k^{(1)} \text{ and } E(Y_{lk}^{(2)}) = W_l e^{(\mathbf{X}\boldsymbol{\beta})_l} \Pi_k^{(2)}.$$

This means that if we define for matrices A and B of equal size

$$\exp(\mathbf{A})_{ij} = e^{A_{ij}} \text{ and } (\mathbf{A} \circ \mathbf{B})_{ij} = A_{ij} B_{ij},$$

then we can write

$$E(\mathbf{y}) = (\mathbf{W} \circ \exp(\mathbf{X}\boldsymbol{\beta})) \otimes \boldsymbol{\Pi},$$

where $\otimes$ denotes the tensor product between two vectors.

Next, we recall the vectors $\tilde{\boldsymbol{\Pi}}^{(1)}$, $\tilde{\boldsymbol{\Pi}}^{(2)} \in \Re^{\mathrm{K}}$, which represent the (unconditional) variances. Define their sums as σ_1^2 and σ_2^2

$$\sum_{k=1}^{K} \tilde{\Pi}_k^{(1)} = \sigma_1^2 \text{ and } \sum_{k=1}^{K} \tilde{\Pi}_k^{(2)} = \sigma_2^2.$$

Also define

$$\tilde{\mathbf{\Pi}} = \begin{pmatrix} \tilde{\mathbf{\Pi}}^{(1)} \\ \tilde{\mathbf{\Pi}}^{(2)} \end{pmatrix}.$$

We model the unconditional variances for loss period l and development period k as

$$\tilde{V}_{lk}^{(1)} := Var(Y_{lk}^{(1)}) = W_l e^{(X\beta)_l} \tilde{\Pi}_k^{(1)} \text{ and } \tilde{V}_{lk}^{(2)} := Var(Y_{lk}^{(2)}) = W_l e^{(X\beta)_l} \tilde{\Pi}_k^{(2)}.$$

Note that we use the same $\mathbf{X}$ and $\boldsymbol{\beta}$ as we did for the expectations. In matrix notation this becomes

$$Cov(\mathbf{y}) = \left((\mathbf{W} \circ \exp(\mathbf{X}\boldsymbol{\beta})) \otimes \tilde{\mathbf{\Pi}}\right)_\Delta.$$

Here we denote the diagonal matrix with the vector $\mathbf{v}$ as its diagonal by $\mathbf{v}_\Delta$. This describes the unconditional distribution of the vector $\mathbf{y}$. We can now use (2.2) to find the conditional distribution of $\mathbf{y}$, which then completely specifies our model.

3.1 Modeling the development curves

For a sensible approach to the estimation problem, it is necessary to limit the number of parameters by assuming a parametric model for the development vectors $\mathbf{\Pi}^{(1)}$, $\mathbf{\Pi}^{(2)}$, $\tilde{\mathbf{\Pi}}^{(1)}$, and $\tilde{\mathbf{\Pi}}^{(2)}$. To explain our method, let us concentrate on one of the arrays, for example $\mathbf{Y}^{(1)}$. We suppose that in loss period l, we expect a total loss of μ_l. Now suppose that the length of the loss period is T time units. The claims occurring in the small interval $[t, t+\Delta t]$, have an effect on the expected loss in the time interval $[s, s+\Delta s]$, $t \le s$ equal to

$$\frac{\mu_l \Delta t}{T} f_{\boldsymbol{\theta}}(s-t)\Delta s,$$

where f_θ is a (possibly negative) function such that

$$\int_0^\infty \mathrm{f}_{\boldsymbol{\theta}}(\mathrm{x})\mathrm{dx} = 1$$

for all possible choices of the parameter vector $\boldsymbol{\theta}$.

In the next section, we will describe a particular family of such functions, which possess some desirable properties. For now, let us note that the total loss over loss period l equals μ_l. Indeed, if $[t_l, t_l + T]$ denotes the loss period l,

$$\frac{1}{T}\int_{t_l}^{t_l+T} \mu_l \int_t^{\infty} f_{\boldsymbol{\theta}}(s-t)dsdt = \mu_l .$$

We are interested in the expected loss from loss period l in development period k. Denote with I_k the interval corresponding to this development period. We see that

$$\Pi_k = \frac{1}{T}\int_{t_l}^{t_l+T} \int_{t_l \oplus I_k \cap [t,\infty)} f_{\boldsymbol{\theta}}(s-t)dsdt$$

$$= \frac{1}{T}\int_0^T \int_{I_k \cap [t,\infty)} f_{\boldsymbol{\theta}}(s-t)dsdt.$$

Usually, the loss and development periods have the same length T. We can choose $T = 1$ so that $I_k = [k-1, k]$. We get

$$\Pi_1 = \int_0^1 \int_t^1 f_{\boldsymbol{\theta}}(s-t)dsdt \tag{3.2}$$

$$\Pi_k = \int_0^1 \int_{k-1}^k f_{\boldsymbol{\theta}}(s-t)dsdt, \ k \geq 2. \tag{3.3}$$

If we define the survival function

$$S_{\boldsymbol{\theta}}(x) = \int_x^{\infty} f_{\boldsymbol{\theta}}(y)dy$$

and the function

$$H_{\boldsymbol{\theta}}(x) = \int_0^x S_{\boldsymbol{\theta}}(y)dy$$

then we can rewrite (3.2) as

$$\Pi_1 = \int_0^1 (1 - S_{\boldsymbol{\theta}}(1-t))dt = 1 - H_{\boldsymbol{\theta}}(1),$$

and (3.3) as

$$\Pi_k = \int_0^1 (S_{\boldsymbol{\theta}}(k-1-t) - S_{\boldsymbol{\theta}}(k-t))dt = 2H_{\boldsymbol{\theta}}(k-1) - H_{\boldsymbol{\theta}}(k) - H_{\boldsymbol{\theta}}(k-2).$$

We conclude that it is useful to choose the functions $f_{\boldsymbol{\theta}}$ in such a way that we can calculate $H_{\boldsymbol{\theta}}$ explicitly. In the next section we will do just that. We conclude this section by mentioning that the development of the variances is modeled in a similar way. In that case we do need to make sure that the functions $f_{\boldsymbol{\theta}}$ are always positive.

3.2 A parametric family of functions

Now, we will introduce a parametric family of functions that meet the requirement of the previous section.

$$\{f(x;\beta,\gamma,\mu,\sigma) : \beta,\gamma,\sigma > 0, \mu \geq 0\},$$

where $x \in [0,\infty)$, These functions all satisfy

- $\int_0^\infty f(x;\beta,\gamma,\mu,\sigma)dx = 1 (\forall \beta,\gamma,\sigma > 0, \mu \geq 0).$
- $f(x;\beta,\gamma,\mu,\sigma) = Cx^{\gamma-1} + o(x^{\gamma-1})(x \downarrow 0)$ for some $C > 0$, depending on the parameters.
- $f(x;\beta,\gamma,\mu,\sigma) = Cx^{-2-\beta} + o(x^{-2-\beta})(x \to \infty)$ for some $C \in \Re$, depending on the parameters.

Furthermore, there exist analytic expressions for both the first and the second primitive of the function $f(.\,;\beta,\gamma,\mu,\sigma)$. Finally, for $\mu > 1$, $f(.\,;\beta,\gamma,\mu,\sigma)$ will have a negative tail.

We will use an auxiliary variable *y* to define our parametric family, and at first ignore the dependence on the scaling parameter σ. Define

$$\mathbf{y}(x) = \int_0^x \left(1 + \left(\frac{t\mathbf{B}\left(\frac{1}{\gamma},\frac{\beta}{\gamma}\right)}{\gamma}\right)^\gamma\right)^{-\frac{1+\beta}{\gamma}} dt$$

where

$$\mathrm{B}(a,b) = \int_0^1 t^{a-1}(1-t)^{b-1}dt$$

is the incomplete regularized beta-function. Now define

$$f(x;\lambda,\beta,\gamma,\mu,1) \overset{\text{def}}{=} (1+\beta)x^{\gamma-1}\left(\mathrm{B}\left(\frac{1}{\gamma},\frac{\beta}{\gamma}\right)/\gamma\right)^{\gamma}\left(1+\left(\frac{x\mathrm{B}\left(\frac{1}{\gamma},\frac{\beta}{\gamma}\right)}{\gamma}\right)^{\gamma}\right)^{-1-\frac{1+\beta}{\gamma}}$$

$$\times(1-\mu y(x)^{\gamma})+\gamma\mu y(x)^{\gamma-1}\left(1+\left(\frac{x\mathrm{B}\left(\frac{1}{\gamma},\frac{\beta}{\gamma}\right)}{\gamma}\right)^{\gamma}\right)^{-\frac{2+2\beta}{\gamma}}.$$

Finally, we include the scale parameter σ so that

$$f(x;\lambda,\beta,\gamma,\mu,\sigma) \overset{def}{=} \frac{1}{\sigma} f(\frac{x}{\sigma};\lambda,\beta,\gamma,\mu,1). \tag{3.4}$$

We verify that

$$H(x;\lambda,\beta,\gamma,\mu,1) = \int_0^x\int_t^\infty f(s;\lambda,\beta,\gamma,\mu,1)dsdt = y(x) - \frac{\mu y(x)^{1+\gamma}}{1+\gamma}$$

and

$$H(x;\lambda,\beta,\gamma,\mu,\sigma) = \sigma H(\frac{x}{\sigma};\lambda,\beta,\gamma,\mu,1).$$

We will now describe the effect of the various parameters on the shape of the development function. The parameter μ is the most interesting parameter. If we choose $\mu \leq 1$, we get a positive density, whose left and right tail behavior is determined by γ and β respectively. As μ approaches 1, the bump around the mode becomes more pronounced. When $\mu > 1$, the density "falls through" the x-axis, only to approach it again as $x \to \infty$; the tail behavior is still determined by γ and β. See Figure 1. Note that from the previous section it follows that

$$\int_0^\infty xf(x;\beta,\gamma,\mu,\sigma)dx = \left(1 - \frac{\mu}{1+\gamma}\right)\sigma$$

The effect of the parameters β and γ is similar to the behavior of these parameters in the parametric family of positive densities we get when we choose $\mu = 0$. The parameter β determines the right tail of the density, whereas γ determines the left-tail (near zero). In Figures 2 and 3 we chose $\mu = 5$.

Figure 1: Behavior of the density when varying μ

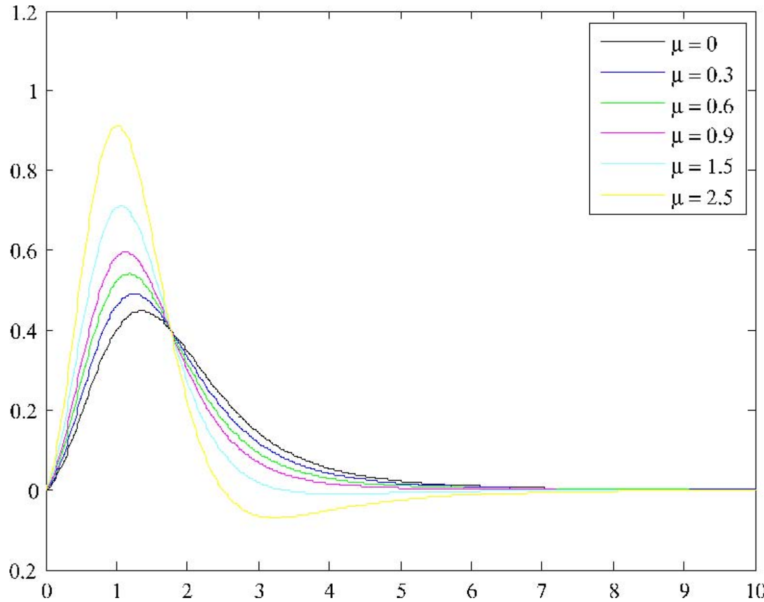
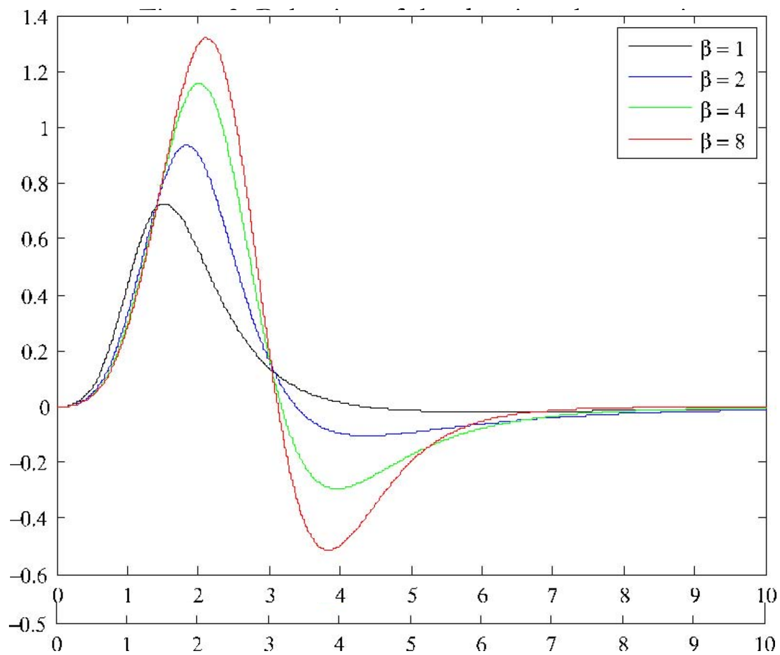


Figure 2: Behavior of the density when varying β



4. AGGREGATE OBSERVATIONS

Often we do not observe all the elements of the vector $\mathbf{y}$ individually, but compounded in various aggregates. For instance, for certain years we may only have records of payments per quarter, while for other years payments per month are available.

Suppose we observe J aggregates. If we assume that different aggregates never involve the same payments, we can introduce a zero-one matrix $\mathbf{S}$ with pair-wise orthogonal rows, of size $J \times 2KL$. Observing various independent sums of the elements of the vector $\mathbf{y}$ then corresponds to $\mathbf{z} = \mathbf{Sy}$.

Conditionally on $\{(\mathbf{Y^{(1)}} - \mathbf{Y^{(2)}})\mathbf{1} = \mathbf{0}\}$, $\mathbf{z}$ has a multivariate normal distribution with mean $\mathbf{SEy}$ and covariance matrix, $\mathbf{S}\sum\mathbf{S'}$ where $\mathbf{\Sigma}$ is given in (2.2). The advantage of choosing a multivariate normal model is very prominent here, since in this case it is still feasible to determine the likelihood of the data $\mathbf{z}$.

5. ESTIMATION AND PREDICTION

We can estimate the parameters of our model by maximizing the likelihood of the data. If we call the vector of parameters $\boldsymbol{\theta}$, then we maximize

$$\mathrm{lik}(\boldsymbol{\theta}) = \mathrm{P}_{\boldsymbol{\theta}}(\mathbf{z} = \mathrm{z} \mid \mathbf{Y^{(1)}} - \mathbf{Y^{(2)}})\mathbf{1} = \mathbf{0}). \tag{5.1}$$

The parameter vector $\boldsymbol{\theta}$ is very high dimensional. Indeed, there are at least 16 parameters describing the (unconditional) means and variances of the $Y_{lk}^{(1)}$ and $Y_{lk}^{(2)}$. Maximizing (5.1) is a delicate affair and must involve some iterative procedure. The speed and success will depend on the algorithm which is used and, perhaps even more importantly, on the starting point. The starting point should be some ad hoc estimator, which is relatively easy to compute but still reasonably close to the true maximum likelihood estimator.

To evaluate the accuracy of our estimates we use standard theory for maximum likelihood estimation. That is, we use the Hessian of the log likelihood at the maximum likelihood estimate to approximate the Fisher information.

Typically, we are not so much interested in the parameters, as we are in a prediction of the reserve. Conditionally on the data and the equality of the row sums, the reserve has a multivariate normal distribution and we can use the conditional expectation as a prediction. The uncertainty in this prediction is a combination of the stochastic uncertainty of the model and the uncertainty in the parameter estimates.

6. SIMULATION

To evaluate the effect of conditioning on the equality of the row-sums, we conduct a simulation experiment. We estimate the reserve with conditioning on equal row-sums, using the run-off tables simultaneously, as described in this paper. We refer to this approach as the *joint* method. For comparison, we also estimated the reserve without conditioning on the row-sums, essentially only using the paid table. We call this approach the *marginal* method. Of course, the marginal method is much easier, as it involves no conditioning. However, result in this section show that the more complicated joint method does produce better results.

We carry out the following simulation experiment. We consider a set of actual insurance data that, for reasons of privacy, we have made anonymous by multiplying with some undisclosed factor. We fit a model using the parametric family of densities described in section 3 for the development curves of the expectation and variance of both the paid and the incurred table. This results in a 19 dimensional parameter $\boldsymbol{\theta}_{\mathbf{0}}$, which contained

- $2 \times 4 = 8$ parameters for the two expectation development curves.
- $2 \times 4 = 8$ parameters for the two variance development curves.
- Three parameters for the exposure (β) to account for two regime changes.

We define: R, the reserve as the sum of future payments and $\hat{R}$, the estimator for R.

For this data set we estimate the reserve $\hat{R}_0$ for the paid table and its variance $\hat{V}_0$ conditioned on the aggregated data and taking into account both the stochastic uncertainty and the parameter uncertainties. We find $\hat{R}_0 = 5.24$ and $\hat{V}_0 = 2.04$.

Next, we simulate two entire tables (paid and incurred) from the multivariate normal model determined by the estimated parameter vector $\boldsymbol{\theta}_0$, and repeat this about 6000 times. By using the estimated parameter vector, we make sure that our simulated data resembles realistic data. Of course, for each simulated data set, we know the "true" reserve R, as the sum of total simulated future payments. Hereafter, the estimated reserve $\hat{\mathrm{R}}$ is based on the simulated data set of historical payments. The error in the estimated reserve, as the difference between R and $\hat{R}$, is compared for the two methods.

6.1 Reserve Estimation

Denote R as the true reserve in a given simulated data set, $\hat{R}_1$ and $\hat{R}_2$ as the estimates for the reserve for the joint and marginal methods, respectively. One of the most important measures for the quality of a prediction is the Mean Squared Error (MSE). Our simulation showed that

$$E(R-\hat{R}_1)^2 = 2.18$$
$$E(R-\hat{R}_2)^2 = 6.90\,.$$

Clearly, by using both tables simultaneously we achieve superior performance. We remark here, that using more simulations would not have changed this conclusion. In Figure 4 we show the convergence of the average of the squared error for the joint method as the number of simulations increases. We see that the average has sufficiently stabilized towards the end. The convergence for the marginal method is very similar.

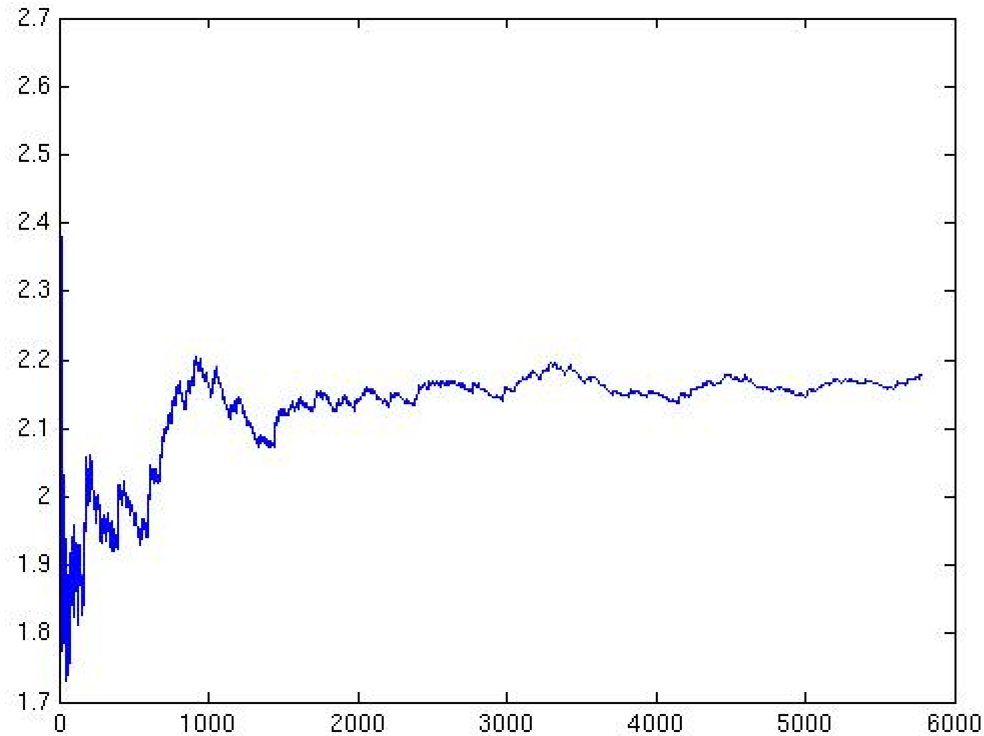


Figure 4: Convergence of the average mean squared error for method 1.

The bias of the estimators is also important

$$E(\hat{R}_1 - R) \;\; = -0.18$$

$$E(\hat{R}_2 - R) \;\; = -0.32$$

We note that the bias of both methods is very small compared to the MSE. It is not surprising that we find a similar bias for both methods, since the expectation structure in both models is the same. Recalling that the MSE consists of the estimator's variance and its squared bias, we conclude that large MSE of the marginal method is an immediate consequence of its inability to correctly estimate this covariance structure.

It is also interesting to see how well both methods do at determining the accuracy of the estimate. We have calculated a conditional variance of the estimated reserve, given the data, taking into account the uncertainty in the parameter estimates. This leads to

$$median\ (\hat{V}_1) = 1.63$$
$$median\ (\hat{V}_2) = 4.28\ .$$

Since the reserve estimates are almost unbiased, these values should be close to the mean-squared errors. This is not the case; both methods underestimate the variance. This is also clearly visible in Figure 5 where we plot the histogram of the estimated variances. The skewness indicates that we frequently underestimate the variance. This is a problem, when we want to estimate percentiles. We address this issue in the next sub-section.

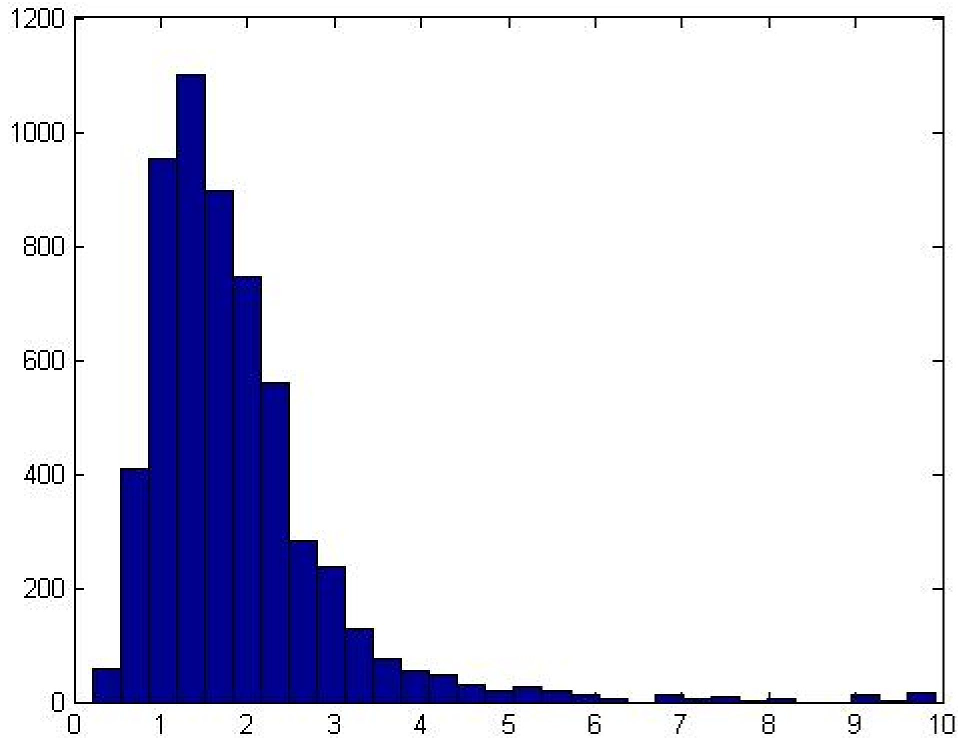


Figure 5: Histogram of the estimated variance for the joint method.

6.2 Estimating Percentiles

For loss reserving it is typically not sufficient to only have a point estimate of the reserve; percentiles are also needed. In this sub-section we discuss why the standard approach to estimating the percentiles does not work well in our case. We also provide an alternative.

The standard way of estimating percentiles is based on the following idea. When we estimate R by $\hat{R}$, and we estimate the variance by $\hat{V}$, we assume that the standardized residuals are approximately standard normal distributed. That is,

$$\frac{R-\hat{R}}{\sqrt{\hat{V}}} \sim N(0,1).$$

Then we can use the percentiles of the standard normal to find approximate percentiles for the reserve. We used this method to estimate the 75% and the 95% percentiles. To verify the results, we look at the percentage of times the true (simulated) reserve was larger than the estimated percentile. This gives

- $P(R > \hat{q}_{75}) = 0.32$ and $P(R > \hat{q}_{95}) = 0.12$ for the joint method.
- $P(R > \hat{q}_{75}) = 0.36$ and $P(R > \hat{q}_{95}) = 0.19$ for the marginal method.

Both methods seem to underestimate the percentiles, and we checked that this effect does not disappear as we increase the number of simulations. This is very troubling as it will lead to overoptimistic loss reserving.

In Figure 6 we plot the histogram of the standardized residuals for the joint method, and note that the distribution is not standard normal at all! Not only is it skewed, but also its mean is 0.24 instead of 0 and its variance is 1.49 instead of 1. This explains why the percentiles are not estimated accurately. The problem originates with the underestimation of the variance we discussed in the previous section, since having a small variance leads to a small percentile.

We conclude that the standard approach to estimating the percentiles does not work well. Therefore, we would like to suggest an alternative approach. The idea is simple: use the distribution of Figure 6 instead of the standard normal to calculate percentiles. This is essentially an application of the bootstrap. This would lead to a 75th percentile of 0.95, (instead of 0.67 for the standard normal) and a 95th percentile of 2.37 (instead of the familiar 1.65). For our original data set with $\hat{R}_0 = 5.24$ and $\hat{V}_0 = 2.04$ this means that the percentiles for the reserve are given by

- $\hat{q}_{75} = 6.60$ and $\hat{q}_{95} = 8.62$ using the bootstrap method.
- $\hat{q}_{75} = 6.20$ and $\hat{q}_{95} = 7.59$ using the normal method.

The relative difference between the two methods becomes more pronounced for higher percentiles, mainly because the relative contribution of the estimated reserve $\hat{R}_0$ diminishes.

Performing the many simulations needed to determine the distribution of the standardized residuals is a substantial computational burden. It took us four days to create Figure 6. In certain applications this is prohibitive.

Although the distribution of Figure 6 is specific to our particular data set, it is certainly conceivable that similar distributions would result from other data sets. Indeed, for data sets concerning similar insurance products this seems plausible at least. This suggests the following approach. We perform the simulations for a number of different data sets with varying characteristics. Then, when confronted with a new data set, we choose the histogram that is most appropriate, and use it instead of the standard normal to calculate percentiles.

Another suggestion to deal with this problem is judging the standardized residuals of the original loss triangle data set, given the parameter estimates. While the kurtosis of these residuals differs from the normality 3-value, the percentiles for loss reserves should be adjusted by taking these percentiles from a *t*-distribution, whereby the degree of freedom depends on the magnitude of the difference for the calculated kurtosis and the value 3.

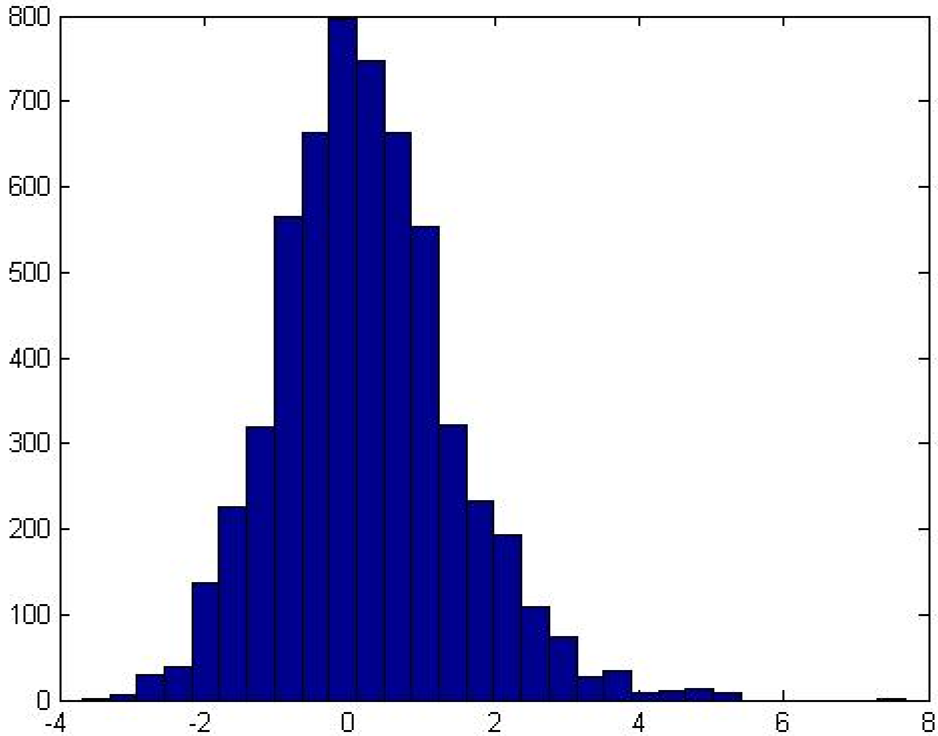
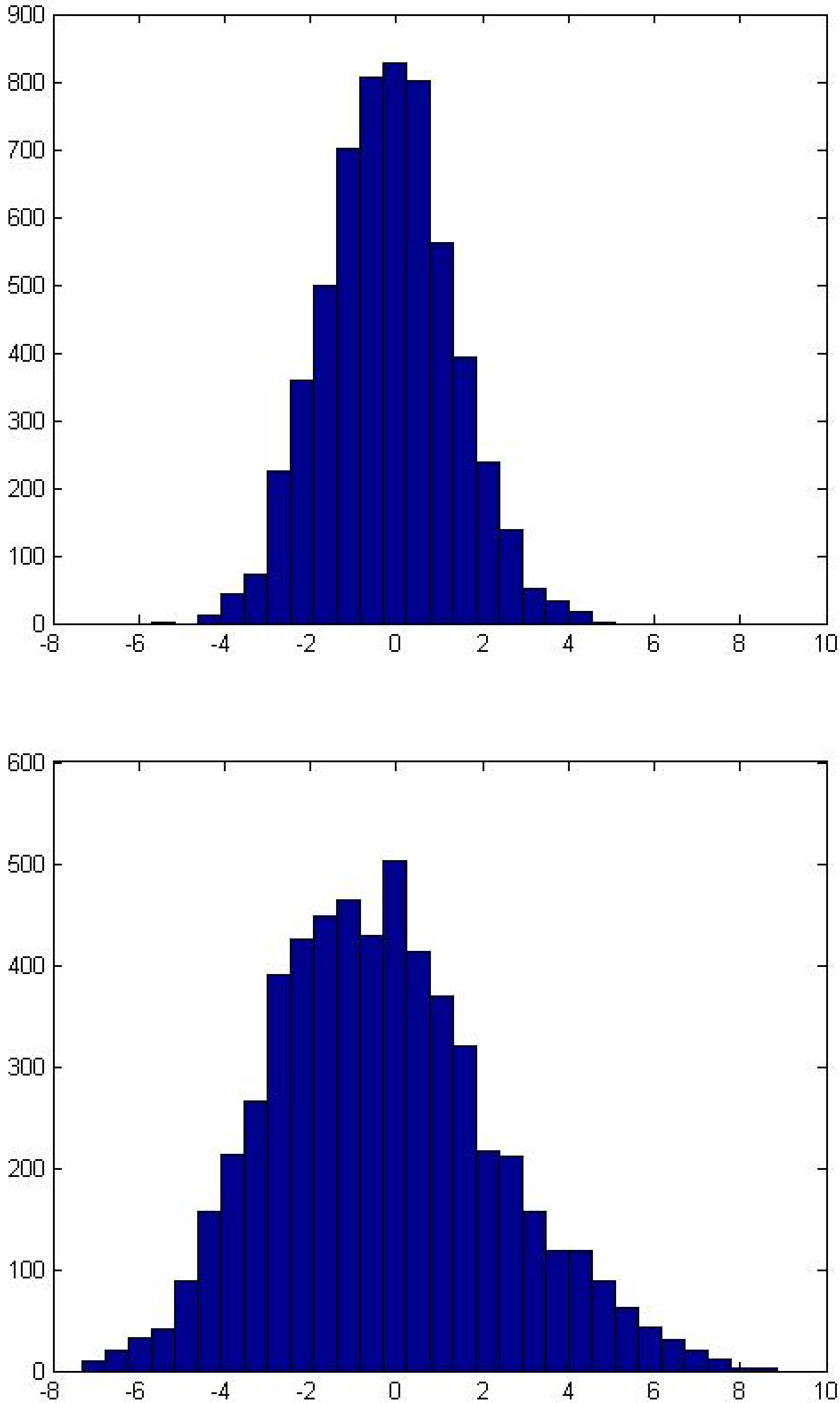


Figure 6: Histogram of the standardized residuals for the joint method.

7. CONCLUSION

The incurred loss of an insurer consists of payments on claims and reserves for claims that have been reported. As all claims are settled eventually, the cumulative paid and incurred losses for a given loss period become equal. On the basis of this observation, we construct a joint model for the paid and incurred loss arrays. In our description, each follows a multivariate normal distribution, conditioned on equality of the total paid and incurred losses for a given loss period. On the basis of this model, we make predictions for future payments.



Figures 7 and 8: Histogram of the difference between the true and estimated reserve based on the joint model (top) and the marginal model (bottom)

A rather technical, but important feature of our model is the use of a new parametric family of functions that are ideally suited for modeling development curves.

We have compared the performance of the joint model of the paid and incurred losses to an approach where we analyze only the paid table. In Figure 7 and 8 we present the results of a simulation study. These figures show histograms of the difference of the true and predicted reserves for both methods. While both methods are approximately unbiased, the one based on the joint model has much smaller variance. A more detailed discussion of this result is found in the previous section, but here we conclude that joint modeling is to be preferred over utilizing only the paid table.

Since the practice of loss reserving also takes the distribution of the reserve into account, we have studied the estimation of percentiles as well. We noted that inference from the normal assumptions does not produce good results. In fact, the results would lead to over-optimistic assessment of economic capital and prudence margins, which is, of course, to be avoided. We have proposed an alternative approach based on the bootstrap. It entails performing many simulations, to replace the assumed standard normal distribution of the standardized residuals with a more accurate description. Carrying out this method requires substantial computational effort, which in practice is only feasible on a highly aggregated level.

APPENDIX A

For ease of reference, we recall a well-known fact about the multivariate normal distribution.

Consider a random vector X, which is distributed according to the multivariate normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. Suppose we partition X into two sub-vectors

$$\mathbf{X} = \begin{pmatrix} \mathbf{X}^{(1)} \\ \mathbf{X}^{(2)} \end{pmatrix}.$$

Correspondingly, we write

$$\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}^{(1)} \\ \boldsymbol{\mu}^{(2)} \end{pmatrix} \quad \text{and} \quad \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11}\boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{22} \end{pmatrix}.$$

Now, if $\det(\boldsymbol{\Sigma}_{22}) > 0$, then the conditional distribution of $\mathbf{X}^{(1)}$ given $\mathbf{X}^{(2)}$ is multivariate normal with mean

$$\boldsymbol{\mu}^{(1)} + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\mathbf{X}^{(2)} - \boldsymbol{\mu}^{(2)})$$

and covariance matrix

$$\boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}.$$

8. REFERENCES

[1] England, P. and R.J. Verral, "Analytic and bootstrap estimates of prediction errors in claims reserving," **1999**, *Insurance: Mathematics and Economics* 25, 281-291.

[2] Mack, T., "Distribution-free calculation of the standard error of chain ladder reserve estimates," **1993**, *ASTIN Bulletin* 23, 213-225.

[3] Quarg, G. and T. Mack,, "Munich chain ladder: A reserving method that reduces the gap between IBNR projections based on paid losses and IBNR projections based on incurred losses," **2004**, *Blätter DGVFM,* Heft 4, 597-630.

[4] Renshaw, A.E. and R.J. Verral, "A stochastic model underlying the chain ladder technique," **1998**, *British Actuarial Journal* 4, 903-923.

[5] Schmidt, K.D., "A bibliography on loss reserving," 2007, www.math.tu-dresden.de/sto/schmidt/dsvm/reserve.pdf

Biographies of Authors

Bouke Posthuma is Founder and President and Consulting Actuary of Posthuma Partners, a Dutch actuarial firm. He was chairman of the ASTIN board in the Netherlands and a member of the International ASTIN Committee. This year, he received the Honorary Member Award of Recognition and Merit from the Actuarial Association of The Netherlands (Actuarieel Genootschap) www.posthuma-partners.nl.

Eric Cator is an Associated Professor at the Faculty of Information Technology and Systems of the Delft Technical University and holds a PhD in mathematics. E.A.Cator@TUDelft.nl.

Wim Veerkamp is a Consulting Actuary with Posthuma Partners and involved in loss reserving. He holds a master in econometrics, a PhD in psychometrics and a minor in actuarial science. He is an associated member of the Actuarial Association of The Netherlands (Actuarieel Genootschap). wim.veerkamp@posthuma-partners.nl.

Erik van Zwet is an Associated Professor at the Mathematical Institute of Leiden University and holds a PhD in mathematics. evanzwet@math.leidenuniv.nl

Loss Development in Workers Compensation in the Presence of Legislative Reform

Frank Schmid

Motivation. Legislative reforms affect loss development patterns in various ways. Some legislative innovations may affect new policy (or accident) years only, while others have diagonal effects as they affect both new and existing claims. Modeling these effects is critical for adequacy in ratemaking and reserving.
Method. Using a Bayesian state-space model, workers compensation triangles are developed subject to the applicable legislative stipulations. Most importantly, this model is capable of accommodating the legislative environment as it evolves over time.
Results. The model is applied to an unidentified state, which experienced a reform cluster in the period 1990/92. The model shows how this reform cluster affects the ultimate loss and the 19th-to-ultimate tail factors.
Conclusions. Ultimate losses are not only dependent on the legislative environment at time of loss, but are also affected by how this legislative environment evolves over time. The statistical model is capable of quantifying the effects of such legislative changes on the loss development pattern.
Availability. The model runs in OpenBUGS 2.2.0 (http://mathstat.helsinki.fi/openbugs/) within the R (www.r-project.org) package BRugs 0.3-3 (http://cran.r-project.org). OpenBUGS is administered by the Department of Mathematics and Statistics of the University of Helsinki, Finland; R is administered by the Technical University of Vienna, Austria. OpenBUGS and R are GNU projects of the Free Software Foundation and, hence, available free of charge.

Keywords. Workers Compensation, Trend and Loss Development, Reserving Methods.

1. INTRODUCTION

Workers compensation is a line of insurance that operates in a legal environment that is subject to frequent and (sometimes) sweeping changes. Such legislative changes affect the loss development patterns in ratemaking and reserving in powerful and complex ways. Traditional loss development models do not acknowledge the specific legal environment in which the losses have been observed, nor are these models capable of incorporating changes in the legal setting into the loss development pattern; as a consequence, these models are not capable of quantifying the impacts of changes to legal stipulations on the ultimate loss or tail factor.

What follows is a Bayesian state-space model of loss development that explicitly accounts for the legal environment in which the losses of a given (indemnity or medical) triangle were observed. Accounting for the legal environment means translating legal stipulations into data inputs, which are then fed into the model. The model is set up to accommodate a wide array of legal changes, among which are changes to the stipulated rates of escalation (for indemnity) and (any) factors that bear on the rate at which incremental payments decay in development net of the calendar-year effect.

1.1 Research Context

A wide array of loss development models have been suggested, some of which are extensions of traditional actuarial methods (mostly related to the chain ladder; see, for instance, Mack [5]), while other models cast loss development into a time series framework (see, for instance, de Jong and Zehnwirth [2], de Jong [1], and Verrall [9]). For overviews on loss development models, see England and Verrall [3] and Taylor [8]. Bayesian modeling of loss development using the software platform BUGS (Bayesian inference Using the Gibbs Sampler) has been pioneered by Scollnik [6,7]. The model presented here draws on Scollnik [7].

1.2 Objective

The objective of the loss development model introduced in this paper is to give the practicing actuary a framework for developing losses in a changing legal environment. By acknowledging changes in pertinent legal stipulations, the model is capable of delivering values for the ultimate loss (and, hence, the tail factor) that are adequate for ratemaking and reserving. Specifically, the model allows for quantification of reform impacts on the ultimate loss and the tail factor.

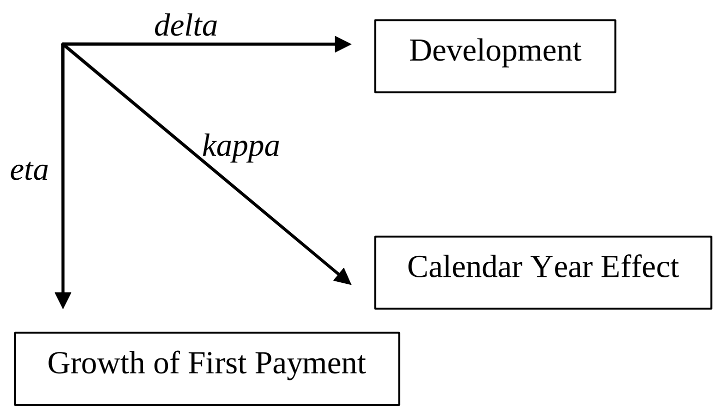
1.3 Outline

The following section first outlines the basic structure of the Bayesian state-space model of loss development and then, in a sub-section, applies this model to an unidentified state. This application details how regulatory information is fed into the model and how the model quantifies the reform impact. Section 3 presents the results of this empirical analysis. Section 4 offers conclusions.

2. BACKGROUND AND METHODS

The Bayesian state-space model of loss development employed in the analysis of legislative reform treats incremental payments as a three-dimensional time series problem. Specifically, the incremental payments are driven by three time processes, which are growth of the first payment, development, and the calendar-year effect; these processes are illustrated in Exhibit 1.

Exhibit 1: Time Processes in Loss Development



The model fits to (the logarithms of) incremental payments and, at the same time, employs a stochastic cumulative sum (cusum) constraint to ensure that, for any development year, the sum of the estimated incremental payments for a given policy (or accident) year add up (approximately) to the observed cumulative payment for that policy (or accident) year.

As an example, consider the stylized triangle displayed in Exhibit 2. Let $y[i,j]$ be the (natural) logarithm of the incremental payment of policy (or accident) year i in development year j, which materializes as a draw from a normal distribution with expected value $b[i,j]$. Then, the expected value of the logarithm of the first payment in the first policy (or accident) year, $b[1,1]$, develops into $b[1,2] = b[1,1] + delta[2] + kappa[1,2]$, where $b[1,2]$ is the expected value of the logarithm of the second incremental payment in the first policy (or accident) year. The parameter $delta[2]$ is the rate of decay (which is expressed as a logarithmic rate of growth) of the calendar-year effect-adjusted incremental payments from development year 1 to development year 2, whereas the term $kappa[1,2]$ is the calendar-year effect (which, again, is expressed as a logarithmic rate of growth) from calendar year 1 to calendar year 2. Note that the calendar-year effect is not restricted to be uniform along a given diagonal—for instance, $kappa[2,3]$ is allowed to differ from $kappa[3,2]$; this is because different types of indemnity claims (which consist of Temporary Total [TT], Permanent Partials [PP], Permanent Totals [PT], and Fatals) may escalate at different rates and the fraction of the various types in the total may change across development years. Finally, for the expected value of the logarithm of the first payment in the second policy (or accident) year, $b[2,1]$, we can write

$b[2,1] = b[1,1] + eta[2]$, where $eta[2]$ (which is again expressed as a logarithmic rate of growth) equals the change in expected values.

Exhibit 2: Stylized Triangle

Policy (or Accident) Year

	Development Year 1	2	3
1	y[1,1]	y[1,2]	y[1,3]
2	y[2,1]	y[2,2]	
3	y[3,1]		

The run-off rate (*delta*) is estimated using a smoothed random walk specification; the smoothing is obtained by scaling the innovation variance with a Gompertz function. The rate of growth of the expected value of the first incremental payment (*eta*) is also estimated using a smoothed random walk; unlike the innovation variance of the run-off rate (which decreases as development progresses), the innovation variance of *eta* is constant. (The smaller the innovation variance, the smoother is the estimated trajectory of growth rates.)

The model draws on expert information in determining the prior for the calendar-year effect, which manifests itself in the growth rate *kappa*. For indemnity benefits inflation, these expert priors are the rates of escalation as stipulated in the law; these stipulated rates of escalation may vary by type of claim. Additionally, the expert priors for the rates of escalation may vary by policy (or accident) years and development years. The expert prior for medical benefits inflation is the rate of growth of the Medical Care component of the CPI (Consumer Price Index; www.bls.gov), M-CPI for short.

The model develops future losses subject to the assumption that the expert priors for the (non-constant) rates of inflation follow random walks, starting at the final observed rates. The purpose of these random walks is to incorporate uncertainty about the future rates of inflation. The innovation variances of these random walks have to be determined by an expert based on the actual behavior of the applicable inflation series. Due to the skewed, lognormal distribution of the incremental payments, greater uncertainty about future rates of inflation (that is, greater innovation variances in the random walks) implies higher expected values of incremental payments and, all else being equal, a larger tail factor.

The model assumes that beyond the final observed development year, the projected run-off rate is the minimum of the final estimated run-off rate (that is, the run-off rate that applies in the final observed development year) and a mortality-based run-off rate. Starting with the final estimated logarithmic run-off rate, this logarithmic mortality-based run-off rate decreases linearly in every development year such that in development year 60, this rate equals the current official (logarithmic) mortality rate for age 80. Beyond age 80 (development years 61 through 70), the (logarithmic) mortality-based run-off rate equals the (logarithm of the) official mortality rate for the applicable age. The mortality information originates from the Social Security Administration (Periodic Life Table, www.ssa.gov). Where indemnity benefits are not granted for life (due to an age limit or an otherwise stipulated restriction in the duration of benefits), the number of payments is reduced accordingly, as detailed in the following section.

For details on the model, see Appendixes 2 and 3; Appendix 4 offers a list of variables. The model was estimated using Markov chain Monte Carlo simulation; for introduction to this estimation technique see, for instance, Gilks, Richardson, and Spiegelhalter [4]. The equations were coded in BUGS and run in R (using BRugs [Version 0.3-3, which utilizes OpenBUGS 2.2.0 beta from February 2006]) with a burn-in of 40,000 iterations, followed by a sample of 40,000 iterations, of which every fourth draw entered the posterior distribution (to mitigate autocorrelation in the Markov chains).

2.1 The Reform Impact of an Unidentified State

This section presents an application of the loss development model for the purpose of studying the impact of legislative reform on the loss development pattern, and the tail factor in particular. The model is applied to a loss triangle of policy year data; the first report of payments of any given policy year comprises 24 months of experience. The policy years in the loss triangle range from 1980 through 2005. The triangle, which is displayed in Exhibit A-1 in Appendix 1, is incomplete due to a missing upper left-hand side triangle, a missing upper right-hand side triangle, and a missing lower left-hand side (single-observation) triangle.

The purpose of the analysis is to study the reform impact in an unidentified state; this state experienced major reforms in workers compensation in the years 1982, 1986, 1990 (effective

September 1), and 1992 (indemnity-related reforms effective May 18, and medical-related benefits reform effective November 1). The 1982 and 1986 reforms are not broken out because the first diagonal in the triangle refers to the year 1988. The reform impact of interest is the one of the 1990/1992 reform cluster; for this purpose, we define the time window 1988-1989 as the pre-reform period, and the window 1993-2005 as the post-reform period. Four of the most significant impacts of the 1990/1992 reforms were (1) the introduction of escalation of indemnity benefits at the rate of the CPI (regardless of the date of the injury) for PT disability claims and PP disability claims in May 1991 (beyond 312 weeks of benefits; indemnity benefits for fatal claims had been escalating at a fixed rate of 4 percent since June 1986); (2) a limitation of the duration of TT disability claims to 52 weeks; (3) closer scrutiny regarding continued eligibility of indemnity benefits; and (4) an indemnity retirement offset that is immediate for accidents past age 55 or, otherwise, sets in five years prior to the official retirement age. Whereas the introduction of a cost-of-living adjustment is captured in the model as a calendar-year effect (as such adjustment started applying to claims of any maturity), the time limitation on TT claims, the increased scrutiny regarding continued eligibility, and the social security offset can be expected to bear on the run-off rate (*delta*). The run-off rate (*delta*) picks up the effect of a social security offset to the extent that such offset kicks in for (older) claimants within the first 20 development years (as these are the development years covered by the data). Yet, because the social security offset may not be fully captured by the run-off rate (due to there being [younger] claimants for whom the offset does not kick in within the 20 observed development years), the model assumes (as an approximation) a 50 percent reduction of the incremental indemnity payments past development year 40. Note that the increased scrutiny regarding continued eligibility of indemnity benefits may spill over into medical benefits, thus causing medical claims to close faster. Hence, we expect the 1990/1992 reform cluster to lead to a faster run-off not only in indemnity but also in medical incremental payments. (Note that although the most significant impacts of the 1990/1992 reform cluster were the indemnity reforms mentioned above, the 1990/1992 reform cluster also included a medical reform in November 1992, as mentioned above.)

Exhibit A-2 in Appendix 1 details the shapes of the pre-reform and post-reform triangles. The area of the pre-reform triangle for which there is data is shaded gray; this area comprises all observations between (and inclusive of) the 1988 diagonal and the 1989 policy year. The post-reform triangle is bordered by a solid line and consists solely of post-1992 diagonals. Note that the model does not fit to (the six) observations between (and inclusive of) the 1990 policy year and the 1992 diagonal, although these observations are assigned to the pre-reform period for the purpose of the post-reform estimation. The pre-reform and post-reform loss development processes are estimated simultaneously. The missing upper left-hand side triangle (diagonals 1980 through 1987) is given its own trajectory of run-off rates, which is the same for both the pre-reform and the post-reform estimation. Finally, for the post-reform estimation, the run-off rates that apply to the diagonals from 1988 through 1992 are allowed to differ from the estimated pre-reform run-off rates.

Although the pre-reform triangle consists only of policy-year data prior to 1990, this triangle includes elements through the 2005 diagonal. To the degree that the 1990/1992 reform cluster affected existing (instead of only new) claims (for instance, by accelerating their closure), the model may underestimate the impact of the reform cluster on the ultimate loss; however, the post-reform ultimate losses (and tail factors) would still be accurate, as argued below. For the data set at hand, the pertinent (future) policy year for ratemaking is 2008.

Unlike the pre-reform triangle, the post-reform triangle consists only of diagonals observed in the pertinent legislative environment. Yet, only in the first column of the post-reform triangle do all observations fall into the post-reform regime. As development time increases, the post-reform triangle phases in observations from the previous legislative setting, as indicated by the step function that defines the post-reform triangle in Exhibit A-2. For instance, in the first development year, all 13 incremental payments (of which the one for policy year 2005 is missing) are from the post-reform period. In the second development year, there are again 13 incremental payments (of which none are missing), but only 12 originate in the post-reform regime; and so on. The progressive phasing in of observations from the prior legislative regime rests on the premise that the run-off rates (but not necessarily the level of payments) of the post-reform regime approach the pre-reform run-off rates as development time advances; this is because the rates of decline of calendar-year effect-adjusted incremental payments deep in development may predominantly be driven by factors immune to the reforms, such as mortality. (It is because the reform may affect the level of payments deep in development [due to its effect on the run-off rates early in development] that the pre-reform run-off rates in the post-reform estimation are allowed to differ from the pre-reform run-off rates in the pre-reform estimation.) If the run-off rates (of the pre-reform policy years) deep in development are indeed immune to the reform, then the model estimates accurately both the pre-reform and post-reform ultimate losses. If, on the other hand, the run-off rates (of the pre-reform policy years) deep in development are affected by the reform, then the model underestimates the reform impact (but still estimates the post-reform ultimate loss accurately because it is the post-reform development pattern that materializes in the post-reform diagonals). But then there is a third situation where the model is not able to quantify the post-reform ultimate loss (as well as the impact of the reform). Such situation arises when the reform affects the run-off deep in development of new claims only, as is the case when a second-injury fund is eliminated. Because the reform takes many years to play out in the data (that is, manifest itself in incremental payments of new claims deep in development), the model is incapable of quantifying such reform impact immediately.

When estimating the loss development model, the pre-reform and post-reform triangles are estimated simultaneously, subject to the constraint that the two triangles have identical calendar-year effects, identical rates of growth of the expected value of the logarithm of the first payment, and identical variances in the measurement equations of the incremental payments. For details on the model, see Appendices 2 and 3.

3. RESULTS

Odd-numbered charts exhibit the indemnity results, whereas even-numbered charts display the results for medical.

Charts 1 and 2 show the indemnity and medical benefits estimated run-off rates (*delta*) along the development year axis—remember that the run-off rates are the rates of growth of the incremental payments, adjusted for the calendar-year effect. As mentioned, the run-off rates beyond the final observed year of development incorporate mortality information. Whereas the displayed run-off rates for medical benefits (Chart 2) describe the trajectory of the run-off rate as employed in the computation of the ultimate loss (and, hence, the tail factor), the run-off trajectory of the incremental indemnity payments (Chart 1) needs adjustment before inputting it into the computation of the ultimate loss or the tail factor; this is because indemnity benefits may not be granted for life, or there may be a social security offset. (If there is an immediate social security offset that applies regardless of the age of the claimant, then such offset is captured by the trajectory of the run-off rate *delta*.) In the unidentified state in question, effective May 1992, a social security offset applies to accidents that happen past age 55 or within five years of the legal retirement age. As a result of this legislative change, the incremental payments for development years 41 through 70 (70 being the final development year) were reduced by 50 percent of what would be projected otherwise.

Charts 3 and 4 present the expert priors (lines with full circles) and the posteriors for the calendar-year effect in the second development year (which is the first year of escalation). Due to this being a policy year triangle, the prior in the displayed second development year comprises 18 months of inflation (which is the time difference between the mid-points of the first 24 months of experience and the subsequent 12 months of experience). Note that, in general, a systematic difference between the expert prior for the calendar-year effect and the (unknown) workers compensation-specific rate of inflation factors into the run-off rate *delta*. Specifically, if for all incremental payments the actual (logarithmic) rate of benefits inflation exceeds the expert prior by a constant c (which may be positive or negative), then such constant will be absorbed by the rate of decay (*delta*) of the calendar-year effect-adjusted incremental payments—in statistical terms, the parameter c is unidentified.

Whereas the prior for medical inflation (Chart 4) is the M-CPI for all policy years, the prior for indemnity escalation (Chart 3) is a weighted average of the legally stipulated rates of escalation (of which the model accommodates two non-zero rates of escalation in addition to the zero rate [no escalation]). For instance, for the second development year, the rate of escalation that applies to a given type of claim (for a given policy or accident year) is weighted by the fraction of (incremental) losses associated with the given type of claim in the first development year. (Note that the fraction of incremental losses that applies to a given type of claim for a given development year is held constant for every policy (or accident) year, as such information is not available for every single policy or accident year.) Before policy year 1984, there was no escalation of indemnity benefits. Then, in policy year 1984, the escalation of fatal claims (at 4 percent), as introduced in June 1986, shows up in the prior (to the extent that this policy year was affected by the legislative change). The weight of such escalation increased in policy year 1985 before reaching (in policy year 1986) the level that corresponds to the fraction of Fatal (incremental) losses in the first development year. This level of escalation then rose again in policy year 1989 when in May 1991 PT claims started escalating at the CPI rate of inflation. This escalation of PT claims reached its full weight (at the fraction of PT incremental losses in the first development year) in policy year 1991. Note that because CPI inflation varies over time, the expert prior for the escalation of indemnity claims shows time variation even after 1991 (as indicated by the slight bumps in the applicable line in Chart 3).

Charts 5 and 6 displays the priors (lines with full circles) and the posteriors for the calendar-year effect of the latest observed diagonal; remember that there are no observations available for the final six values of the latest diagonal, which is why for these values the posterior equals the prior. Again, note that the first value on the diagonal comprises 18 months of inflation. For medical benefits, the expert prior for the calendar-year effect (which is the M-CPI; Chart 6) is uniform along the diagonal, except for the first value, which comprises inflation of a longer time period. For indemnity benefits, the expert prior for the rate of escalation (as determined by the pertinent legal stipulations) varies along the diagonal (beyond the initial change caused by switching from 18 months of inflation to 12 months of inflation); this is because diagonals span several development years. As a given set of indemnity claims develops, the proportions of incremental payments going to the various claim types (TT, PP, PT, and fatal) change; if these claim types escalate at different rates, then the expert prior for the escalation of the total of incremental payments within a given calendar year (diagonal) varies by development year. As mentioned, fatal claims escalate at four percent and PT claims escalate at the rate of CPI inflation; because the fraction of these claims is small in the first development year, the expert prior for the rate of escalation embedded in the total incremental payments in the second development year is close to zero. As development progresses, the fractions of incremental payments that apply to these two types of claim increases, as indicated by the rising line (full circles) in Chart 5 for development years 2 through 6. After 312 weeks of benefits, PP claims start escalating at the rate of CPI inflation. With TT claims having expired (or technically behaving like PT or PP claims), all claims escalate from development year 6 onward. (Fatal claims keep escalating at the stipulated four percent, whereas all other claims escalate at the CPI rate of inflation.)

Charts 7 and 8 show for $1 of initial (that is, first report) payment, kernel density estimates for the impact of the reform-induced change in the run-off rate (*delta*) on the ultimate loss for (the future) policy year 2008; remember that the first year comprises 24 months of development. Note that the payments are adjusted for the calendar-year effect; otherwise, studying the reform-induced difference in the ultimate loss would require choosing a specific pre-reform reference year (because of the time variation of the rate of inflation). Breaking out the reform impact on medical benefits is straightforward as for medical benefits, legislative reforms generally feed into the run-off rate *delta*. (Remember that any systematic difference between the workers compensation-specific medical inflation and M-CPI inflation are captured by the run-off rate *delta*; hence, any changes to the difference between these two inflation rates will be reflected in changes to *delta*.) Breaking out the reform impact on the ultimate loss of indemnity is more demanding than isolating such impact on the ultimate loss of medical; this is because legislative changes may not only change the run-off rate but also affect the stipulated rate of escalation, age limit for benefits, duration of benefits, or social security offset. The reform impact on the ultimate loss in indemnity, as depicted in Chart 7, is adjusted for the calendar-year effect, which means that the legislative changes to the applicable rates of escalation are not captured. Of course, the impact of the change in escalation can be broken out as well, but this requires choosing a specific reference year, as the CPI rate of inflation varies over time. (Alternatively, the ultimate losses of the various policy years [per $1 of initial payment] could be presented in a chart similar to Charts 9 and 10, which display the tail factors by policy year, while fully accounting for reform impacts.) As mentioned, to the extent that the 1990/92 reform cluster led to faster closing of existing claims and this way affected the run-off rates of post-1992 diagonals for pre-1990 policy years, the reform impact displayed in Charts 7 and 8 may be understated; this is because, even though the post-reform losses are accurately estimated, the "as-if-pre-reform" post-reform ultimate losses may be understated. Most interestingly, Chart 8 shows that the 1990/92 reform cluster indeed reduced the ultimate loss for medical (per $1 of initial payment), thus pointing to a faster run-off of medical payments due to increased scrutiny regarding continued eligibility for indemnity payments. As mentioned, the 1990/1992 reform cluster pertained mainly to indemnity benefits, but there was also a medical benefits reform, which occurred in November 1992.

Charts 9 and 10 exhibit the 19th-to-ultimate tail factors, differentiated by pre-reform and post-reform period; the post-reform period includes the future policy year (2008) of interest to ratemaking. The displayed tail factors rest on two alternative concepts. The first concept ("Tail Factors Based on b") computes the tail factors based on the estimated data-generating process. The second concept ("Tail Factors Based on *y.hat*") computes the tail factor based on the estimated incremental payments. Generally, for future policy (or accident) years, depending on the case, the two concepts generate the same number. The tail factor trajectory that stretches over the entire analyzed time period offers actual pre-reform tail factors (to the left of the left-most vertical separator) and "as-if-pre-reform" post-reform tail factors (to the right of the right-most vertical separator). The vertical differences between the "as-if-pre-reform" post-reform tail factors and the actual post-reform tail factors gauge the (full) reform impact. As argued above, to the extent that the reform cluster affected post-1992 diagonals for pre-1990 policy years, the "as-if-pre-reform" post reform tail factor may be understated.

Charts 11 and 12 offer a demonstration of how sensitive tail factors are to the rate of inflation that applies to the pertinent future policy year 2008. For indemnity, this rate of inflation is the rate of growth of the CPI, which is the (post-reform) stipulated rate of escalation for PP claims (after 312 weeks of benefits) and PT claims; the rate of escalation of fatal claims is kept at four percent. For medical, the rate of inflation is the M-CPI. Note that, due to the convexity of the tail factor in the rate of inflation, greater variability in the rate of inflation entails larger tail factors when averaged across policy years.

Charts 13 through 18 are diagnostic tools. These charts gauge how well the model has been calibrated; they display by policy year (Charts 13 and 14), development year (Charts 15 and 16) and calendar year (Charts 17 and 18) the difference between the log incremental payments predicted by the data-generating process (*b*) and the actual log incremental payments (*y*); the solid line indicates the median difference. Early in development, the solid lines in Charts 15 and 16 must be close to zero; late in development, these lines may turn jagged as outliers (in the percentage difference between observed and predicted payments) become more likely. The diagnostic Charts 13 through 18 signify that the model is well calibrated (as the median differences [solid lines] show no persistent departure from the zero line); in particular, the calendar-year effect (Charts 17 and 18) is properly captured.

Charts 19 and 20 are another set of diagnostic tools. These charts inform about data outliers and may serve as data quality indicators. The charts display by policy year the difference between the actual log cumulative payments (*z*) and the fitted log cumulative payments (*z.hat*) along the development year time axis. Based on experience, values within the interval (-0.005; 0.005) indicate that the model is able to replicate the underlying data. Values outside this interval but within the interval (-0.01; 0.01) have to be considered outliers. Values outside the interval (-0.01; 0.01) must be considered data points of poor quality.

Chart 1: Indemnity: Trajectory for delta (Run-off Rate, Calendar-Year Effect-Adjusted); "9": Pre-Reform; "8": Post-Reform

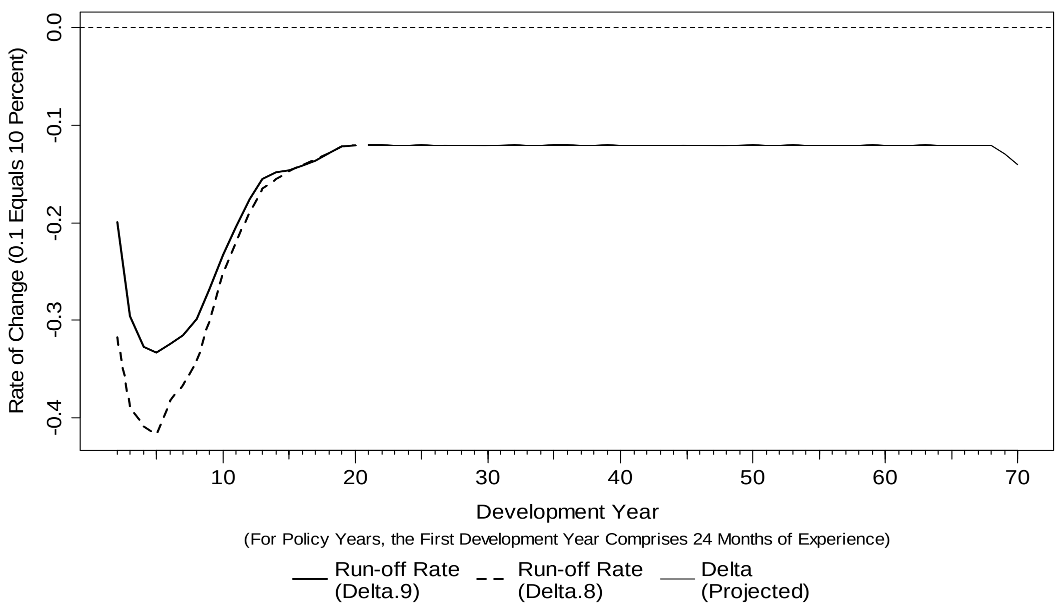


Chart 2: Medical: Trajectory for delta (Run-off Rate, Calendar-Year Effect-Adjusted); "9": Pre-Reform; "8": Post-Reform

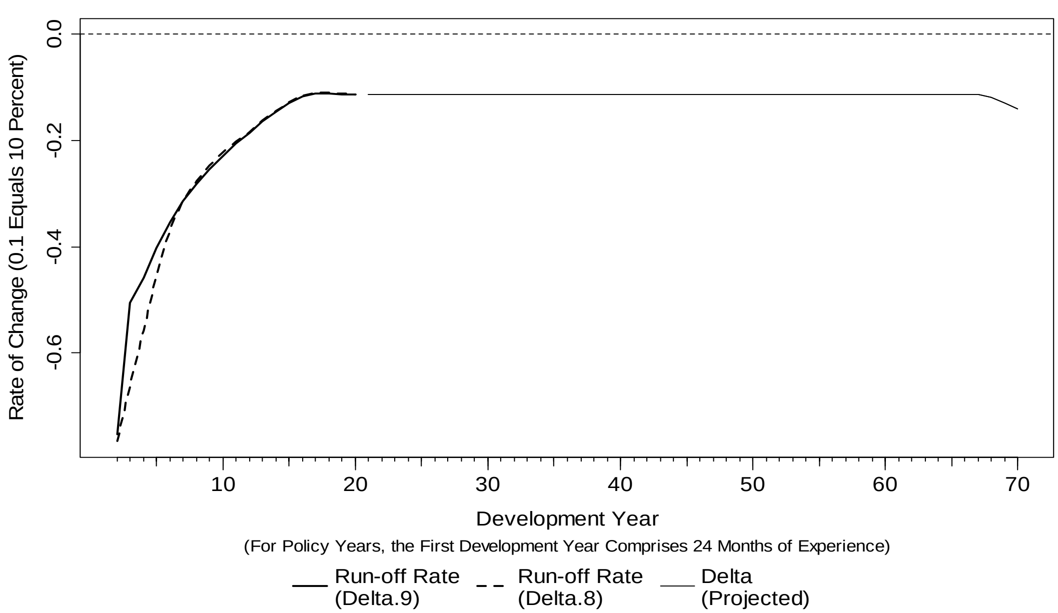


Chart 3: Indemnity: Calendar-Year Effect, Second Development Year

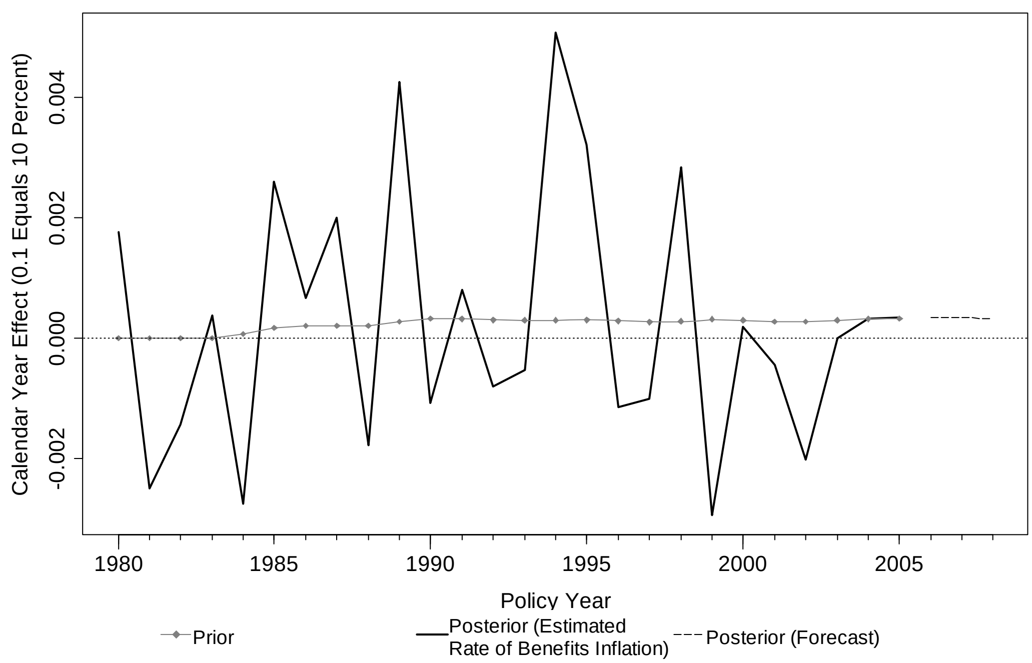


Chart 4: Medical: Calendar-Year Effect, Second Development Year

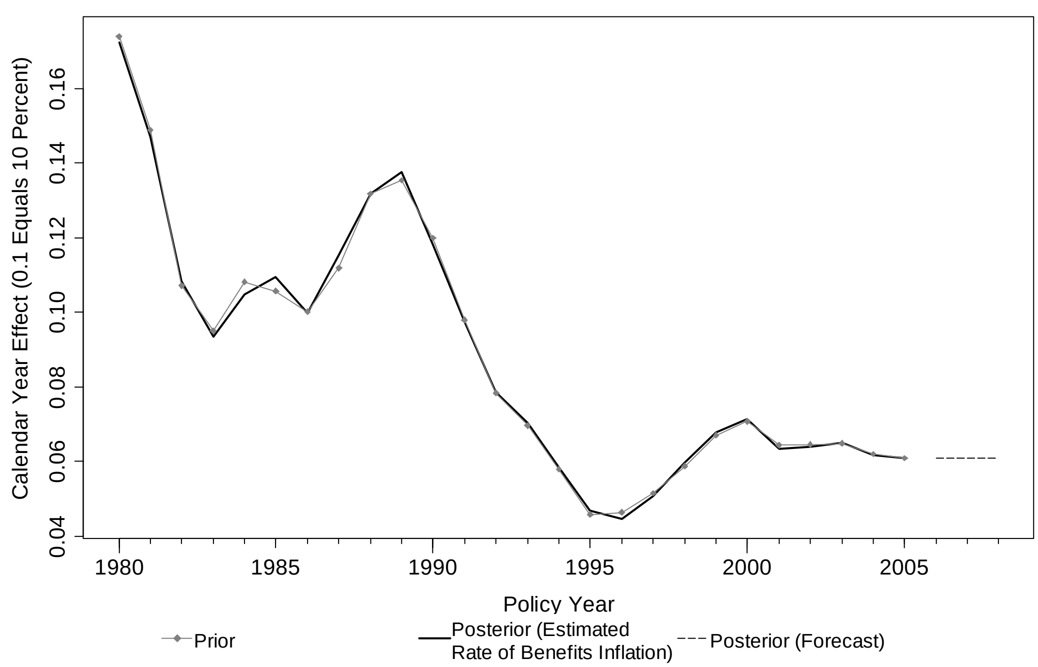


Chart 5: Indemnity: Calendar-Year Effect, Final Diagonal

Chart 6: Medical: Calendar-Year Effect, Final Diagonal

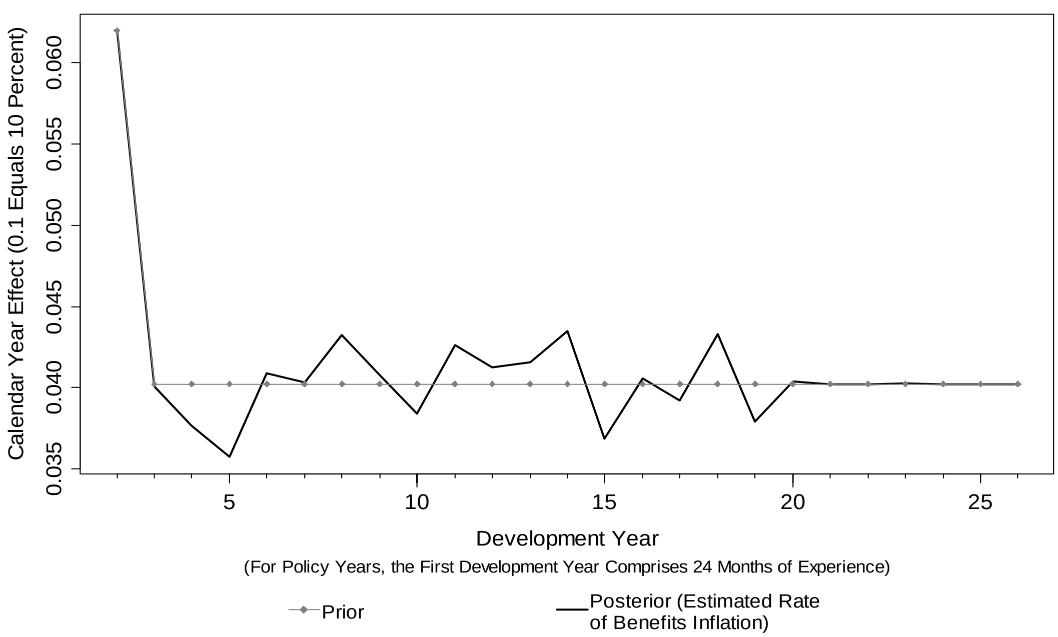


Chart 7: Indemnity: Reform Impact on the Ultimate Loss per $1 of First Report Payment (Adjusted for Calendar-Year Effect); Kernel Density Estimation

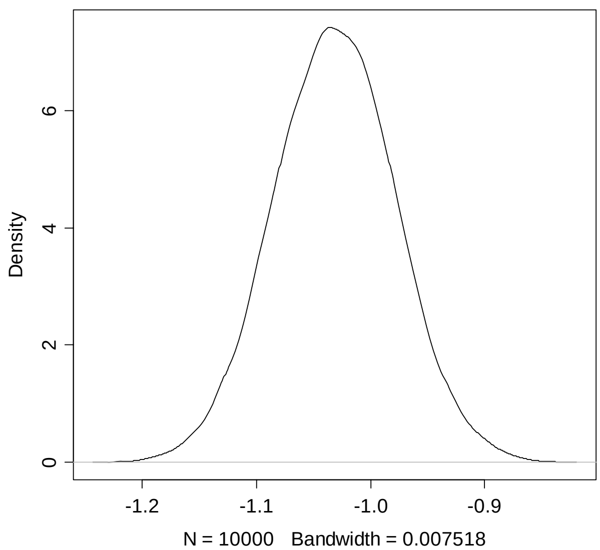


Chart 8: Medical: Reform Impact on the Ultimate Loss per $1 of First Report Payment (Adjusted for Calendar-Year Effect); Kernel Density Estimation

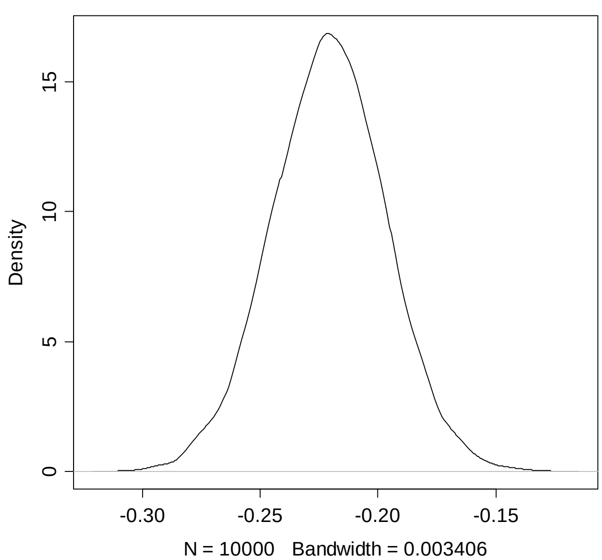


Chart 9: Indemnity: Tail Factor (Vertical Separators Border Reform Cluster)

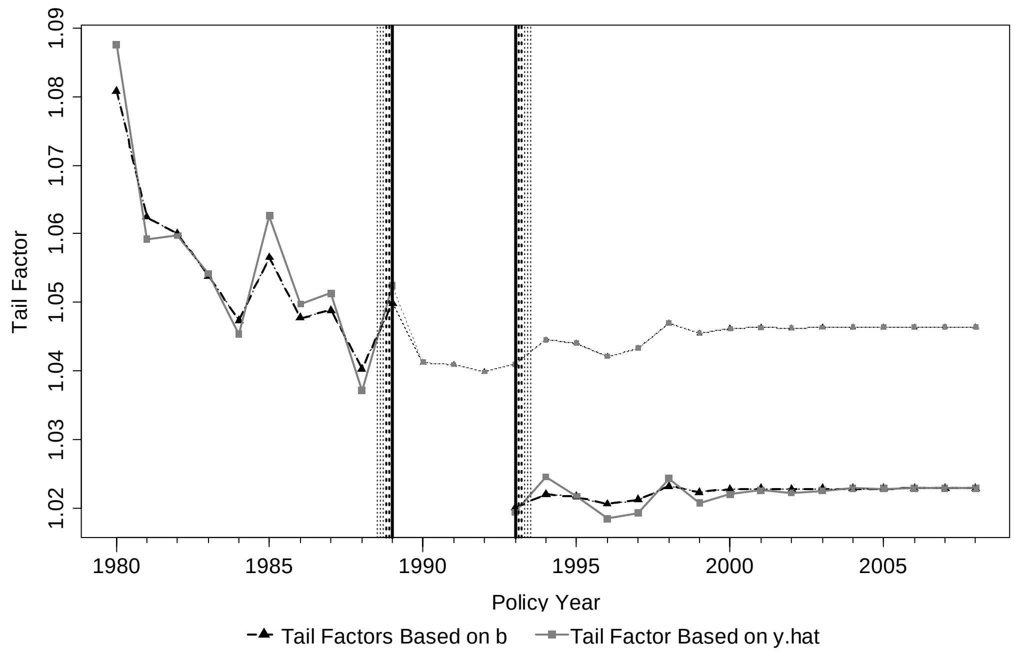


Chart 10: Medical: Tail Factor (Vertical Separators Border Reform Cluster)

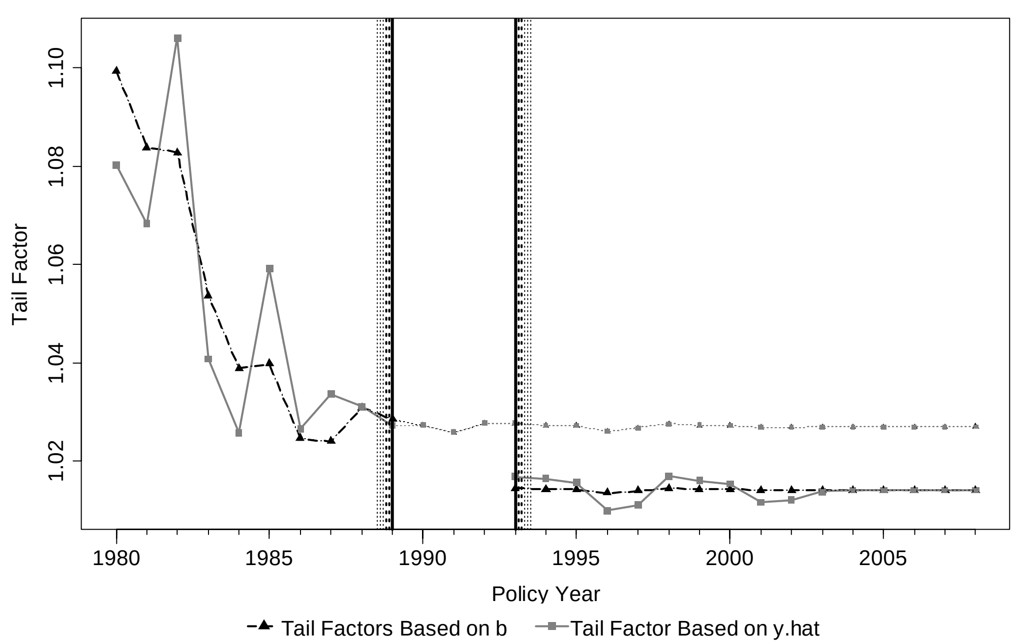


Chart 11: Indemnity: Sensitivity of Tail Factor to Official Rate of Inflation (CPI) for Policy Year 2008

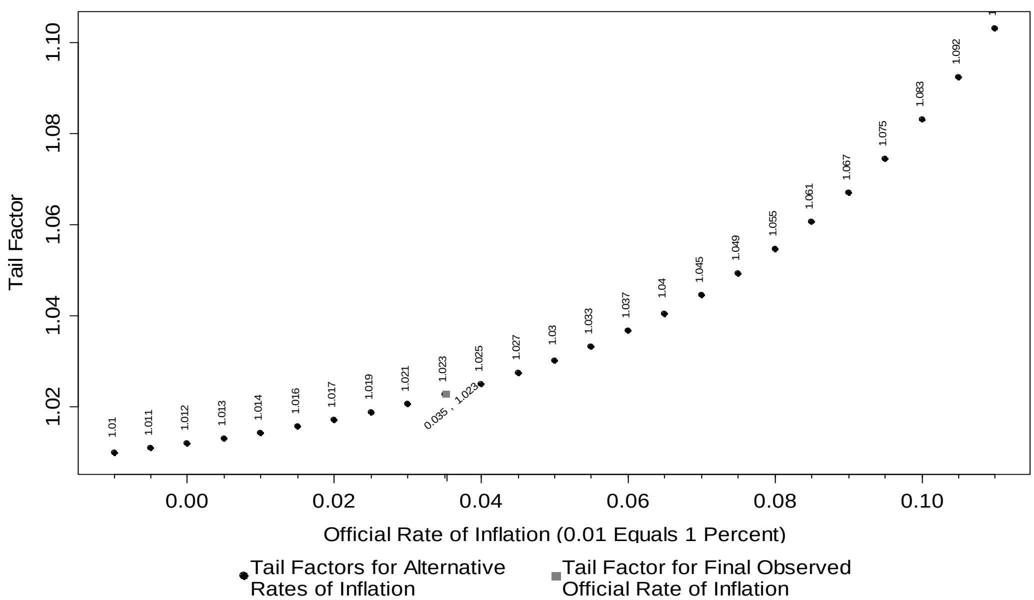


Chart 12: Medical: Sensitivity of Tail Factor to Official Rate of Inflation (M-CPI) for Policy Year 2008

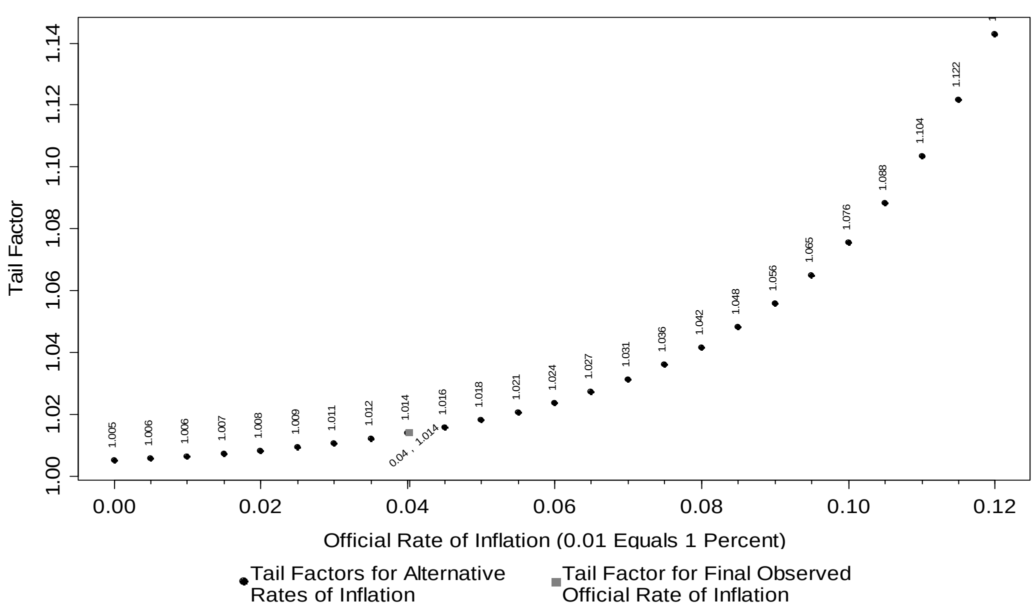


Chart 13: Indemnity: Difference between Actual Observations (*y*) and Estimated Process (*b*) by Policy Year, Post-Reform

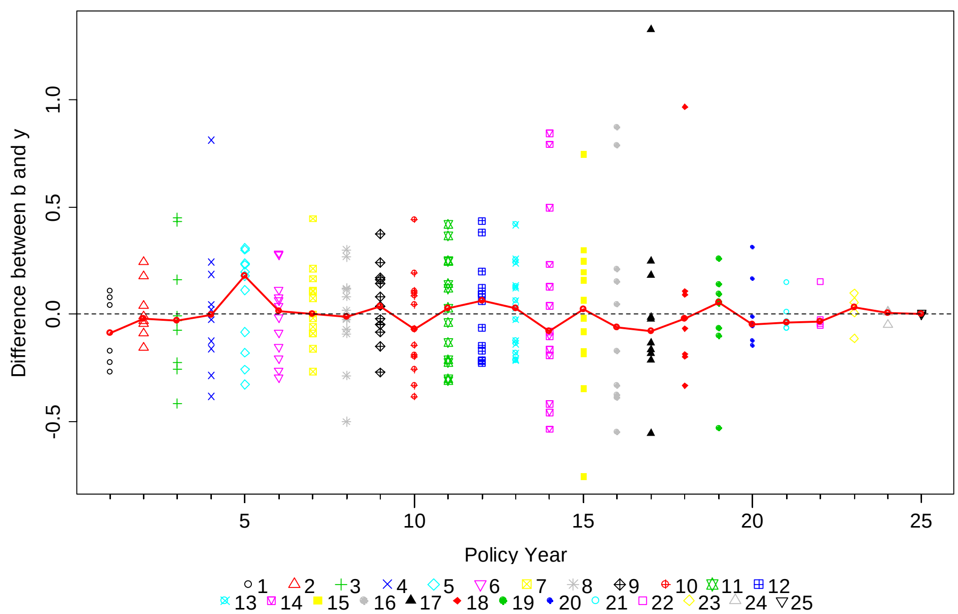


Chart 14: Medical: Difference between Actual Observations (*y*) and Estimated Process (*b*) by Policy Year, Post-Reform

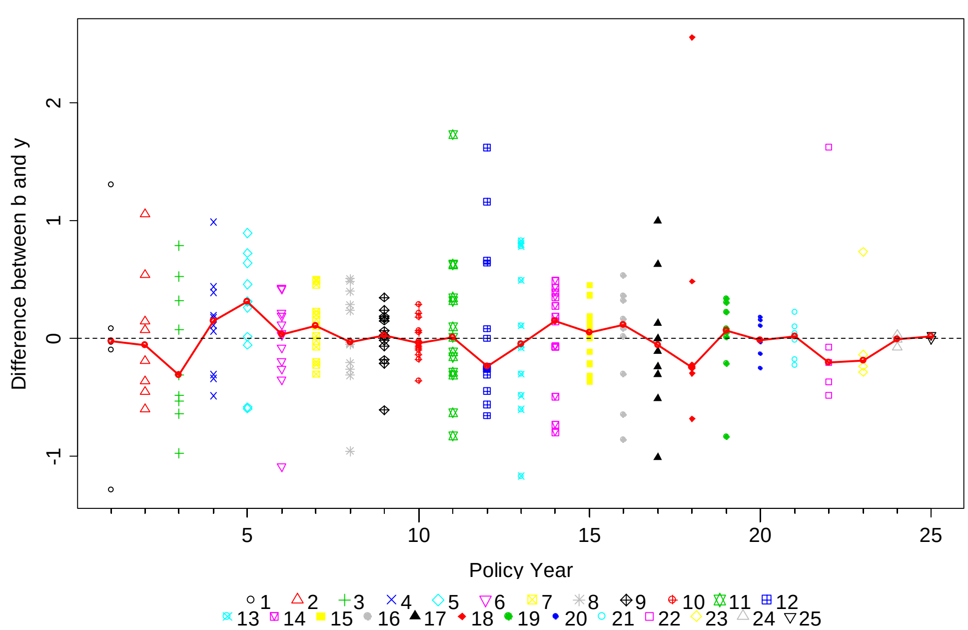


Chart 15: Indemnity: Difference between Actual Observations (*y*) and Estimated Process (*b*) by Development Year, Post-Reform

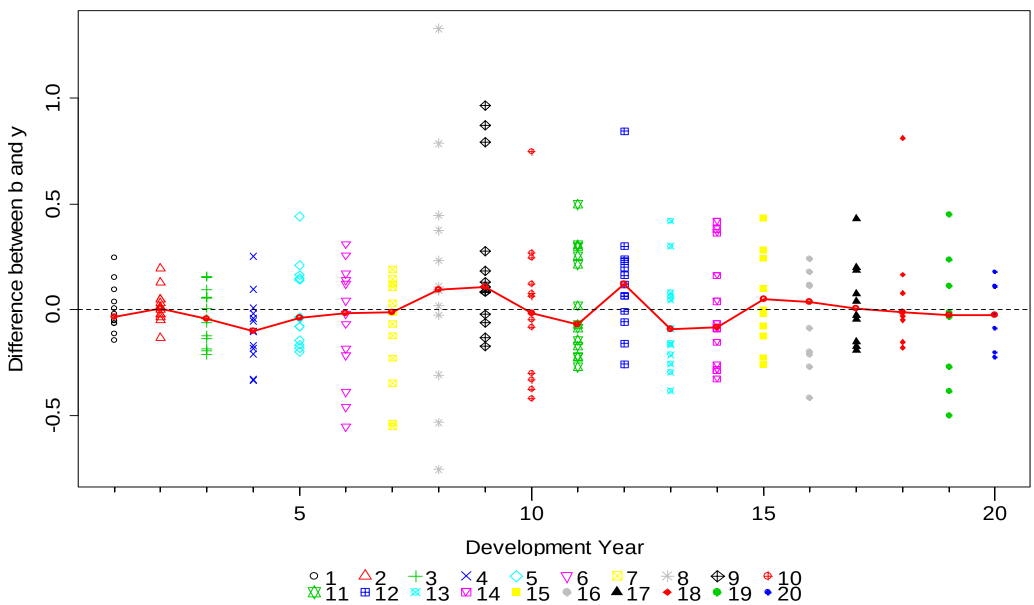


Chart 16: Medical: Difference between Actual Observations (*y*) and Estimated Process (*b*) by Development Year, Post-Reform

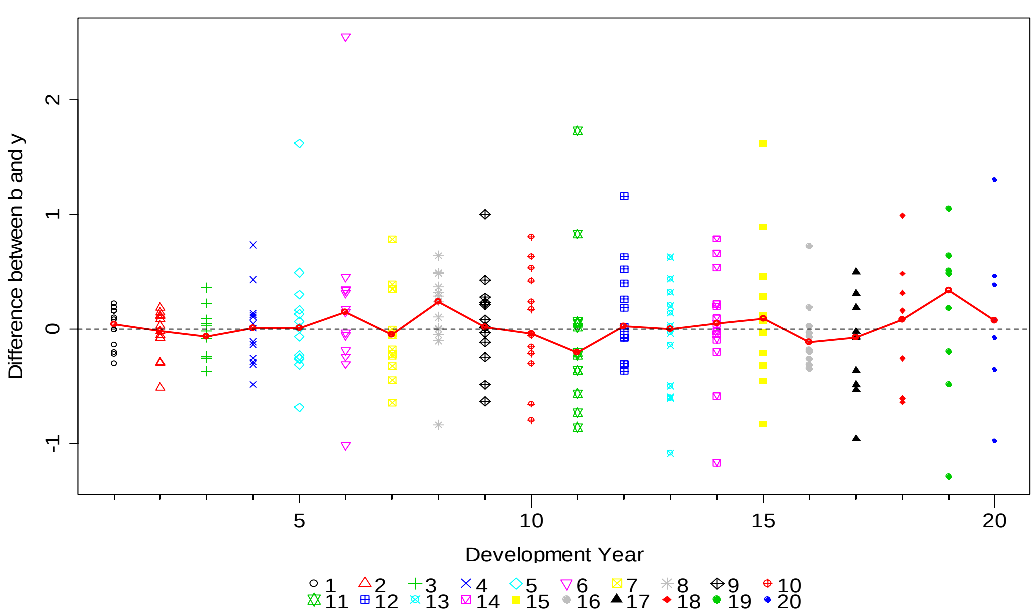


Chart 17: Indemnity: Difference between Actual Observations (*y*) and Estimated Process (*b*) by Diagonal (Calendar Year), Post-Reform

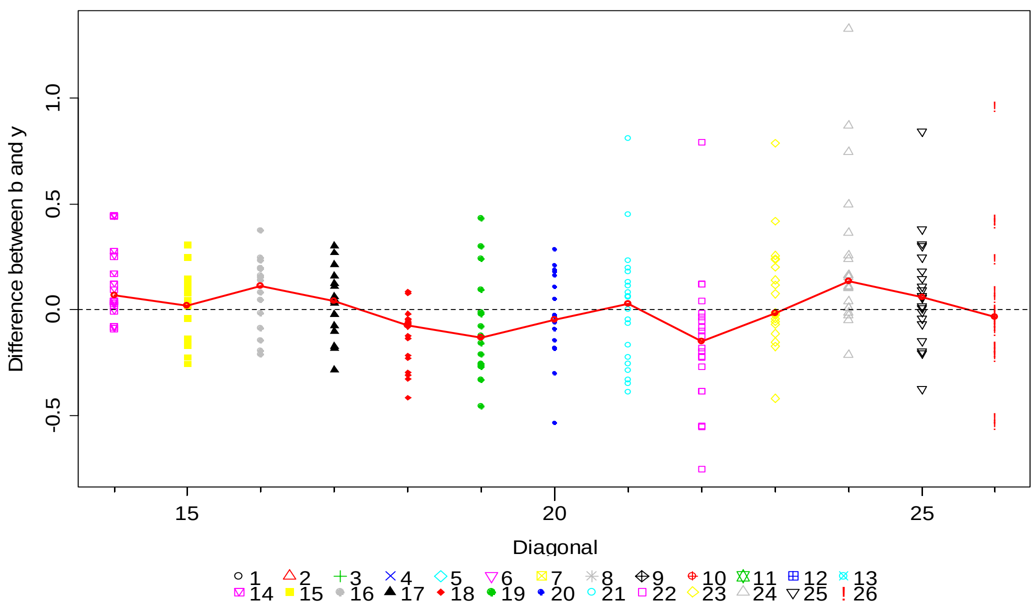


Chart 18: Medical: Difference between Actual Observations (*y*) and Estimated Process (*b*) by Diagonal (Calendar Year), Post-Reform

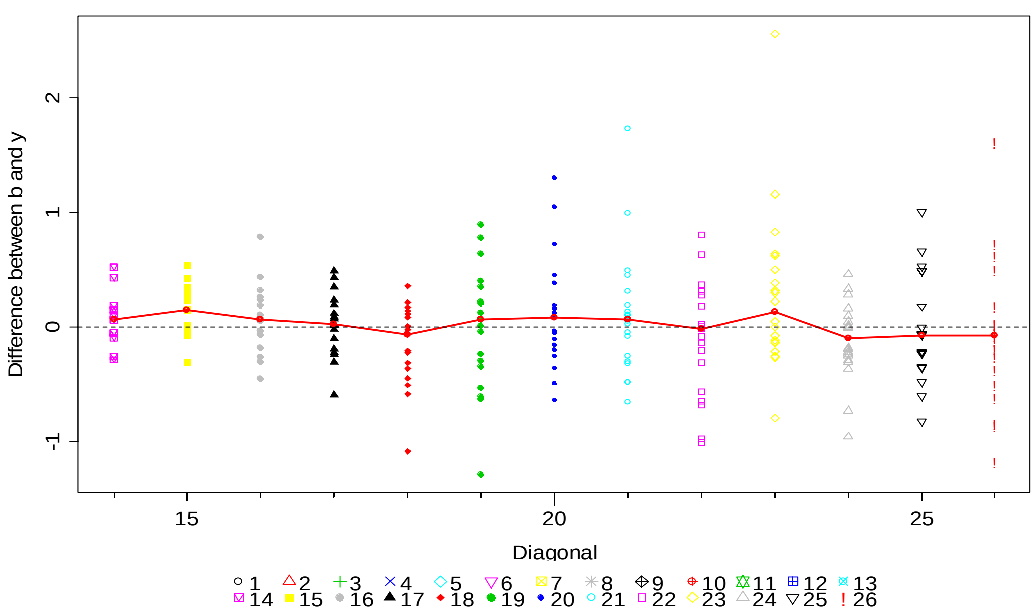


Chart 19: Indemnity: Actual Log Cumulative minus Predicted Log Cumulative Payments, Post-Reform

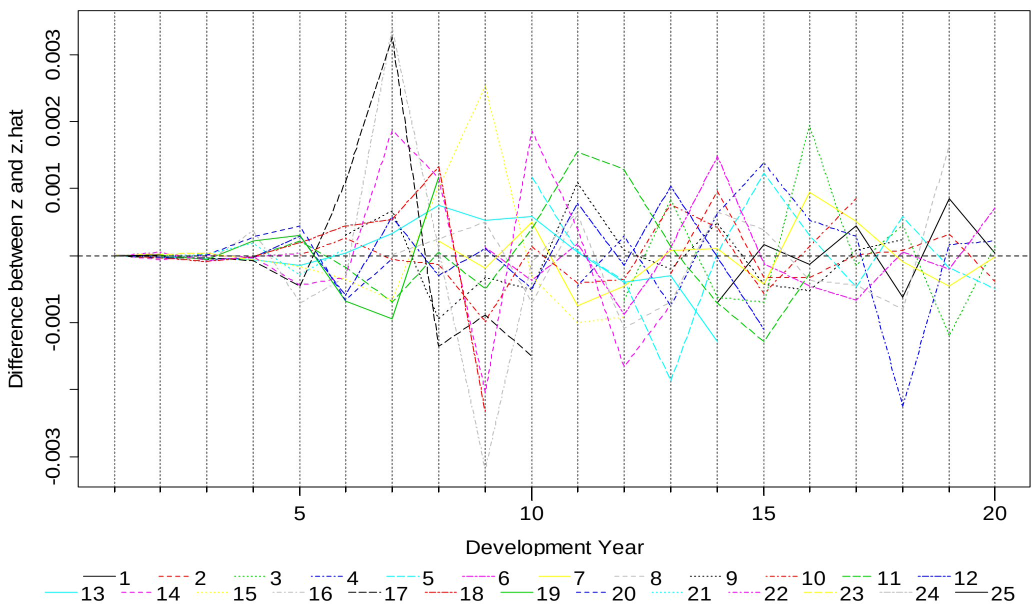
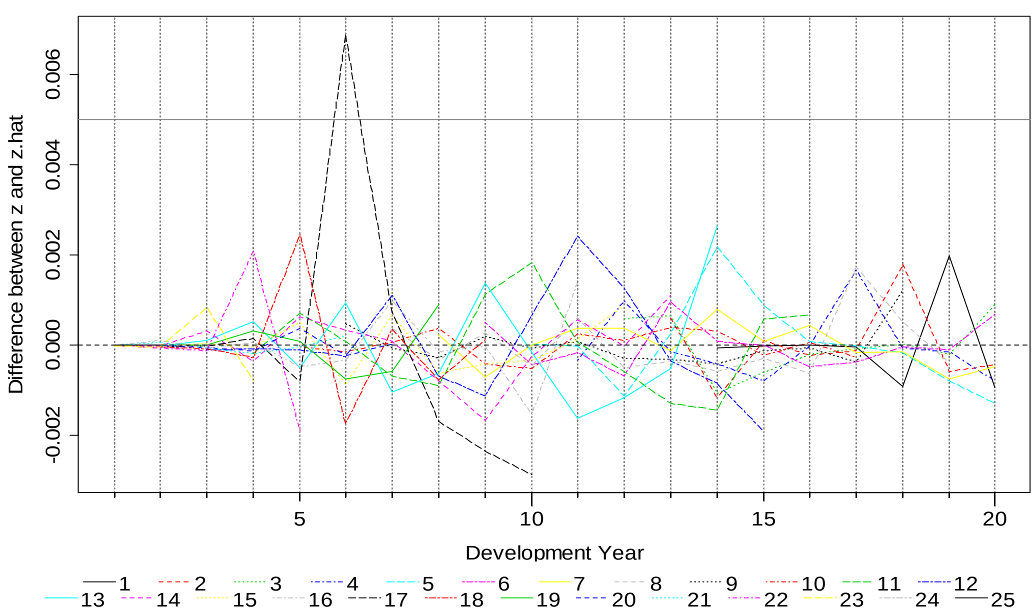


Chart 20: Medical: Actual Log Cumulative minus Predicted Log Cumulative Payments, Post-Reform



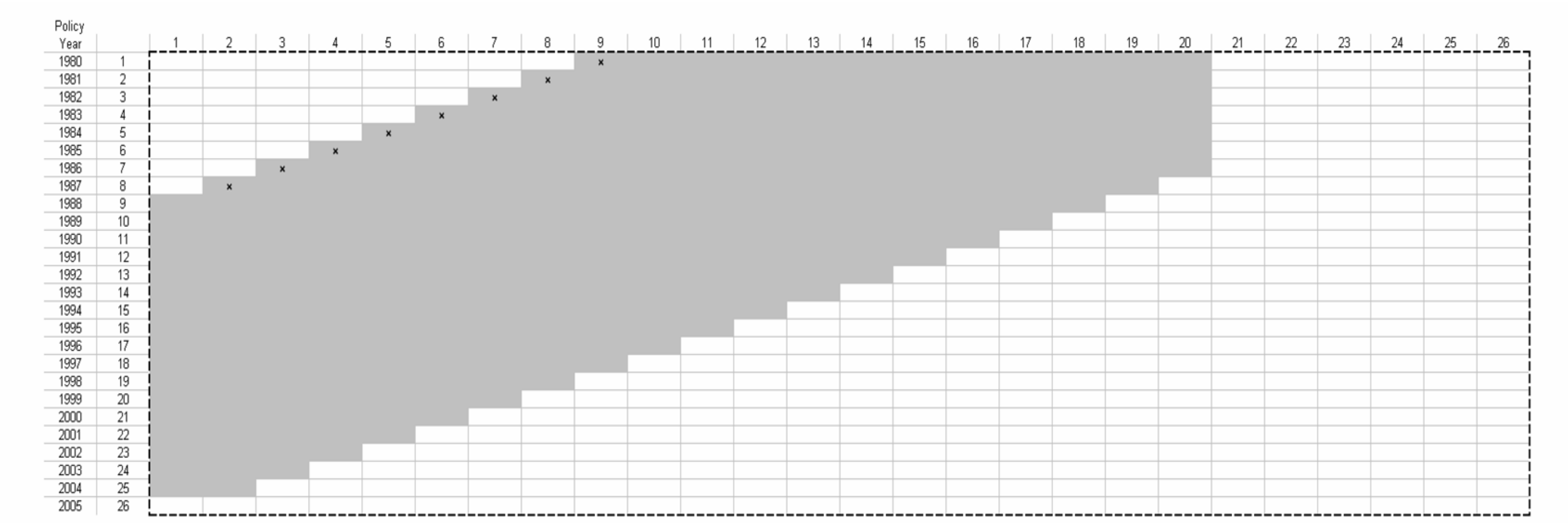
4. CONCLUSIONS

A loss development model has been presented that explicitly accounts for the legislative environment that applies to the time period during which the losses have been observed. Most importantly, the model accommodates changes in the legislative environment, which may be multi-faceted, having either diagonal (calendar year) or horizontal (policy year) effects (or both). The application of the model to an unidentified state demonstrates how, due to its high degree of flexibility, the model is capable of accommodating complex changes to loss development patterns. Further, the model is able to break out and quantify individual aspects of the legislative reform, such as calendar-year effects versus changes to the (calendar-year effect-adjusted) run-off.

Most interesting to the practicing actuary is the ability of the model to incorporate expert information as Bayesian priors in the estimation process. As shown, such expert priors may be legally stipulated rates of escalation (for indemnity) or information on medical price inflation at large (where more detailed information on the inflation embedded in medical benefits is unavailable).

Appendix 1

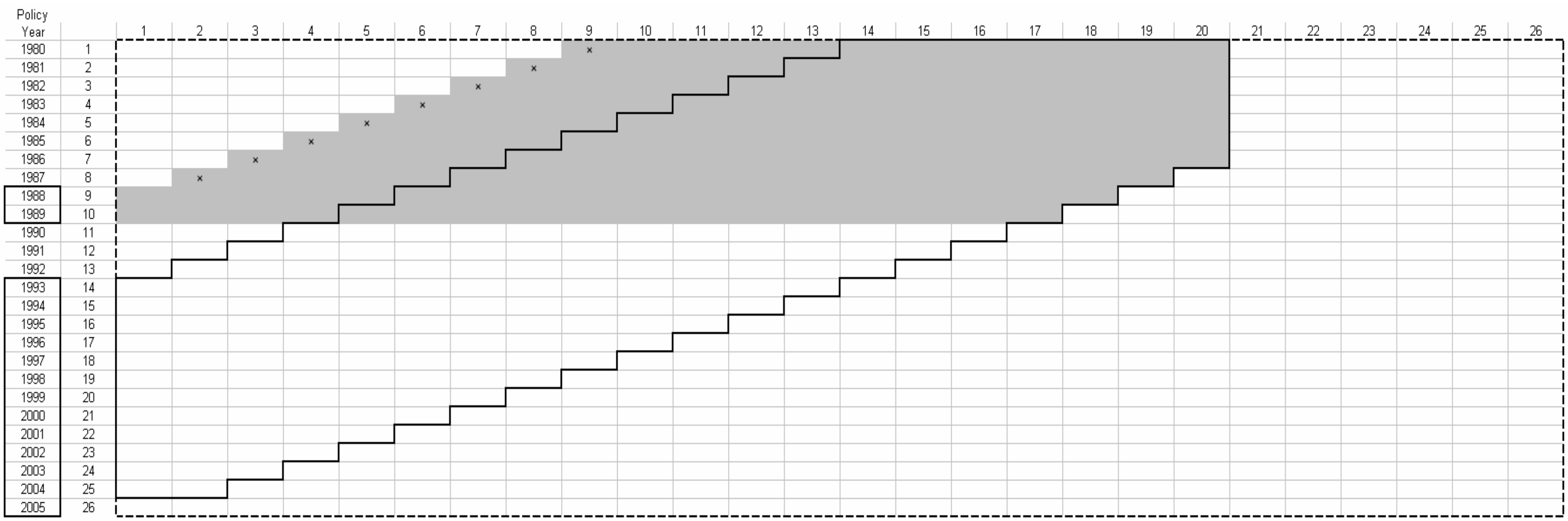
Exhibit A-1: Loss Triangle Template, Indemnity and Medical



Note: Available payments are shaded gray. For the cells marked by the symbol ×, only cumulative (but no incremental) payments are available.

Appendix 1, *cont.'d*

Exhibit A-2: Loss Triangle Template, Pre-Reform and Post-Reform



Note: The payments constituting the pre-reform triangle are shaded gray; the payments forming the post-reform triangle are framed by a solid line. For the cells marked by the symbol ×, only cumulative (but no incremental) payments are available.

Appendix 2: Pre-Reform Model (Model Type 9)

$$y_{i,j} \sim \mathrm{N}(b_{i,j,9}, \sigma_y^2) \begin{cases} i = 1,\ldots, rg-1,\ j = cg-i+2,\ldots, cf-i+1 \\ i = 2,\ldots, rf,\ j = cf-i+2,\ldots, cf \\ i = rf+1,\ldots, rh_9,\ j = cf-i+2,\ldots, c-i+1 \\ i = cf+2,\ldots, rh_9,\ j = 1,\ldots, c-i+1 \\ i = rg,\ldots, rh_9,\ j = 1,\ldots, cf-i+1 \end{cases} \tag{A2-1}$$

$$cs.mean_{i,j,9} \begin{cases} = \hat{z}_{i,j,9} - z_{i,j} \quad \text{for} \begin{cases} i = 1,\ldots, rg-1,\ j = cg-i+1,\ldots, cf-i+1 \\ i = 2,\ldots, rf,\ j = cf-i+2,\ldots, cf \\ i = rf+1,\ldots, rh_9,\ \ j = cf-i+2,\ldots, c-i+1 \\ i = cf+2,\ldots, rh_9,\ \ j = 1,\ldots, c-i+1 \\ i = rg,\ldots, rh_9,\ \ j = 1,\ldots, cf-i+1 \end{cases} \\ = mv_{\mu,i,j,9} \quad \text{for} \begin{cases} i = 1,\ldots, rg-1,\ j = 1,\ldots, cg-i \\ i = 1,\ldots, rf-1,\ j = cf+1,\ldots, c-i+1 \\ i = rg_9+1,\ldots, r,\ j = 1,\ldots, c-i+1 \\ i = 2,\ldots, r,\ j = c-i+2,\ldots, c \end{cases} \end{cases} \tag{A2-2}$$

$$\hat{y}_{i,j,9} = \mathrm{N}(b_{i,j,9}, \sigma_y^2)\ ,\ i, j = 1,\ldots, c \tag{A2-3}$$

$$\hat{z}_{i,j,9} = \log\left(\sum_{k=1}^{j} \exp\left(\hat{y}_{i,k,9}\right)\right)\ ,\ i, j = 1,\ldots, c \tag{A2-4}$$

$$\mathbf{mv}_9 \sim \mathrm{N}(\mathbf{mv}_{\mu,9}, \mathbf{\Omega}_1) \tag{A2-5}$$

$$\mathbf{mv}_{\mu,9} \sim \mathrm{N}(\mathbf{mv}_9, \mathbf{\Omega}_2) \tag{A2-6}$$

$$cs_{i,j,9} = 0\ ,\ i, j = 1,\ldots, c \tag{A2-7}$$

$$\mathbf{cs}_{i,9}' \sim \mathrm{N}(\mathbf{cs.mean}_9', \mathbf{\Omega}_1)\ ,\ i = 1,\ldots, r \tag{A2-8}$$

$$b_{r_h,1,9} \sim \mathrm{N}(y_{r_h,1,9}, \sigma_{b.init}^2) \tag{A2-9}$$

$$b_{i,1,9} = b_{i+1,1,9} - \Lambda_{i+1,1,9} \;,\; i = 1,\ldots,r_{h,9} - 1 \tag{A2-10}$$

$$\Lambda_{i+1,1,9} = \eta_{i+1} \;,\; i = 1,\ldots,r_{h,9} - 1 \tag{A2-11}$$

$$\eta_i \sim \mathrm{N}(\eta_{i-1}, \sigma_\eta^2) \;,\; i = 3,\ldots,r \tag{A2-12}$$

$$\eta_2 \sim \mathrm{N}(0, \sigma_\eta^2) \tag{A2-13}$$

$$b_{i,1,9} = b_{i-1,1,9} + \Lambda_{i,1,9} \;,\; i = r_{h,9} + 1,\ldots,r \tag{A2-14}$$

$$\Lambda_{i,1,9} = \eta_{i,1} \;,\; i = r_{h,9} + 1,\ldots,r \tag{A2-15}$$

$$b_{i,j,9} = b_{i,j-1,9} + \Lambda_{i,j,9} \;,\; i = 1,\ldots,r-1,\; j = 2,\ldots,c-i+1 \tag{A2-16}$$

$$\Lambda_{i,j,9} = \delta_{j,9.pre} + \kappa_{i,j} \;,\; i = 1,\ldots,r_{g,9} - 1,\; j = 2,\ldots,c_{g,9} - i + 1 \tag{A2-17}$$

$$\Lambda_{i,j,9} = \delta_{j,9} + \kappa_{i,j} \;,\; i = 1,\ldots,r_{g,9} - 1,\; j = c_{g,9} - i + 2,\ldots,c - i + 1 \tag{A2-18}$$

$$\Lambda_{i,j,9} = \delta_{j,9} + \kappa_{i,j} \;,\; i = r_{g,9},\ldots,r,\; j = 2,\ldots,c - i + 1 \tag{A2-19}$$

$$\delta_{j,9} \sim \mathrm{N}(\delta.prior_j, \sigma_{\delta,2}^2) \;,\; j = 2,3 \tag{A2-20}$$

$$\delta_{j,9} \sim \mathrm{N}(\delta_{j-1,9}, \sigma_{\delta,j}^2) \;,\; j = 4,\ldots,cf \tag{A2-21}$$

$$\delta_{j,9} \sim \mathrm{N}(\delta_{cf}, \sigma_{\delta,1}^2) \;,\; j = cf + 1,\ldots,c \tag{A2-22}$$

$$\sigma^2_{\delta,j} = \sigma^2_{\delta,1} \cdot 10^{-\alpha+\alpha \cdot e^{-\beta \cdot e^{-\gamma \cdot (j-1)}}} \text{ , } j = 4,...,cf;\ \alpha,\beta,\gamma > 0 \qquad \text{(A2-23)}$$

$$\sigma^2_{\delta,2} \text{ , } \sigma^2_{b.init} \text{ large} \qquad \text{(A2-24)}$$

$$\sigma^2_{\delta,1} \text{ small} \qquad \text{(A2-25)}$$

$$\kappa_{i,j} \sim \mathrm{N}(\mu_{i,j}, \sigma^2_{\kappa}) \text{ , } i = 1,...,r;\ j = 2,...,c \qquad \text{(A2-26)}$$

$$\mu_{i,j} = \lambda_{1,j} \cdot \pi_{1,i+j} + \lambda_{2,j} \cdot \pi_{2,i+j} \text{ , } i = 1,...,r;\ j = 2,...,c,\ \lambda_{1,j} + \lambda_{2,j} \le 1 \qquad \text{(A2-27)}$$

where y, and $\hat{y}$ are the observed and estimated logarithmic incremental payments, respectively. For negative incremental payments, the corresponding values of y are coded as missing values. The indexes i and j indicate policy (or accident) and development years, respectively; $r = c$ signifies the number of years in the loss triangle. The parameter c_f signifies the column with the final value for the cumulative (and incremental) payment in the first $r_f - 1$ rows, where the first $r_f - 1$ rows are those affected by the cut-off in reported development. The parameter c_g signifies the first column that has a value for the cumulative payment in the first row; note that the first incremental payment in this row is located in column $c_g + 1$. The parameter $r_g (= c_g)$ indicates the first row that has a value for the cumulative (and thus incremental) payment in the first column.

The parameter $r_{g,9} (= c_{g,9})$ indicates the row (column) with the first pre-reform incremental payment in the first column (row). If there was no structural break prior to the reform of interest, then $r_{g,9} (= c_{g,9}) = r_g (= c_g)$. Conversely, if there was such a possible structural break, then the parameter $r_{g,9} (= c_{g,9})$ indicates the first row (column) with an incremental payment in the first column (row) that belongs to the post-structural-break pre-reform period.

Equation (A2-1) fits the observations of the logarithmic incremental payments to a normal distribution. Equation (A2-2) defines the deviation of the estimated logarithm of the cumulative payment ($\hat{z}_{i,j,9}$, where the index 9 indicates pre-reform) in policy (or accident) year j and development year i from and the observed logarithm of the cumulative payment ($z_{i,j}$); this deviation is denoted $cs.mean_{i,j}$, where cs stands for cumulative sum. Equation (A2-3) simulates the predicted values of the logarithmic incremental payments; these predicted values feed into the estimated logarithmic cumulative payments in Equation (A2-4). Where such cumulative sum does not exist (to the right of the final diagonal, up to the final observed development year), $cs.mean_{i,j}$ is replaced by a draw from a multivariate distribution, $mv_{\mu,i,j,9}$, as shown in Equation (A2-6). Specifically, the row vector $\mathbf{cs.mean}_i$ comprises the differences between the predicted and observed logarithmic cumulative payments of row i for those columns for which observed logarithmic cumulative payments are available; for all other columns, the elements of $\mathbf{cs.mean}_i$ are taken from a vector of (expected) values that generates a multivariate normal distribution of the same variance as the one that $\mathbf{cs.mean}_i$ is fitted to. The covariance matrices $\mathbf{\Omega}_{1,2}^{-1}$ are modeled on Wishart distributions. Equations (A2-5) and (A2-6) generate a distribution the $mv_{\mu,i,j,9}$ can be drawn from; the distributions of the observed and the generated values of $\mathbf{cs.mean}_i$ share the same covariance matrix, $\mathbf{\Omega}_1^{-1}$. Equation (A2-7) stipulates that the observed differences between the logarithms of the observed and estimated cumulative payments be zero, on average. Equation (A2-8) represents the cumulative sum (cusum) constraint. This stochastic constraint ensures that, for every cell of the loss triangle, the sum of estimated incremental payments lines up (approximately) with the observed cumulative payment. The cusum constraint also serves as a means of interpolating between incremental payments when there is a missing value (due to a negative incremental payment).

Equation (A2-9) initializes for the upper-left hand side region (where no observations are available for the first incremental payment) the first logarithmic increment payment on the first logarithmic incremental payment of the first row for which such a payment is available (denoted as row r_h).

Equations (A2-10-), (A2-11), and (A2-14) through (A2-19) describe the process displayed in Exhibit 1. Equation (A2-12) describes the random walk of *eta*, and Equation (A2-13) its starting value. Equation (A2-21) describes the random walk of *delta*, and Equation (A2-20) describes how the first two values of delta are estimated before the random walk sets in, whereas Equation (A2-22) details how delta is extrapolated into the future after the random walk ends with the final observed development year. Equation (A2-23) describes a Gompertz function for the innovation variance of the random walk of *delta*; this innovation variance approaches the variance displayed in Equation (A-25). The variance for estimating the first two values of *delta* (that is, before the random walk sets in) is shown in Equation (A2-24). Finally, Equations (A2-26 and A2-27) detail how the calendar-year effect is estimated using an expert prior on the rate of escalation (indemnity) and inflation (medical).

The model has two layers of noise, which implies that there are two predicted values (for each observed value of incremental payment). First, there is the variable b, which aggregates the three processes (run-off in development, growth of expected value of first payment, and calendar-year effect). Second, there is the variable $y.hat$, which is a draw from a normal distribution, the expected value of which is b. Where there are no observations (the run-off triangle is squared, the tail is estimated, and future policy or accident years are forecast), the variable $y.hat$ corresponds to the expected value, b. The variable $y.hat$ gauges the ability of the model to replicate the observed incremental payments.

The variables $\pi_{i,j}$ $(i=1,2)$ are expert priors for (logarithmic) rates of inflation, which may vary by policy (or accident) and development years. (For policy years, the first prior in any given policy (or accident) year comprises inflation for a period of 18 months, this being the time difference between the mid-point of the initial 24 months of experience and the subsequent 12-month period.) The model accommodates two non-zero rates of inflation, differentiated by type of claim; this is important for indemnity claims (but irrelevant for medical claims). Thus, the prior for the calendar-year effect in any given development year, j, is a weighted average of three (one zero and two non-zero) expert rates of inflation, the weights being the fractions of dollars in incremental payments that apply to up to two differently inflating claim types in development year $j-1$, $\lambda_{k,j}$, $k=1,2$ (while a third claim type may inflate at a zero rate). If there is only one claim type (as is the case for medical claims) or all claim types escalate at the same rate, then $\pi_{2,j}$ and $\lambda_{2,j}$ equal zero for all j, and $\lambda_{1,j}$ equals 1 for all j.

Specifically, for indemnity, the expert prior for the (logarithmic) calendar-year effect equals the official (logarithmic) rate of inflation relevant to the cost-of-living adjustment, weighted by the fractions of incremental dollars that have been paid on escalating claims in the development year $j-1$, λ_j. The official rate of inflation pertinent to cost-of-living adjustment may be the rate of growth of the state-level average weekly wage (as measured by the Quarterly Census of Employment and Wages, QCEW, http://www.bls.gov) or the U.S. CPI (Consumer Price Index, http://www.bls.gov), depending on the applicable legislative provision; we apply an observation and implementation lag of 14 months. The expert inflation prior for medical benefits is the (contemporaneous logarithmic) rate of growth of the Medical Care component of the U.S. CPI.

The QCEW average weekly wage is calculated as the ratio of the total wage bill for the calendar year, summed up over four quarterly values, and then divided by the average employment for the calendar year; this average employment for the calendar year is calculated from 12 monthly numbers. The Medical Care component of the CPI is the published annual calendar year number.

It is important to note that the rate of growth of the expected value of the first incremental payment (η) is specified in nominal terms, which means that the rate of inflation is not broken out. As a consequence, the mentioned inflation modeling applies solely to the way the incremental payments inflate in development but has no bearing on the how the first incremental payment inflates from one policy (or accident) year to the next.

The chosen set of hyper-parameters of the prior distributions has been calibrated to incremental payments, the logarithm of which fall into the range of 7 to 11; the incremental (and cumulative) payments of the loss triangle that is to be analyzed have to be normalized accordingly. With such normalization, the chosen set of hyper-parameters accommodates any sufficiently well-behaved triangle. As a consequence, the final calibration of the model when applied to a loss triangle is done solely by choosing the three parameters of the Gompertz function, with one exception; this exception concerns the variance of the rate of growth of the expected value of the first payment, as exhibited in Equations (A2-12, 13). For triangles with a high degree of variation in the rate of growth of the first incremental payment (such as percentage point differences in the higher double digits), a larger variance is needed. Further, the parameters of the Gompertz function need to be chosen. This Gompertz function serves the purpose of smoothing the run-off rate δ by means of controlling the innovation variance of the random walk. The Gompertz function accommodates convex, concave, and "S"-shaped trajectories of this variance. The first parameter of the Gompertz function, α, determines the upper asymptote; the parameter β is (roughly) a horizontal shift parameter, and the parameter γ determines the rate of the growth (that is, the steepness and curvature). The choice of the parameters β and γ is ultimately a matter of judgment, especially for small triangles. Several diagnostic charts have been developed (as discussed in the body of the text) that assist in this choice.

Note that the pre-reform and post-reform models have all variances in common; further, the two models have a common calendar-year effect and common rates of growth of the expected value of the first payment. For all scalar variances in the model, there are gamma distributions used as priors.

Appendix 3: Post-Reform Model (Model Type 8)

$$y_{i,j} \sim \mathrm{N}(b_{i,j,8}, \sigma_y^2) \begin{cases} i=1,\ldots,rg_8-1,\ j=cg_8-i+1,\ldots,cf-i+1 \\ i=2,\ldots,rf,\ j=cf-i+2,\ldots,cf \\ i=rf+1,\ldots,cf+1,\ j=cf-i+2,\ldots,c-i+1 \\ i=cf+2,\ldots,rh_8,\ j=1,\ldots,c-i+1 \\ i=rg_8,\ldots,cf,\ j=1,\ldots,cf-i+1 \end{cases} \tag{A3-1}$$

$$cs_{i,j,8} \begin{cases} =\hat{z}_{i,j,8}-z_{i,j} & \text{for} \begin{cases} i=1,\ldots,rg_8-1,\ j=cg_8-i+1,\ldots,cf-i+1 \\ i=2,\ldots,rf,\ j=cf-i+2,\ldots,cf \\ i=rf+1,\ldots,cf+1,\ \ j=cf-i+2,\ldots,c-i+1 \\ i=cf+2,\ldots,rh_8,\ \ j=1,\ldots,c-i+1 \\ i=rg_8,\ldots,cf,\ \ j=1,\ldots,cf-i+1 \end{cases} \\ =mv_{\mu,i,j,8} & \text{for} \begin{cases} i=1,\ldots,rg_8-1,\ j=1,\ldots,cg_8-i \\ i=1,\ldots,rf-1,\ j=cf+1,\ldots,c-i+1 \\ i=rh_8+1,\ldots,r,\ j=1,\ldots,c-i+1 \\ i=2,\ldots,r,\ j=c-i+2,\ldots,c \end{cases} \end{cases} \tag{A3-2}$$

$$\hat{y}_{i,j,8} = \mathrm{N}(b_{i,j,8}, \sigma_y^2)\ ,\ i,j=1,\ldots,c \tag{A3-3}$$

$$\hat{z}_{i,j,8} = \log\left(\sum_{k=1}^{j} \exp\left(\hat{y}_{i,k,8}\right)\right)\ ,\ i,j=1,\ldots,c \tag{A3-4}$$

$$\mathbf{mv}_8 \sim \mathrm{N}(\mathbf{mv}_{\mu,8}, \mathbf{\Omega}_1) \tag{A3-5}$$

$$\mathbf{mv}_{\mu,8} \sim \mathrm{N}(\mathbf{mv}_8, \mathbf{\Omega}_2) \tag{A3-6}$$

$$cs_{i,j,8} = 0\ ,\ i,j=1,\ldots,c \tag{A3-7}$$

$$\mathbf{cs}_{i,8}' \sim \mathrm{N}(\mathbf{cs.mean}_8', \mathbf{\Omega}_1)\ ,\ i=1,\ldots,c \tag{A3-8}$$

$$b_{r_h,1,8} \sim \mathrm{N}(y_{r_h,1,8}, \sigma^2_{b.init}) \tag{A3-9}$$

$$b_{i,1,8} = b_{i+1,1,8} - \Lambda_{i+1,1,8} \ , \ i = 1,\ldots,r_{h,8} - 1 \tag{A3-10}$$

$$\Lambda_{i+1,1,8} = \eta_{i+1} \ , \ i = 1,\ldots,r_{h,8} - 1 \tag{A3-11}$$

$$b_{i,1,8} = b_{i-1,1,8} + \Lambda_{i,1,8} \ , \ i = r_{h,8} + 1,\ldots,r \tag{A3-12}$$

$$\Lambda_{i,1,8} = \eta_i \ , \ i = r_{h,8} + 1,\ldots,r \tag{A3-13}$$

$$b_{i,j,8} = b_{i,j-1,8} + \Lambda_{i,j,8} \ , \ i = 1,\ldots,r-1, \ j = 2,\ldots,c-i+1 \tag{A3-14}$$

$$\Lambda_{i,j,8} = \delta_{j,9.pre} + \kappa_{i,j} \ , \ i = 1,\ldots,r_{g,9} - 1, \ j = 2,\ldots,c_{g,9} - i + 1 \tag{A3-15}$$

$$\Lambda_{i,j,8} = \delta_{j,89} + \kappa_{i,j} \ , \ i = 1,\ldots,r_{g,9} - 1, \ j = c_{g,9} - i + 2,\ldots,c_{g,8} - i + 1 \tag{A3-16}$$

$$\Lambda_{i,j,8} = \delta_{j,89} + \kappa_{i,j} \ , \ i = r_{g,9},\ldots,r_{g,8} - 1, \ j = 2,\ldots,c_{g,8} - i + 1 \tag{A3-17}$$

$$\Lambda_{i,j,8} = \delta_{j,8} + \kappa_{i,j} \ , \ i = 1,\ldots,r_{g,8} - 1, \ j = c_{g,8} - i + 2,\ldots,c - i + 1 \tag{A3-18}$$

$$\Lambda_{i,j,8} = \delta_{j,8} + \kappa_{i,j} \ , \ i = r_{g,8},\ldots,r, \ j = 2,\ldots,c - i + 1 \tag{A3-19}$$

$$\delta_{j,8} \sim \mathrm{N}(\delta.prior_j, \sigma^2_{\delta,2}) \ , \ j = 2,3 \tag{A3-20}$$

$$\delta_{j,8} \sim \mathrm{N}(\delta_{j-1,8}, \sigma^2_{\delta,j}) \ , \ j = 4,\ldots,cf \tag{A3-21}$$

$$\delta_{j,8} \sim \mathrm{N}(\delta_{cf}, \sigma^2_{\delta,1}) \ , \ j = cf + 1,\ldots,c \tag{A3-22}$$

$$\delta.diff_j = \delta_{j,8f} - \delta_{j,9} \ , \ j = 4,\ldots,cf \tag{A3-23}$$

$$\delta.diff.z_j \sim \mathrm{N}(\delta.diff_j, \sigma^2_{\delta,j}) \ , \ j = 4,\ldots,cf \tag{A3-24}$$

$$\delta.diff.z_j = 0 \ , \ j = 4,\ldots,cf \tag{A3-25}$$

$$\kappa_{i,j} \sim \mathrm{N}(\mu_{i,j}, \sigma^2_{\kappa}) \ , \ i = 1,\ldots,r; \ j = 2,\ldots,c \tag{A3-26}$$

$$\mu_{i,j} = \lambda_{1,j} \cdot \pi_{1,i+j} + \lambda_{2,j} \cdot \pi_{2,i+j} \ , \ i = 1,\ldots,r; \ j = 2,\ldots,c, \ \lambda_{1,j} + \lambda_{2,j} \le 1 \tag{A3-27}$$

The parameter $r_{g,8} (= c_{g,8})$ indicates the row (column) of the first post-reform incremental payment in the first column (row). Equations (A3-23, 24, and 25) define the convergence constraint for the run-off rates of the pre-and post-reform triangles; this constraint becomes tighter as development progresses. Note that the pre-reform run-off rates of the post-reform triangle are allowed to differ from the run-off rates of the pre-reform triangle (except for the $\delta_{j,9.pre}$ area). For the definitions of the variables parameters, see Appendix 2. Further, see Appendix 4 for a complete list of variables.

Acknowledgment

Thanks to Carolyn Bergh and Jon Evans for comments, and to Chris Laws, Jose Ramos, and Manuel de la Guardia for research assistance.

5. REFERENCES

[1] De Jong, Piet, "Forecasting Runoff Triangles," *North American Actuarial Journal* 10, 28-38, 2006.

[2] De Jong, Piet, and Ben Zehnwirth, "Claims Reserving, State-Space Models and the Kalman Filter," *Journal of the Institute of Actuaries* 110, 157-181, 1983.

[3] England, Peter, and Richard Verrall, "Stochastic Claims Reserving in General Insurance," *Journal of the Institute of Actuaries* 129, 1-76, 2002.

[4] Gilks, Wally R., Sylvia Richardson, and David J. Spiegelhalter, *Markov Chain Monte Carlo in Practice,* (Boca Raton, FL: Chapman & Hall/CRC, 1996).

[5] Mack, Thomas, "Which Stochastic Model is Underlying the Chain Ladder Method?" *Insurance, Mathematics and Economics* 15(2/3), 133-138, 1994.

[6] Scollnik, David, "Bayesian Reserving Models Inspired by Chain Ladder Methods and Implemented Using WinBUGS," http://math.ucalgary.ca/~scollnik/balducci/Papers/resnaaj.pdf, April 2004.

[7] Scollnik, David, "Implementation of Four Models for Outstanding Liabilities in WinBUGS: A discussion of a paper by Ntzoufras and Dellaportas," *North American Actuarial Journal* 6(1):128-136, 2002.

[8] Taylor, Gregory, *Loss Reserving: An Actuarial Perspective*, (Boston: Kluwer, 2000).

[9] Verrall, Richard, "Modeling Claims Run-Off Triangles with Two-Dimensional Time Series," *Scandinavian Actuarial Journal* 129-138, 1989.

Abbreviations and notations

BUGS, Bayesian inference Using the Gibbs Sampler
CPI, Consumer Price Index
MCMC, Markov Chain Monte Carlo (Simulation)
M-CPI, Medical Care Component of the CPI
NCCI, National Council on Compensation Insurance
QCEW, Quarterly Census of Employment and Wages
PP, Permanent Partial (Claims)
PT, Permanent Total (Claims)
TT, Temporary Total (Claims)

Biography of the Author

Frank Schmid, Dr. habil., is a Director and Senior Economist at the National Council on Compensation Insurance.

Loss Development in Workers Compensation in the Presence of Legislative Reform

Frank Schmid

Motivation. Legislative reforms affect loss development patterns in various ways. Some legislative innovations may affect new policy (or accident) years only, while others have diagonal effects as they affect both new and existing claims. Modeling these effects is critical for adequacy in ratemaking and reserving.
Method. Using a Bayesian state-space model, workers compensation triangles are developed subject to the applicable legislative stipulations. Most importantly, this model is capable of accommodating the legislative environment as it evolves over time.
Results. The model is applied to an unidentified state, which experienced a reform cluster in the period 1990/92. The model shows how this reform cluster affects the ultimate loss and the 19th-to-ultimate tail factors.
Conclusions. Ultimate losses are not only dependent on the legislative environment at time of loss, but are also affected by how this legislative environment evolves over time. The statistical model is capable of quantifying the effects of such legislative changes on the loss development pattern.
Availability. The model runs in OpenBUGS 2.2.0 (http://mathstat.helsinki.fi/openbugs/) within the R (www.r-project.org) package BRugs 0.3-3 (http://cran.r-project.org). OpenBUGS is administered by the Department of Mathematics and Statistics of the University of Helsinki, Finland; R is administered by the Technical University of Vienna, Austria. OpenBUGS and R are GNU projects of the Free Software Foundation and, hence, available free of charge.

Keywords. Workers Compensation, Trend and Loss Development, Reserving Methods.

1. INTRODUCTION

Workers compensation is a line of insurance that operates in a legal environment that is subject to frequent and (sometimes) sweeping changes. Such legislative changes affect the loss development patterns in ratemaking and reserving in powerful and complex ways. Traditional loss development models do not acknowledge the specific legal environment in which the losses have been observed, nor are these models capable of incorporating changes in the legal setting into the loss development pattern; as a consequence, these models are not capable of quantifying the impacts of changes to legal stipulations on the ultimate loss or tail factor.

What follows is a Bayesian state-space model of loss development that explicitly accounts for the legal environment in which the losses of a given (indemnity or medical) triangle were observed. Accounting for the legal environment means translating legal stipulations into data inputs, which are then fed into the model. The model is set up to accommodate a wide array of legal changes, among which are changes to the stipulated rates of escalation (for indemnity) and (any) factors that bear on the rate at which incremental payments decay in development net of the calendar-year effect.

1.1 Research Context

A wide array of loss development models have been suggested, some of which are extensions of traditional actuarial methods (mostly related to the chain ladder; see, for instance, Mack [5]), while other models cast loss development into a time series framework (see, for instance, de Jong and Zehnwirth [2], de Jong [1], and Verrall [9]). For overviews on loss development models, see England and Verrall [3] and Taylor [8]. Bayesian modeling of loss development using the software platform BUGS (Bayesian inference Using the Gibbs Sampler) has been pioneered by Scollnik [6,7]. The model presented here draws on Scollnik [7].

1.2 Objective

The objective of the loss development model introduced in this paper is to give the practicing actuary a framework for developing losses in a changing legal environment. By acknowledging changes in pertinent legal stipulations, the model is capable of delivering values for the ultimate loss (and, hence, the tail factor) that are adequate for ratemaking and reserving. Specifically, the model allows for quantification of reform impacts on the ultimate loss and the tail factor.

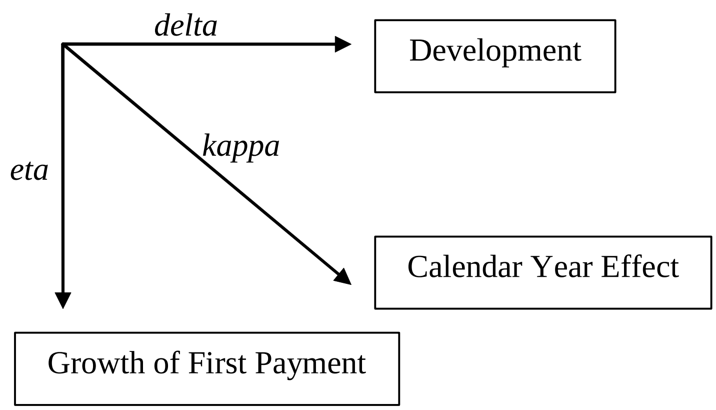
1.3 Outline

The following section first outlines the basic structure of the Bayesian state-space model of loss development and then, in a sub-section, applies this model to an unidentified state. This application details how regulatory information is fed into the model and how the model quantifies the reform impact. Section 3 presents the results of this empirical analysis. Section 4 offers conclusions.

2. BACKGROUND AND METHODS

The Bayesian state-space model of loss development employed in the analysis of legislative reform treats incremental payments as a three-dimensional time series problem. Specifically, the incremental payments are driven by three time processes, which are growth of the first payment, development, and the calendar-year effect; these processes are illustrated in Exhibit 1.

Exhibit 1: Time Processes in Loss Development



The model fits to (the logarithms of) incremental payments and, at the same time, employs a stochastic cumulative sum (cusum) constraint to ensure that, for any development year, the sum of the estimated incremental payments for a given policy (or accident) year add up (approximately) to the observed cumulative payment for that policy (or accident) year.

As an example, consider the stylized triangle displayed in Exhibit 2. Let $y[i, j]$ be the (natural) logarithm of the incremental payment of policy (or accident) year i in development year j, which materializes as a draw from a normal distribution with expected value $b[i, j]$. Then, the expected value of the logarithm of the first payment in the first policy (or accident) year, $b[1,1]$, develops into $b[1,2] = b[1,1] + delta[2] + kappa[1,2]$, where $b[1,2]$ is the expected value of the logarithm of the second incremental payment in the first policy (or accident) year. The parameter $delta[2]$ is the rate of decay (which is expressed as a logarithmic rate of growth) of the calendar-year effect-adjusted incremental payments from development year 1 to development year 2, whereas the term $kappa[1,2]$ is the calendar-year effect (which, again, is expressed as a logarithmic rate of growth) from calendar year 1 to calendar year 2. Note that the calendar-year effect is not restricted to be uniform along a given diagonal—for instance, $kappa[2,3]$ is allowed to differ from $kappa[3,2]$; this is because different types of indemnity claims (which consist of Temporary Total [TT], Permanent Partials [PP], Permanent Totals [PT], and Fatals) may escalate at different rates and the fraction of the various types in the total may change across development years. Finally, for the expected value of the logarithm of the first payment in the second policy (or accident) year, $b[2,1]$, we can write

$b[2,1] = b[1,1] + eta[2]$, where $eta[2]$ (which is again expressed as a logarithmic rate of growth) equals the change in expected values.

Exhibit 2: Stylized Triangle

Policy (or Accident) Year	Development Year: 1	2	3
1	y[1,1]	y[1,2]	y[1,3]
2	y[2,1]	y[2,2]	
3	y[3,1]		

The run-off rate (*delta*) is estimated using a smoothed random walk specification; the smoothing is obtained by scaling the innovation variance with a Gompertz function. The rate of growth of the expected value of the first incremental payment (*eta*) is also estimated using a smoothed random walk; unlike the innovation variance of the run-off rate (which decreases as development progresses), the innovation variance of *eta* is constant. (The smaller the innovation variance, the smoother is the estimated trajectory of growth rates.)

The model draws on expert information in determining the prior for the calendar-year effect, which manifests itself in the growth rate *kappa*. For indemnity benefits inflation, these expert priors are the rates of escalation as stipulated in the law; these stipulated rates of escalation may vary by type of claim. Additionally, the expert priors for the rates of escalation may vary by policy (or accident) years and development years. The expert prior for medical benefits inflation is the rate of growth of the Medical Care component of the CPI (Consumer Price Index; www.bls.gov), M-CPI for short.

The model develops future losses subject to the assumption that the expert priors for the (non-constant) rates of inflation follow random walks, starting at the final observed rates. The purpose of these random walks is to incorporate uncertainty about the future rates of inflation. The innovation variances of these random walks have to be determined by an expert based on the actual behavior of the applicable inflation series. Due to the skewed, lognormal distribution of the incremental payments, greater uncertainty about future rates of inflation (that is, greater innovation variances in the random walks) implies higher expected values of incremental payments and, all else being equal, a larger tail factor.

The model assumes that beyond the final observed development year, the projected run-off rate is the minimum of the final estimated run-off rate (that is, the run-off rate that applies in the final observed development year) and a mortality-based run-off rate. Starting with the final estimated logarithmic run-off rate, this logarithmic mortality-based run-off rate decreases linearly in every development year such that in development year 60, this rate equals the current official (logarithmic) mortality rate for age 80. Beyond age 80 (development years 61 through 70), the (logarithmic) mortality-based run-off rate equals the (logarithm of the) official mortality rate for the applicable age. The mortality information originates from the Social Security Administration (Periodic Life Table, www.ssa.gov). Where indemnity benefits are not granted for life (due to an age limit or an otherwise stipulated restriction in the duration of benefits), the number of payments is reduced accordingly, as detailed in the following section.

For details on the model, see Appendixes 2 and 3; Appendix 4 offers a list of variables. The model was estimated using Markov chain Monte Carlo simulation; for introduction to this estimation technique see, for instance, Gilks, Richardson, and Spiegelhalter [4]. The equations were coded in BUGS and run in R (using BRugs [Version 0.3-3, which utilizes OpenBUGS 2.2.0 beta from February 2006]) with a burn-in of 40,000 iterations, followed by a sample of 40,000 iterations, of which every fourth draw entered the posterior distribution (to mitigate autocorrelation in the Markov chains).

2.1 The Reform Impact of an Unidentified State

This section presents an application of the loss development model for the purpose of studying the impact of legislative reform on the loss development pattern, and the tail factor in particular. The model is applied to a loss triangle of policy year data; the first report of payments of any given policy year comprises 24 months of experience. The policy years in the loss triangle range from 1980 through 2005. The triangle, which is displayed in Exhibit A-1 in Appendix 1, is incomplete due to a missing upper left-hand side triangle, a missing upper right-hand side triangle, and a missing lower left-hand side (single-observation) triangle.

The purpose of the analysis is to study the reform impact in an unidentified state; this state experienced major reforms in workers compensation in the years 1982, 1986, 1990 (effective

September 1), and 1992 (indemnity-related reforms effective May 18, and medical-related benefits reform effective November 1). The 1982 and 1986 reforms are not broken out because the first diagonal in the triangle refers to the year 1988. The reform impact of interest is the one of the 1990/1992 reform cluster; for this purpose, we define the time window 1988-1989 as the pre-reform period, and the window 1993-2005 as the post-reform period. Four of the most significant impacts of the 1990/1992 reforms were (1) the introduction of escalation of indemnity benefits at the rate of the CPI (regardless of the date of the injury) for PT disability claims and PP disability claims in May 1991 (beyond 312 weeks of benefits; indemnity benefits for fatal claims had been escalating at a fixed rate of 4 percent since June 1986); (2) a limitation of the duration of TT disability claims to 52 weeks; (3) closer scrutiny regarding continued eligibility of indemnity benefits; and (4) an indemnity retirement offset that is immediate for accidents past age 55 or, otherwise, sets in five years prior to the official retirement age. Whereas the introduction of a cost-of-living adjustment is captured in the model as a calendar-year effect (as such adjustment started applying to claims of any maturity), the time limitation on TT claims, the increased scrutiny regarding continued eligibility, and the social security offset can be expected to bear on the run-off rate (*delta*). The run-off rate (*delta*) picks up the effect of a social security offset to the extent that such offset kicks in for (older) claimants within the first 20 development years (as these are the development years covered by the data). Yet, because the social security offset may not be fully captured by the run-off rate (due to there being [younger] claimants for whom the offset does not kick in within the 20 observed development years), the model assumes (as an approximation) a 50 percent reduction of the incremental indemnity payments past development year 40. Note that the increased scrutiny regarding continued eligibility of indemnity benefits may spill over into medical benefits, thus causing medical claims to close faster. Hence, we expect the 1990/1992 reform cluster to lead to a faster run-off not only in indemnity but also in medical incremental payments. (Note that although the most significant impacts of the 1990/1992 reform cluster were the indemnity reforms mentioned above, the 1990/1992 reform cluster also included a medical reform in November 1992, as mentioned above.)

Exhibit A-2 in Appendix 1 details the shapes of the pre-reform and post-reform triangles. The area of the pre-reform triangle for which there is data is shaded gray; this area comprises all observations between (and inclusive of) the 1988 diagonal and the 1989 policy year. The post-reform triangle is bordered by a solid line and consists solely of post-1992 diagonals. Note that the model does not fit to (the six) observations between (and inclusive of) the 1990 policy year and the 1992 diagonal, although these observations are assigned to the pre-reform period for the purpose of the post-reform estimation. The pre-reform and post-reform loss development processes are estimated simultaneously. The missing upper left-hand side triangle (diagonals 1980 through 1987) is given its own trajectory of run-off rates, which is the same for both the pre-reform and the post-reform estimation. Finally, for the post-reform estimation, the run-off rates that apply to the diagonals from 1988 through 1992 are allowed to differ from the estimated pre-reform run-off rates.

Although the pre-reform triangle consists only of policy-year data prior to 1990, this triangle includes elements through the 2005 diagonal. To the degree that the 1990/1992 reform cluster affected existing (instead of only new) claims (for instance, by accelerating their closure), the model may underestimate the impact of the reform cluster on the ultimate loss; however, the post-reform ultimate losses (and tail factors) would still be accurate, as argued below. For the data set at hand, the pertinent (future) policy year for ratemaking is 2008.

Unlike the pre-reform triangle, the post-reform triangle consists only of diagonals observed in the pertinent legislative environment. Yet, only in the first column of the post-reform triangle do all observations fall into the post-reform regime. As development time increases, the post-reform triangle phases in observations from the previous legislative setting, as indicated by the step function that defines the post-reform triangle in Exhibit A-2. For instance, in the first development year, all 13 incremental payments (of which the one for policy year 2005 is missing) are from the post-reform period. In the second development year, there are again 13 incremental payments (of which none are missing), but only 12 originate in the post-reform regime; and so on. The progressive phasing in of observations from the prior legislative regime rests on the premise that the run-off rates (but not necessarily the level of payments) of the post-reform regime approach the pre-reform run-off rates as development time advances; this is because the rates of decline of calendar-year effect-adjusted incremental payments deep in development may predominantly be driven by factors immune to the reforms, such as mortality. (It is because the reform may affect the level of payments deep in development [due to its effect on the run-off rates early in development] that the pre-reform run-off rates in the post-reform estimation are allowed to differ from the pre-reform run-off rates in the pre-reform estimation.) If the run-off rates (of the pre-reform policy years) deep in development are indeed immune to the reform, then the model estimates accurately both the pre-reform and post-reform ultimate losses. If, on the other hand, the run-off rates (of the pre-reform policy years) deep in development are affected by the reform, then the model underestimates the reform impact (but still estimates the post-reform ultimate loss accurately because it is the post-reform development pattern that materializes in the post-reform diagonals). But then there is a third situation where the model is not able to quantify the post-reform ultimate loss (as well as the impact of the reform). Such situation arises when the reform affects the run-off deep in development of new claims only, as is the case when a second-injury fund is eliminated. Because the reform takes many years to play out in the data (that is, manifest itself in incremental payments of new claims deep in development), the model is incapable of quantifying such reform impact immediately.

When estimating the loss development model, the pre-reform and post-reform triangles are estimated simultaneously, subject to the constraint that the two triangles have identical calendar-year effects, identical rates of growth of the expected value of the logarithm of the first payment, and identical variances in the measurement equations of the incremental payments. For details on the model, see Appendices 2 and 3.

3. RESULTS

Odd-numbered charts exhibit the indemnity results, whereas even-numbered charts display the results for medical.

Charts 1 and 2 show the indemnity and medical benefits estimated run-off rates (*delta*) along the development year axis—remember that the run-off rates are the rates of growth of the incremental payments, adjusted for the calendar-year effect. As mentioned, the run-off rates beyond the final observed year of development incorporate mortality information. Whereas the displayed run-off rates for medical benefits (Chart 2) describe the trajectory of the run-off rate as employed in the computation of the ultimate loss (and, hence, the tail factor), the run-off trajectory of the incremental indemnity payments (Chart 1) needs adjustment before inputting it into the computation of the ultimate loss or the tail factor; this is because indemnity benefits may not be granted for life, or there may be a social security offset. (If there is an immediate social security offset that applies regardless of the age of the claimant, then such offset is captured by the trajectory of the run-off rate *delta*.) In the unidentified state in question, effective May 1992, a social security offset applies to accidents that happen past age 55 or within five years of the legal retirement age. As a result of this legislative change, the incremental payments for development years 41 through 70 (70 being the final development year) were reduced by 50 percent of what would be projected otherwise.

Charts 3 and 4 present the expert priors (lines with full circles) and the posteriors for the calendar-year effect in the second development year (which is the first year of escalation). Due to this being a policy year triangle, the prior in the displayed second development year comprises 18 months of inflation (which is the time difference between the mid-points of the first 24 months of experience and the subsequent 12 months of experience). Note that, in general, a systematic difference between the expert prior for the calendar-year effect and the (unknown) workers compensation-specific rate of inflation factors into the run-off rate *delta*. Specifically, if for all incremental payments the actual (logarithmic) rate of benefits inflation exceeds the expert prior by a constant *c* (which may be positive or negative), then such constant will be absorbed by the rate of decay (*delta*) of the calendar-year effect-adjusted incremental payments—in statistical terms, the parameter *c* is unidentified.

Whereas the prior for medical inflation (Chart 4) is the M-CPI for all policy years, the prior for indemnity escalation (Chart 3) is a weighted average of the legally stipulated rates of escalation (of which the model accommodates two non-zero rates of escalation in addition to the zero rate [no escalation]). For instance, for the second development year, the rate of escalation that applies to a given type of claim (for a given policy or accident year) is weighted by the fraction of (incremental) losses associated with the given type of claim in the first development year. (Note that the fraction of incremental losses that applies to a given type of claim for a given development year is held constant for every policy (or accident) year, as such information is not available for every single policy or accident year.) Before policy year 1984, there was no escalation of indemnity benefits. Then, in policy year 1984, the escalation of fatal claims (at 4 percent), as introduced in June 1986, shows up in the prior (to the extent that this policy year was affected by the legislative change). The weight of such escalation increased in policy year 1985 before reaching (in policy year 1986) the level that corresponds to the fraction of Fatal (incremental) losses in the first development year. This level of escalation then rose again in policy year 1989 when in May 1991 PT claims started escalating at the CPI rate of inflation. This escalation of PT claims reached its full weight (at the fraction of PT incremental losses in the first development year) in policy year 1991. Note that because CPI inflation varies over time, the expert prior for the escalation of indemnity claims shows time variation even after 1991 (as indicated by the slight bumps in the applicable line in Chart 3).

Charts 5 and 6 displays the priors (lines with full circles) and the posteriors for the calendar-year effect of the latest observed diagonal; remember that there are no observations available for the final six values of the latest diagonal, which is why for these values the posterior equals the prior. Again, note that the first value on the diagonal comprises 18 months of inflation. For medical benefits, the expert prior for the calendar-year effect (which is the M-CPI; Chart 6) is uniform along the diagonal, except for the first value, which comprises inflation of a longer time period. For indemnity benefits, the expert prior for the rate of escalation (as determined by the pertinent legal stipulations) varies along the diagonal (beyond the initial change caused by switching from 18 months of inflation to 12 months of inflation); this is because diagonals span several development years. As a given set of indemnity claims develops, the proportions of incremental payments going to the various claim types (TT, PP, PT, and fatal) change; if these claim types escalate at different rates, then the expert prior for the escalation of the total of incremental payments within a given calendar year (diagonal) varies by development year. As mentioned, fatal claims escalate at four percent and PT claims escalate at the rate of CPI inflation; because the fraction of these claims is small in the first development year, the expert prior for the rate of escalation embedded in the total incremental payments in the second development year is close to zero. As development progresses, the fractions of incremental payments that apply to these two types of claim increases, as indicated by the rising line (full circles) in Chart 5 for development years 2 through 6. After 312 weeks of benefits, PP claims start escalating at the rate of CPI inflation. With TT claims having expired (or technically behaving like PT or PP claims), all claims escalate from development year 6 onward. (Fatal claims keep escalating at the stipulated four percent, whereas all other claims escalate at the CPI rate of inflation.)

Charts 7 and 8 show for $1 of initial (that is, first report) payment, kernel density estimates for the impact of the reform-induced change in the run-off rate (*delta*) on the ultimate loss for (the future) policy year 2008; remember that the first year comprises 24 months of development. Note that the payments are adjusted for the calendar-year effect; otherwise, studying the reform-induced difference in the ultimate loss would require choosing a specific pre-reform reference year (because of the time variation of the rate of inflation). Breaking out the reform impact on medical benefits is straightforward as for medical benefits, legislative reforms generally feed into the run-off rate *delta*. (Remember that any systematic difference between the workers compensation-specific medical inflation and M-CPI inflation are captured by the run-off rate *delta*; hence, any changes to the difference between these two inflation rates will be reflected in changes to *delta*.) Breaking out the reform impact on the ultimate loss of indemnity is more demanding than isolating such impact on the ultimate loss of medical; this is because legislative changes may not only change the run-off rate but also affect the stipulated rate of escalation, age limit for benefits, duration of benefits, or social security offset. The reform impact on the ultimate loss in indemnity, as depicted in Chart 7, is adjusted for the calendar-year effect, which means that the legislative changes to the applicable rates of escalation are not captured. Of course, the impact of the change in escalation can be broken out as well, but this requires choosing a specific reference year, as the CPI rate of inflation varies over time. (Alternatively, the ultimate losses of the various policy years [per $1 of initial payment] could be presented in a chart similar to Charts 9 and 10, which display the tail factors by policy year, while fully accounting for reform impacts.) As mentioned, to the extent that the 1990/92 reform cluster led to faster closing of existing claims and this way affected the run-off rates of post-1992 diagonals for pre-1990 policy years, the reform impact displayed in Charts 7 and 8 may be understated; this is because, even though the post-reform losses are accurately estimated, the "as-if-pre-reform" post-reform ultimate losses may be understated. Most interestingly, Chart 8 shows that the 1990/92 reform cluster indeed reduced the ultimate loss for medical (per $1 of initial payment), thus pointing to a faster run-off of medical payments due to increased scrutiny regarding continued eligibility for indemnity payments. As mentioned, the 1990/1992 reform cluster pertained mainly to indemnity benefits, but there was also a medical benefits reform, which occurred in November 1992.

Charts 9 and 10 exhibit the 19th-to-ultimate tail factors, differentiated by pre-reform and post-reform period; the post-reform period includes the future policy year (2008) of interest to ratemaking. The displayed tail factors rest on two alternative concepts. The first concept ("Tail Factors Based on b") computes the tail factors based on the estimated data-generating process. The second concept ("Tail Factors Based on *y.hat*") computes the tail factor based on the estimated incremental payments. Generally, for future policy (or accident) years, depending on the case, the two concepts generate the same number. The tail factor trajectory that stretches over the entire analyzed time period offers actual pre-reform tail factors (to the left of the left-most vertical separator) and "as-if-pre-reform" post-reform tail factors (to the right of the right-most vertical separator). The vertical differences between the "as-if-pre-reform" post-reform tail factors and the actual post-reform tail factors gauge the (full) reform impact. As argued above, to the extent that the reform cluster affected post-1992 diagonals for pre-1990 policy years, the "as-if-pre-reform" post reform tail factor may be understated.

Charts 11 and 12 offer a demonstration of how sensitive tail factors are to the rate of inflation that applies to the pertinent future policy year 2008. For indemnity, this rate of inflation is the rate of growth of the CPI, which is the (post-reform) stipulated rate of escalation for PP claims (after 312 weeks of benefits) and PT claims; the rate of escalation of fatal claims is kept at four percent. For medical, the rate of inflation is the M-CPI. Note that, due to the convexity of the tail factor in the rate of inflation, greater variability in the rate of inflation entails larger tail factors when averaged across policy years.

Charts 13 through 18 are diagnostic tools. These charts gauge how well the model has been calibrated; they display by policy year (Charts 13 and 14), development year (Charts 15 and 16) and calendar year (Charts 17 and 18) the difference between the log incremental payments predicted by the data-generating process (*b*) and the actual log incremental payments (*y*); the solid line indicates the median difference. Early in development, the solid lines in Charts 15 and 16 must be close to zero; late in development, these lines may turn jagged as outliers (in the percentage difference between observed and predicted payments) become more likely. The diagnostic Charts 13 through 18 signify that the model is well calibrated (as the median differences [solid lines] show no persistent departure from the zero line); in particular, the calendar-year effect (Charts 17 and 18) is properly captured.

Charts 19 and 20 are another set of diagnostic tools. These charts inform about data outliers and may serve as data quality indicators. The charts display by policy year the difference between the actual log cumulative payments (*z*) and the fitted log cumulative payments (*z.hat*) along the development year time axis. Based on experience, values within the interval (-0.005; 0.005) indicate that the model is able to replicate the underlying data. Values outside this interval but within the interval (-0.01; 0.01) have to be considered outliers. Values outside the interval (-0.01; 0.01) must be considered data points of poor quality.

Chart 1: Indemnity: Trajectory for delta (Run-off Rate, Calendar-Year Effect-Adjusted); "9": Pre-Reform; "8": Post-Reform

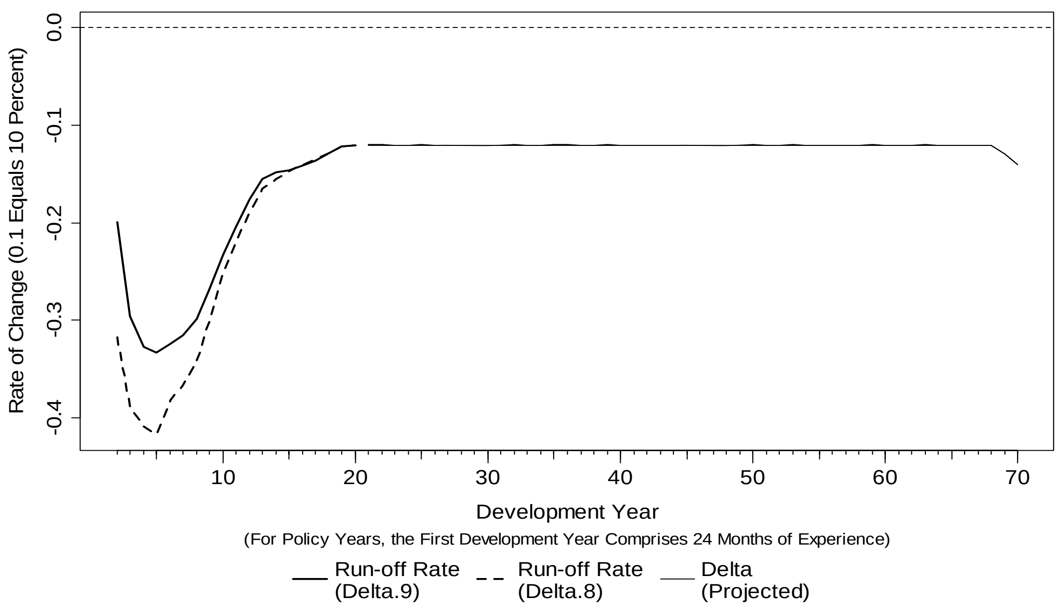


Chart 2: Medical: Trajectory for delta (Run-off Rate, Calendar-Year Effect-Adjusted); "9": Pre-Reform; "8": Post-Reform

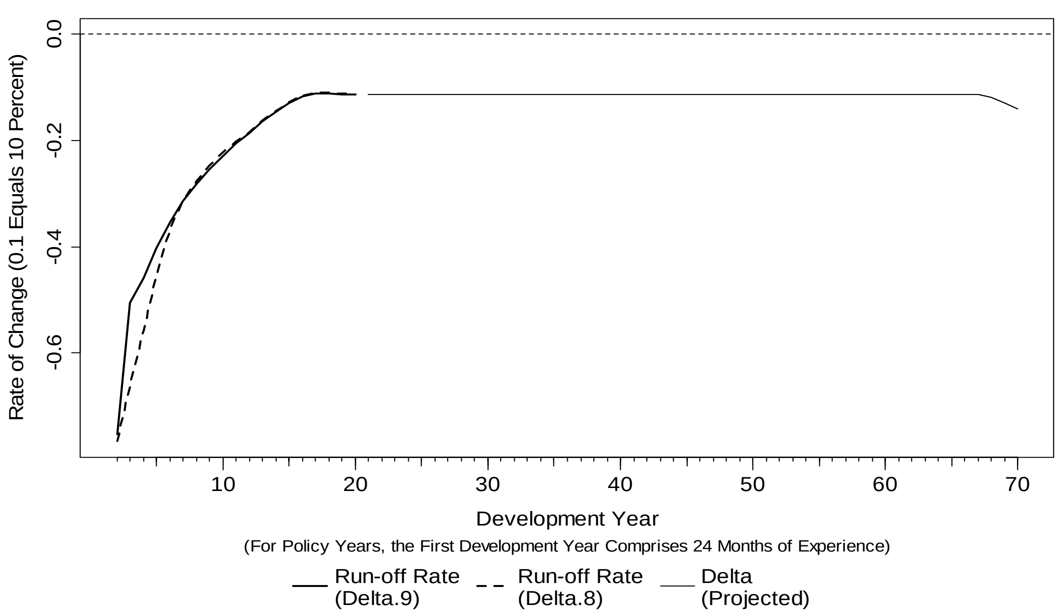


Chart 3: Indemnity: Calendar-Year Effect, Second Development Year

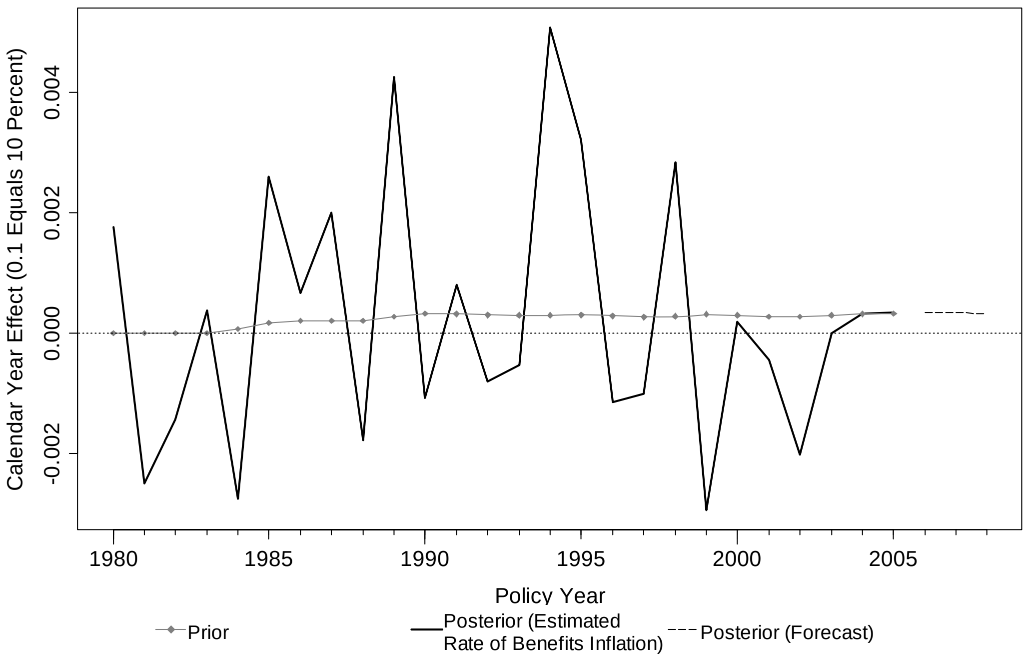


Chart 4: Medical: Calendar-Year Effect, Second Development Year

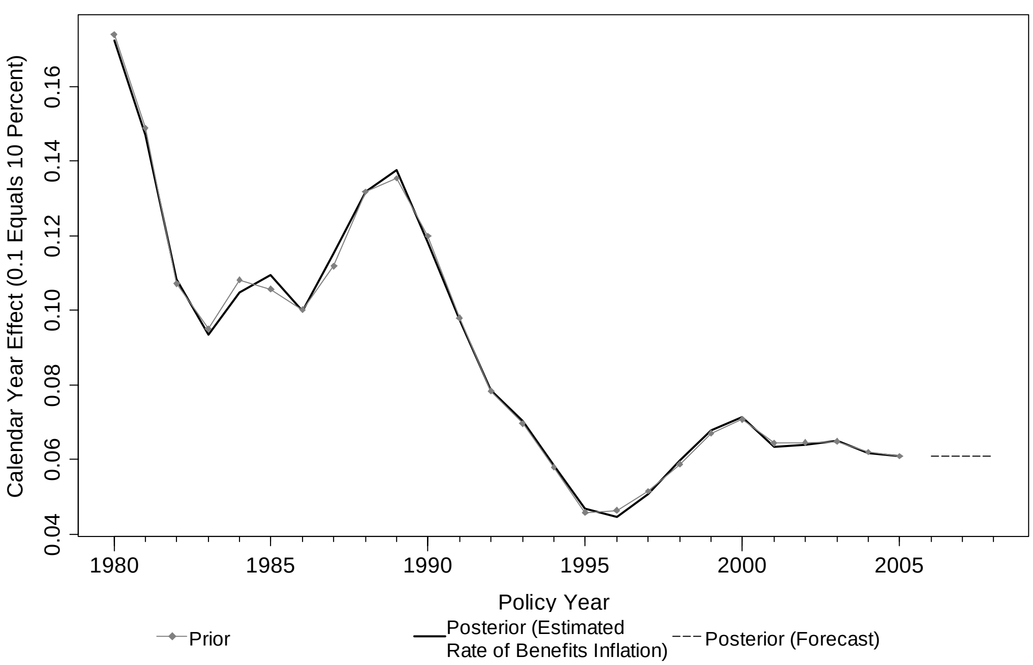


Chart 5: Indemnity: Calendar-Year Effect, Final Diagonal

Chart 6: Medical: Calendar-Year Effect, Final Diagonal

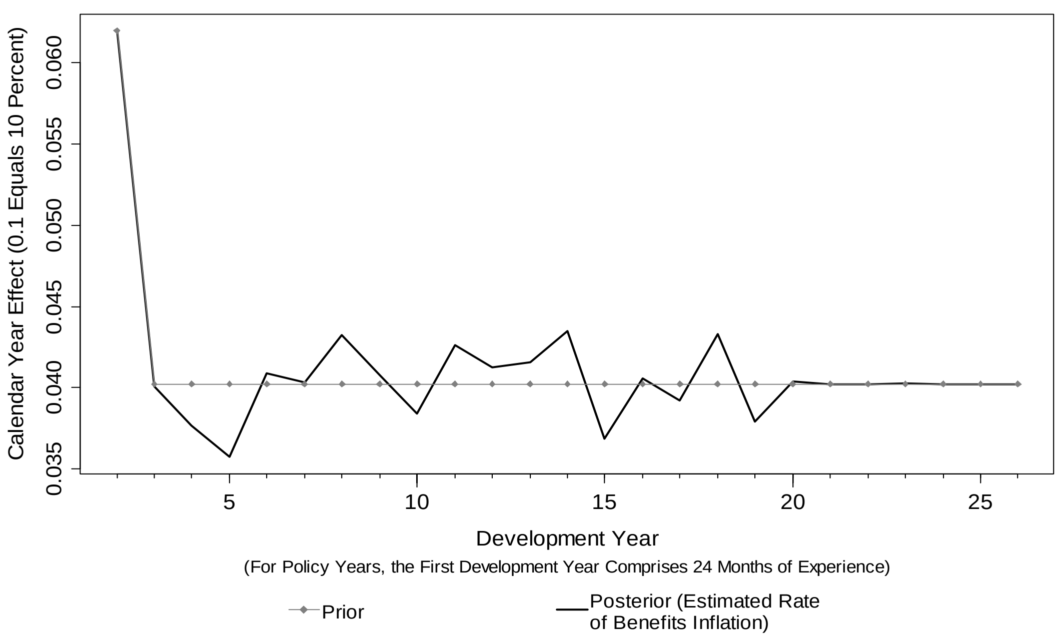


Chart 7: Indemnity: Reform Impact on the Ultimate Loss per $1 of First Report Payment (Adjusted for Calendar-Year Effect); Kernel Density Estimation

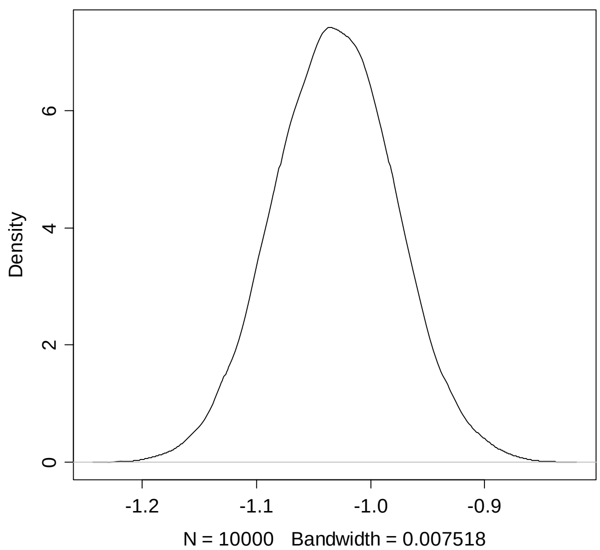


Chart 8: Medical: Reform Impact on the Ultimate Loss per $1 of First Report Payment (Adjusted for Calendar-Year Effect); Kernel Density Estimation

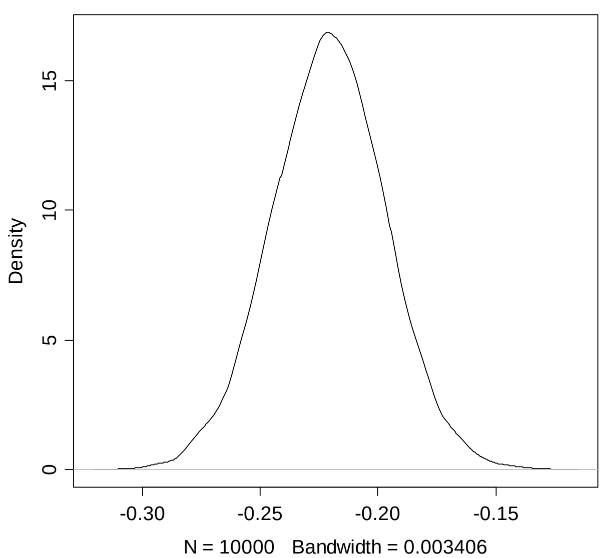


Chart 9: Indemnity: Tail Factor (Vertical Separators Border Reform Cluster)

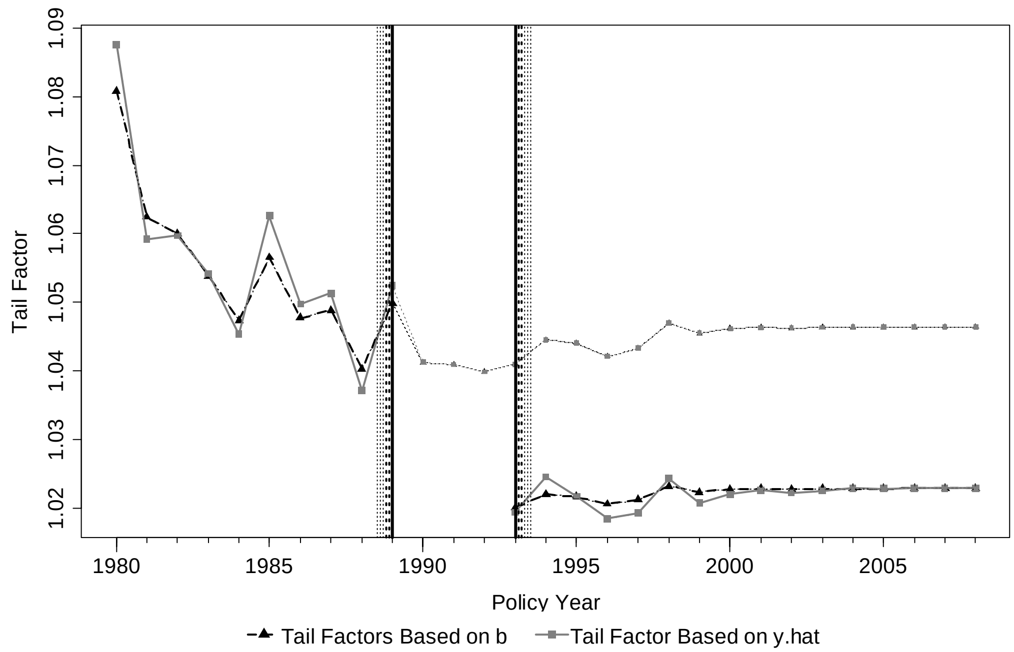


Chart 10: Medical: Tail Factor (Vertical Separators Border Reform Cluster)

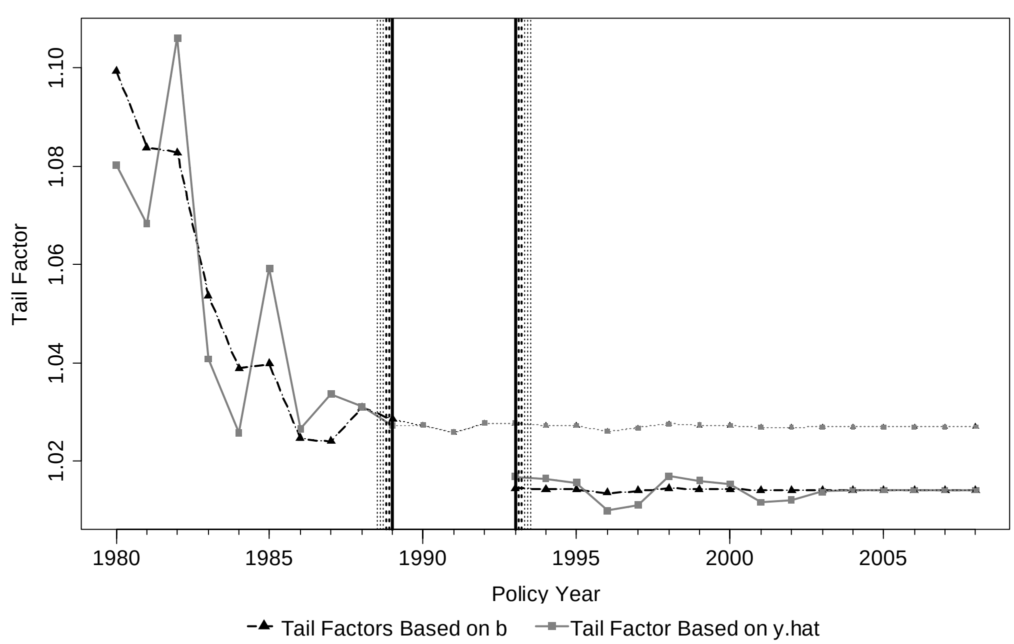


Chart 11: Indemnity: Sensitivity of Tail Factor to Official Rate of Inflation (CPI) for Policy Year 2008

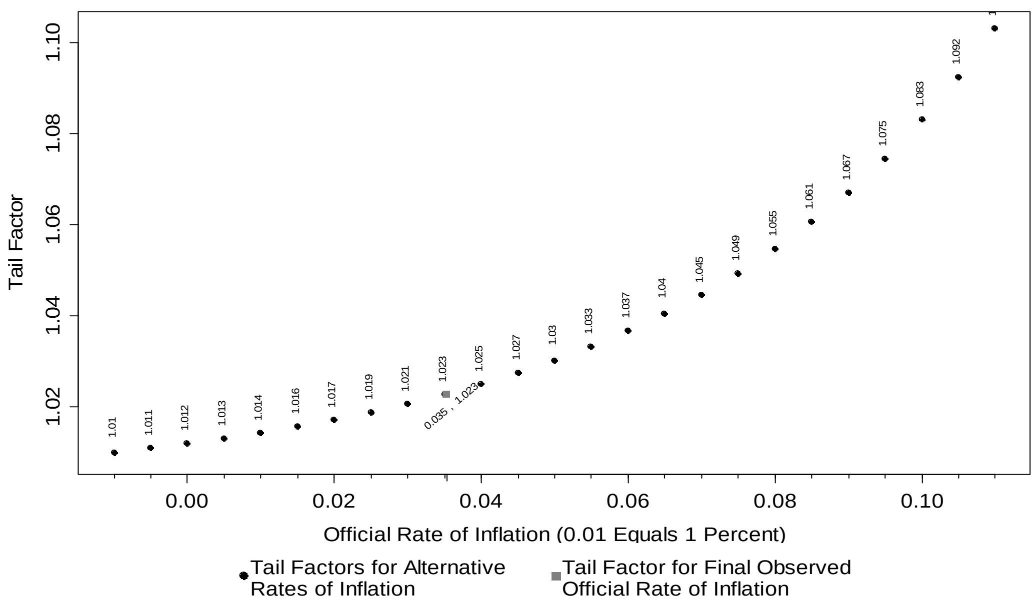


Chart 12: Medical: Sensitivity of Tail Factor to Official Rate of Inflation (M-CPI) for Policy Year 2008

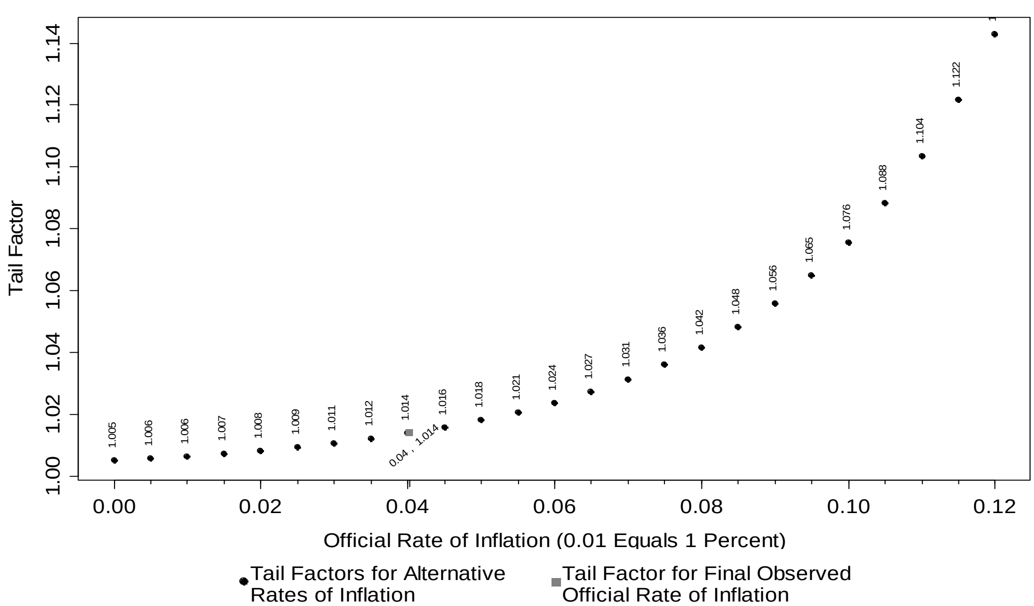


Chart 13: Indemnity: Difference between Actual Observations (*y*) and Estimated Process (*b*) by Policy Year, Post-Reform

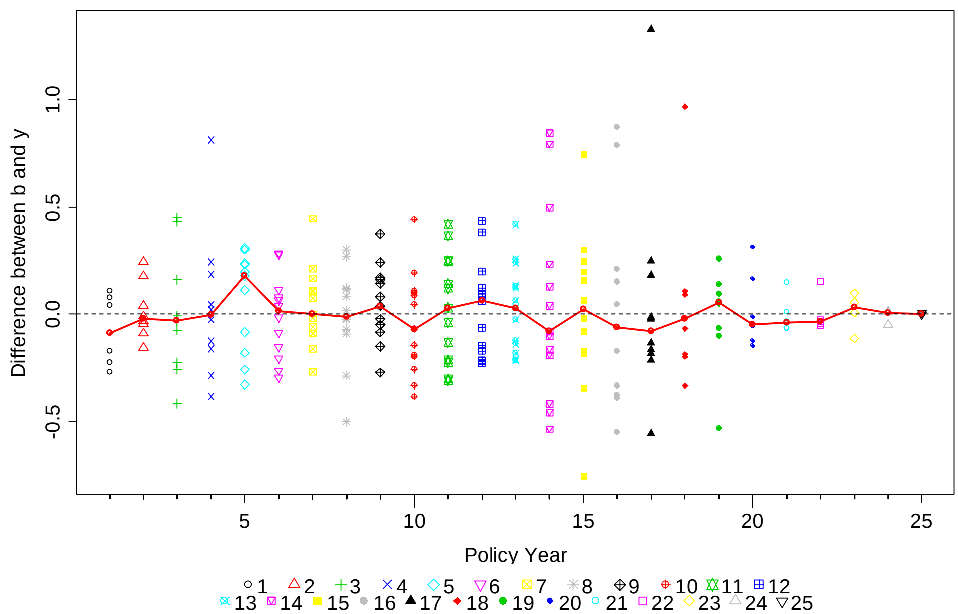


Chart 14: Medical: Difference between Actual Observations (*y*) and Estimated Process (*b*) by Policy Year, Post-Reform

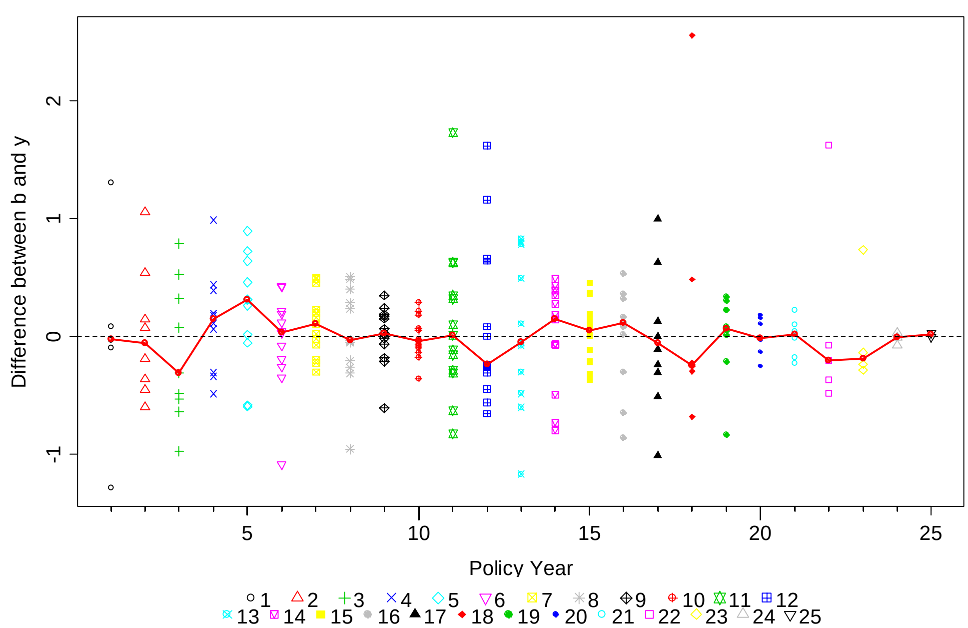


Chart 15: Indemnity: Difference between Actual Observations (*y*) and Estimated Process (*b*) by Development Year, Post-Reform

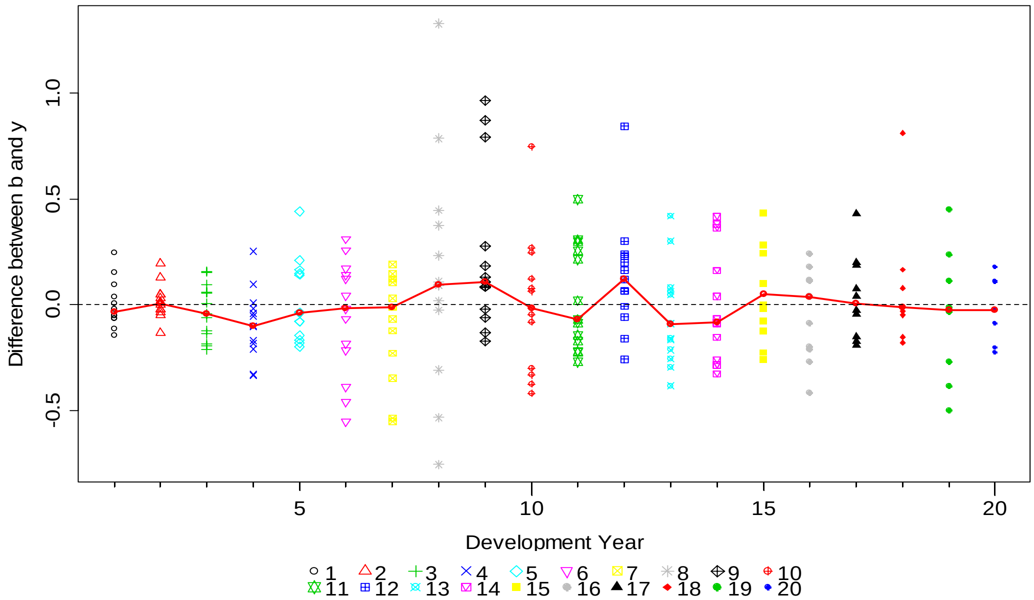


Chart 16: Medical: Difference between Actual Observations (*y*) and Estimated Process (*b*) by Development Year, Post-Reform

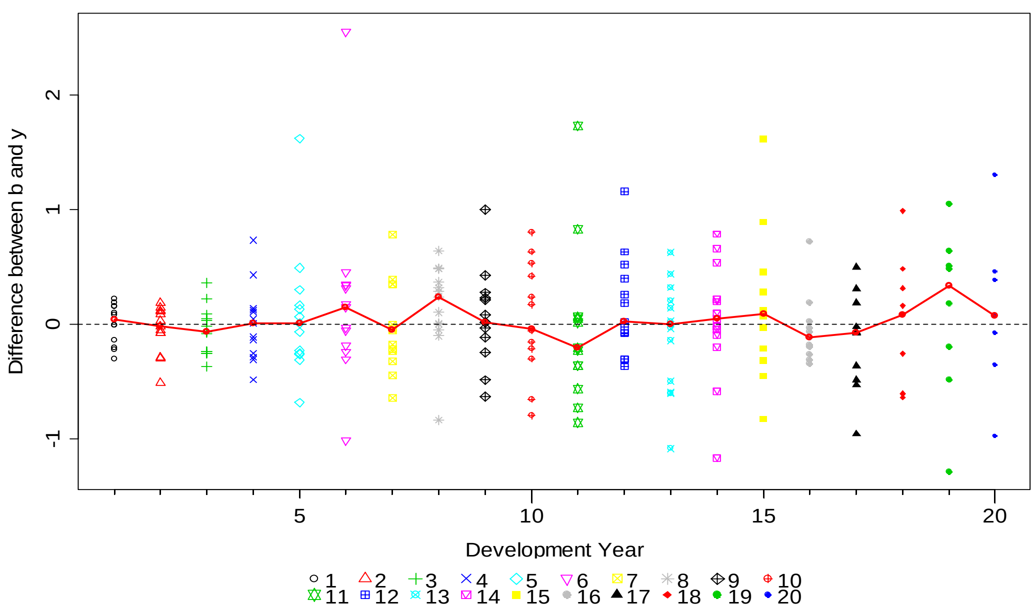


Chart 17: Indemnity: Difference between Actual Observations (*y*) and Estimated Process (*b*) by Diagonal (Calendar Year), Post-Reform

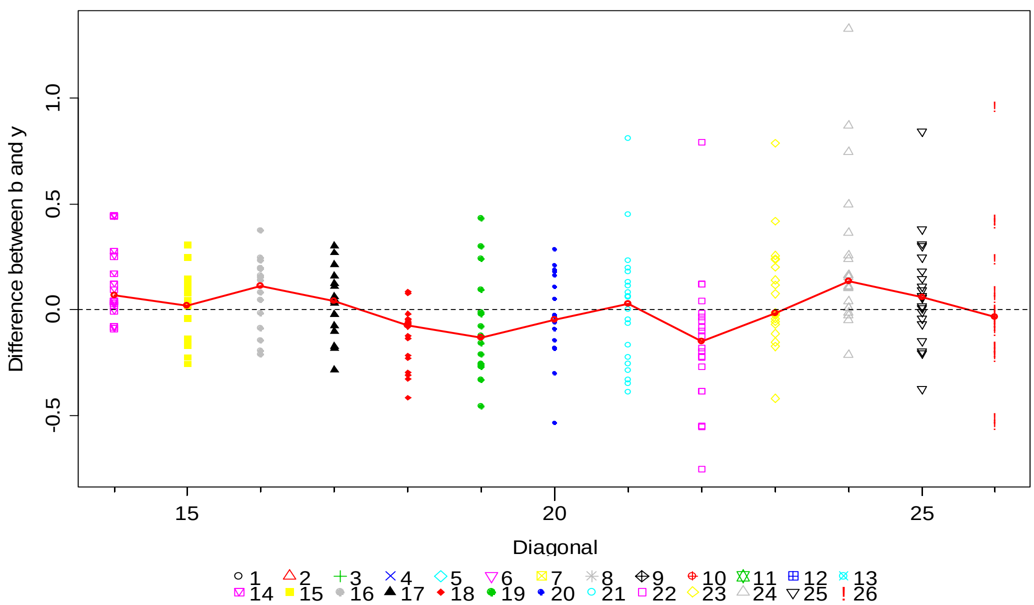


Chart 18: Medical: Difference between Actual Observations (*y*) and Estimated Process (*b*) by Diagonal (Calendar Year), Post-Reform

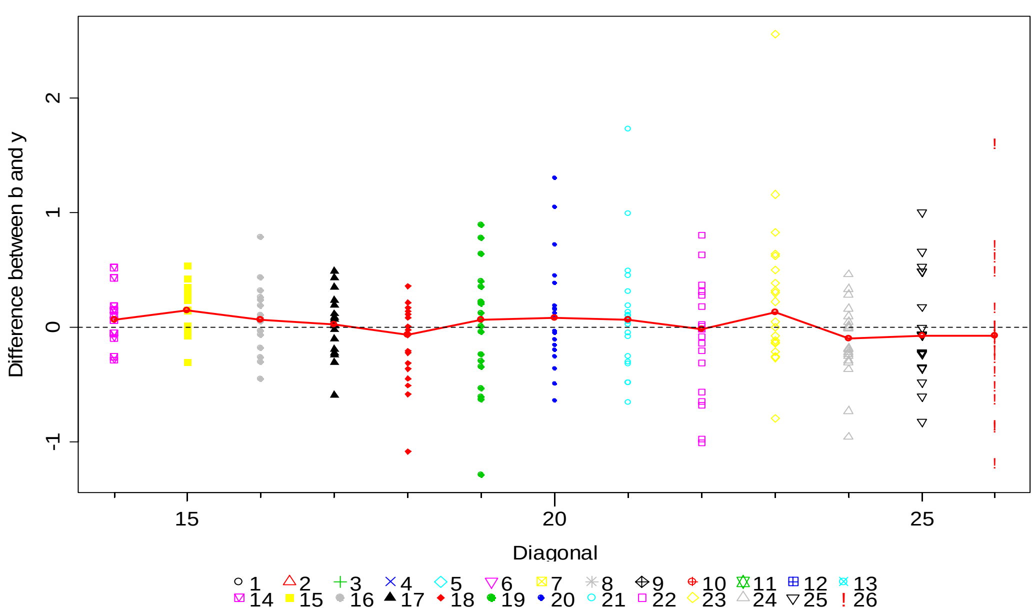


Chart 19: Indemnity: Actual Log Cumulative minus Predicted Log Cumulative Payments, Post-Reform

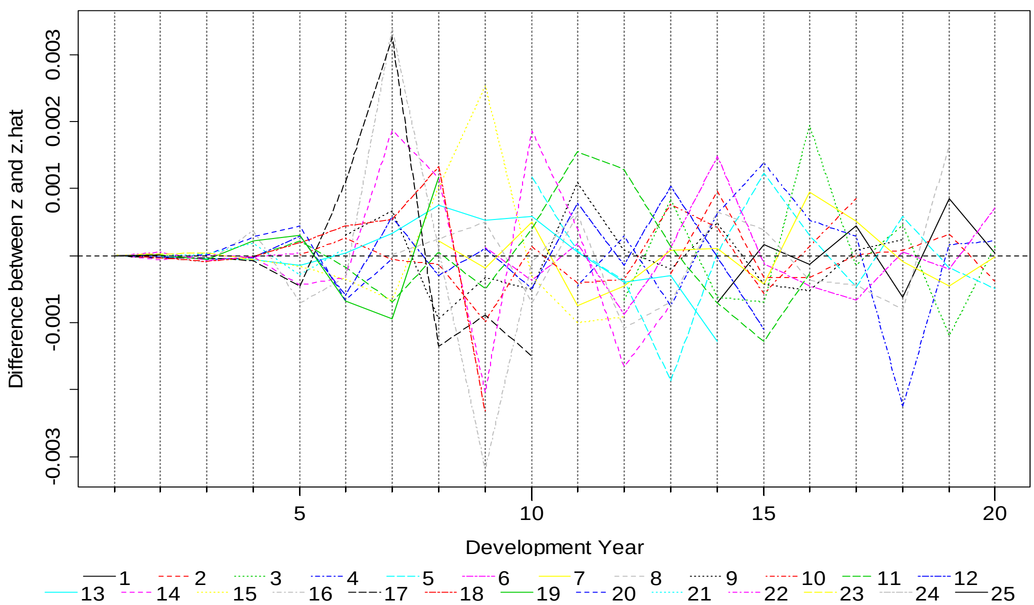
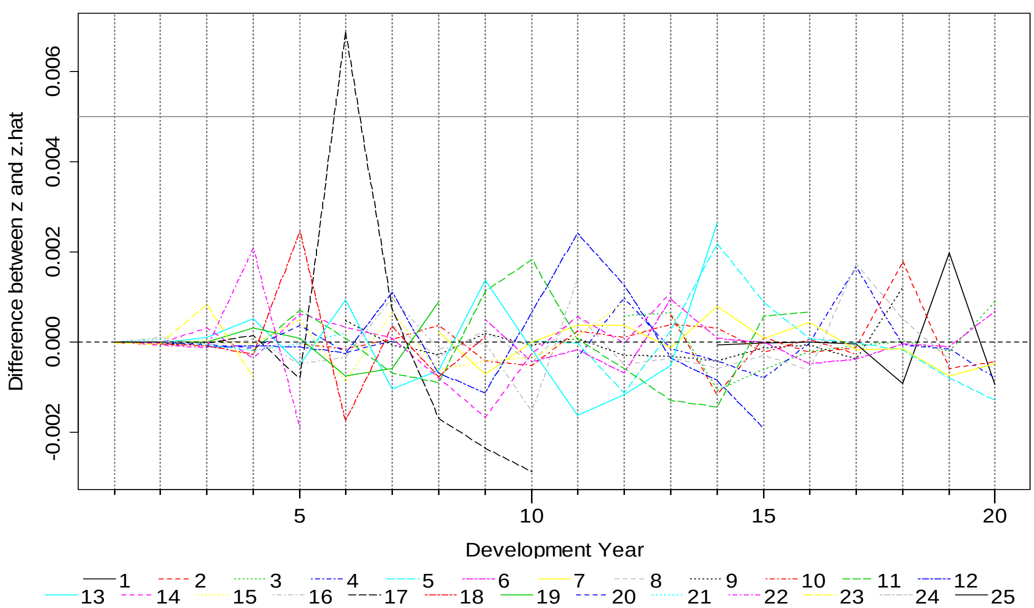


Chart 20: Medical: Actual Log Cumulative minus Predicted Log Cumulative Payments, Post-Reform



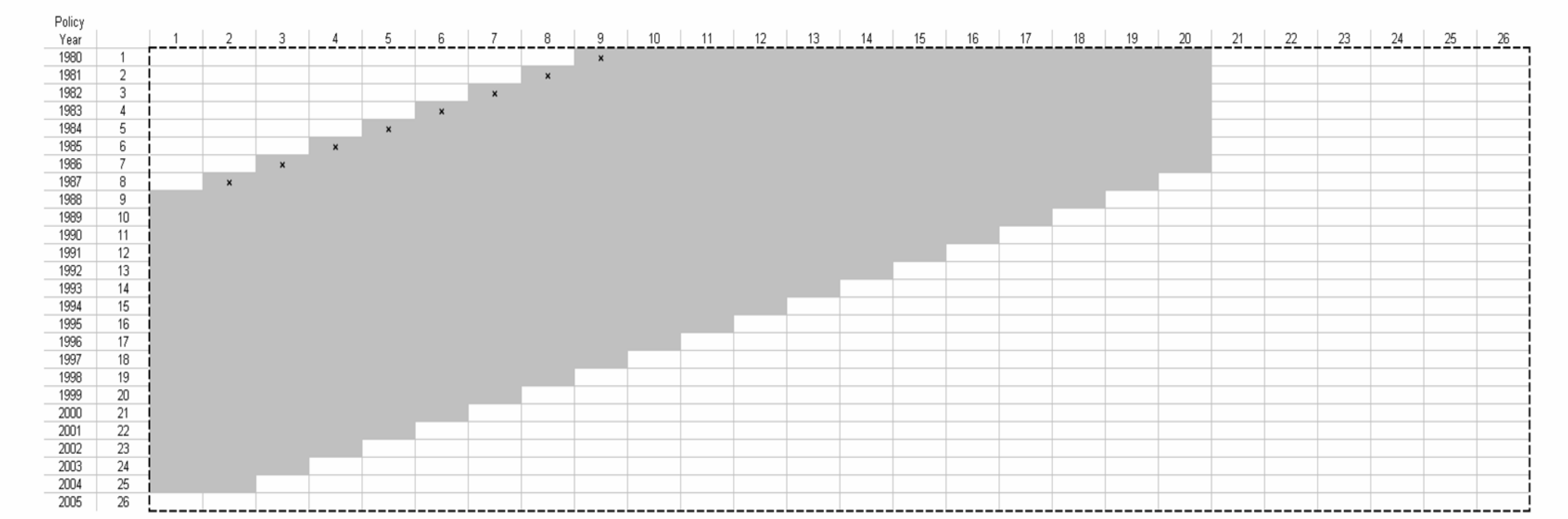
4. CONCLUSIONS

A loss development model has been presented that explicitly accounts for the legislative environment that applies to the time period during which the losses have been observed. Most importantly, the model accommodates changes in the legislative environment, which may be multi-faceted, having either diagonal (calendar year) or horizontal (policy year) effects (or both). The application of the model to an unidentified state demonstrates how, due to its high degree of flexibility, the model is capable of accommodating complex changes to loss development patterns. Further, the model is able to break out and quantify individual aspects of the legislative reform, such as calendar-year effects versus changes to the (calendar-year effect-adjusted) run-off.

Most interesting to the practicing actuary is the ability of the model to incorporate expert information as Bayesian priors in the estimation process. As shown, such expert priors may be legally stipulated rates of escalation (for indemnity) or information on medical price inflation at large (where more detailed information on the inflation embedded in medical benefits is unavailable).

Appendix 1

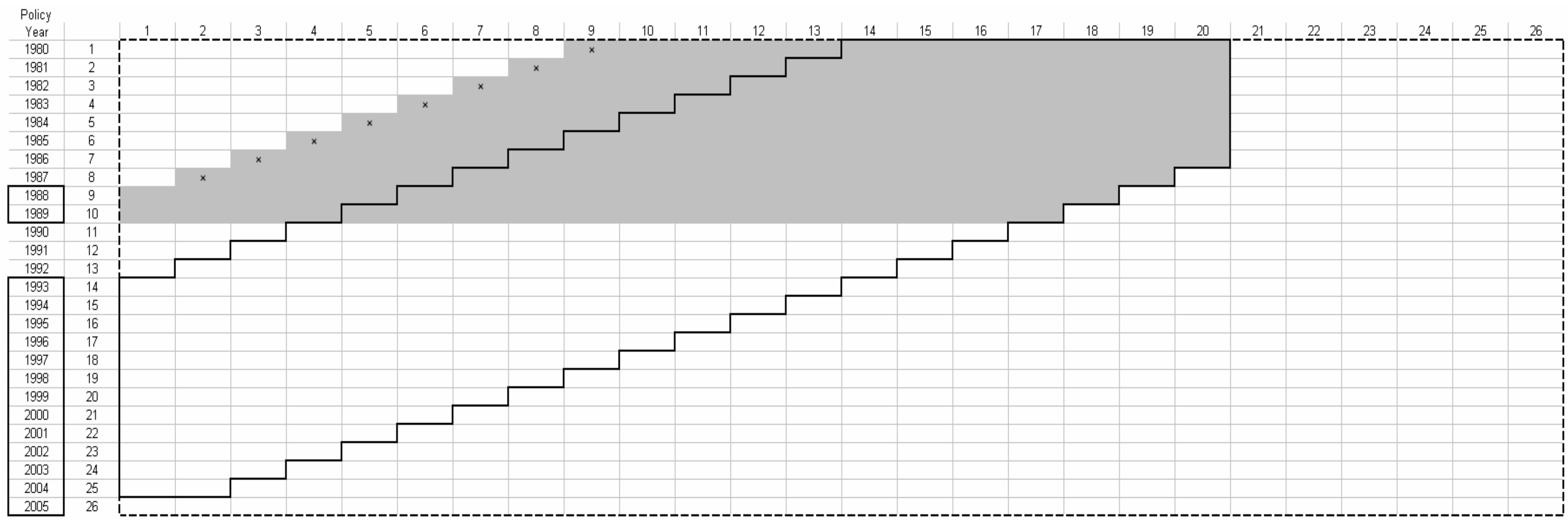
Exhibit A-1: Loss Triangle Template, Indemnity and Medical



Note: Available payments are shaded gray. For the cells marked by the symbol ×, only cumulative (but no incremental) payments are available.

Appendix 1, *cont.'d*

Exhibit A-2: Loss Triangle Template, Pre-Reform and Post-Reform



Note: The payments constituting the pre-reform triangle are shaded gray; the payments forming the post-reform triangle are framed by a solid line. For the cells marked by the symbol ×, only cumulative (but no incremental) payments are available.

Appendix 2: Pre-Reform Model (Model Type 9)

$$y_{i,j} \sim \mathrm{N}(b_{i,j,9}, \sigma_y^2) \begin{cases} i = 1,\ldots,rg-1,\ j = cg-i+2,\ldots,cf-i+1 \\ i = 2,\ldots,rf,\ j = cf-i+2,\ldots,cf \\ i = rf+1,\ldots,rh_9,\ j = cf-i+2,\ldots,c-i+1 \\ i = cf+2,\ldots,rh_9,\ j = 1,\ldots,c-i+1 \\ i = rg,\ldots,rh_9,\ j = 1,\ldots,cf-i+1 \end{cases} \tag{A2-1}$$

$$cs.mean_{i,j,9} \begin{cases} = \hat{z}_{i,j,9} - z_{i,j} \quad \text{for} \begin{cases} i = 1,\ldots,rg-1,\ j = cg-i+1,\ldots,cf-i+1 \\ i = 2,\ldots,rf,\ j = cf-i+2,\ldots,cf \\ i = rf+1,\ldots,rh_9,\ \ j = cf-i+2,\ldots,c-i+1 \\ i = cf+2,\ldots,rh_9,\ \ j = 1,\ldots,c-i+1 \\ i = rg,\ldots,rh_9,\ \ j = 1,\ldots,cf-i+1 \end{cases} \\ = mv_{\mu,i,j,9} \quad \text{for} \begin{cases} i = 1,\ldots,rg-1,\ j = 1,\ldots,cg-i \\ i = 1,\ldots,rf-1,\ j = cf+1,\ldots,c-i+1 \\ i = rg_9+1,\ldots,r,\ j = 1,\ldots,c-i+1 \\ i = 2,\ldots,r,\ j = c-i+2,\ldots,c \end{cases} \end{cases} \tag{A2-2}$$

$$\hat{y}_{i,j,9} = \mathrm{N}(b_{i,j,9}, \sigma_y^2)\ ,\ i, j = 1,\ldots,c \tag{A2-3}$$

$$\hat{z}_{i,j,9} = \log\left(\sum_{k=1}^{j} \exp\left(\hat{y}_{i,k,9}\right)\right)\ ,\ i, j = 1,\ldots,c \tag{A2-4}$$

$$\mathbf{mv}_9 \sim \mathrm{N}(\mathbf{mv}_{\mu,9}, \mathbf{\Omega}_1) \tag{A2-5}$$

$$\mathbf{mv}_{\mu,9} \sim \mathrm{N}(\mathbf{mv}_9, \mathbf{\Omega}_2) \tag{A2-6}$$

$$cs_{i,j,9} = 0\ ,\ i, j = 1,\ldots,c \tag{A2-7}$$

$$\mathbf{cs}_{i,9}' \sim \mathrm{N}(\mathbf{cs.mean}_9', \mathbf{\Omega}_1)\ ,\ i = 1,\ldots,r \tag{A2-8}$$

$$b_{r_h,1,9} \sim \mathrm{N}(y_{r_h,1,9}, \sigma_{b.init}^2) \tag{A2-9}$$

$$b_{i,1,9} = b_{i+1,1,9} - \Lambda_{i+1,1,9} \;,\; i = 1,\ldots,r_{h,9} - 1 \tag{A2-10}$$

$$\Lambda_{i+1,1,9} = \eta_{i+1} \;,\; i = 1,\ldots,r_{h,9} - 1 \tag{A2-11}$$

$$\eta_i \sim \mathrm{N}(\eta_{i-1}, \sigma_\eta^2) \;,\; i = 3,\ldots,r \tag{A2-12}$$

$$\eta_2 \sim \mathrm{N}(0, \sigma_\eta^2) \tag{A2-13}$$

$$b_{i,1,9} = b_{i-1,1,9} + \Lambda_{i,1,9} \;,\; i = r_{h,9} + 1,\ldots,r \tag{A2-14}$$

$$\Lambda_{i,1,9} = \eta_{i,1} \;,\; i = r_{h,9} + 1,\ldots,r \tag{A2-15}$$

$$b_{i,j,9} = b_{i,j-1,9} + \Lambda_{i,j,9} \;,\; i = 1,\ldots,r-1,\; j = 2,\ldots,c-i+1 \tag{A2-16}$$

$$\Lambda_{i,j,9} = \delta_{j,9.pre} + \kappa_{i,j} \;,\; i = 1,\ldots,r_{g,9} - 1,\; j = 2,\ldots,c_{g,9} - i + 1 \tag{A2-17}$$

$$\Lambda_{i,j,9} = \delta_{j,9} + \kappa_{i,j} \;,\; i = 1,\ldots,r_{g,9} - 1,\; j = c_{g,9} - i + 2,\ldots,c-i+1 \tag{A2-18}$$

$$\Lambda_{i,j,9} = \delta_{j,9} + \kappa_{i,j} \;,\; i = r_{g,9},\ldots,r,\; j = 2,\ldots,c-i+1 \tag{A2-19}$$

$$\delta_{j,9} \sim \mathrm{N}(\delta.prior_j, \sigma_{\delta,2}^2) \;,\; j = 2,3 \tag{A2-20}$$

$$\delta_{j,9} \sim \mathrm{N}(\delta_{j-1,9}, \sigma_{\delta,j}^2) \;,\; j = 4,\ldots,cf \tag{A2-21}$$

$$\delta_{j,9} \sim \mathrm{N}(\delta_{cf}, \sigma_{\delta,1}^2) \;,\; j = cf + 1,\ldots,c \tag{A2-22}$$

$$\sigma^2_{\delta,j} = \sigma^2_{\delta,1} \cdot 10^{-\alpha+\alpha\cdot e^{-\beta\cdot e^{-\gamma\cdot(j-1)}}} \quad , \; j = 4,\ldots,cf\,;\; \alpha,\beta,\gamma > 0 \tag{A2-23}$$

$$\sigma^2_{\delta,2} \;,\; \sigma^2_{b.init} \text{ large} \tag{A2-24}$$

$$\sigma^2_{\delta,1} \text{ small} \tag{A2-25}$$

$$\kappa_{i,j} \sim \mathrm{N}(\mu_{i,j}, \sigma^2_{\kappa}) \;,\; i = 1,\ldots,r;\; j = 2,\ldots,c \tag{A2-26}$$

$$\mu_{i,j} = \lambda_{1,j} \cdot \pi_{1,i+j} + \lambda_{2,j} \cdot \pi_{2,i+j} \;,\; i = 1,\ldots,r;\; j = 2,\ldots,c,\; \lambda_{1,j} + \lambda_{2,j} \le 1 \tag{A2-27}$$

where y, and $\hat{y}$ are the observed and estimated logarithmic incremental payments, respectively. For negative incremental payments, the corresponding values of y are coded as missing values. The indexes i and j indicate policy (or accident) and development years, respectively; $r = c$ signifies the number of years in the loss triangle. The parameter c_f signifies the column with the final value for the cumulative (and incremental) payment in the first $r_f - 1$ rows, where the first $r_f - 1$ rows are those affected by the cut-off in reported development. The parameter c_g signifies the first column that has a value for the cumulative payment in the first row; note that the first incremental payment in this row is located in column $c_g + 1$. The parameter $r_g (= c_g)$ indicates the first row that has a value for the cumulative (and thus incremental) payment in the first column.

The parameter $r_{g,9} (= c_{g,9})$ indicates the row (column) with the first pre-reform incremental payment in the first column (row). If there was no structural break prior to the reform of interest, then $r_{g,9} (= c_{g,9}) = r_g (= c_g)$. Conversely, if there was such a possible structural break, then the parameter $r_{g,9} (= c_{g,9})$ indicates the first row (column) with an incremental payment in the first column (row) that belongs to the post-structural-break pre-reform period.

Equation (A2-1) fits the observations of the logarithmic incremental payments to a normal distribution. Equation (A2-2) defines the deviation of the estimated logarithm of the cumulative payment ($\hat{z}_{i,j,9}$, where the index 9 indicates pre-reform) in policy (or accident) year j and development year i from and the observed logarithm of the cumulative payment ($z_{i,j}$); this deviation is denoted $cs.mean_{i,j}$, where cs stands for cumulative sum. Equation (A2-3) simulates the predicted values of the logarithmic incremental payments; these predicted values feed into the estimated logarithmic cumulative payments in Equation (A2-4). Where such cumulative sum does not exist (to the right of the final diagonal, up to the final observed development year), $cs.mean_{i,j}$ is replaced by a draw from a multivariate distribution, $mv_{\mu,i,j,9}$, as shown in Equation (A2-6). Specifically, the row vector $\mathbf{cs.mean}_i$ comprises the differences between the predicted and observed logarithmic cumulative payments of row i for those columns for which observed logarithmic cumulative payments are available; for all other columns, the elements of $\mathbf{cs.mean}_i$ are taken from a vector of (expected) values that generates a multivariate normal distribution of the same variance as the one that $\mathbf{cs.mean}_i$ is fitted to. The covariance matrices $\mathbf{\Omega}_{1,2}^{-1}$ are modeled on Wishart distributions. Equations (A2-5) and (A2-6) generate a distribution the $mv_{\mu,i,j,9}$ can be drawn from; the distributions of the observed and the generated values of $\mathbf{cs.mean}_i$ share the same covariance matrix, $\mathbf{\Omega}_1^{-1}$. Equation (A2-7) stipulates that the observed differences between the logarithms of the observed and estimated cumulative payments be zero, on average. Equation (A2-8) represents the cumulative sum (cusum) constraint. This stochastic constraint ensures that, for every cell of the loss triangle, the sum of estimated incremental payments lines up (approximately) with the observed cumulative payment. The cusum constraint also serves as a means of interpolating between incremental payments when there is a missing value (due to a negative incremental payment).

Equation (A2-9) initializes for the upper-left hand side region (where no observations are available for the first incremental payment) the first logarithmic increment payment on the first logarithmic incremental payment of the first row for which such a payment is available (denoted as row r_h).

Equations (A2-10-), (A2-11), and (A2-14) through (A2-19) describe the process displayed in Exhibit 1. Equation (A2-12) describes the random walk of *eta*, and Equation (A2-13) its starting value. Equation (A2-21) describes the random walk of *delta*, and Equation (A2-20) describes how the first two values of delta are estimated before the random walk sets in, whereas Equation (A2-22) details how delta is extrapolated into the future after the random walk ends with the final observed development year. Equation (A2-23) describes a Gompertz function for the innovation variance of the random walk of *delta*; this innovation variance approaches the variance displayed in Equation (A-25). The variance for estimating the first two values of *delta* (that is, before the random walk sets in) is shown in Equation (A2-24). Finally, Equations (A2-26 and A2-27) detail how the calendar-year effect is estimated using an expert prior on the rate of escalation (indemnity) and inflation (medical).

The model has two layers of noise, which implies that there are two predicted values (for each observed value of incremental payment). First, there is the variable b, which aggregates the three processes (run-off in development, growth of expected value of first payment, and calendar-year effect). Second, there is the variable $y.hat$, which is a draw from a normal distribution, the expected value of which is b. Where there are no observations (the run-off triangle is squared, the tail is estimated, and future policy or accident years are forecast), the variable $y.hat$ corresponds to the expected value, b. The variable $y.hat$ gauges the ability of the model to replicate the observed incremental payments.

The variables $\pi_{i,j}$ $(i = 1,2)$ are expert priors for (logarithmic) rates of inflation, which may vary by policy (or accident) and development years. (For policy years, the first prior in any given policy (or accident) year comprises inflation for a period of 18 months, this being the time difference between the mid-point of the initial 24 months of experience and the subsequent 12-month period.) The model accommodates two non-zero rates of inflation, differentiated by type of claim; this is important for indemnity claims (but irrelevant for medical claims). Thus, the prior for the calendar-year effect in any given development year, j, is a weighted average of three (one zero and two non-zero) expert rates of inflation, the weights being the fractions of dollars in incremental payments that apply to up to two differently inflating claim types in development year $j-1$, $\lambda_{k,j}$, $k = 1,2$ (while a third claim type may inflate at a zero rate). If there is only one claim type (as is the case for medical claims) or all claim types escalate at the same rate, then $\pi_{2,j}$ and $\lambda_{2,j}$ equal zero for all j, and $\lambda_{1,j}$ equals 1 for all j.

Specifically, for indemnity, the expert prior for the (logarithmic) calendar-year effect equals the official (logarithmic) rate of inflation relevant to the cost-of-living adjustment, weighted by the fractions of incremental dollars that have been paid on escalating claims in the development year $j-1$, λ_j. The official rate of inflation pertinent to cost-of-living adjustment may be the rate of growth of the state-level average weekly wage (as measured by the Quarterly Census of Employment and Wages, QCEW, http://www.bls.gov) or the U.S. CPI (Consumer Price Index, http://www.bls.gov), depending on the applicable legislative provision; we apply an observation and implementation lag of 14 months. The expert inflation prior for medical benefits is the (contemporaneous logarithmic) rate of growth of the Medical Care component of the U.S. CPI.

The QCEW average weekly wage is calculated as the ratio of the total wage bill for the calendar year, summed up over four quarterly values, and then divided by the average employment for the calendar year; this average employment for the calendar year is calculated from 12 monthly numbers. The Medical Care component of the CPI is the published annual calendar year number.

It is important to note that the rate of growth of the expected value of the first incremental payment (η) is specified in nominal terms, which means that the rate of inflation is not broken out. As a consequence, the mentioned inflation modeling applies solely to the way the incremental payments inflate in development but has no bearing on the how the first incremental payment inflates from one policy (or accident) year to the next.

The chosen set of hyper-parameters of the prior distributions has been calibrated to incremental payments, the logarithm of which fall into the range of 7 to 11; the incremental (and cumulative) payments of the loss triangle that is to be analyzed have to be normalized accordingly. With such normalization, the chosen set of hyper-parameters accommodates any sufficiently well-behaved triangle. As a consequence, the final calibration of the model when applied to a loss triangle is done solely by choosing the three parameters of the Gompertz function, with one exception; this exception concerns the variance of the rate of growth of the expected value of the first payment, as exhibited in Equations (A2-12, 13). For triangles with a high degree of variation in the rate of growth of the first incremental payment (such as percentage point differences in the higher double digits), a larger variance is needed. Further, the parameters of the Gompertz function need to be chosen. This Gompertz function serves the purpose of smoothing the run-off rate δ by means of controlling the innovation variance of the random walk. The Gompertz function accommodates convex, concave, and "S"-shaped trajectories of this variance. The first parameter of the Gompertz function, α, determines the upper asymptote; the parameter β is (roughly) a horizontal shift parameter, and the parameter γ determines the rate of the growth (that is, the steepness and curvature). The choice of the parameters β and γ is ultimately a matter of judgment, especially for small triangles. Several diagnostic charts have been developed (as discussed in the body of the text) that assist in this choice.

Note that the pre-reform and post-reform models have all variances in common; further, the two models have a common calendar-year effect and common rates of growth of the expected value of the first payment. For all scalar variances in the model, there are gamma distributions used as priors.

Appendix 3: Post-Reform Model (Model Type 8)

$$y_{i,j} \sim \mathrm{N}(b_{i,j,8}, \sigma_y^2) \begin{cases} i = 1,\ldots, rg_8 - 1,\ j = cg_8 - i + 1,\ldots, cf - i + 1 \\ i = 2,\ldots, rf,\ j = cf - i + 2,\ldots, cf \\ i = rf + 1,\ldots, cf + 1,\ j = cf - i + 2,\ldots, c - i + 1 \\ i = cf + 2,\ldots, rh_8,\ j = 1,\ldots, c - i + 1 \\ i = rg_8,\ldots, cf,\ j = 1,\ldots, cf - i + 1 \end{cases} \tag{A3-1}$$

$$cs_{i,j,8} \begin{cases} = \hat{z}_{i,j,8} - z_{i,j} & \text{for} \begin{cases} i = 1,\ldots, rg_8 - 1,\ j = cg_8 - i + 1,\ldots, cf - i + 1 \\ i = 2,\ldots, rf,\ j = cf - i + 2,\ldots, cf \\ i = rf + 1,\ldots, cf + 1,\ \ j = cf - i + 2,\ldots, c - i + 1 \\ i = cf + 2,\ldots, rh_8,\ \ j = 1,\ldots, c - i + 1 \\ i = rg_8,\ldots, cf,\ \ j = 1,\ldots, cf - i + 1 \end{cases} \\ = mv_{\mu,i,j,8} & \text{for} \begin{cases} i = 1,\ldots, rg_8 - 1,\ j = 1,\ldots, cg_8 - i \\ i = 1,\ldots, rf - 1,\ j = cf + 1,\ldots, c - i + 1 \\ i = rh_8 + 1,\ldots, r,\ j = 1,\ldots, c - i + 1 \\ i = 2,\ldots, r,\ j = c - i + 2,\ldots, c \end{cases} \end{cases} \tag{A3-2}$$

$$\hat{y}_{i,j,8} = \mathrm{N}(b_{i,j,8}, \sigma_y^2)\ ,\ i, j = 1,\ldots, c \tag{A3-3}$$

$$\hat{z}_{i,j,8} = \log\left(\sum_{k=1}^{j} \exp\left(\hat{y}_{i,k,8}\right)\right)\ ,\ i, j = 1,\ldots, c \tag{A3-4}$$

$$\mathbf{mv}_8 \sim \mathrm{N}(\mathbf{mv}_{\mu,8}, \mathbf{\Omega}_1) \tag{A3-5}$$

$$\mathbf{mv}_{\mu,8} \sim \mathrm{N}(\mathbf{mv}_8, \mathbf{\Omega}_2) \tag{A3-6}$$

$$cs_{i,j,8} = 0\ ,\ i, j = 1,\ldots, c \tag{A3-7}$$

$$\mathbf{cs}_{i,8}{}' \sim \mathrm{N}(\mathbf{cs.mean}_8{}', \mathbf{\Omega}_1)\ ,\ i = 1,\ldots, c \tag{A3-8}$$

$$b_{r_h,1,8} \sim \mathrm{N}(y_{r_h,1,8}, \sigma^2_{b.init}) \tag{A3-9}$$

$$b_{i,1,8} = b_{i+1,1,8} - \Lambda_{i+1,1,8} \;,\; i = 1,\ldots,r_{h,8} - 1 \tag{A3-10}$$

$$\Lambda_{i+1,1,8} = \eta_{i+1} \;,\; i = 1,\ldots,r_{h,8} - 1 \tag{A3-11}$$

$$b_{i,1,8} = b_{i-1,1,8} + \Lambda_{i,1,8} \;,\; i = r_{h,8} + 1,\ldots,r \tag{A3-12}$$

$$\Lambda_{i,1,8} = \eta_i \;,\; i = r_{h,8} + 1,\ldots,r \tag{A3-13}$$

$$b_{i,j,8} = b_{i,j-1,8} + \Lambda_{i,j,8} \;,\; i = 1,\ldots,r-1,\; j = 2,\ldots,c-i+1 \tag{A3-14}$$

$$\Lambda_{i,j,8} = \delta_{j,9.pre} + \kappa_{i,j} \;,\; i = 1,\ldots,r_{g,9} - 1,\; j = 2,\ldots,c_{g,9} - i + 1 \tag{A3-15}$$

$$\Lambda_{i,j,8} = \delta_{j,89} + \kappa_{i,j} \;,\; i = 1,\ldots,r_{g,9} - 1,\; j = c_{g,9} - i + 2,\ldots,c_{g,8} - i + 1 \tag{A3-16}$$

$$\Lambda_{i,j,8} = \delta_{j,89} + \kappa_{i,j} \;,\; i = r_{g,9},\ldots,r_{g,8} - 1,\; j = 2,\ldots,c_{g,8} - i + 1 \tag{A3-17}$$

$$\Lambda_{i,j,8} = \delta_{j,8} + \kappa_{i,j} \;,\; i = 1,\ldots,r_{g,8} - 1,\; j = c_{g,8} - i + 2,\ldots,c - i + 1 \tag{A3-18}$$

$$\Lambda_{i,j,8} = \delta_{j,8} + \kappa_{i,j} \;,\; i = r_{g,8},\ldots,r,\; j = 2,\ldots,c - i + 1 \tag{A3-19}$$

$$\delta_{j,8} \sim \mathrm{N}(\delta.prior_j, \sigma^2_{\delta,2}) \;,\; j = 2,3 \tag{A3-20}$$

$$\delta_{j,8} \sim \mathrm{N}(\delta_{j-1,8}, \sigma^2_{\delta,j}) \;,\; j = 4,\ldots,cf \tag{A3-21}$$

$$\delta_{j,8} \sim \mathrm{N}(\delta_{cf}, \sigma^2_{\delta,1}) \ , \ j = cf + 1, \ldots, c \tag{A3-22}$$

$$\delta.diff_j = \delta_{j,8f} - \delta_{j,9} \ , \ j = 4, \ldots, cf \tag{A3-23}$$

$$\delta.diff.z_j \sim \mathrm{N}(\delta.diff_j, \sigma^2_{\delta,j}) \ , \ j = 4, \ldots, cf \tag{A3-24}$$

$$\delta.diff.z_j = 0 \ , \ j = 4, \ldots, cf \tag{A3-25}$$

$$\kappa_{i,j} \sim \mathrm{N}(\mu_{i,j}, \sigma^2_{\kappa}) \ , \ i = 1, \ldots, r; \ j = 2, \ldots, c \tag{A3-26}$$

$$\mu_{i,j} = \lambda_{1,j} \cdot \pi_{1,i+j} + \lambda_{2,j} \cdot \pi_{2,i+j} \ , \ i = 1, \ldots, r; \ j = 2, \ldots, c, \ \lambda_{1,j} + \lambda_{2,j} \le 1 \tag{A3-27}$$

The parameter $r_{g,8} (= c_{g,8})$ indicates the row (column) of the first post-reform incremental payment in the first column (row). Equations (A3-23, 24, and 25) define the convergence constraint for the run-off rates of the pre-and post-reform triangles; this constraint becomes tighter as development progresses. Note that the pre-reform run-off rates of the post-reform triangle are allowed to differ from the run-off rates of the pre-reform triangle (except for the $\delta_{j,9.pre}$ area). For the definitions of the variables parameters, see Appendix 2. Further, see Appendix 4 for a complete list of variables.

Acknowledgment

Thanks to Carolyn Bergh and Jon Evans for comments, and to Chris Laws, Jose Ramos, and Manuel de la Guardia for research assistance.

5. REFERENCES

[1] De Jong, Piet, "Forecasting Runoff Triangles," *North American Actuarial Journal* 10, 28-38, 2006.
[2] De Jong, Piet, and Ben Zehnwirth, "Claims Reserving, State-Space Models and the Kalman Filter," *Journal of the Institute of Actuaries* 110, 157-181, 1983.
[3] England, Peter, and Richard Verrall, "Stochastic Claims Reserving in General Insurance," *Journal of the Institute of Actuaries* 129, 1-76, 2002.
[4] Gilks, Wally R., Sylvia Richardson, and David J. Spiegelhalter, *Markov Chain Monte Carlo in Practice,* (Boca Raton, FL: Chapman & Hall/CRC, 1996).
[5] Mack, Thomas, "Which Stochastic Model is Underlying the Chain Ladder Method?" *Insurance, Mathematics and Economics* 15(2/3), 133-138, 1994.
[6] Scollnik, David, "Bayesian Reserving Models Inspired by Chain Ladder Methods and Implemented Using WinBUGS," http://math.ucalgary.ca/~scollnik/balducci/Papers/resnaaj.pdf, April 2004.
[7] Scollnik, David, "Implementation of Four Models for Outstanding Liabilities in WinBUGS: A discussion of a paper by Ntzoufras and Dellaportas," *North American Actuarial Journal* 6(1):128-136, 2002.
[8] Taylor, Gregory, *Loss Reserving: An Actuarial Perspective*, (Boston: Kluwer, 2000).
[9] Verrall, Richard, "Modeling Claims Run-Off Triangles with Two-Dimensional Time Series," *Scandinavian Actuarial Journal* 129-138, 1989.

Abbreviations and notations

BUGS, Bayesian inference Using the Gibbs Sampler
CPI, Consumer Price Index
MCMC, Markov Chain Monte Carlo (Simulation)
M-CPI, Medical Care Component of the CPI
NCCI, National Council on Compensation Insurance
QCEW, Quarterly Census of Employment and Wages
PP, Permanent Partial (Claims)
PT, Permanent Total (Claims)
TT, Temporary Total (Claims)

Biography of the Author

Frank Schmid, Dr. habil., is a Director and Senior Economist at the National Council on Compensation Insurance.

Corporate Governance and the Loss Reserving Process

E. Daniel Thomas, FCAS, MAAA
Marc F. Oberholtzer, FCAS, MAAA
Timothy Landick, FCAS, MAAA

Abstract: Since the implementation at year-end 2004 of requirements under the Sarbanes-Oxley Act of 2002, many publicly traded property/casualty insurance companies have benefited from improved corporate governance surrounding the loss reserving process. However, the degree of improvement and resultant benefit has varied widely by company. While some have embraced the value of having stronger controls, others have viewed these requirements as resulting in significant additional process with only minimal benefit. The authors believe there are significant benefits to having strong corporate governance surrounding the loss reserving process. This paper defines key principles surrounding a well-controlled loss reserving process, and provides an evaluation framework to identify and prioritize opportunities for improvement. The areas addressed in this paper go beyond reserving approaches and data quality to consider the role of management, oversight by the board of directors and audit committee, documentation surrounding the reserve setting process, and financial statement disclosures.

Keywords: Governance, loss reserves, data quality, Sarbanes-Oxley, SOX, Model Audit Rule, Section 404, ASOP 41, ASOP 43, audit committee, controls, gold standard, framework.

1. INTRODUCTION

Pursuant to the Sarbanes-Oxley Act of 2002, publicly held insurance companies are required to have processes and controls surrounding the financial reporting function. U.S. statutory reporting is expected to be subject to a similar requirement in the near future under the Model Audit Rule.

For property/casualty insurers, the estimating and recording of unpaid losses and loss expenses represents a significant part of the financial reporting process. Over the past several years, some insurance companies have taken great strides toward establishing a well-controlled environment surrounding their loss reserving process. Other companies have implemented a lesser degree of control, although generally sufficient to accomplish the requirements for management's Section 302 and Section 404 certifications and to receive an unqualified external audit opinion.

The authors have experience dealing with many types of companies, including large multinational insurance and reinsurance companies. Based on our experiences, we have developed a set of key principles that define a well-controlled reserving process. We have also described a continuum to measure a company's process and overall maturity for each of these key elements relative to an ideal and well-controlled reserving process.

The key principles and maturity framework examples described in this paper are not intended to be exhaustive in nature. Rather, these principles and examples are intended to be illustrative, designed to encourage readers and company management to think more broadly about the issues surrounding their reserving function.

1.1 Research Context

Based on our review of the CAS Research Taxonomy, the focus areas of the casualty actuarial science that this paper is addressing is I. Actuarial Applications and Methodologies, subtitles A. Accounting and Reporting, and I. Reserving. Since this paper focuses on the corporate governance and controls surrounding the loss reserving and financial reporting process, we have not assigned this paper to further subcategories under these areas.

In addition, based on our experiences and based on our viewing of the CAS Web Site for papers related to corporate governance and loss controls, we identified the following paper as existing literature that is relevant to this topic: "Sarbanes-Oxley Section 404 Internal Controls and Actuarial Processes," by Leslie R. Marlo and G. Chris Nyce in Casualty Actuarial Society *Discussion Paper Program*, 2006. While that paper addresses the requirements under Sarbanes-Oxley Section 404, the focus of this paper is on strengthening the corporate governance and control environment beyond the basic requirements of Section 404 to that of an optimal or ideal framework.

1.2 Objective

While Sarbanes-Oxley implemented significant additional requirements, the extent to which companies have developed sound processes and controls around their loss reserving process has varied significantly. This paper will describe the benefits of embracing a strong corporate governance model. In addition, we will define, in principle, best practices associated with a loss reserving process and a framework by which a loss reserving process can be measured using specific considerations to identify and prioritize opportunities for improvement.

1.3 Outline

This paper contains the following sections:

Section 2 describes at a high level the basic steps related to loss reserve controls that companies are required to take under Sarbanes-Oxley, the importance of corporate governance surrounding the loss reserving process, and the value of striving to have a best practices reserving process.

Section 3 defines a best practice, or "gold standard," reserving process, i.e., the characteristics of a company that has a best practices process for each of the fundamental elements of a reserving process.

Section 4 describes a framework for measuring the development or maturity of a reserving process for a company against the Gold Standard described in Section 3.

2. SARBANES-OXLEY AND GOVERNANCE SURROUNDING LOSS RESERVES

For many publicly held insurance companies in the United States, the Sarbanes-Oxley Act of 2002 was effective beginning year-end 2004, initiating a new era in management's obligations surrounding the financial reporting process.

2.1 Processes and Key Controls

This section provides a brief overview of the basic elements of a controlled loss reserving environment. Marlo and Nyce [1] provide a more detailed description of the requirements under Sarbanes-Oxley Section 404.

There are several key steps that management of a company complete when assessing their loss reserving process and control framework. These steps include (1) documenting the loss reserving process, typically including a narrative accompanied by a flowchart, (2) identifying significant risks within the loss reserving process, (3) identifying or implementing appropriately designed "key controls" to mitigate those risks, and (4) evaluating and testing the key controls to ensure they are designed appropriately and are operating effectively. The overarching goal of these steps is to ensure that appropriate controls exist over the financial statement balances.

Documentation of the reserving process includes the key steps that are used by management throughout the entire process, from the retrieval of raw system data for use in actuarial methods to the reserve amounts recorded on the financial statements. These steps would typically include the retrieval of claims data, the reconciliation of that data to financial records, the development of actuarial estimates, management's review and approval of recorded amounts considering the actuarial estimates, and the recording of the approved amounts in the financial statements. Many loss reserving processes have several subprocesses; each of these needs to be documented as well.

Once the loss reserving process is described in a comprehensive manner, the next step is for management to identify risks inherent in that process. These risks include, but are not limited to, the following:

- Claim data is inaccurate or incomplete or both
- Reserving methods or assumptions or both are inappropriate or unreasonable
- Spreadsheet errors are not identified
- Recorded amounts are not reflective of management's best estimate

Once the risks are identified, management then identifies or implements appropriately designed key controls to mitigate the risk of financial statement errors that could be caused by the identified risks. Such key controls may include the following:

- Reconciliation of claim data to financial records of company
- Peer review of actuarial methods and assumptions
- Technical review of analysis and spreadsheets
- Management review and approval of recorded reserve amounts

Once the key controls are established, management routinely tests the key controls for operating effectiveness (i.e., that the control is operating as intended). The effectiveness in the design and operation of these key controls is the cornerstone to having a well-controlled loss reserving process.

2.2 Documentation

Documentation plays an important role in a well-controlled loss reserving process in a number of ways.

Consistent with Actuarial Standard of Practice 41, "Actuarial Communications," (ASOP 41), actuaries are required to maintain documentation of their work in a manner that is sufficient for another actuary practicing in the same field to have the ability to evaluate the methods, assumptions, and judgments used in the loss reserving process.

Documentation also plays an important role in evidencing that a key control was executed. Clear and comprehensive documentation further allows management, their auditor, or another party to evaluate whether the control was executed as intended, i.e., to assess the operating effectiveness of that key control.

Documentation also is maintained to support that the amounts recorded in the financial statements reflect management's best estimate, particularly in cases when management's best estimates differ from actuarial estimates.

2.3 Our Observations

Based on our experiences supporting external audits under the requirements of Sarbanes-Oxley, we believe there is significant value in embracing a strong corporate governance model surrounding the loss reserving process. Some of the benefits include the following:

- Greater efficiency in operation, allowing for more efficient internal reserve reviews and reduced disruption from staff rotation and turnover
- Greater understanding by senior management, the audit committee, and the board of directors of the reserving process
- Reduced risk of reserve misstatement and decreased likelihood of reserve surprises
- Few or no deficiencies in controls
- Smoother interaction with external parties, facilitating a more effective and more efficient external audit and regulatory exam

Companies that operate with a minimum level of controls tend to struggle more often in the areas listed above. Turnover of staff in the loss reserving area tends to cause significant inefficiencies, disruption, and risk. Unexpected loss reserving issues tend to happen more frequently, in some cases each quarter, which leads to frustrated senior management and board members. External audits and regulatory exams tend to be more inefficient, time-consuming, and costly.

3. DEFINING A "BEST PRACTICES" RESERVING PROCESS

What does it mean to have a strong corporate governance model surrounding the loss reserving process? For purposes of describing this we have organized the loss reserving process into eight key elements:

1. Management and board involvement
2. Actuarial staffing and expertise
3. Data quality and reliability
4. General reserving approach
5. Reserving methodology
6. Documentation of reserving process

7. Use of external actuaries
8. Financial statement disclosures

For each of these key elements, we describe below a high-level summary of the characteristics that a best practices, or "gold standard," company would likely have. The examples cited are intended to be illustrative in nature, designed to encourage readers and company management to think more broadly about the issues surrounding the reserving function.

3.1 Management and Board Involvement

Gold standard companies have senior management and audit committees that are strongly committed to the loss reserving process, including the associated financial reporting. Senior management's commitment is demonstrated by the following:

- Prioritizing and committing necessary resources to the reserving process (e.g., staffing, computer systems, etc.);
- Minimizing potential conflicts of interest; (e.g., ensuring sufficient segregation of duties between the reserving and pricing actuaries);
- Understanding the reserving approaches, methods and key assumptions, and challenging these as warranted; and
- Proactively monitoring changes in reserve estimates and understanding the reasons for those changes through internal management reporting.

Senior management formalizes its oversight of the loss reserve process by initiating a reserve committee or equivalent management group. The reserve committee is comprised of key management stakeholders in the reserving process (e.g., finance, underwriting, claims), and is collectively responsible for determining the recorded reserve levels. As such, the committee is governed by a formalized process including a committee charter, and conclusions of the committee are documented and executed (see Section 3.7, Documentation of Reserving Process, for further discussion).

The lead reserving actuary presents the internal reserve package to the reserve committee on a quarterly basis or more frequently. The package includes supporting information sufficient for the reserve committee to make informed judgments and draw conclusions (e.g., support for key reserving assumptions, documentation of changes to key reserving assumptions, changes in indicated ultimate losses by class of business, schedules of loss reserve runoff/accuracy of prior estimates). The package also includes internal or industry benchmarks, some of which are "traditional" to actuarial work, while others may be common to financial reporting or investor analyst research.

The audit committee actively oversees the reserve-setting process by monitoring and evaluating the policies and principles surrounding reserve setting, the internal controls over the reserving process, and the transparency of related disclosures. In this oversight role, the audit committee meets regularly with internal actuaries, reserve committee members, external actuaries, and the external auditors. For a more detailed discussion of what information audit committees could reasonably expect to receive from their actuaries, refer to the report published in September 2007 by the American Academy of Actuaries' Committee on Property and Liability Financial Reporting, titled "An Overview for Audit Committee Members of P/C Insurers: Effective Use of Actuarial Expertise."

3.2 Actuarial Staffing and Expertise

With regard to the internal actuarial loss reserving function, gold standard companies have the following qualities:

- The loss reserving function is staffed by credentialed professionals (e.g., members of the Casualty Actuarial Society and American Academy of Actuaries) who adhere to continuing education requirements. The actuarial staff is encouraged to participate in relevant professional meetings and seminars, and a program supporting professional advancement (e.g., actuarial student program) exists.
- Staffing levels are of sufficient quantity and quality to allow for comprehensive, timely review of the relevant reserving components, and duties are segregated such that separate actuarial individuals are responsible for the primary analysis function, technical review, and supervisory peer review.
- Reserving personnel are independent of those responsible for underwriting and pricing the business; nevertheless, the reserving personnel consider key metrics evaluated in the pricing department (e.g., pricing or rate monitoring processes, expected loss ratios) and relevant items from other departments (e.g., changes in the mix of business, changes in claims settlement objectives, changes in emphasis on legal challenges).
- Reserving personnel have the requisite experience in the specific classes of business assigned to them. In addition, reserving personnel understand the financial reporting standards related to reserves and recognize the specific areas of the reserve process external auditors are required to

evaluate.

- The lead reserving actuary takes ownership over all reserve estimates, even for areas where the primary analysis may not reside in the actuarial department (e.g., asbestos and environmental, catastrophe reserves).
- Inefficiencies from staffing turnover in the actuarial department are minimized by a loss reserving process that is well organized, comprehensively documented, and properly executed. Documentation allows individuals new to the company's reserving process to understand the reserving methods, key assumptions, and historical conclusions.

3.3 Data Quality and Reliability

With respect to actuarial data, there are several consistent themes with companies exhibiting best practices:

- Loss, premium and other actuarial data are usable for estimation purposes as they are captured and contained in the company's systems, facilitating the reconciliation of data used in actuarial analysis to information published in financial statements.
- Computer systems are capable of capturing data in sufficient quality and detail needed for actuarial review. While highly complex claims or unusual coverages present a greater challenge in this regard, the difficulty in estimating liabilities for such exposures makes this capability that much more important.
- Manual data processing, which is subject to backlogs and higher error rates, is minimal or nonexistent. Where manual processing is necessary, adequately documented and controlled procedures are in place to ensure the accuracy and completeness of manual entries.
- The data for actuarial analysis are available in a timely manner for actuarial review and management consideration in the current period's financial results.
- Managing general agent (MGA) and third-party administrator (TPA) interfaces are well controlled and regularly monitored to ensure that data is properly and timely incorporated into the loss reserving process.
- As needed, computer systems permit functional currencies to be accurately recorded and translated at historical or constant exchange rates, as appropriate, for aggregation with other data

for actuarial analysis.

In summary, a company following the Gold Standard has system-generated data directly usable in the actuarial estimation process, and such data is captured in the detail necessary for an actuary to apply a wide variety of actuarial methods.

3.4 General Reserving Approach

Best practices surrounding the general reserving approach involve a number of items such as the frequency of reviews, gross/ceded/net analyses, reasonableness checks and the use of software.

With regard to frequency of actuarial evaluation, there are several key themes of gold standard companies:

- For companies adhering to quarterly reporting requirements, the actuarial reserve evaluation process is performed and finalized on a quarterly basis and in a timely manner before final management decisions are made as to reserves and other financial statement items.
- For relatively straightforward classes of business (i.e., short-tail classes that lend themselves to traditional actuarial methods), full reviews are completed each quarter using data evaluated as of the quarter-ending date (i.e., not on a quarterly lag). For companies with classes of business where the size, complexity and/or long-tail nature of the exposures prohibit a comprehensive review in this time frame, reserve reviews are completed with one quarter lag and are coupled with a rigorous actual-versus-expected analysis for the most recent quarter.
- For nontraditional exposures that may not be suited to traditional actuarial methods (e.g., asbestos, pollution or directors and officers coverages), full reviews are completed at least once per year, with key monitoring statistics using current data considered during the quarterly reporting process.

Gold standard companies have the same rigor of analyses for the reserves prepared gross of reinsurance as they do on a net of reinsurance basis. Further, gross and net analyses, or another combination such as gross and ceded analyses, are prepared concurrently and the results compared for reasonableness. The impact the reserving process has on other financial statement items associated with actuarially determined processes (e.g., reinsurance recoverable, adjustable ceding commissions, additional premiums) is also considered concurrently at this point in the process.

If reserves reflect a discount for the time value of money, the key approaches and assumptions used to calculate such discounted amounts are consistent with the actuarial analysis underlying the selection of the ultimate undiscounted amounts.

Standard outputs from the reserve estimation process include reasonability checks and analytical or diagnostic metrics. These metrics may include loss ratios by accident year, various frequency and severity statistics, or other measures that are helpful to facilitate an understanding of the key drivers of the reserve estimates.

Gold standard companies use consistent and standardized reserving software that has been developed either internally or externally. Such software is well controlled (e.g., protected from inadvertent changes, planned modifications are thoroughly tested and documented, data inputs are separate from calculation modules) but typically contains sufficient flexibility to allow users to apply new methods, if desired. Ad hoc spreadsheets are rare exceptions, but are used with appropriate end-user controls when the flexibility of such a tool is necessary to improve the quality of the estimates. Further, the reserving software facilitates the actuary's documentation of their considerations for assumptions or judgments that deviate from a guideline. Manual hand-offs/transfers (e.g., "copy, paste, value") are negligible to the process.

3.5 Reserving Methodology

Actuaries following the Gold Standard prepare their reserve estimates in a manner consistent with guidance provided by Actuarial Standard of Practice No. 43, "Property/Casualty Unpaid Claim Estimates" (ASOP 43). ASOP 43 provides guidance for many topics surrounding the loss reserve estimation process.

In addition, a gold standard reserving process uses the most suitable methods available for a given circumstance, not just those that are the easiest to apply. Key assumptions are vetted among claims, underwriting, and actuarial management to ensure an appropriate level of exchange of approaches and viewpoints. Further, where multiple business units and/or multiple locations are involved, dedicated teams are built to form a broader or global approach to evaluating consistent parameters of reserving models (e.g., development tail factors, loss trend rates, reserve positions taken on special complex claims) or for exposures that tend to be insured and reinsured globally (e.g., directors and officers, catastrophe reinsurance, high excess clash covers, aviation, etc.).

Finally, the reserving actuaries interact closely with underwriters and pricing actuaries to obtain appropriate price monitoring information as inputs into the reserve estimation process. Considerations should include an evaluation of how a company establishes rate level adequacy, the quality of systems, reports, and documentation of policies regarding the level of discretionary pricing available to the underwriter, the degree of data accuracy and completeness within the price monitoring reports, and the extent of exposure analysis and pricing evaluations within the underwriting audit process.

3.6 Documentation of Reserving Process

Gold standard companies document their reserving process, from the data used in the actuarial analysis, which is reconciled to the financial records of the company, through the compilation and actuarial analysis, and, ultimately, to management's review and approval of amounts recorded in the financial statements. The documentation contains supporting analysis and calculations in sufficient detail for another actuary practicing in the area to follow, consistent with ASOP 41. Additional documentation exists to demonstrate the execution and operating effectiveness of peer review and other controls.

Companies following the Gold Standard record management's best estimate and appropriately document it as such. The recorded amount may or may not equal the internal actuarial indication (or third-party actuarial indication, if there is no internal actuarial indication). In circumstances where the recorded amount equals the actuarial indication, then a record is made by management actively supporting the actuarial indication is its best estimate. In circumstances where the recorded amount does not equal the actuarial indication, then a record is made by management that qualitatively and quantitatively supports, as appropriate, why the recorded amount represents a better amount than the actuarial indication. Further, care is taken to ensure that the recorded amount is still considered to be a reasonable actuarial estimate. Management's record supporting the recorded amounts is both understandable and consistent in principle across reporting periods.

3.7 Use of External Actuaries

Gold standard companies periodically engage third-party actuaries to perform corroborative reserve analyses. Company management understands that the third-party is typically independent and, therefore, is expected to provide a more objective assessment. In addition, third-party actuaries often provide unique information and expertise that may not otherwise be available to company employees, especially with respect to unusual exposures (e.g., asbestos and environmental ground-up reserve analyses).

Company management is engaged throughout the third-party review to understand the reserving methodologies and key assumptions. Companies that do not employ internal reserving actuaries will review the third-party reserve indications, appropriately challenge these indications, consider the results in the reserve-setting process, and document the resulting conclusions even if no changes are made to recorded amounts. For companies that employ internal reserving actuaries but also engage a third-party reserving actuary, the third-party indications are reviewed, meaningful differences between the internal and external indications are understood and documented, and management considers these differences in its reserve setting process with appropriate documentation on the conclusions reached.

The frequency and breadth of third-party reserve analyses depends upon the nature of the liabilities (i.e., long-tail versus short-tail, level of complexity), the perceived value of an independent estimate, and the additional information and/or expertise that the third-party can bring. Companies with more complex exposures have third-party reviews completed no less frequently than once per year. Appropriate controls exist over the data provided to the third-party for analysis.

In addition, input and related advice are regularly sought from the external auditor's actuaries, including but not limited to views on reserve adequacy, effectiveness of controls over the reserving process, and ideas on how to improve efficiency in the reserving process and effectiveness of the financial reporting disclosures.

3.8 Financial Statement Disclosures

Gold standard companies continuously benchmark their financial statement and Management's Discussion and Analysis of Financial Condition and Results of Operations (MD&A) disclosures with the SEC's evolving views on financial statement transparency. In particular, such companies provide clear and understandable disclosures regarding:

- The process management undertakes to determine its recorded reserves;
- The description of management's process for adjusting the liability for unpaid claims and claim adjustment expenses to an amount that is different than the actuarial indication, including the method used to determine the adjustment, the amount of the adjustment and the specific reasons why the adjustment is necessary;
- Either reserve ranges or other key reserve sensitivity metrics or both that provide transparency as to the uncertainty in the estimates, with adequate characterization of the range or metrics provided;
- Presentations of accident year data that are consistent with the underlying actuarial analysis and management's best estimates, regardless of whether the underlying data were analyzed on an accident year, report year, policy year, underwriting year, or calendar year basis;

- Explanations regarding the amounts and reasons for prior period development, even if increases (or decreases) are offset with decreases (or increases) in other lines. Further, the amounts of development attributable to true claims development, premium development, accretion of discount or foreign exchange are determined, presented separately, and appropriately characterized;
- Other information that may useful (e.g., global loss development triangles);

With each of the above items, gold standard companies have controls in place and documentation supporting their disclosures in the same amount of rigor as for the financial statement amounts for loss reserves.

4. MEASURING A LOSS RESERVING PROCESS USING THE MATURITY FRAMEWORK

From our experience, we believe most companies do not operate at the optimal level defined in Section 3, at least not in all of the eight components. Further, we believe that many companies are at different levels of "maturity" as it relates to the individual eight components described above. For example, a company may be very strong with Management and Board Involvement, but not as strong with Data Quality.

4.1 Maturity Framework

To compare each component of a company's process relative to the optimal level defined in Section 3, we consider a maturity framework, in which we assess if the company's process is operating at one of four levels: minimal, developing, accomplished, or optimal. These levels are defined as follows:

Minimal—operating near or at the minimum level needed for management to complete their attestation and for its external auditors to complete their audit.

Developing—reserving process not well standardized, significant changes exist from period to period—the process runs smoothly some of the time but is inefficient or ineffective at other times; numerous gaps and shortcuts exist.

Accomplished—reserving process is well standardized—generally smooth, efficient, and timely; however, some gaps and shortcuts still exist, which are noticeable on occasion.

Optimal—the component operates near or at the optimal level described in Section 3.

After performing an unbiased, objective assessment of a reserving process, the actuaries, management, and the audit committee could work together to identify specific opportunities to improve current processes and then develop appropriate action plans to achieve a stronger corporate governance model.

4.2 Measuring the Loss Reserving Process

To measure a component of the loss reserving process against the maturity framework's levels described above, one approach would be to ask simple questions and develop answers that would correspond to a given maturity level. Several examples of these questions and answers are provided below.

4.2.1 Question - Management and Board Involvement

How committed is senior management to maintaining strong corporate governance over the loss reserving process?

Minimal	Senior management voices commitment, but their actions are vague. Personnel resources tend to be overwhelmed. Systems are often either old outdated or both. Management challenges actuarial results occasionally, but generally only when results are unfavorable.
Developing	Senior management voices commitment and its actions are clear in certain spots. Typically, resources are moderately strained and there is room for improvement. Management challenges results at times – favorable or unfavorable – but is not consistent in its method and process.
Accomplished	Senior management voices commitment and its actions are clear in most areas. Resources are at acceptable levels in all but isolated spots. Management challenges results regularly and understands the process but does not attempt to understand the details.
Optimal	Senior management strongly committed to loss reserving processes; regularly demonstrated by prioritizing and committing necessary resources, by minimizing potential conflicts of interest, by ensuring they understand and challenge reserving approaches, methods, and key assumptions, as warranted.

4.2.2 Question—Actuarial Staffing and Expertise

Are appropriate staffing levels supporting the loss reserving process?

Minimal Staffing levels allow for only annual or semi-annual review; detailed for some lines, high-level review for others. The same individuals often have multiple functions; e.g., one individual might be responsible for the primary analysis function, a self-technical review, and self-peer review.

Developing Staffing levels allow for quarterly review in some areas but are stretched in others—only semi-annual or annual reviews are completed in these areas. Reserve reviews are typically detailed in nature, with some exceptions. Duties are more segregated, although some control functions, such as formal technical review, might not exist.

Accomplished Staffing levels, roles, and responsibilities are sufficient in quality and quantity in most areas; however, several gaps still exist, often in highly specialized areas.

Optimal Staffing levels are of sufficient quantity and quality to allow for comprehensive, timely review of the relevant reserving components, and duties are segregated such that separate individuals are responsible for the primary analysis function, technical review, and supervisor peer review.

4.2.3 Question - Data Quality and Reliability

Many large and complex companies have data quality issues and system limitations; how do these limitations affect the reserving process?

Minimal Actuarial data (e.g., loss, premium) is not captured in sufficient detail for purposes of actuarial analysis for many lines of business, creating difficulties in directly reconciling actuarial data to the financial statements. Complexities of the business have outgrown system capabilities or systems tend to be outdated. Manual "work-arounds" are relatively routine, some of which have effective controls.

Developing Actuarial data is not captured in sufficient detail for purposes of actuarial analysis for some lines of business. Certain systems may be outdated, but the problem is not pervasive. Manual processing with effective controls is common.

Accomplished Actuarial data may not be captured in all cases in sufficient detail for purposes of actuarial analysis, but the problem is generally isolated. System limitations are minor.

Optimal Actuarial data is captured in sufficient detail for purposes of actuarial analysis, allowing for relatively easy reconciliation of the actuarial data to the financial statements. Systems capabilities dovetail with actuarial needs; manual processing is minimal or non-existent.

4.2.4 Question - Documentation of Reserving Process

How complete and comprehensive is the documentation surrounding the actuarial loss reserve estimation process?

Minimal No consolidated report or standard process exists. Actuarial calculations are part of the documentation, and are sometimes accompanied with a memorandum describing the methods and assumptions. Analyses are performed by multiple departments and are not summarized at the reporting segment and/or consolidated level.

Developing No consolidated report exists, although the reserving process is reasonably standardized. Actuarial calculations in final form exist, and typically include an explanatory memorandum as part of the documentation. Analyses are still performed by multiple departments and might be summarized at a high level at the reporting segment or consolidated level or both.

Accomplished While no consolidated, stand-alone report exists, such reports do exist for certain divisions within the company/segment. Results are summarized in some form at the reporting segment or consolidated level or both. Written documentation adequately describes the process, key assumptions, and findings.

Optimal Documentation is standardized and self-contained in a report and clearly leads from the data used in the actuarial analysis (reconciled to the financial records), through the compilation and decision-making process and, ultimately, to the amounts recorded in the financial statements.

4.2.5 Question - Financial Statement Disclosures

Are disclosures in publicly available information describing the company's loss reserving process and uncertainty in reserves effective?

Minimal	The disclosures regarding the reserve estimation process are vague and do not represent clearly the underlying process. The disclosures related to uncertainty surrounding the recorded reserves are overly simplistic and do not explain the relationship of the uncertainty in the actuarial estimates to the resulting risk of reserve variability.
Developing	The disclosures related to reserve estimation generally represent the process used by the company to establish reserves. The disclosures related to the reserve range are understandable, yet rather general and only minimally address the company's particular risks and variability.
Accomplished	The disclosures accurately and clearly describe the process used to establish reserves. If ranges or other metrics are provided, the information is meaningful and generally relates to the company's particular characteristics.
Optimal	The disclosures are clear on the process used to establish reserves and why management chose its particular estimate. Significant differences between recorded amounts and internal actuarial indications, if any, are provided and the reasons for such differences are appropriately described. Reserve ranges or other quantitative measures of variability are provided and described in an understandable manner to a non-actuary.

5. CONCLUSIONS

There are significant advantages to having a strong corporate governance environment and an optimally controlled reserving process. Loss reserves are typically the most significant and uncertain item on a property/casualty insurance company's balance sheet. A reserving process functioning at an optimal level has strong internal controls with few or no deficiencies, reduced risk of reserve misstatement, high-quality documentation of the actuarial analysis, and appropriate management support for the recorded amounts. These factors result in a more effective and efficient external audit, as well as a significantly reduced likelihood of issues arising from the audit of the recorded amounts or testing of internal controls. The benefits go beyond financial reporting, as a strong control environment allows senior management and the audit committee to make better informed company decisions on underwriting, capital allocation, and other business decisions.

Although some companies have a general sense of opportunities to improve upon current practices, few companies have systematically studied the whole actuarial reserving process to assess their current practices in relation to an ideal, best practices reserving process. A complete assessment would identify opportunities and facilitate management's prioritization of key areas to help their company reap the rewards of a stronger corporate governance model.

6. REFERENCES

[1] Marlo, Leslie R., and G. Chris Nyce, "Sarbanes-Oxley Section 404 Internal Controls and Actuarial Processes," Casualty Actuarial Society *Discussion Paper Program*, 2006, 37-65.

Abbreviations and notations

ASOP, Actuarial Standard of Practice
MGA, Managing General Agent
TPA, Third Party Administrator
SEC, Securities and Exchange Commission
MD&A, Management Discussion and Analysis

Biographies of the Authors

E. Daniel Thomas is a principal at PricewaterhouseCoopers LLP in New York. He is a Fellow of the Casualty Actuarial Society (FCAS) and a member of the American Academy of Actuaries (MAAA). He supports audits of property/casualty insurance and reinsurance companies on matters of loss reserves, loss reserve governance process reviews, and reviews of managing general agency practices. He also co-authored "An Investigation of Practical Matters Related to Implementing Fair Value Accounting for Property/Casualty Loss Reserves," which is published in *Fair Value of P&C Liabilities: Practical Implications* by the Casualty Actuarial Society.

Marc Oberholtzer is a director at PricewaterhouseCoopers LLP in Philadelphia. He is a Fellow of the Casualty Actuarial Society (FCAS) and a member of the American Academy of Actuaries (MAAA). He supports audits of property/casualty insurance companies on matters of loss reserves, risk transfer in reinsurance contracts, and other financial reporting matters. He also provides consulting services, including actuarial process reviews, actuarial opinions, merger and acquisitions, and litigation support. He is active in various insurance industry activities and currently serves as the chairperson on the American Academy of Actuaries' Committee on Property and Liability Financial Reporting.

Timothy Landick is a director at PricewaterhouseCoopers LLP in Philadelphia. He is a Fellow of the Casualty Actuarial Society (FCAS) and a member of the American Academy of Actuaries (MAAA). He supports audits of property/casualty insurance companies on matters of loss reserves, design and operating effectiveness of actuarial controls, and other financial reporting matters. He also provides actuarial consulting services, including reserve estimation, legislative analysis, and actuarial process review.

Distribution and Value of Reserves Using Paid and Incurred Triangles

Gary G. Venter, FCAS, MAAA

Abstract

Many loss reserving models are over-parameterized yet ignore calendar-year (diagonal) effects. Venter [1] illustrates techniques to deal with these problems in a regression environment. Venter [2] explores distributional approaches for the residuals. Gluck [3] shows that systematic effects can increase the reserve runoff ranges by more than would be suggested by models fitted to the triangle data alone. Quarg and Mack [4] show how to get more information into the reserve estimates by jointly using paid and incurred data.

This paper uses the basic idea and data from [4] and the methods of [1] to build simultaneous regression models of the paid and incurred data, including diagonal effects and eliminating non-significant parameters. Then alternative distributions of the residuals are compared in order to find an appropriate residual distribution. To get a runoff distribution, parameter and process uncertainty are simulated from the fitted model. The methods of Gluck [3] are then applied to recognize further effects of systematic risk.

Once the final runoff distribution is available, a possible application is estimating the market value pricing of the reserves. Here this is illustrated using probability transforms, as in Wang [5].

Keywords. Reserving Methods; Reserve Variability; Uncertainty and Ranges, Fair Value, Probability Transforms, Bootstrapping and Resampling Methods, Generalized Linear Modeling.

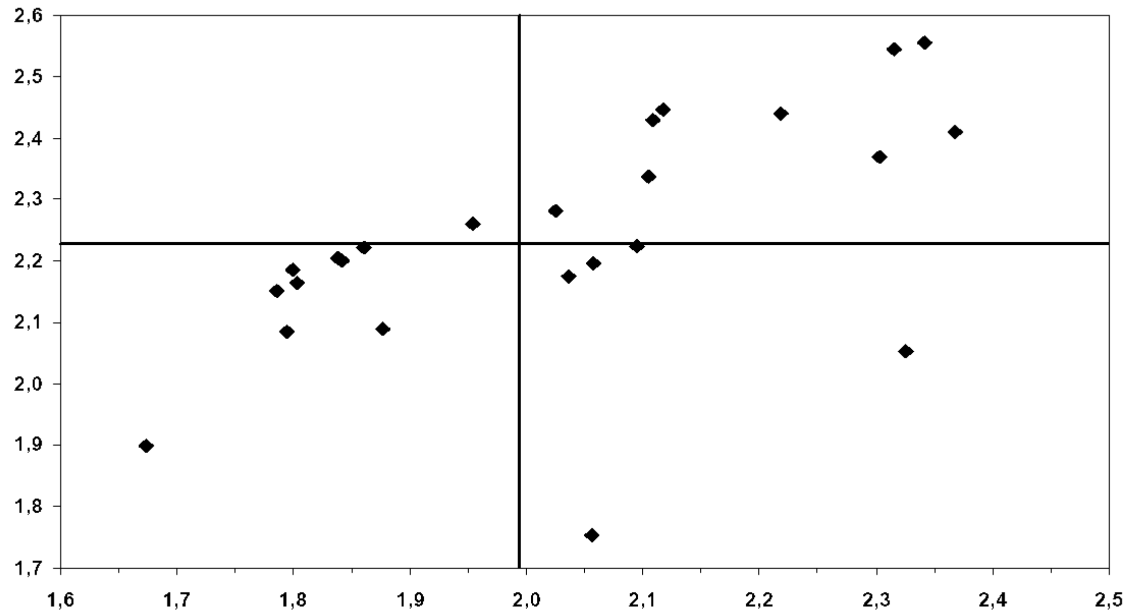
1 INTRODUCTION

Actuaries have used many methods for reconciling reserve estimates from paid and incurred triangles for decades, but formal modeling of paid and incurred simultaneously appears to have begun with Halliwell [6]. His approach was to fit regression models to both data triangles with constraints on the coefficients of both models. More recently Quarg and Mack [4] argue that a high paid-to-incurred ratio for an accident year/lag combination is suggestive of higher-than-average incurred development and lower-than-average paid development in the next period. For instance, some paid factors compared to incurred/paid ratios from http://www.actuaries.org/ASTIN/Colloquia/Zurich/Mack_presentation.pdf are reproduced in Figure 1. In [4] the development factors for paid and incurred are adjusted using these ratios. The formulas are available in Mack [7], who also provides a comparable adjustment for multiplicative cross-classified models.

Verdier and Klinger [8] suggest a modified scheme that recognizes that the impact of the incurred/paid ratios reduces in later stages of development. They also calculate the variance of the result, and suggest a multi-line extension. Jedlicka [9] studies alternative estimation procedures, and outlines a model of paid and unpaid losses instead of paid and incurred, where unpaid = incurred – paid.

Figure 1 – Individual Paid Development Factors as a Function of Incurred/Paid Ratio

The current paper models paid and incurred triangles using a regression framework, but does not



fix the explanatory variables in advance. Rather it is left to the modeler to decide, based on regression diagnostics, which variables best explain each triangle's observations. The regressions are set up with incremental losses as the dependent variable, since these are the new elements that need explanation at each lag. Previous incurred, paid, and unpaid losses, cumulative or incremental, are allowed as independent variables for both triangles. Also diagonal dummies are allowed, in case there are diagonal (i.e., calendar year) effects in the triangles. None of the papers cited above include calendar-year effects, although these are common in development triangles.

Regression modeling is both an art and a science. It is not a model, but a way to build models. Here it is applied to building models of loss development triangles, but many of the issues are more general. The key issue in building regression models is what variables to include. With generalized linear models, another issue becomes what distribution best describes the residuals, and non-linear functions of the regression result become possible.

One criterion for evaluating regression models is the significance of the variables. Typically significance at the 5% level is sought, which is often close to requiring that the estimate be at least twice its standard error. Sometimes this is relaxed a bit, perhaps to the 10% level. Another useful statistic is the standard error of the regression. That incorporates a penalty for additional parameters, so can increase when insignificant variables are added. Usually variables significant at even the 10% level will improve the standard error. The adjusted-R^2 is similarly penalized but it can be difficult to tell if a slight increase is worthwhile. Also there seems to be some ambiguity as to how it is defined for no-constant regressions, which are common in reserve analysis.

These ideas are used to build and apply regression models for reserves when both paid and incurred triangles are available. The data for the continuing example for this paper consists of Tables 1 and 2 from Quarg and Mack [4], and Table 3, which is their difference.

Table 1 – Paid Cumulative Losses

Acc. / Dev	0	1	2	3	4	5	6
0	576	1804	1970	2024	2074	2102	2131
1	866	1948	2162	2232	2284	2348	
2	1412	3758	4252	4416	4494		
3	2286	5292	5724	5850			
4	1868	3778	4648				
5	1442	4010					
6	2044						

Table 2 – Incurred Cumulative Losses

Acc. / Dev	0	1	2	3	4	5	6
0	978	2104	2134	2144	2174	2182	2174
1	1844	2552	2466	2480	2508	2454	
2	2904	4354	4698	4600	4644		
3	3502	5958	6070	6142			
4	2812	4882	4852				
5	2642	4406					
6	5022						

Quarg and Mack suggest that using paid and incurred triangles together can help reconcile their differences and improve the reserve estimates from both. In his discussion at the 2003 ASTIN Colloquium in Berlin, Mack suggested that this could also be done in a regression setting, where both the paid and incurred losses could be used in the regressions for either. This paper follows up on that suggestion, also incorporating the methods of Venter [1] to eliminate statistically insignificant variables and to incorporate any diagonal effects that may be in the data. Alternative distributions for the residuals are also fit. These are the topics of section 2.

Table 3 – Losses Estimated to Be Unpaid at Year-End

Acc. / Dev	**0**	**1**	**2**	**3**	**4**	**5**	**6**
0	402	300	164	120	100	80	43
1	978	604	304	248	224	106	
2	1492	596	446	184	150		
3	1216	666	346	292			
4	944	1104	204				
5	1200	396					
6	2978						

Section 3 addresses the issue of runoff ranges arising from the models developed in section 2. Section 4 widens the runoff ranges to include systematic risk, as discussed by Gluck [3]. Section 5 discusses uses for the resulting distribution, and in particular proposes a method to use the runoff distribution to estimate the value of the reserves.

2 BUILDING MODELS

2.1 Exploratory Analysis

To paraphrase Yogi Berra, you can see a lot about your data just by observing it. The starting point of building a regression model is to explore the relationships that may be in the data. This is what makes this approach difficult to reduce to a strict algorithm, however. Some of the steps that can be used in looking at paid and incurred development triangles are outlined below.

Modeling paid losses as a function of paid and incurred could also include using unpaid losses as an explanatory variable, as unpaid is just the difference between incurred and paid. The first step in this analysis is to look at the data and explore relationships that may exist.

One thing that stands out in the unpaid triangle is that the lag 0 loss for the most recent year is more than double that of any previous year. The incurred is also at an unprecedented level, but the paid is not. That raises a question as to whether or not the latest year represents a significant increase in exposure, or is just an unusual fluctuation. The paid chain ladder estimate for ultimate for year 6 is 6128, compared to 8429 for incurred development, or a difference of 2301. Usually an analyst would know more about the business reasons for such a difference. For instance, there could have been a significant increase in premium volume, or a major loss event, or, on the other hand, a change in reserving methodology that does not affect paid losses. Without such background, only historical data patterns can be used on this point, even though it is quite a bit out of the range of historical observations.

The paid losses could be modeled as a function of the previous incurred, paid, or outstanding, or some combination of those. Here the incremental paid losses at each lag are modeled, as that is the new information at that lag. To start the analysis, the correlations of the paid losses with the previous unpaid and the previous cumulative paid and incurred for the continuing example are shown in Table 4.

Table 4 – Correlation % for Incremental Paid Losses with Previous Cumulative Losses

Incremental Paid at Lag with:	**Incurred**	**Paid**	**Unpaid**
Paid at Lag 1	88	84	70
Paid at Lag 2	68	57	92

Table 4 shows that the lag1 incremental paids correlate most strongly with the previous incurreds, while at lag 2, the correlation is strongest with previous unpaid. At later lags (not shown) the unpaid continue to be strong predictors of the next incremental payments, and interestingly enough, after a few years the percentage of unpaid that is paid in the next year is fairly steady, as shown in Table 5, which is calculated as the sum of paid divided by the sum of previous unpaid column by column. The high factor at lag 1 reflects the continuing reporting of claims after lag 0. The similar factors after lag 2 suggest that only one parameter will be needed for the later lags.

Table 5 – Average Paid at Each Lag as Factor Times Previous Unpaid (Sum/Sum)

Lag	**1**	**2**	**3**	**4**	**5**	**6**
Percent	1.95	0.67	0.33	0.33	0.28	0.36

Since unpaid losses are a strong predictor of the next period's paid losses, a model for projecting future unpaid is needed to fill out the triangles. Unpaid can be calculated from models for incurred and paid losses, or could be modeled directly, say by expressing expected unpaid as a factor times previous unpaid.

In the continuing example, the unpaid losses at lags 1 and 2 have a stronger correlation with previous cumulative paid losses than with previous incurred losses (61% vs. 52% for lag 1 and 47% vs. 42% for lag 2). Preliminary regressions indicated that for lag 1, current incremental paid was significant, but not for lag 2. Also a constant term was significant for lag 1.

For the later lags, the fairly constant ratio of paid to previous unpaid would suggest the same for unpaid to previous unpaid. Due to changing incurred development, this was fairly noisy, however. Table 6 shows the unpaid losses at each later lag as a percentage of the unpaid at the previous lag. There is no clear trend, so it may work to model these all as a single constant percentage, especially given that pattern for paid losses.

Table 6 – Unpaid Losses as a Percent of Previous Unpaid

Acc. / Dev	**3**	**4**	**5**	**6**
0	73	83	80	54
1	82	90	47	
2	41	82		
3	84			

2.2 Regression Analysis

The entire triangle can be set up as a single large regression analysis, either for paid or unpaid losses. This is in effect a series of regressions combined into a single error structure. As an example, for paid losses, the dependent and independent variables for a trial regression are shown in Table 7. This is to explore the structure of the data, but depending on the patterns in the residuals other regressions may be needed to find better models. A reasonable starting point is ordinary multiple regression, which assumes constant variance of the residuals (homoscedasticity). Even though the residuals are not likely to be constant here, as the small increments at the end of the triangle will probably have smaller residuals, such heteroscedasticity usually does not affect the regression coefficients much, although it does affect the overall predictive error distribution.

The coefficients for the three variables in this regression are: 0.818, 0.696, and 0.325, with standard errors of 0.033, 0.131, and 0.264. Thus the first two variables are significant but the third is not. Even though all the lags have similar ratios of paid to previous unpaid on average, the individual ratios are enough different to reduce the significance.

Table 7 – Dependent and Independent Variables for Paid Regression

Paid Increments	**Previous Incurred**	**Previous Unpaid**	**Previous Unpaid**
1228	978	0	0
1082	1844	0	0
2346	2904	0	0
3006	3502	0	0
1910	2812	0	0
2568	2642	0	0
166	0	300	0
214	0	604	0
494	0	596	0
432	0	666	0
870	0	1104	0
54	0	0	164
70	0	0	304
164	0	0	446
126	0	0	346
50	0	0	120
52	0	0	248
78	0	0	184
28	0	0	100
64	0	0	224
29	0	0	80

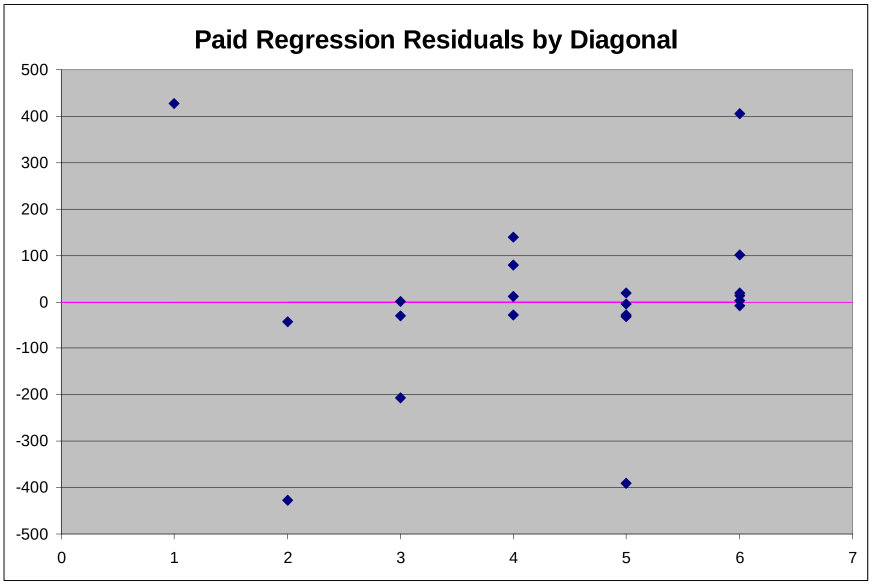
There are also diagonal effects in the residuals. The j^{th} diagonal is the one with row number plus column number = j. It also has j elements. The sum of the residuals by diagonal and number of positive residuals are in Table 8. The residuals by diagonal are graphed in Figure 2. Each diagonal can be seen to be quite biased.

Table 8 – Sum of Residuals and Number of Positive Residuals by Diagonal

Diagonal	**1**	**2**	**3**	**4**	**5**	**6**
Sum	427.5	–470.2	–236.8	200.9	–437.3	532.8
# > 0	1	0	1	3	1	5

There appear to be strong diagonal effects, coming in pairs of years, so offsetting each other over time. Dummy variables can be put in to model diagonal effects. Putting in dummies that are 0 or 1 would give additive effects for each diagonal – essentially adding or subtracting a positive constant for each cell on the diagonal. However, because the incremental paids are of such different sizes, some scaling of diagonal effects would be desirable. For modeling calendar-year effects, it is often more convenient to work with logs of losses, so the effects are automatically multiplicative, as in Barnett and Zehnwirth [10].

Figure 2



Here another method was used to create scaling in the diagonal effects. Since there is only one positive independent observation for each dependent observation in Table 7, and the independent and dependent variables all scale in a similar way, setting the dummy for each dependent variable equal to the positive independent variable would have a scaling effect. Also making a single dummy for each pair of diagonals, with opposite signs on the two diagonals, would reduce the number of variables, possibly without harming the goodness of fit. The matrix of dependent and independent variables for this is in Table 9.

For instance, the variable "*d* 6 – 5" is the dummy variable for diagonals 5 and 6. The observations in that column consist of the value of the independent variable for diagonal 6, its negative for diagonal 5, and 0 elsewhere. The (positive) coefficient for this variable will thus produce a reduction in the fitted values for diagonal 5 and an increase in the values for diagonal 6. This will not be an additive constant, but will be to a large extent scaled to the value of the increment being fitted. The other diagonal dummies work the same way.

The coefficients (not shown) come out quite similar to those for the regression with no diagonal elements, but now all are significant. The standard error of the regression has gone down from 206.6 for the regression without the diagonals to 73.4 with the diagonal dummies. The standard error is penalized for the number of variables, so is a good test to see if adding a variable is helpful. Sometimes regression modelers will keep in a variable that is only weakly significant if it improves the overall standard error.

Table 9 – Dependent and Independent Losses for Paid Regression with Diagonal Pairs

Paid	Incurred	Unpaid	Unpaid	*d*6 - 5	*d*4 - 3	*d*1 - 2
1228	978	0	0	0	0	978
1082	1844	0	0	0	0	-1844
2346	2904	0	0	0	-2904	0
3006	3502	0	0	0	3502	0
1910	2812	0	0	-2812	0	0
2568	2642	0	0	2642	0	0
166	0	300	0	0	0	-300
214	0	604	0	0	-604	0
494	0	596	0	0	596	0
432	0	666	0	-666	0	0
870	0	1104	0	1104	0	0
54	0	0	164	0	-164	0
70	0	0	304	0	304	0
164	0	0	446	-446	0	0
126	0	0	346	346	0	0
50	0	0	120	0	120	0
52	0	0	248	-248	0	0
78	0	0	184	184	0	0
28	0	0	100	-100	0	0
64	0	0	224	224	0	0
29	0	0	80	80	0	0

Separating the diagonal dummies into individual variables for each diagonal did not help the standard error except in the case of diagonals 1 and 2. Putting in individual diagonal elements for them dropped the overall standard error to 63.3. The coefficients are in Table 10. The coefficients for diagonals 1 and 2 can be seen to be quite different in magnitude, so combining them into a single variable gives a worse fit.

Table 10 – Paid Regression Model

Parameter	Estimated	St dev	*t*	Pr(>|*t*|)
Incurred 0	0.8286	0.0107	77.341	0.0000
Unpaid 1	0.6619	0.0406	16.309	0.0000
Unpaid 2 - 5	0.3342	0.0808	4.1340	0.0012
Diagonal 6 – 5	0.1378	0.0155	8.9102	0.0000
Diagonal 4 – 3	0.0326	0.0138	2.3682	0.0341
Diagonal 2	-0.2384	0.0355	-6.7189	0.0000
Diagonal 1	0.4270	0.0656	6.5056	0.0000

Without going into so much detail, a similar process for fitting a model to the unpaid losses led to a regression with independent variables the previous cumulative paid and current paid for lag 1 (with a constant term). Just previous cumulative paid was the explanatory variable for lag 2, and a single variable of previous unpaid was used for the later lags. This means that for lags beyond 2, the expected unpaid was modeled as a constant percentage (here estimated as 66.15%) of the previous unpaid.

The only significant diagonal is diagonal 3, which was modeled with a dummy variable similar to those in Table 9. The problem is that lag 1 itself is a multiple regression with two explanatory variables, so to define the diagonal dummy the rule used for any row was to give it the largest value among the explanatory variables in that row if the row is on diagonal 3, and zero otherwise. The overall standard error of the regression is 77.0. Dropping the diagonal 3 dummy increases the standard error to 92.6, so the dummy helps a good deal. The coefficients and other statistics are in Table 11. Diagonal 4 is not significant but improves the standard error slightly to 76.7. In the end this was not included in the model.

In this model, the high incurred losses for accident year 6 at development 0 will affect the projected paid at development 1, which will go into the estimated unpaid at development 2 and so on. However this is not as dramatic an effect as in the chain ladder, where the high incurred losses in the lower left corner would be multiplied by a large cumulative factor.

Table 11 – Unpaid Regression Model

Parameter	Est value	St dev	*t* student	Prob(>|*t*|)
Paid Cum 0	0.8215	0.1036	7.9316	0.0000
Paid Incrm 1	-0.5436	0.0864	-6.2889	0.0000
Constant 1	522.68	96.860	5.3963	0.0001
Paid Cum 1	0.0766	0.0098	7.8092	0.0000
Unpaid 2 - 5	0.6615	0.0983	6.7315	0.0000
Diagonal 3	0.0800	0.0281	2.8501	0.0128

2.3 Distribution of Residuals

Various distributions can be fit to the selected models by MLE. Typically the distributions are parameterized so that the mean is one of the parameters, and for each cell that is fit as a function of the covariates. All the other parameters of the distribution are constant across all the cells. However for many distributions it can work just as well if some parameter not the mean is a function of the covariates, and the other parameters are still constant.

Typically in generalized linear models, the residuals are modeled as members of the exponential family. These distributions are characterized by expressing the variance of each cell as a function of its mean, often as proportional to the p^{th} power of the cell's mean. However the skewness of the distributions also grows with p, which is not always in accord with the data. In Appendix 1, several distributions are discussed which give the variance of each cell as a multiple of the p^{th} power of the cell's mean by making p one of the parameters of the distribution. Then even with the same value of p, the different distributions can still have heavier or lighter tails, as indicated for instance by skewness.

The Weibull can be used as well, but it is more difficult to adjust its mean-variance relationship as it involves gamma functions, so a p-version was not fit. But the Weibull is an interesting possible residual distribution as it can be fairly heavy-tailed or lighter tailed than the normal, or even negatively skewed, depending on the parameters. It is most easily expressed by its survival function $1 - F(x) = S(x) = \exp[-(x/b)^c]$, and $E[X^j] = b^j(j/c)!$, where $y!$ is short for $\Gamma(1+y)$. The skewness is negative for c above about 3.6. The variance is proportional to the square of the mean, so p is always 2. The regression fit the b for each cell, not the mean.

Table 12 shows the results of fitting several distributions to the paid model. For this data, moving to less skewed distributions increases p and at the same time improves the fit (as measured by log-likelihood, which is equivalent to any of the information criteria such as AIC as all the distributions have the same number of parameters, except the Weibull, which has one fewer but has the best fit anyway). The Weibull, with $c = 7.437$ has skewness of -0.50.

Table 12 – Paid Model Distribution Fits

	p	**– Ln L**	**Skew**
Lognormal-p	1.50	111.94	> 3CV
Gamma-p	1.57	111.23	2CV
ZMCSP-p	1.60	110.52	CV
Normal-p	1.61	109.88	0
Weibull	2	108.76	-0.50

The similar, somewhat abbreviated, results for the model of unpaid losses are in Table 13. That Weibull has $c = 6.037$ and skewness -0.38. These two models will be used to project paid and unpaid losses. This does not imply that the Weibull is better in general. Other data could give quite different distributions.

Table 13 – Unpaid Model Distribution Fits

	p	**–Ln L**	**Skew**
ZMCSP-p	1.96	113.30	CV
Normal-p	2.03	112.93	0
Weibull	2	111.88	-0.38

With a different distribution of residuals, the coefficients for previous unpaid, etc. change a bit from the usual regressions. For the Weibull, the larger cells have higher variances, so higher residuals are not penalized so much there, but the fits are now closer for the smaller cells. For the paid regression this ends up with the diagonal 6 – 5 and diagonal 4 – 3 parameters almost the same. Forcing these to be the same reduces the number of parameters by one but barely affects the loglikelihood, so this change was made. This is done by making a single dummy variable that is the sum of the d 6 – 5 and d 4 – 3 variables in Table 9. As mentioned above, the Weibull fit was for the b parameter, not the mean, so the coefficients have to be multiplied by $(1/c)!$ to get their effect on the mean. Table 14 shows the resulting coefficients for the two models.

Table 14 – Weibull Models' Estimated Covariate Parameters

Paid Parameter	Estimate	Unpaid Parameter	Estimate
Incurred 0	0.7811	**Paid Cum 0**	0.7358
Unpaid 1	0.6854	**Paid Incr 1**	-0.4275
Unpaid 2 - 5	0.3306	**Constant 1**	388.41
Diagonal 6–5+4–3	0.0339	**Paid Cum 1**	0.0908
Diagonal 2	-0.1873	**Unpaid 2 - 5**	0.7234
Diagonal 1	0.3971	**Diagonal 3**	0.0525

The projected mean incurred in Table 15 agrees closely in total with Quarg and Mack [4] except for year 6, for which they are about 1000 higher. Their model seems to give more emphasis to the incurred value for that year than to the paid. This model leans more toward believing the paid, but still ends up higher than year 3, which had more paid at 0. The average of the paid and incurred CL estimates is 7279, halfway between this model and [4]'s.

Table 15 – Completing the Square

Incurred	0	1	2	3	4	5	6
0	978	2104	2134	2144	2174	2182	2174
1	1844	2552	2466	2480	2508	2454	2460
2	2904	4354	4698	4600	4644	4652	4658
3	3502	5958	6070	6142	6158	6169	6177
4	2812	4882	4852	4863	4871	4877	4881
5	2642	4406	4646	4665	4679	4690	4697
6	5022	6182	6656	6685	6707	6722	6733

3 RUNOFF RANGES

The sum of the Weibull estimates in the bottom triangle may be close to being normally distributed, but simulation is usually required to get a good handle on the actual distribution of the runoff losses. The simulation can be divided into parameter risk and process risk components. Distributions for the regression coefficients and the Weibull *c*s can be estimated by either the Fisher information matrix or the bootstrap, as detailed below. Here the information matrix method was used. Parameters can be simulated from the estimated distributions of the parameters, and then the runoff losses can be simulated from the Weibull distributions for each cell.

At the parameter values that maximize the likelihood function, the derivative of the negative log-likelihood (NLL) with respect to each parameter should be zero, but the second derivatives should be positive. This just means that the likelihood surface is flat at the minimum NLL but is curved upwards, which is usual for a minimum value. The mixed second partial derivatives could be anything, however. As detailed in the actuarial exams, the Fisher information matrix is the matrix of all the second derivatives and mixed second partials of the NLL with respect to the parameters. Thus if there are n parameters, it is an nxn matrix. Its matrix inverse is an estimate of the covariance matrix of the parameters.

Bootstrapping could be done by resampling with replacement from the normalized residuals of the fitted triangles to generate new triangles, and refitting the models. Each resampled triangle would give a new set of fitted parameters for the paid and unpaid models. The table of parameters that results from doing this many times would be the estimated empirical parameter distribution. For these models this would probably give some correlation to some of the parameters across the paid and unpaid models, which are uncorrelated under the information matrix method since they come from different models. Also the dependent and independent variables would change with each resampling, which could end up with more parameter diversity as well.

In Tables 7 and 9, label the dependent variables y_j for j = 1, …, 21, and label the corresponding independent variables $x_{i,j}$. In the final models i ranges from 1 to 6. Call the covariate parameters β_i, i = 1, …, 6. The Weibull b parameter for each dependent variable is $b_j = \sum_{i=1}^{6} \beta_i x_{i,j}$. Then the derivative of b_j with respect to β_i is just $x_{i,j}$. Thus, the derivative of NLL = $-\sum_j \ln f(y_j)$ with respect to β_i is $\partial NLL / \partial \beta_i = -\sum_j x_{i,j} \, \partial \ln f(y_j) / \partial b_j$. Similarly, $\partial^2 NLL / \partial \beta_i \partial \beta_k = -\sum_j x_{i,j} x_{k,j} \, \partial^2 \ln f(y_j) / \partial {b_j}^2$ and

$\partial^2 NLL \big/ \partial\beta_i \partial c = -\sum_j x_{i,j} \, \partial^2 \ln f(y_j) \big/ \partial b_j \partial c$ The nice thing about these formulas is that the dependence on *i* or *k* is only in the *x* factor. The rest is just a single column (function of *j*) that comes right from the Weibull.

The Weibull formulas (suppressing *j*) are:

$$\partial \ln f(y) \big/ \partial b = \frac{c}{b}\left[\left(\frac{y}{b}\right)^c - 1\right]$$

$$\partial^2 \ln f(y) \big/ \partial b \partial c = \frac{1}{b}\left[\left(\frac{y}{b}\right)^c \left(1 + c\ln\left(\frac{y}{b}\right)\right) - 1\right]$$

$$\partial^2 \ln f(y) \big/ \partial b^2 = \frac{c}{b^2}\left[1 - (1+c)\left(\frac{y}{b}\right)^c\right]$$

$$\partial^2 \ln f(y) \big/ \partial c^2 = -\frac{1}{c} - \left(\frac{y}{b}\right)^c \left(\ln\left(\frac{y}{b}\right)\right)^2$$

These give the parameter standard deviations and correlation matrices in Tables 16 - 19.

Table 16 – Paid Parameters and Standard Deviations

Paid	**Inc 0**	**Unpd 1**	**Unpd 2-5**	**Diag 6543**	**Diag 2**	**Diag 1**	***c***
Parameters	0.832	0.730	0.352	0.036	-0.200	0.423	7.427
Standard dev	0.050	0.052	0.016	0.014	0.069	0.176	1.392
Ratio	16.70	14.08	22.31	2.49	-2.87	2.40	5.33

Table 17 – Unpaid Parameters and Standard Deviations

Unpaid	**Pd Cum 0**	**Pd Inc 1**	**Const**	**Pd Cum 1**	**Unpd 2-5**	**Diag 3**	***c***
Parameters	0.793	-0.461	418.5	0.098	0.780	0.057	6.037
Standard dev	0.145	0.100	102.9	0.008	0.042	0.022	1.148
Ratio	5.48	-4.62	4.07	11.64	18.55	2.52	5.26

Table 18 – Paid Correlation Matrix

1	0.17	0.00	-0.12	-0.24	-0.28	0.11
0.17	1	0.00	-0.19	-0.62	-0.05	0.14
0.00	0.00	1	0.19	0.01	0.00	0.26
-0.12	-0.19	0.19	1	0.13	0.03	-0.03
-0.24	-0.62	0.01	0.13	1	0.07	-0.08
-0.28	-0.05	0.00	0.03	0.07	1	-0.03
0.11	0.14	0.26	-0.03	-0.08	-0.03	1

Table 19 – Unpaid Correlation Matrix

1	-0.86	0.00	0.02	0.01	-0.03	0.06
-0.86	1	-0.49	0.00	-0.01	-0.05	-0.03
0.00	-0.49	1	-0.02	-0.01	0.07	-0.03
0.02	0.00	-0.02	1	0.07	-0.29	0.29
0.01	-0.01	-0.01	0.07	1	-0.04	0.22
-0.03	-0.05	0.07	-0.29	-0.04	1	-0.09
0.06	-0.03	-0.03	0.29	0.22	-0.09	1

Two simulation steps were done with these parameters. First the parameters were simulated, then the Weibull losses were simulated for each cell in the projected lower triangles.

To simulate the parameters, MLE parameters are asymptotically multivariate normal with the derived correlation matrices. However with small samples like these, the normal approximation might not hold. Simulation experiments have found that lognormal distributions are more realistic for small samples. As one example, the simple Pareto shape parameter, given a known location parameter, is inverse gamma distributed. This gets normal-like for large samples, but is heavier-tailed, as is the lognormal, for small samples. For the lognormal assumption, the absolute value of negative parameters could be assumed to be lognormal. One advantage of the lognormal over the normal is that the simulated parameters will not change signs from the mean parameter, even for remote points in their distributions.

For these reasons the lognormal was used here. To simulate the multivariate lognormal, the correlation matrices were input into a normal copula, and then lognormal marginal distributions applied. This maintains the Kendall's tau and rank correlation, but not the linear correlation, of the parameters. The needed reserve position, calculated as ultimate incurred less current incurred, from the mean parameters is 6212. 10,000 simulations had mean runoff of 6203, with a standard deviation of 801. This gives a CV of 13%. The coefficient of skewness is 3.4%, which is closer to a normal distribution than the 26% for a gamma with the same CV. Thus, the normal might provide a reasonable approximation in this case.

4 OTHER SYSTEMATIC RISK

Loss reserves are subject to inflation and trends in the lawsuit environment that happen between occurrence and payment. Some degree of such trend may be built into the accident year level changes, but this is hardly a full reflection of the risk. The average level of future inflation built into the projections could be off, and in addition there are likely to be year-to-year changes in inflation, perhaps correlated one year to the next. Usually the data in the triangle itself is not sufficient to estimate these systematic risks, so they have to be superimposed afterwards.

An internal study of historical variability in trends and actual runoff, based on U.S. annual statement data for a number of companies and inflation variability, suggested that there is quite a bit more variability in actual runoff than standard reserving models would predict. Also Wright [11] found in a simulation test that runoff ranges from typical methods tend to be too narrow. Gluck [3] proposes ways to incorporate systematic risk elements into insurer financial models in general and loss reserve runoff risk in particular. The model used below is roughly consistent with his approach but the numerical values are for illustration only.

In a single simulation of the runoff, the simulated value for losses paid in accident year w at lag d, and thus in calendar year $w+d$, is multiplied by a simulated factor $H_{w,d}$ given by:

$$H_{w,d} = BD^{w+d-n}E_{w+d}\text{, where:}$$

B is a mean 1 factor for all calendar years that can be thought of as frequency risk; a normal distribution with a standard deviation of 10% is assumed in the example.

D is a lognormal mode 1 draw for all calendar years in the simulation to represent an overall trend error that compounds; n is the last diagonal in the data; a standard deviation of 2% is used for D.

E_{w+d} is generated from an AR-1 model, to represent (ii). The process for E is as follows: The X_is are independent $N(0, \sigma^2)$ random draws, and $\rho \in [0,1]$ is the autocorrelation coefficient. Let $t_1 = X_1$, and $t_{i+1} = \rho t_i + X_{i+1}$. Then $E_{w+d} = \exp\left(\sum_{j=1}^{w+d-n} t_j\right)$. The values $\sigma = 2.5\%$ and $\rho = 70\%$ are used in the numerical example.

Using lognormal mode 1 factors gives an increase in the mean reserve. Actually something similar happens in just multiplying normal mean 1 factors that are positively correlated. This is justified in a few ways. First, an error in trend would compound and the effects on each year would be correlated. Second, new claim types and other superimposed changes tend to have an upward drift. Third, many reserve models, including this one, do not project ongoing calendar-year trends, but these often do affect open claims from previous years.

In a sense, this approach incorporates a degree of model risk, in addition to process and parameter risk. For instance, it is difficult in the fitting to distinguish calendar-year trends from upward and downward individual calendar-year gyrations. This is a model-risk issue. Even if the fitted model has calendar-year trends in it, there is still a question of which trend to project going forward. Thus putting trends into the model does not automatically solve the model-risk problem, and systematic projection risk still needs to be incorporated. The simulated distributions, with annual discount factor 0.96 (a rate of 4 $^1/_6$ %), are in Table 20.

Table 20 – Simulated Moments and Percentiles of Runoff Distribution

Probability	Model	+ Systematic	Discounted
0.4%	4,089	3,766	3,507
1.0%	4,359	4,054	3,760
5.0%	4,881	4,614	4,289
10.0%	5,174	4,930	4,586
25.0%	5,665	5,506	5,119
50.0%	6,203	6,186	5,746
75.0%	6,734	6,941	6,441
90.0%	7,228	7,663	7,107
95.0%	7,515	8,101	7,507
99.0%	8,078	9,056	8,396
99.6%	8,389	9,714	8,924
Mean	6,203	6,258	5,808
Std. Dev.	801	1,076	991
CV	0.13	0.17	0.17
Skewness	0.03	0.40	0.38

Including systematic risk at this level slightly increased the mean, but approximately doubled the variance, increasing the spread both upward and downward. Even the discounted losses with systematic risk were higher than the original model above the 95th percentile. Both systematic risk distributions were slightly more skewed than the gamma with the same CV, which would have skewness of about 0.34, but were less skewed than the lognormal or inverse Gaussian. Possible distributional assumptions that could match these distributions are discussed in Appendix 2.

If several lines are being modeled, it would be reasonable to assume that the systematic risk elements were highly correlated across lines, since they arise largely from external influences. Thus even if the development patterns themselves are not highly correlated, including systematic risk could produce a higher correlation.

5 VALUE OF RESERVES

Once a distribution of reserves has been estimated, what do you do with it? One application is to estimate the financial value of the reserves, which could be useful for market-value accounting or valuation of the entire company. There are a few alternatives for how to do this. For instance, in Australia insurers post the 75th percentile of the distribution as the balance-sheet value. To try to get to a market value, there are two prevailing financial theories: the capital asset pricing model (CAPM) and its generalizations, and arbitrage-free pricing. Typically CAPM-like approaches price only the systematic risk, while arbitrage-free pricing looks at the whole distribution of possible outcomes. There are also traditional actuarial pricing methods, like mean plus a percentage of standard deviation.

Balance-sheet items need to be additive, as users of financial statements like to add and subtract assets and liabilities. If there really were a market for reserve risk, prices would be additive also. Otherwise traders could buy risk, pool it, and sell it for no-risk profits. Arbitrage-free and CAPM prices are additive, and standard deviation loads can be made additive, as shown below. These methods can be subdivided at will and still maintain additivity. Lines can be allocated by state and accident year, and summed to by-state totals, etc. Here only methods that use the entire distribution will be used, but having a model for systematic risk would allow using CAPM-type approaches as well.

Arbitrage-free pricing uses probability transforms of possible events, putting more weight on adverse outcomes, and takes the transformed mean as the price. This is where having a distribution of reserve runoff could be applied. One well-known transform is the Wang [5] transform. This transform applies to the survival function $S(x) = 1 - F(x)$ to produce a transformed survival function $S^*(x)$. In its original form it just translated normal percentiles, so $S^*(x) = \Phi[\lambda+\Phi^{-1}(S(x))]$, where Φ is the standard normal distribution function. A bit different form, first suggested by John Major, is

$S^*(x) = Q_\nu[\lambda+\Phi^{-1}(S(x))]$, where Q_ν is the t distribution with ν degrees of freedom. This puts more weight into the tails of the distribution.

The original normal-normal version will be called the NN transform here. The NN transform moves the probability away from the lower percentiles towards the higher percentiles. The Wang transform generally does this as well, but the heavier t tails can also put more probability into the extreme left tail, even after translating by λ. This is a stronger effect with lower values of ν, because the t approaches the normal for high value of ν. In this transform ν does not have to be an integer, as the beta distribution can be used to calculate the t even for non-integer degrees of freedom. Using the function betadist as defined in Excel, the calculation is $Q_\nu(x) = ½ + ½$ sign(x)betadist[$x^2/(\nu+x^2)$, ½, $\nu/2$].

Once the transformed events probabilities are calculated, the value of the reserves are estimated as the transformed mean. The mean of each accident year can be calculated with the same transformed probabilities. Then the resulting accident-year values add up to the total value. If several lines are being done simultaneously, the transform is done on the aggregate loss probabilities. That gives probabilities for each simulated scenario, then they are applied to the losses for each line and accident year in that scenario. Thus, any correlations gets into the overall value and the individual line values reflect the correlations.

Another transform with theoretical and empirical support (for example, see Venter [12]) is the Esscher transform. While the Wang transform is defined on the aggregate distribution, the Esscher transform is defined on the density or discrete probability function. For density $g(x)$, the transformed density with parameter c is defined by $g^*(x) = g(x)\exp(cx/EX)/\ E[\exp(cX/EX)]$. This transform depends on the distribution being transformed, as the transform at x depends on x. The Wang transform, on the other hand, depends only on $S(x)$. Thus with given parameters ν and λ, any simulation of 10,000 equally-likely events will get the same transformed probabilities.

Since the Wang transform is done on the survival function, a couple of steps are needed to apply it to the scenario probabilities from a simulation. Some of these are a bit arbitrary. The survival function at the k^{th} simulation here is calculated as $k/10{,}001$. This keeps the survival function in the range (0,1), although there are other ways to do that. Then S^* has to be translated back to individual scenario probabilities. To do this, the lowest point was considered to represent the range from zero to half the way, in probability, between it and the second lowest point. Thus, it was assigned probability $1 - [S^*(x_1)+S^*(x_2)]/2$. Then the next point gets the average between the next two midpoints, or $[S^*(x_3) - S^*(x_1)]/2$, etc. Finally the last point gets $[S^*(x_{10,000})+S(x_{9999})]/2$. This forces the probabilities to sum to 1.

The standard deviation loading can be allocated with the Euler method. This method was used by Patrik, Burnegger, and Rüegg [13] for capital allocation and Venter, Major, and Kreps [14] used it for allocation of risk measures, and showed the steps needed to apply Euler's work to random variables. The use for risk measures can be used to allocate standard deviation loading as well. The general approach for a risk measure $\rho(Y)$, where $Y = X_1+\ldots+X_n$ is to allocate to X_k $r(X_k) = \lim_{\varepsilon \to 0} \frac{\rho(Y) - \rho(Y - \varepsilon X_k)}{\varepsilon}$.

The numerator is the reduction in the risk measure from ceding a quota share of ε of X_k. Then $r(X_k)$ is the reduction in $\rho(Y)$ from an incremental reduction in X_k scaled up by ε. Basically it is treating every increment of X_k as the last in. The result of Euler is that the sum of the allocations over all the Xs is the whole risk measure $\rho(Y)$ in the case ρ is homogeneous of degree 1, i.e., $\rho(aY) = a\rho(Y)$. When $\rho(Y)$ = standard deviation(Y), the allocation is shown in [13] to be $r(X_k) = \mathrm{Cov}(X_k,Y)/\rho(Y)$. This is not based on the standard deviations of each component, but rather the component's contribution to the standard deviation of Y.

Even a loading based on a percentile of the distribution can be allocated in this manner. The p^{th} percentile can be expressed as $E[Y|F(y) = p]$. Then the marginal allocation is shown in [13] to be $E[X_k|F(y) = p]$. In a simulation, this would be the value of X_k for the simulation where the probability of Y is p. However this is not a very stable allocation, and in practice the average of simulations for a range around that simulation is used. This is then not truly an allocation of the percentile but an allocation of a range around it, sometimes called blurred value of risk.

All of the pricing measures discussed have a free parameter or two which have to be set to something. In practice some market benchmarking can help establish this. Unlimited portfolio transfers are not usually available, but a limit of twice the mean may be. That is over the 99.99^{th} percentile for this simulation, so may be a good approximation for unlimited. An internal study a few years ago found that many reinsurance treaties are priced at the mean plus one-third to one-half of a standard deviation. Taking the Wang parameters of $\nu = 10$ and $\lambda = 0.47$ gives a loading of close to half a standard deviation, with a discounted market value of 6303.8. This is the 70.7^{th} percentile of the discounted distribution, so would produce a profit slightly more than 70% of the time. Another benchmark is how much capital above the premium could be kept at return of 15%, and the probability level of that capital. In this case, with a profit load of 495, that would be 2807 of capital, which with the premium would get to the 99.93rd percentile. That would be a fairly safe capital level. Thus this is a reasonable value by some benchmarks. Whether or not it is reasonable, in fact, would require more benchmarking against actual deals, however.

The same price would come from an Esscher transform with c = 2.6612. The resulting values for the individual accident years from each methodology are shown in Table 21.

Table 21 – Value of Discounted Reserves by Accident Year for Several Methodologies

	1	2	3	4	5	6	total
Esscher transform	106.2	154.8	311.0	220.5	658.7	4,852.6	6,303.8
Wang transform	106.3	154.9	311.1	221.4	662.3	4,847.6	6,303.8
Standard dev.	106.2	154.6	310.6	220.1	657.4	4,854.9	6,303.8
Percentile alloc	107.1	156.4	316.3	218.4	646.1	4,859.4	6,303.8
Mean discnted	102.1	147.9	294.9	208.8	620.7	4,434.0	5,808.5
Mean undiscnted	112.4	164.6	330.9	235.8	694.9	4,718.9	6,257.7
Cov Disc w total	8,039	13,230	31,058	22,325	72,840	834,834	982,326
Cor Disc w total	43%	40%	43%	42%	50%	97%	100%

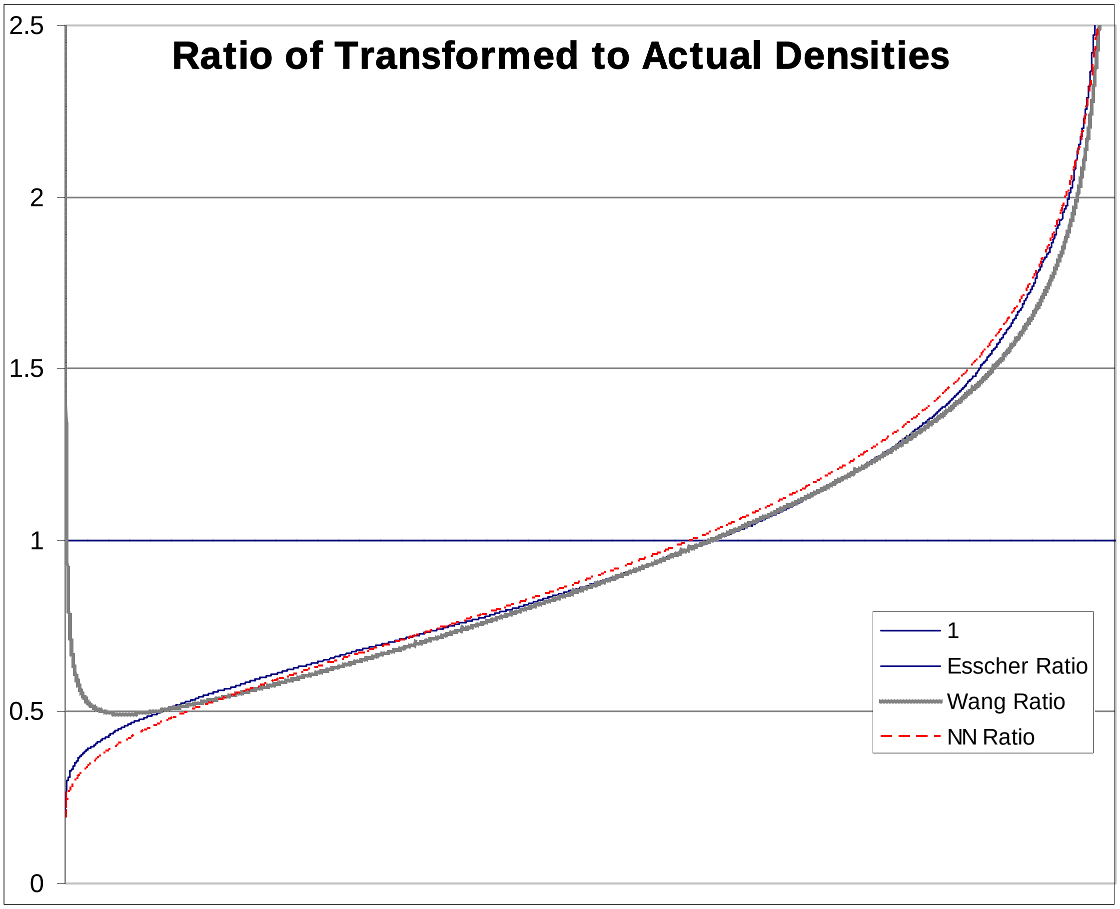
The allocations are all quite similar, but the percentile allocation is slightly different than the others. The Esscher and standard deviation values are closest overall. The percentile allocation is actually the average of 101 simulations centered at the actual percentile value adjusted slightly to balance to the mean.

The ratios of transformed to actual probabilities for the Esscher and Wang transforms are graphed in Figure 3, along with the NN version of the Wang transform, which matches the transformed mean by setting $\lambda = 0.485$.

The Wang transform strengthens both tails, but with ν as high as 10, the left tail strengthening is not great. It also strengthens the right tail quite a bit. For the Esscher, Wang, and NN transforms, the ratios at the second to highest value are 8.4, 13.8, and 5.3. For most of the range, however, the transforms are fairly similar.

The additive methods reviewed here give similar allocations in this case. When some components are more heavy-tailed, there can be greater differences. The importance is in using some kind of additive approach. Further benchmarking would be necessary to see which make most market sense.

Figure 3 – Ratios of Transformed to Actual Probabilities



6 SUMMARY AND FURTHER POSSIBILITIES

A growing body of research is finding that paid and incurred losses can help predict each other. Here a regression approach was used to model paid and unpaid losses, with earlier paid, incurred, or unpaid losses all available as independent variables. It is not asserted that the best possible regression was found. Using paid-to-incurred ratios as independent variables could potentially be useful, for instance. In fact, after the first few lags, unpaid losses were significant in predicting future paid and unpaid, consistent with the suggestion of Jedlicka [9]. Coefficient and overall standard errors were the key regression diagnostics used for evaluating models. Barnett and Zehnwirth [10] recommend residual plots as well, which can be useful but sometimes require regression experience to evaluate.

Once a reasonable regression model was found, MLE was used to evaluate other residual distributions. This requires a non-linear optimization routine. Weibull residuals with slight negative skewness and variance proportional to the mean squared maximized the likelihood. This is a bit surprising, as if you think of the cells as compound frequency-severity processes, positive skewness and variance proportional to a lower power of the mean would be more anticipated.

Once in the world of non-linear optimization, other models become possible as well. For instance, the cell means could be linear functions of independent variables times diagonal effects times some power of the paid-to-incurred ratio, possibly with the power declining for later lags as in Verdier and Klinger [8], plus an additive residual. However this kind of modeling would not readily be able to take advantage of the ease of linear regression software for exploratory analysis. If the observations are all positive, as would be likely with paid and unpaid data, the regression steps could be done in logs, then a multiplicative model with additive residuals fit later if needed for a good residual distribution. Another modeling approach worth pursuing is the idea of Mack [7] to look and paid and incurred development in cross-classified multiplicative models.

The information matrix from MLE was used to estimate parameter uncertainty, with the selection of a lognormal distribution of parameters due to the small sample sizes. Bootstrapping is certainly an alternative here, and may be preferable in that it can pick up possible correlations among the parameters of the two different models. In fact Liu and Verrall [15] have already used bootstrapping for the model of Quarg and Mack. Bootstrapping for the model here would be a bit more computationally intensive than usual, due to the non-linear multivariate optimization at each step, but would have another advantage in that the choice of normal vs. lognormal parameter errors would not be needed.

Systematic risk, including model risk, is clearly an issue in reserve modeling, and historical loss development volatility has been substantial, with even more variability than standard models might suggest. This was reflected here with selected distributions, but should be studied in a more formal way. Similarly, reserve value was illustrated with some rough benchmarks, but more research into the market value of loss reserve risk is called for. The methods of transformed distributions and Euler for producing additive market-values were illustrated. In this case they were not so different, once an overall market value was established.

REFERENCES

[1] Venter, Gary. 2007a. "Refining Reserve Runoff Ranges," Casualty Actuarial Society *E-Forum*, Summer 2007.

[2] Venter, Gary. 2007b. "Generalized Linear Models beyond the Exponential Family with Loss Reserve Applications," CAS *E-Forum*, Summer 2007.

[3] Gluck, Spencer, "A Multiline Risk Factor Model," Paper presented at the 2007 ASTIN Colloquium, Or-

lando, Florida, http://www.actuaries.org/ASTIN/Colloquia/Orlando/Papers_EN.html

[4] Quarg, Gerhard and Thomas Mack. 2004. "Munich chain-ladder—A reserving method that reduces the gap between IBNR projections based on paid losses and IBNR projections based on incurred losses." *Blätter DGVFM,* Heft 4, 597–630.

[5] Wang, Shaun. 2004. "Cat Bond Pricing Using Probability Transforms," Geneva Papers: Etudes et Dossiers, special issue on "Insurance and the State of the Art in Cat Bond Pricing," January 2004, No. 278:19-29. http://www.gsm.pku.edu.cn/stat/public_html/ifirm/reports/Cat%20bond%20pricing%20using%20probability%20transforms%2001-2004.pdf

[6] Halliwell, Leigh J. 1997. "Conjoint Prediction of Paid and Incurred Losses," *CAS Forum*, Summer, 241-380.

[7] Mack, Thomas. 2005. "Recent Developments in Claims Reserving," Presented at the 2005 ASTIN Colloquium, Zurich, Switzerland, http://www.actuaries.org/ASTIN/Colloquia/Zurich/Mack_presentation.pdf

[8] Verdier, Bertrand and Artur Klinger. 2005. "JAB Chain: A model-based calculation of paid and incurred loss development factors," Presented at the 2005 ASTIN Colloquium, Zurich, Switzerland, http://www.actuaries.org/ASTIN/Colloquia/Zurich/Verdier_Klinger.pdf.

[9] Jedlicka, Petr. 2007. "Various extensions based on Munich Chain Ladder method," Presented at the 2007 ASTIN Colloquium, Orlando, Florida, http://www.actuaries.org/ASTIN/Colloquia/Orlando/Presentations/Jedlicka.pdf.

[10] Barnett, Glen and Ben Zehnwirth. 2000. "Best Estimates for Reserves," *Proceedings of the Casualty Actuarial Society* 87:245-321.

[11] Wright, Thomas. 2007. Chapter 9 and Appendix B of "Report of "Best Estimate and Reserve Uncertainty" Working Party (chair: Lis Gibson)," GIRO.

[12] Venter, Gary. 2004 "Discussion of 'Distribution-Based Pricing Formulas Are Not Arbitrage-Free,'" *Proceedings of the Casualty Actuarial Society* 91:25-32, http://www.casact.org/pubs/proceed/proceed04/04014-2.pdf.

[13] Patrik, Gary, Stefan Burnegger, and Marcel Beat Rüegg. 1999. "The Use Of Risk Adjusted Capital To Support Business Decision-Making," *CAS Forum*, Summer, 243-391.

[14] Venter, Gary, John Major, and Rodney Kreps. 2006. "Marginal Decomposition of Risk Measures" ASTIN Bulletin 36(2):375-413.

[15] Liu, Huijuan, and Richard Verrall 2007. "A Dynamic Approach for Predictive Distributions of Reserves Using Paid and Incurred Claims," Presented at the 2007 IME Conference, Piraeus, Greece, http://stat.unipi.gr/ime2007/abstracts.php.

Biography of the Author

Gary Venter is managing director at Guy Carpenter, LLC. He has an undergraduate degree in philosophy and mathematics from the University of California and an MS in mathematics from Stanford University. He has previously worked at Fireman's Fund, Prudential Reinsurance, NCCI, Workers Compensation Reinsurance Bureau, and Sedgwick Re, some of which still exist in one form or another. At Guy Carpenter, Gary develops risk management and risk modeling methodology for application to insurance companies. He also teaches graduate seminars in loss modeling at Columbia University.

917.937.3277 gary.g.venter@guycarp.com

Appendix 1 – *p*-Distributions

In Venter [2] parameters are added to standard distributions to specify the relationship of variance and mean when used with covariates. Typically a distribution will be re-parameterized so that the mean is a parameter, and that parameter will be a function of the covariates. The other parameters will be constant for all the observations. Many distributions can be parameterized so that the variance of each observation will be proportional to any desired power of the mean. That power parameter can then be estimated by MLE to get an idea of how the residuals' variances relate to their means for a given data set. Not all the distributions below are in [2], and some are parameterized a bit differently here. Although they can each produce a variance proportional to any desired power of the mean, they differ in other shape features, such as skewness. Sometimes the over-dispersed Poisson (ODP) is defined as any distribution where the variance is proportional to the mean. Under that definition, any of these distributions can be an ODP just by taking $p=1$. However they will differ in other shape characteristics. The distributions below are by increasing skewness.

Normal-*p*

The normal distribution is typically parameterized with mean μ and variance σ^2. Introducing two new parameters k and p, it can be re-parameterized just by setting $\sigma^2 = k\mu^p$. It then has log density $\ln f(x) = -½\ln(2\pi k\mu^p) - (x-\mu)^2/(2k\mu^p)$. With k and p constant across all observations, each observation's variance will just be k times its mean raised to the p. The skewness is 0.

ZMCSP-*p*

The zero-modified continuous scaled Poisson, as discussed in Venter [2] and Mack [15], is the Poisson distribution function extended to the positive reals, plus a scaling factor, with the probability at 0 set to the value needed to bring the entire probability to 1. It has variance close to proportional to the mean and skewness close to the coefficient of variance (CV), which is the ratio of standard deviation to mean. It is a continuous form of ODP that retains much of the shape of the Poisson distribution. The density can be written as:

$$f(x) = e^{-k\mu^{2-p}}\left(k\mu^{2-p}\right)^{kx\mu^{1-p}} k\mu^{1-p} / \Gamma\left(1 + kx\mu^{1-p}\right).$$

For large means, the mean, variance, and skewness are very close to μ, μ^p/k, and CV. For smaller means, a small adjustment is needed. See [2] for details.

Tweedie

The Tweedie distribution has p between 1 and 2, and the skewness is pCV. It is actually a special case of the Poisson-gamma aggregate distribution with the frequency and severity means coordinated. Starting with a Poisson in λ and a gamma in θ and α, introduce new parameters p, μ, and k with $p = 1+1/(\alpha+1)$, $\lambda = k\mu^{2-p}$, and the severity mean $\alpha\theta = \mu^{p-1}/k$. Then the aggregate mean is μ. Since α is positive, p is between 1 and 2, so both the frequency and severity means are increasing functions of μ. Thus a higher overall mean in a cell is a combination of a higher frequency mean with a higher severity mean. The aggregate variance turns out to be $\mu^p/[(2-p)k]$. Fitting by MLE is discussed in [2].

Gamma-*p*

The gamma distribution is usually parameterized $F(x,\theta,\alpha) = \Gamma(x/\theta;\, \alpha)$ with the incomplete gamma function Γ. This has mean $\alpha\theta$ and variance $\alpha\theta^2$. To get the mean to be a parameter, set $F(x,\mu,\alpha) = \Gamma(x\alpha/\mu;\, \alpha)$. Then the variance is μ^2/α and μ is still a scale parameter. For the gamma-p, take $F(x;\mu,k,p) = \Gamma[x/(k\mu^{p-1});\, \mu^{2-p}/k]$, which has mean μ and variance $k\mu^p$, with skewness = 2CV.

Lognormal-*p*

The usual parameterization of the lognormal is: $F(x;\mu,\sigma) = N\left(\frac{\ln(x)-\mu}{\sigma}\right)$. This has mean $e^{\mu+\sigma^2/2}$ and variance $e^{2\mu+\sigma^2}\left(e^{\sigma^2}-1\right)$. Now reparameterize with three parameters p, m and s:

$$F(x;m,s,p) = N\left(\frac{\ln\left((x/m)\sqrt{1+s^2m^{p-2}}\right)}{\sqrt{\ln\left(1+s^2m^{p-2}\right)}}\right).$$

This has mean m, variance s^2m^p, and skewness 3CV+CV^3, where CV = $sm^{p/2-1}$. Here μ has been replaced by $\ln\left(\frac{m}{\sqrt{1+s^2m^{p-2}}}\right)$ and σ^2 by $\ln(1+s^2m^{p-2})$.

Appendix 2 – Possible Distributions for Simulations

Some of the work on loss reserve risk is on moments only, so having simulated distributions can provide a test of different parameterized distributions. In this case there are three simulated distributions: the original model, that plus systematic risk, and that discounted. The CV for the first is 13% and for the other two is 17%. The skewnesses are 3%, 40%, and 38%, respectively. For two-parameter distributions, the CV and skewness are often determined by only one of the parameters, so they become functions of each other as well. The skewness for CVs of 13% and 17% for a few common distributions is shown in Table A2-1.

Table A2-1 – Skewness for CVs of 13% and 17%

Distribution CV:	**13.0%**	**17.0%**
Normal	0	0
Weibull	-45.6%	-60.1%
Poisson	13.0%	17.0%
Gamma	26.0%	34.0%
Inverse Gaussian	39.0%	51.0%
Lognormal	39.2%	51.5%

The skewness for the original model is closest to, but higher than, that of the normal. For the other models, it is closest to but somewhat higher than that of the gamma.

A convenient distribution for matching three moments is the shifted gamma. $X - a$ is gamma in θ and β, so $EX = a + \theta\beta$, $\mathrm{Var}X = \theta^2\beta$, and skewness $= 2\beta^{-1/2}$. For a positively skewed distribution it is always possible to solve for the three parameters in terms of these moments, but the shift can be negative, giving positive probability to negative values of X. If the skewness is 2 or greater, the gamma has its mode at zero, and the density declines from there, which may not be a realistic shape in some cases. Then perhaps a shifted-lognormal or power-transformed beta or gamma may work better. In terms of the moments the parameters are: $\beta = (2/\mathrm{skw})^2$; then $\theta = \mathrm{Var}/\beta = \mathrm{stdev}*\mathrm{skw}/2$; and $a = \mathrm{mean} - \theta\beta$. The parameters for the three distributions simulated are in Table A2-2.

Table A2-2 – Shifted gamma parameters for simulated distributions

	Model	**+Systematic**	**Discounted**
a	-41,581.7	910.01	580.25
θ	13.424	216.58	187.89
β	3559.6	24.69	27.83

The parameters for the original model look strange, but in fact the probability of a negative result is less than 10^{-15}. The shifted-gamma probabilities for selected percentiles of the simulated distributions are shown in Table A2-3.

Table A2-3 – Shifted gamma probabilities for simulated percentiles

Probability	**Model**	**+Systematic**	**Discounted**
0.40%	0.38%	0.30%	0.32%
1.00%	1.00%	0.93%	0.92%
5.00%	4.84%	4.94%	4.96%
10.00%	9.88%	9.94%	10.01%
25.00%	25.19%	25.28%	25.30%
50.00%	50.23%	50.02%	49.99%
75.00%	74.75%	75.05%	75.09%
90.00%	89.92%	89.84%	89.92%
95.00%	94.84%	94.67%	94.73%
99.00%	98.98%	98.96%	99.02%
99.60%	99.65%	99.72%	99.69%

The fits are fairly good between 1% and 99%, with a little fading off in both far tails. This is not a given from matching three moments, because other distributions matching the same moments could have fairly different shapes. The shifted gamma may or may not fit as well to other development triangle runoff distributions.

Robustifying Reserving

Gary G. Venter, FCAS, MAAA, and Dumaria R. Tampubolon, Ph.D.

Abstract

Robust statistical procedures have a growing body of literature and in actuarial applications have been applied in loss severity fitting. Here an introduction of robust methods is made for loss reserving. In particular, following Tampubolon [1], reserve models for a development triangle are compared based on the sensitivity of the reserve estimates to changes in individual data points. This is then related to the generalized degrees of freedom used by the model at each point.

Keywords. Loss reserving; regression modeling; robust, generalized degrees of freedom.

All models are wrong, but some are useful.
Christian Dior (or maybe George E. P. Box)

0 INTRODUCTION

The idea of this paper is simple: for models of a loss development triangle, look at the derivative of the loss reserve with respect to each data point. All else being equal, models that are highly sensitive to a few particular observations are less preferred than ones that are not. This is supported by the fact that individual cells can be highly unstable. This general approach, based on Tampubolon [1], is along the lines of robust statistics, so some background into robust statistics will be the starting point. Published models on three data sets will be tested by this methodology. For two of them, unsuspected problems with the previously best-fitting models are found, leading to improved models.

The sensitivity of the reserve estimate to individual points is related to the power of those points to pull the fitted model towards them. This can be measured by what Ye [2] calls generalized degrees of freedom (GDF). For a model and fitting procedure, the GDF at each point is defined as the derivative of the fitted point with respect to the observed point. If any change in a sample point is matched by the same change in the fitted, the model and fitting procedure are giving that point full control over its fit, so a full degree of freedom is used. GDF does not fully explain the sensitivity of the reserve to a point, as the position of the point in the triangle also gives it more or less power to change the reserve estimate, but it adds some insight into that.

Section 1 provides some background into robust analysis and section 2 shows some previous applications to actuarial problems. These help to place the current proposal into perspective in that literature. Sections 3, 4, and 5 apply this approach to some published development models. Section 6 concludes.

1 ROBUST METHODS IN GENERAL

Classical statistics takes a model structure and tries to optimize the fit of data to the model under the assumption that the data is in fact generated by the process postulated in the model. But in many applied situations, the model is a convenient simplification of a more complex process. In this case the optimality of estimation methods like maximum likelihood (MLE) may no longer hold. In fact a few observations that do not arise from the model assumptions can sometimes significantly distort the estimated parameters when standard techniques are used. For instance, Tukey [3] gives examples where even small deviations from the assumed model can greatly reduce the optimality properties. Robust statistics looks for estimation methods that in one way or another can insulate the estimates from such distortions.

Perhaps the simplest such procedure is to identify and exclude outliers. Sometimes outliers clearly arise from some other process than the model being estimated, and it may even be clear when current conditions are likely to generate such outliers, so that the model can then be adjusted. If the parameter estimates are strongly influenced by such outliers, and the majority of the observations are not consistent with those estimates, it is reasonable to exclude the outliers and just be cautious about when to use the model.

An example is provided by models of the US one-month Treasury bill rates at monthly intervals. Typical models postulate that the volatility of the rate is higher when the rate itself is higher. Often the volatility is proposed to be proportional to the p^{th} power of the rate. The question is – what is p? One model, the CIR or Cox, Ingersoll, Ross model, takes $p = ½$. Other models postulate p as 1 or even 1.5, and others try to estimate p as a parameter. An analysis by Dell'Aquila et al. [4] found that when using traditional methods, the estimate of p is very sensitive to a few observations in the 1979-82 period, when the US Federal Reserve bank was experimenting with monetary policy. Including that period in the data, models with p=1.5 cannot be rejected, but excluding that period finds that $p = ½$ works just fine. That period also experienced very high values of the interest rate itself, so their analysis suggests that using $p = ½$ unless the interest rate is unusually high would make sense.

A key tool in robust statistics is the identification of influential observations, using the influence function defined by Hampel [5]. This looks at statistics calculated from a sample, such as estimated parameters, as functionals of the random variables that are sampled. The influence function for the statistic at any observation is a functional derivative of the statistic with respect to the observed point. In practice, analysts often use what is called the empirical influence. For instance, Bilodeau [6] suggests calculating that at each sample point as the sample size times the decrease (which may be negative) in the statistic from excluding the point from the sample. That is, the influence is n times [statistic with full sample minus statistic excluding the point]. If the statistic is particularly sensitive to a single or a few observations, that calls its accuracy into question. The gross error sensitivity (GES) is defined as the maximum absolute value of the influence function across the sample.

The effect on the statistic of small changes in the influential observations is also a part of robust analysis, as these effects should not be too large either. If each observation has a substantial randomness to it, the random component of influential observations would be having a disproportionate impact on the statistic. The approach used below in the loss reserving case is to identify observations for which small changes have large impacts on the reserve estimate.

Exclusion is not the only option for dealing with outliers. Estimation procedures that use but limit the influence of the outliers are also an important element of robust statistics. Also finding alternative models that are not dominated by a few influential points and estimating them by traditional means can be an outcome of a robust analysis. In the interest rate case, a model with one p parameter for October 1979 through September 1982 and another elsewhere does this. Finding alternative models with less influence from a few points is what we will be attempting in the reserve analysis.

2 ROBUST METHODS IN INSURANCE

Several papers on applying robust analysis to fitting loss severity distributions have appeared in recent years. For instance, Brazauskas and Serfling [7] focus on estimation of the simple Pareto tail parameter α assuming that the scale parameter b is known. In this notation the survival function is $S(x) = (b/x)^{\alpha}$. They compare several estimators of α, such as MLE, matching moments or percentiles, etc. One of their tests is the asymptotic relative efficiency (ARE) of the estimate compared to MLE, which is the factor which when applied to the sample size would give the sample size needed for MLE to give the same asymptotic estimation error. Due to the asymptotic efficiency of MLE, these factors are never greater than unity, assuming the sample is really from that Pareto distribution.

The problem is, however, that the sample might not be simple Pareto. Even then, however, you would not want to identify and eliminate outliers: whatever process is generating the losses would be expected to continue, so no losses can be ignored.[1] The usual approach to this problem is thus to find alternative estimators that have low values of the GES and high values of ARE. Brazauskas and Serfling [7] suggest estimators they call generalized medians (GM). The k^{th} generalized median is the median of all MLE estimators of subsets of size k of the original data. That can be fairly calculation-intensive, however, even with $k = 3, 4$, or 5.

Finkelstein et al. [8] define an estimator they call the probability integral transform statistic (PITS) which is quite a bit easier to calculate but not quite as robust as the GM. It has a tuning parameter t in (0,1) to control the trade-off between efficiency and robustness. Since $(b/x)^{\alpha}$ is a probability, it should be distributed uniform [0,1]. Thus $(b/x)^{t\alpha}$ should be distributed like a uniform raised to the t power. The average of these over a sample is known to have expected value $1/(t+1)$, so the PITS estimator is the value of β for which the average of $(b/x)^{t\beta}$ over the sample is $1/(t+1)$. This is a single-variable root-finding exercise. Finklestein et al. give values of the ARE and GES for the GM and PITS estimators, shown in Table 1. A simulation suggests that the GES for MLE for $\alpha = 1$ is about 3.9, and since its ARE is 1.0 by definition, PITS at 0.94 ARE is not worthwhile in this context. In general the generalized median estimators are more robust by this measure.

Other robust severity studies include Brazauskas and Serfling [9] who use GM estimation for both parameters of the simple Pareto, Gather, and Schultze [10] who show that the best GES for the exponential is the median scaled to be unbiased, but this has low ARE, and Serfling [11] who applies GM to the lognormal distribution.

Table 1: Comparative efficiency and robustness of two robust estimators of Pareto α

ARE	GM-k	PITS-t	GM-GES	PITS-GES
0.88	3	0.531	2.27α	2.88α
0.92	4	0.394	2.60α	3.54α
0.94	5	0.324	2.88α	4.08α

[1] A related problem is contamination of large losses by a non-recurring process. The papers on robust severity also address this, but it is a somewhat different topic than fitting a simple model to a complex process.

3 ROBUST APPROACH TO LOSS DEVELOPMENT

Omitting points from loss development triangles can sometimes lead to strange results, and not every development model can be automatically extended to deal with this, so instead of calculating the influence function for development models, we look at the sensitivity of the reserve estimate to changes in the cells of the development triangle, as in Tampubolon [1]. In particular, we define the impact of a cell on the reserve estimate under a particular development methodology as the derivative of the estimate with respect to the value in the cell. We do this for the incremental triangle, so a small change in a cell affects all subsequent cumulative values for the accident year. This seems to make more sense than looking at the derivative with respect to cumulative cells, whose changes would not continue into the rest of the triangle.

If you think of a number in the triangle as its mean plus a random innovation, the derivative with respect to the random innovation would be the same as that with respect to the total, so a high impact of a cell would imply a high impact of its random component as well. Thus models with some cells having high impacts would be less desirable. One measure of this is the maximum impact of any cell, which would be analogous to the GES, but we will also look at the number of cells with impacts above various thresholds in absolute value.

This is just a toe in the water of robust analysis of loss development. We are not proposing any robust estimators, and will stick with MLE or possibly quasi-likelihood. Rather we are looking at the impact function as a model selection and refinement tool. It can be used to compare competing models of the same development triangle, and it can identify problems with models that can guide a search for more robust alternatives. This is similar to finding models that work for the entire history of interest rate changes and are not too sensitive to any particular points.

To help interpret the impact function, we will also look at the generalized degrees of freedom (gdf) at each point. This is defined as the derivative of the fitted value with respect to the observed value. If this is near 1, the point's initial degree of freedom has essentially been used up by the model. The gdf is a measure of how much a point is able to pull the fitted value towards itself. Part of the impact of a point is this power to influence the model, but where it appears in the triangle also can influence the estimated reserve. Just like with the impact function, high values of the gdf would be a detriment.

For the chain ladder (CL) model, some observations can be made in general. All three corners of the triangle have high impact. The lower left corner is the initial value of the latest accident year, and the full cumulative development applies to it. Since this point does not affect any other calculations, its impact is the development factor, which can sometimes be substantial. The upper right corner usually produces a development factor which, though small, applies to all subsequent accident years, so its impact can also be substantial. When there is only one year at ultimate, this impact is the ratio of the sum of all accident years not yet at ultimate, developed to the penultimate lag, to the penultimate cumulative value for the oldest accident year. The upper left corner is a bit strange in that its impact is usually negative. Increasing it will increase the cumulative loss at every lag, without affecting future incrementals, so every incremental-to-previous-cumulative ratio will be reduced. The points near the upper right corner also tend to have high impact, and those near the upper left tend to have negative impact, but the lower left point often stands alone in its high impact.

The GDFs for CL are readily calculated when factors are sums of incrementals over sums of previous cumulatives. The fitted value at a cell is the factor applied to the previous cumulative, so its derivative is its previous cumulative times the derivative of the factor with respect to the cell value. But that derivative is just the reciprocal of the sum of the previous cumulatives, so the gdf for the cell is its previous cumulative over the sum. Thus these GDFs sum down a column to unity, so each development factor uses up a total gdf of 1.0. Essentially each factor uses 1 degree of freedom, agreeing with standard analysis. The average gdf in a column is thus the reciprocal of the number of observations in that column. Thus the upper right cell uses 1 gdf, the previous column's cells use ½ each on average, etc. Thus the upper right cells have high GDFs and high impact.

We will use ODP to refer to the cross-classified development model in which each cell mean is modeled as a product of a row parameter and a column parameter, the variance of the cell is proportional to its mean, and the parameters are estimated by quasi-likelihood. It is well known that this model gives the same reserve estimate as CL. Thus if you change a cell slightly, the changed triangle will give the same reserve under ODP and CL. Thus the impacts of each cell under ODP will be the same as those of CL. The GDFs will not be the same, however, as the fitted values are not the same for the two models. The CL fitted value is the factor times the previous cumulative, whereas the ODP cumulative fitted values are backed down from the latest diagonal by the development factors, and then differenced to get the incremental fitted. It is possible to write down the resulting GDFs explicitly, but it is probably easier to calculate them numerically.

It may be fairly easy to find models that reduce the impact of the upper right cells. Usually the development factors at those points are not statistically significant. Often the development is small and random, and is not correlated with the previous cumulative values. In such cases, it may be reasonable to model a number of such cells as a simple additive constant. Since several cells go into the estimation of this constant, the impact of some of them is reduced. Alternatively the factors in that region may follow some trends, linear or not, that can be used to express them with a small number of parameters. Again this would limit the impact of some of the cells.

The lower left point is more difficult to deal with in a CL-like model. One alternative is a Cape Cod-type model, where every accident year has the same mean level. This can arise, for instance, if there is no growth in the business, but also can be seen when the development triangle consists of on-level loss ratios, which have been adjusted to eliminate known differences among the accident years. In this type of model, all the cells go into estimating the level of the last accident year, so the lower left cell has much less impact. This reduction in the impact of the random component of this cell is a reason for using on-level triangles.

The next three sections illustrate these concepts using development triangles from the actuarial literature. The impacts and GDFs are calculated for various models fit to these triangles. The impacts are calculated by numerical derivatives, as are the GDFs except for those for the CL, which have been derived above.

4 A DEVELOPMENT-FACTOR EXAMPLE

4.1 Chain Ladder

Table 2 is a development triangle used in Venter [12]. Note that the first two accident years are developed all the way to the end of the triangle, at lag 11. Table 3 shows the impact of each cell on the reserve estimate using the usual sum/sum development factors. In the CL model an explicit formula can be derived for these impacts, but it is easier to do the derivatives numerically, simply by adding a small value to each cell separately and recalculating the estimated reserve to get the change in reserve for the derivative.

Table 2: Incremental Loss Development Triangle

L0	L1	L2	L3	L4	L5	L6	L7	L8	L9	L10	L11
11,305	18,904	17,474	10,221	3,331	2,671	693	1,145	744	112	40	13
8,828	13,953	11,505	7,668	2,943	1,084	690	179	1,014	226	16	616
8,271	15,324	9,373	11,716	5,634	2,623	850	381	16	28	558	
7,888	11,942	11,799	6,815	4,843	2,745	1,379	266	809	12		
8,529	15,306	11,943	9,460	6,097	2,238	493	136	11			
10,459	16,873	12,668	9,199	3,524	1,027	924	1,190				
8,178	12,027	12,150	6,238	4,631	919	435					
10,364	17,515	13,065	12,451	6,165	1,381						
11,855	20,650	23,253	9,175	10,312							
17,133	28,759	20,184	12,874								
19,373	31,091	25,120									
18,433	29,131										
20,640											

Table 3: Impact of CL

	L0	L1	L2	L3	L4	L5	L6	L7	L8	L9	L10	L11
AY0	-1.21	-0.34	0.04	0.39	0.73	1.10	1.48	1.85	2.46	3.35	4.61	7.31
AY1	-1.21	-0.34	0.04	0.39	0.73	1.10	1.48	1.85	2.46	3.35	4.61	7.31
AY2	-1.17	-0.29	0.08	0.44	0.78	1.14	1.53	1.89	2.51	3.39	4.66	
AY3	-1.15	-0.27	0.10	0.46	0.80	1.16	1.55	1.91	2.53	3.41		
AY4	-1.14	-0.27	0.11	0.46	0.80	1.17	1.56	1.92	2.54			
AY5	-1.10	-0.23	0.15	0.50	0.84	1.21	1.59	1.96				
AY6	-1.07	-0.20	0.18	0.53	0.87	1.24	1.62					
AY7	-1.03	-0.16	0.22	0.57	0.91	1.28						
AY8	-0.95	-0.08	0.30	0.65	0.99							
AY9	-0.73	0.14	0.52	0.87								
AY10	-0.31	0.57	0.95									
AY11	0.70	1.58										
AY12	4.95											

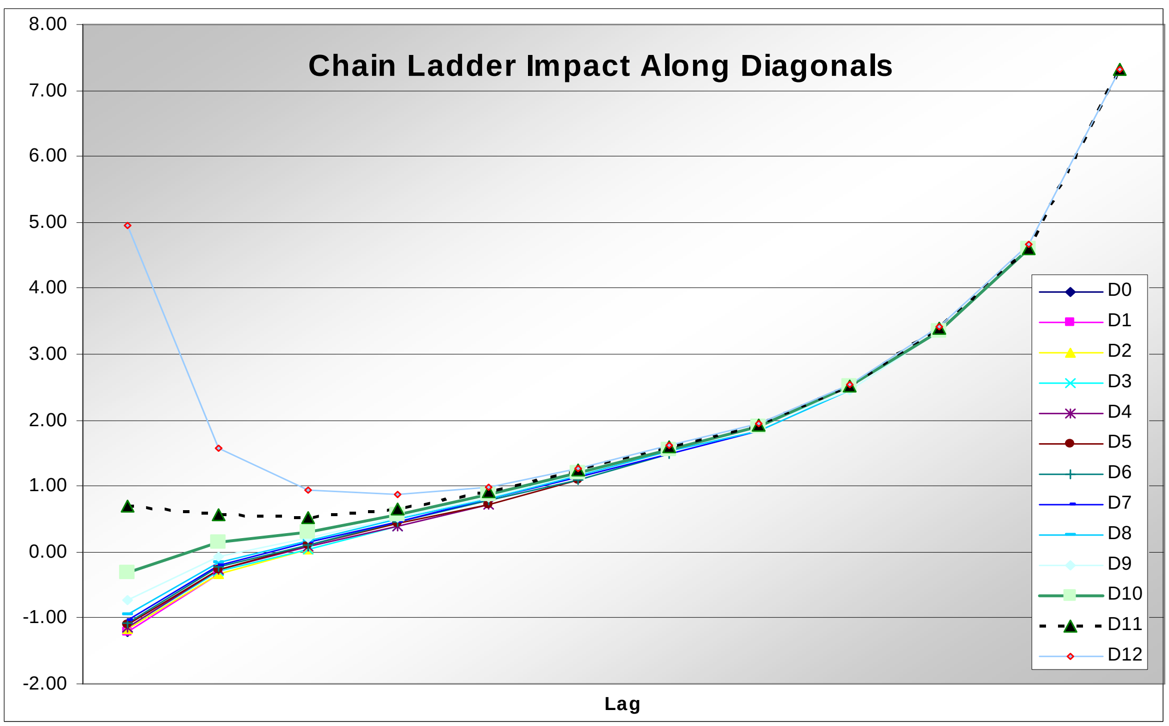
As discussed, the impacts are highest in the upper right and lower left corners, and the upper left has negative impact. The impacts increase moving to the right and down. The last four columns and the lower left point have impacts above 2, and six points have impacts above 4. Table 4 shows the GDFs for the chain ladder using the formula previous cumulative /sum previous cumulatives derived in Section 3. L0's GDFs are shown as identically 1.0. Like the impact function, these increase going to the right after lag 0. Within each column the sizes depend on the volume of the year.

Table 4: GDFs of CL

	L0	L1	L2	L3	L4	L5	L6	L7	L8	L9	L10	L11
AY0	1.0	0.080	0.093	0.114	0.133	0.151	0.177	0.201	0.245	0.306	0.394	0.581
AY1	1.0	0.063	0.070	0.082	0.097	0.110	0.128	0.145	0.174	0.221	0.285	0.419
AY2	1.0	0.059	0.073	0.079	0.103	0.124	0.147	0.167	0.202	0.250	0.321	
AY3	1.0	0.056	0.061	0.076	0.089	0.106	0.128	0.148	0.177	0.223		
AY4	1.0	0.061	0.073	0.086	0.104	0.126	0.149	0.168	0.202			
AY5	1.0	0.074	0.084	0.096	0.113	0.130	0.149	0.170				
AY6	1.0	0.058	0.062	0.077	0.089	0.106	0.123					
AY7	1.0	0.074	0.086	0.098	0.123	0.146						
AY8	1.0	0.084	0.100	0.134	0.149							
AY9	1.0	0.122	0.141	0.158								
AY10	1.0	0.138	0.156									
AY11	1.0	0.131										
AY12	1.0											

Figure 1 graphs the impacts by lag along the diagonals of the triangle. After the first four lags, the impacts are almost constant across diagonals.

Figure 1: Impact of Chain Ladder by Diagonal



4.2 Regression Model

Venter [12] fit a regression model to this triangle, keeping the first five development factors but including an additive constant. The constant also represents development beyond lag 5. By stretching out the incremental cells to be fitted into a single column **Y**, this was put into the form of a linear model $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, which assumes a normal distribution of residuals with equal variance (homoscedasticity) across cells. **X** has the previous cumulative for the corresponding incrementals, with zeros to pad out the columns, a column of 1's for the constant. There were also diagonal (calendar year) effects in the triangle. Two diagonal dummy variables were included in **X**, one with 1s for observations on the 4^{th} diagonal and 0 elsewhere, and one equal to 1 on the 5^{th}, 8^{th}, and 10^{th} diagonals, -1 on the 11^{th} diagonal, and 0 elsewhere. The diagonals are numbered starting at 0, so the 4^{th} is the one beginning with 8,529 and the 10th starts with 19,373. The variance calculation used a heteroscedasticity correction. This model with eight parameters fit the data better than the development factor model with 11 parameters. Here we are only addressing the robustness properties, however.

Table 5 gives the impact function for this model. It is clear that the large impacts on the right side have been eliminated by using the constant instead of factors to represent late development. The effects of the diagonal dummies can also be seen, especially in the right of the triangle. Now only 1 point has impact above 2, and above 4.

Table 5: Impact of Regression Model

	L0	L1	L2	L3	L4	L5	L6	L7	L8	L9	L10	L11
AY0	-1.36	0.02	0.42	0.67	0.10	0.87	1.35	1.35	0.97	1.35	0.97	1.73
AY1	-1.56	0.22	0.66	-0.04	0.67	1.28	1.35	0.97	1.35	0.97	1.73	1.35
AY2	-1.53	0.52	-0.39	0.38	1.02	1.27	0.97	1.35	0.97	1.73	1.35	
AY3	-0.51	-0.64	0.15	0.78	1.07	0.90	1.35	0.97	1.73	1.35		
AY4	-1.24	-0.31	0.45	0.76	0.64	1.27	0.97	1.73	1.35			
AY5	-1.38	0.11	0.47	0.32	1.00	0.89	1.73	1.35				
AY6	-1.61	0.22	0.18	0.80	0.68	1.66	1.35					
AY7	-0.89	-0.36	0.35	0.24	1.34	1.25						
AY8	-1.34	0.00	-0.12	0.87	0.94							
AY9	0.29	-0.44	0.61	0.57								
AY10	-0.18	0.66	0.43									
AY11	1.11	1.04										
AY12	4.31											

Table 6 shows the GDFs for the regression model. For regression models the GDFs for the observations in the **Y** vector are known to be calculable as the diagonal of the "hat" matrix, where hat = $\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$, e.g., see Ye [2]. However in development triangles, changing an incremental value also changes subsequent cumulatives, so the **X** matrix is a function of lags of **Y**. This requires the derivatives to be done numerically. The total of these, excluding lag 0, is 8.02, which is a bit above the usual number of parameters, due to the exceptions to normal linear models. Compared to the CL, the GDFs are lower for lag 6 onward, but are somewhat higher along the modeled diagonals. They are especially high for diagonal 4, which is short and gets its own parameter.

Table 6: GDFs of Regression Model

	L0	L1	L2	L3	L4	L5	L6	L7	L8	L9	L10	L11
AY0	1.0	0.071	0.089	0.127	0.352	0.195	0.034	0.034	0.055	0.034	0.055	0.076
AY1	1.0	0.047	0.056	0.305	0.107	0.099	0.034	0.055	0.034	0.055	0.076	0.034
AY2	1.0	0.046	0.299	0.084	0.098	0.123	0.055	0.034	0.055	0.076	0.034	
AY3	1.0	0.297	0.067	0.058	0.074	0.107	0.034	0.055	0.076	0.034		
AY4	1.0	0.064	0.056	0.072	0.120	0.128	0.055	0.076	0.034			
AY5	1.0	0.062	0.073	0.110	0.118	0.149	0.076	0.034				
AY6	1.0	0.040	0.067	0.061	0.095	0.140	0.034					
AY7	1.0	0.082	0.075	0.118	0.182	0.172						
AY8	1.0	0.077	0.134	0.212	0.207							
AY9	1.0	0.198	0.239	0.246								
AY10	1.0	0.245	0.253									
AY11	1.0	0.192										
AY12	1.0											

Figure 2: Impact of Regression Model by Diagonal

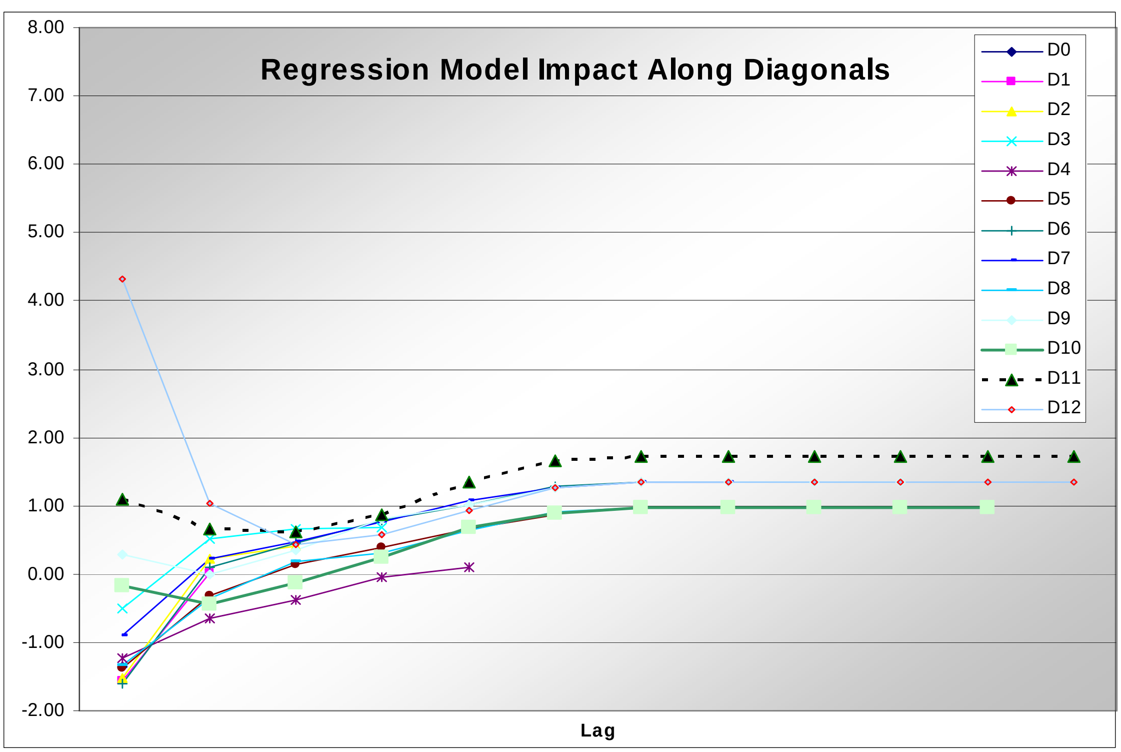


Figure 2 graphs the impacts. Note that due to the diagonal effects, diagonal 11 has higher impact than diagonal 12 after the first two lags.

4.3 Square Root Regression Model

As a correction for heteroscedasticity, regression courses sometimes advise dividing both **Y** and **X** by the square root of **Y**, row by row. This makes the model $\mathbf{Y}^{1/2} = (\mathbf{X}/\mathbf{Y}^{1/2})\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, where the $\boldsymbol{\varepsilon}$ are IID mean zero normals. Then $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Y}^{1/2}\boldsymbol{\varepsilon}$, so now the variance of the residuals is proportional to **Y**. This sounds like a fine idea, but it is a catastrophe from a robust viewpoint. Table 7 shows the impact function. There are 12 points with impact over 2, 7 with impact over 4, 5 with impact over 10, and 3 with impact over 25.

Table 7: Impact of Square Root Regression Model

	L0	L1	L2	L3	L4	L5	L6	L7	L8	L9	L10	L11
AY0	-0.94	-0.08	0.16	0.68	0.15	0.56	0.01	0.00	0.01	0.38	4.57	15.61
AY1	-1.06	-0.10	0.28	-0.30	2.19	1.86	0.01	0.15	0.00	0.15	10.21	0.01
AY2	-0.58	0.12	-0.09	0.20	0.68	0.39	0.01	0.03	28.26	3.09	0.02	
AY3	-0.20	-0.50	0.13	0.66	0.69	0.27	0.00	0.11	0.00	32.67		
AY4	-0.90	-0.15	0.33	0.41	0.59	0.56	0.03	0.14	37.14			
AY5	-1.28	-0.36	0.17	0.37	2.05	2.87	0.00	0.00				
AY6	-1.20	-0.09	0.01	0.77	0.71	2.34	0.02					
AY7	-1.02	-0.18	0.36	0.23	0.76	1.97						
AY8	-0.86	-0.07	-0.01	1.23	0.46							
AY9	-0.91	-0.06	0.59	1.02								
AY10	-0.45	0.48	0.89									
AY11	0.50	1.46										
AY12	4.56											

Part of the problem is that the equation $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Y}^{1/2}\boldsymbol{\varepsilon}$ is not what you would really want. The residual variance should be proportional to the mean, not the observation. This setup gives the small observations small variance, and so the ability to pull the model towards them. But the observations might be small because of a negative residual, with a higher expected value. So this formulation gives the small values too much influence.

Table 8 shows the related GDFs. It is unusual here that some points have GDFs greater than 1. A small change in the original value can make a greater change in the fitted value, but due to the non-linearity the fitted value is still a ways from the data point. The sum of the GDFs is 13.0, which is sometimes interpreted as the implicit number of parameters.

4.4 Gamma-*p* Residuals

Venter [13] fits the same regression model, but by maximum likelihood with gamma-*p* residuals. The gamma-*p* is a gamma distribution, but each cell is modeled to have the variance proportional to the same power *p* of the mean. This models the cells with smaller means as having smaller variances, but the effect is not as extreme as in the square root regression, where the variance is proportional to the observation, not its expected value.

Table 8: GDFs of Square Root Regression Model

	L0	**L1**	**L2**	**L3**	**L4**	**L5**	**L6**	**L7**	**L8**	**L9**	**L10**	**L11**
AY0	1.0	0.082	0.078	0.129	0.697	0.074	0.000	0.000	0.000	0.010	0.356	1.102
AY1	1.0	0.071	0.076	0.230	0.227	0.175	0.000	0.004	0.000	0.011	0.720	0.000
AY2	1.0	0.053	0.287	0.031	0.074	0.042	0.000	0.001	2.199	0.218	0.000	
AY3	1.0	0.201	0.045	0.081	0.064	0.025	0.000	0.008	0.000	0.906		
AY4	1.0	0.051	0.077	0.061	0.066	0.061	0.002	0.010	1.030			
AY5	1.0	0.076	0.102	0.089	0.252	0.322	0.000	0.000				
AY6	1.0	0.073	0.045	0.104	0.072	0.221	0.001					
AY7	1.0	0.069	0.103	0.053	0.106	0.249						
AY8	1.0	0.075	0.051	0.246	0.068							
AY9	1.0	0.117	0.193	0.208								
AY10	1.0	0.145	0.166									
AY11	1.0	0.144										
AY12	1.0											

Table 9: Impact of Gamma-*p* Residual Model

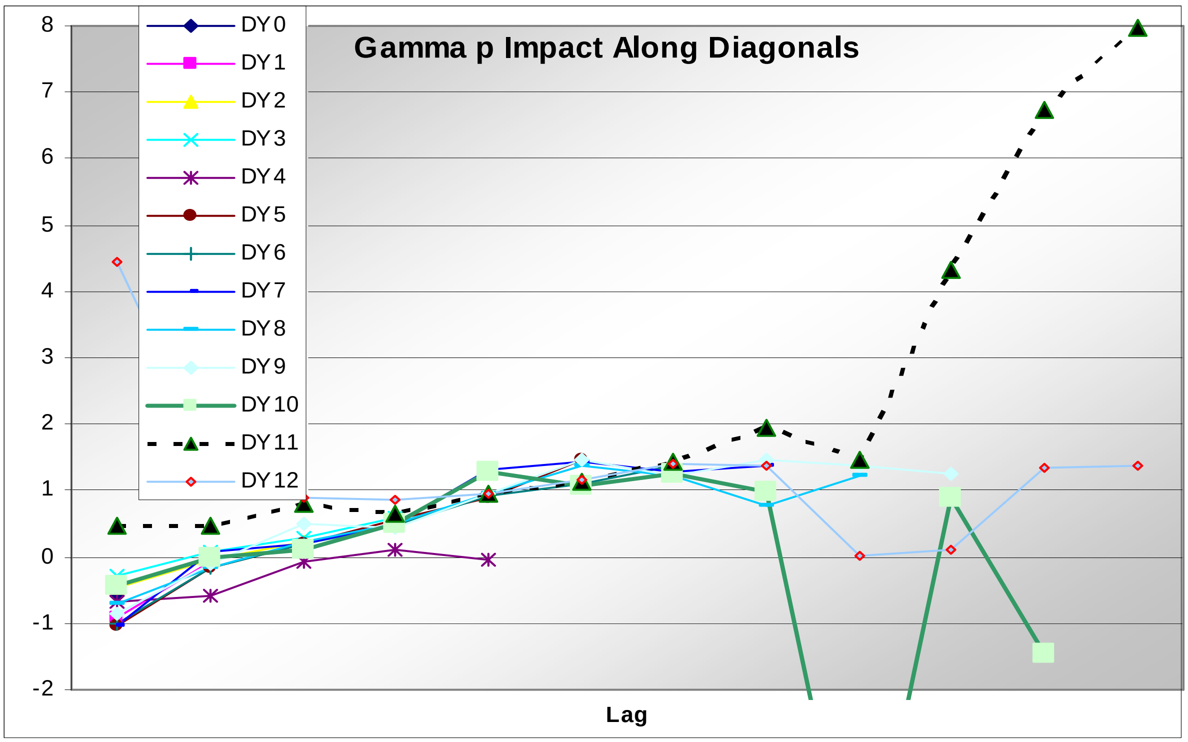
	L0	**L1**	**L2**	**L3**	**L4**	**L5**	**L6**	**L7**	**L8**	**L9**	**L10**	**L11**
AY0	-0.59	-0.07	0.24	0.59	-0.03	1.47	1.37	1.37	1.23	1.25	-1.45	7.97
AY1	-0.90	-0.05	0.28	0.10	0.90	1.11	1.30	0.77	1.37	0.91	6.73	1.36
AY2	-0.46	0.08	-0.07	0.56	0.94	1.43	1.22	1.45	-5.62	4.33	1.35	
AY3	-0.29	-0.58	0.21	0.47	1.31	1.37	1.21	0.98	1.47	0.10		
AY4	-0.68	-0.15	0.19	0.51	0.94	1.48	1.24	1.96	0.02			
AY5	-1.04	-0.18	0.20	0.49	0.96	1.07	1.43	1.38				
AY6	-1.00	0.09	0.22	0.45	1.28	1.13	1.41					
AY7	-1.02	-0.18	0.50	0.50	0.95	1.17						
AY8	-0.71	-0.12	0.12	0.66	0.96							
AY9	-0.85	-0.02	0.80	0.86								
AY10	-0.44	0.48	0.88									
AY11	0.46	1.45										
AY12	4.43											

In this case, *p* was found to be 0.71. The impacts are shown in Table 9 and graphed in Figure 3. It is clear that these are not nearly as dramatic as the square root regression, but worse than the regular regression, and perhaps comparable to the chain ladder. Diagonals 10 and 11 can be seen to have a few significant impacts. These are at points with small observations that are also on modeled diagonals. Even with the variance proportional to a power of the expected value, these points still have a strong pull. The GDFs are in Table 10.

Figure 3: Impact of Gamma-*p* Residual Model

Table 10: GDFs of Gamma-*p* Residual Model

	L0	L1	L2	L3	L4	L5	L6	L7	L8	L9	L10	L11
AY0	1.0	0.082	0.078	0.129	0.697	0.074	0.000	0.000	0.000	0.010	0.356	1.102
AY1	1.0	0.071	0.076	0.230	0.227	0.175	0.000	0.004	0.000	0.011	0.720	0.000
AY2	1.0	0.053	0.287	0.031	0.074	0.042	0.000	0.001	2.199	0.218	0.000	
AY3	1.0	0.201	0.045	0.081	0.064	0.025	0.000	0.008	0.000	0.906		
AY4	1.0	0.051	0.077	0.061	0.066	0.061	0.002	0.010	1.030			
AY5	1.0	0.076	0.102	0.089	0.252	0.322	0.000	0.000				
AY6	1.0	0.073	0.045	0.104	0.072	0.221	0.001					
AY7	1.0	0.069	0.103	0.053	0.106	0.249						
AY8	1.0	0.075	0.051	0.246	0.068							
AY9	1.0	0.117	0.193	0.208								
AY10	1.0	0.145	0.166									
AY11	1.0	0.144										
AY12	1.0											



Again this is less dramatic than for the square root regression, but the small points on the modeled diagonals still have high GDFs. The total of these is 11.3, which is still fairly high. This is somewhat troublesome, as the gamma-*p* model fit the residuals quite a bit better than did the standard regression. The fact that the problems center on small observations on the modeled diagonals suggests that additive diagonal effects may not be appropriate for this data. They do fit into the mold of a generalized linear model, but that is not too important when fitting by MLE anyway. As an alternative, the same model but with the diagonal effects as multiplicative factors was fit. The multiplicative diagonal model can be written:

$$\mathbf{EY = X[,1{:}6]\beta[1{:}6]*\beta[7]^{X[,7]}*\beta[8]^{X[,8]}},$$

which means that the first six columns of **X** are multiplied by the first six parameters, which includes the constant term, and then the last two diagonal parameters are factors raised to the power of the last two columns of **X**. These are now the diagonal dummies, which are 0, 1, or –1. Thus the same diagonals are higher and the same lower, but now proportionally instead of by an additive constant. It turns out that this model actually fits better, with a negative loglikelihood of 625, compared to 630 for the generalized linear model. This solves the robustness problems as well. The impacts are in Table 11, the GDFs in Table 12, and the impacts are graphed in Figure 4.

Table 11: Impact of Gamma-*p* Multiplicative Model

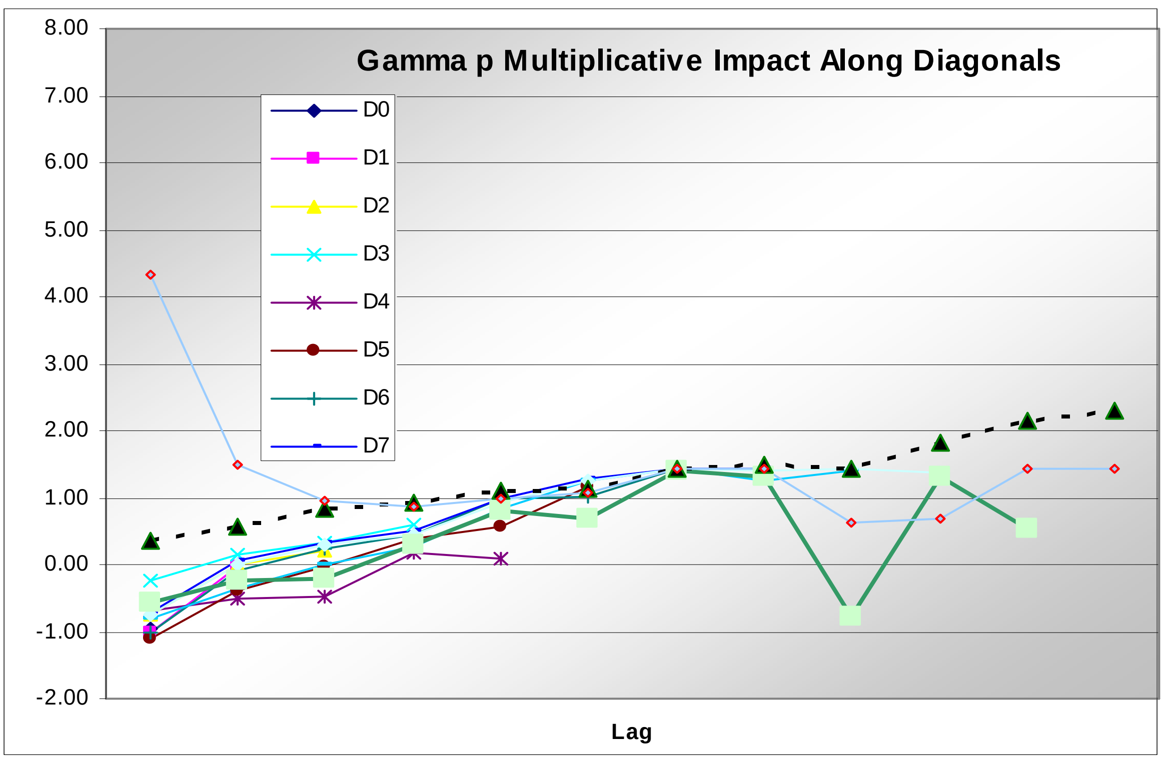
	L0	L1	L2	L3	L4	L5	L6	L7	L8	L9	L10	L11
AY0	-0.94	-0.03	0.22	0.58	0.09	1.16	1.43	1.43	1.42	1.36	0.55	2.31
AY1	-1.02	0.00	0.32	0.17	0.56	1.02	1.43	1.26	1.43	1.30	2.14	1.43
AY2	-0.74	0.15	-0.46	0.39	0.98	1.30	1.42	1.42	-0.78	1.82	1.42	
AY3	-0.25	-0.50	-0.02	0.46	0.97	1.26	1.43	1.33	1.43	0.69		
AY4	-0.68	-0.39	0.23	0.51	0.83	1.26	1.39	1.50	0.64			
AY5	-1.09	-0.10	0.33	0.26	0.93	0.69	1.43	1.43				
AY6	-1.02	0.05	0.00	0.45	0.79	1.12	1.42					
AY7	-0.72	-0.37	0.31	0.29	1.11	1.07						
AY8	-0.81	-0.01	-0.21	0.92	0.99							
AY9	-0.76	-0.25	0.85	0.88								
AY10	-0.58	0.56	0.94									
AY11	0.35	1.50										
AY12	4.34											

Diagonal 11 still has more impact than the others, but this barely exceeds 2.0 at the maximum. The sum of the GDFs is 8.67. There are eight parameters for the cell means but two more for the gamma-*p*. It has been a question whether or not to count those two in determining the number of parameter used in the fitting. The answer to that from the gdf analysis is basically to count each of those as 1/3 in this case. Here the robust analysis has uncovered a previously unobserved problem with the generalized linear model, and lead to an improvement.

Table 12: GDFs of Gamma-*p* Multiplicative Model

	L0	L1	L2	L3	L4	L5	L6	L7	L8	L9	L10	L11
AY0	1.0	0.079	0.087	0.125	0.323	0.136	0.034	0.033	0.038	0.040	0.093	0.074
AY1	1.0	0.063	0.069	0.191	0.210	0.132	0.034	0.048	0.033	0.046	0.066	0.034
AY2	1.0	0.053	0.410	0.079	0.085	0.068	0.038	0.035	0.175	0.050	0.034	
AY3	1.0	0.361	0.105	0.070	0.071	0.063	0.033	0.044	0.031	0.101		
AY4	1.0	0.107	0.070	0.067	0.111	0.084	0.040	0.034	0.106			
AY5	1.0	0.079	0.094	0.158	0.185	0.276	0.030	0.033				
AY6	1.0	0.066	0.106	0.081	0.104	0.117	0.035					
AY7	1.0	0.143	0.093	0.127	0.108	0.200						
AY8	1.0	0.080	0.200	0.220	0.102							
AY9	1.0	0.355	0.281	0.208								
AY10	1.0	0.316	0.196									
AY11	1.0	0.163										
AY12	1.0											

Figure 4: Impact of Gamma-*p* Multiplicative Model



5 A MULTIPLICATIVE FIXED-EFFECTS EXAMPLE

A multiplicative fixed-effects model is one where the cell means are products of fixed factors from rows, columns, and perhaps diagonals. The most well-known is the ODP model discussed in section 3, where there is a factor for each row, interpreted as estimated ultimate, a factor for each column, interpreted as fraction of ultimate for that column, and the variance of each cell is a fixed factor times its mean. This model if estimated by MLE gives the same reserve estimates as the chain ladder and so the same impacts for each cell, but the GDFs are different, due to the different fitted values.

The triangle for this example comes from Taylor-Ashe (1983) and is shown in Table 13. The CL = ODP impacts are in Table 14 and are graphed in Figure 5.

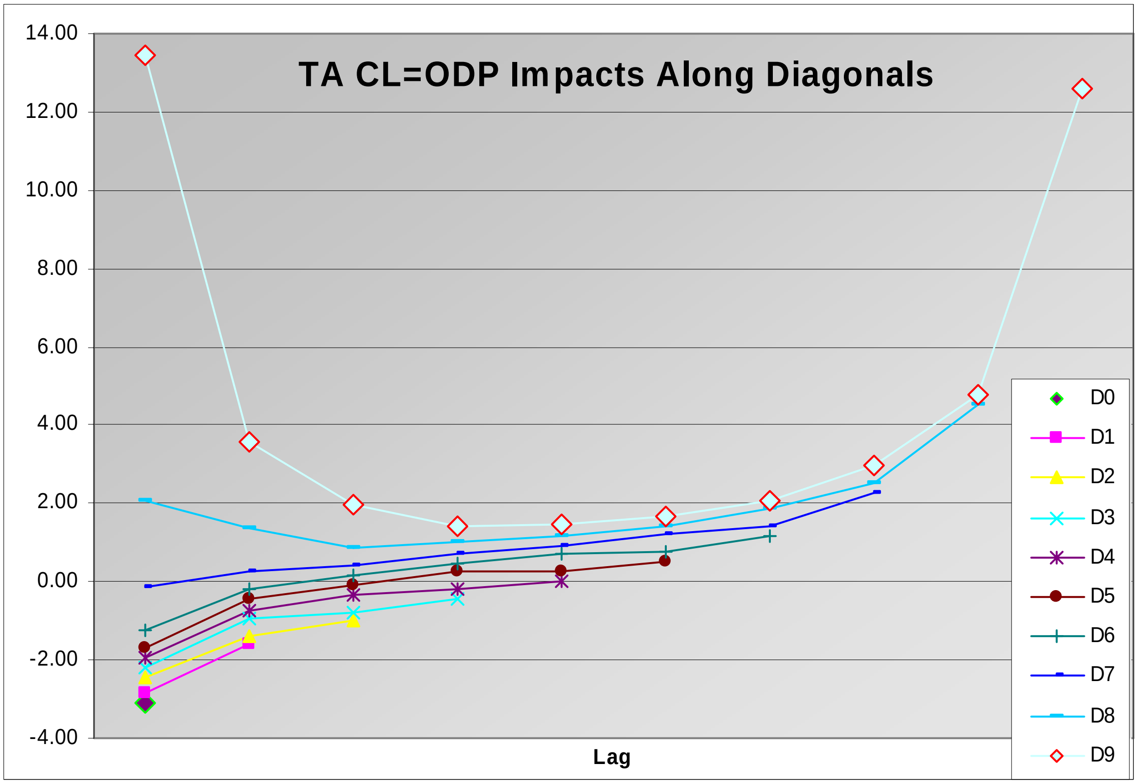
Table 13: Incremental Triangle Taylor-Ashe (1983)

Lag 0	L1	L2	L3	L4	L5	L6	L7	L8	L9
357,84	766,940	610,542	482,940	527,32	574,39	146,34	139,95	227,22	67,94
352,11	884,021	933,894	1,183,28	445,74	320,99	527,80	266,17	425,04	
290,50	1,001,79	926,219	1,016,65	750,81	146,92	495,99	280,40		
310,60	1,108,25	776,189	1,562,40	272,48	352,05	206,28			
443,16	693,190	991,983	769,488	504,85	470,63				
396,13	937,085	847,498	805,037	705,96					
440,83	847,631	1,131,39	1,063,26						
359,48	1,061,64	1,443,37							
376,68	986,608								
344,01									

Table 14: Impact of CL = ODP on TA

	L0	L1	L2	L3	L4	L5	L6	L7	L8	L9
AY0	-3.11	-1.62	-1.01	-0.45	0.01	0.51	1.16	2.27	4.54	12.59
AY1	-2.87	-1.38	-0.77	-0.20	0.25	0.76	1.40	2.51	4.78	
AY2	-2.43	-0.93	-0.33	0.24	0.69	1.20	1.85	2.95		
AY3	-2.21	-0.72	-0.11	0.45	0.91	1.41	2.06			
AY4	-1.95	-0.46	0.15	0.71	1.17	1.67				
AY5	-1.67	-0.18	0.43	0.99	1.45					
AY6	-1.25	0.25	0.85	1.42						
AY7	-0.14	1.35	1.96							
AY8	2.07	3.57								
AY9	13.45									

Figure 5: Impact of CL = ODP on TA



Because the development factors are higher, the impacts are higher than in the previous example. Even though it is a smaller triangle, 14 points have impacts with absolute values over 2, 4 are over 4, and 2 are over 12. The CL GDFs are in Table 15. These sum to 9, excluding the first column, and are fairly high on the right where there are few observations per column. The ODP GDFs are in Table 16. These sum to 19, and are fairly high near the upper right and lower left corners.

Table 15: GDFs of CL on TA

	L0	**L1**	**L2**	**L3**	**L4**	**L5**	**L6**	**L7**	**L8**	**L9**
AY0	1.0	0.108	0.110	0.115	0.120	0.153	0.208	0.272	0.423	1.0
AY1	1.0	0.106	0.121	0.144	0.182	0.211	0.258	0.365	0.577	
AY2	1.0	0.087	0.126	0.147	0.175	0.222	0.259	0.363		
AY3	1.0	0.093	0.138	0.146	0.204	0.224	0.275			
AY4	1.0	0.133	0.111	0.141	0.157	0.189				
AY5	1.0	0.119	0.130	0.145	0.162					
AY6	1.0	0.132	0.126	0.161						
AY7	1.0	0.108	0.139							
AY8	1.0	0.113								
AY9	1.0									

Table 16: GDFs of ODP on TA

	L0	L1	L2	L3	L4	L5	L6	L7	L8	L9
AY0	0.154	0.261	0.273	0.295	0.229	0.224	0.253	0.301	0.459	1.0
AY1	0.186	0.295	0.308	0.333	0.276	0.281	0.325	0.400	0.612	
AY2	0.187	0.300	0.312	0.338	0.278	0.282	0.324	0.398		
AY3	0.188	0.304	0.317	0.344	0.280	0.282	0.323			
AY4	0.184	0.309	0.322	0.348	0.275	0.271				
AY5	0.197	0.331	0.346	0.374	0.293					
AY6	0.221	0.375	0.391	0.423						
AY7	0.284	0.498	0.519							
AY8	0.370	0.747								
AY9	1.0									

The GDFs can be used to allocate the total degrees of freedom of the residuals of $n - p$. The n is allocated 1 to each observation, and the p can be set to the gdf of each observation. This would give a residual degree of freedom to each observation which could be used in calculating a standardized residual that takes into account how the degrees of freedom vary among observations.

Venter [12] looked at reducing the number of parameters in this model by setting parameters equal if they are not significantly different, and using trends, like linear trends between parameters. Also diagonal effects were introduced. The result was a model where each cell mean is a product of its row, column, and diagonal factors. There are six parameters overall. For the rows there are three parameters, for high, medium, and low accident years. Accident year 0 is low, year 7 is high, year 6 is the average of the medium and high levels, and all other years are medium. There are 2 column factors: high and low. Lags 1, 2, and 3 are high, lag 4 is an average of high and low, lag 0 and lags 5 to 8 are low, and lag 9 is 1 minus the sum of the other lags. Finally there is one diagonal parameter c. Diagonals 4 and 6 have factors 1+c, lag 7 has factor 1 – c, and all the other diagonals have factor 1.

With just six parameters this model actually provides a better fit to the data than the 19 parameter model. The combining of parameters does not degrade the fit much, and adding diagonal effects improves the fit. An improved fit over that in Venter [12] was found by using a gamma-p distribution with $p = ½$ so the variance of each cell is proportional to the square root of its mean. The impacts and GDFs of this model are shown in Tables 17 and 18, and the impacts are graphed in Figure 6, this time along accident years.

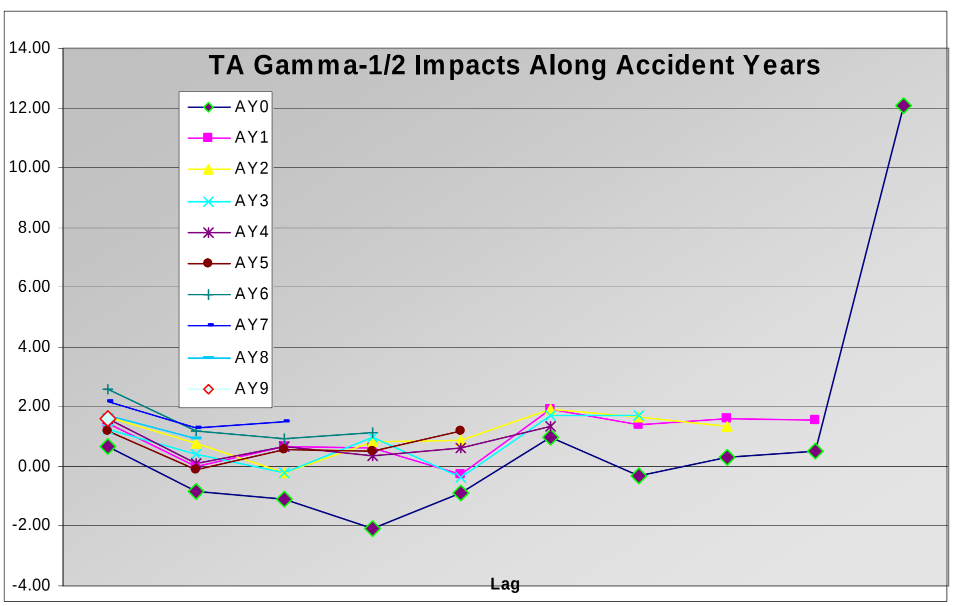
Table 17: Impact of 6-Parameter Gamma-½ on TA

	L0	L1	L2	L3	L4	L5	L6	L7	L8	L9
AY0	0.65	-0.82	-1.08	-2.07	-0.87	0.97	-0.32	0.33	0.53	12.06
AY1	1.45	-0.02	0.68	0.60	-0.25	1.90	1.40	1.61	1.57	
AY2	1.64	0.75	-0.19	0.84	0.90	1.93	1.66	1.36		
AY3	1.26	0.43	-0.21	0.97	-0.36	1.70	1.71			
AY4	1.62	0.08	0.67	0.37	0.63	1.35				
AY5	1.19	-0.11	0.57	0.51	1.17					
AY6	2.56	1.19	0.91	1.13						
AY7	2.18	1.27	1.49							
AY8	1.72	0.92								
AY9	1.59									

Table 18: GDFs of 6-Parameter Gamma-½ on TA

	L0	L1	L2	L3	L4	L5	L6	L7	L8	L9
AY0	0.046	0.152	0.211	0.288	0.150	0.017	0.248	0.095	0.082	0.938
AY1	0.044	0.031	0.057	0.155	0.014	0.115	0.018	0.051	0.043	
AY2	0.055	0.041	0.134	0.027	0.102	0.114	0.046	0.049		
AY3	0.045	0.078	0.062	0.028	0.181	0.052	0.064			
AY4	0.078	0.057	0.026	0.119	0.011	0.037				
AY5	0.037	0.147	0.083	0.032	0.026					
AY6	0.254	0.200	0.095	0.100						
AY7	0.111	0.527	0.250							
AY8	0.047	0.031								
AY9	0.047									

Figure 6: Impact of Gamma-½ on TA



The impacts are now all quite well contained except for one point – the last point in AY0. Possibly because AY0 gets its own parameter, lag 9 influences the level of the other lags' parameters, and this is a small point with a small variance, this model only slightly reduces the high level of impact that point has in ODP. The same thing can be seen in the GDFs as well, where this point has slightly less than a whole gdf. The points on AY7 and the modeled diagonals also have relatively high GDFs, as do some small cells. The total of the GDFs is 6.14. There are six parameters affecting the means, plus one for the variance of the gamma. That one can affect the fit slightly, so counting it as 1/7$^{\text{th}}$ of a parameter seems reasonable.

In an attempt to solve the problem of the upper-right point, an altered model was fit: lag 9 gets half of the paid in the low years. This can be considered a trend to 0 for lag 10. Making the lags sum to 1.0 now eliminates a parameter, so there are five. The NLL is slightly worse, at 722.40 vs. 722.36, but that is worth saving a parameter. The robustness is now much better, with only two impacts above 2.0, the largest being 2.35.

6 PAID AND INCURRED EXAMPLE

Venter [15], following Quarg and Mack [16], builds a model for simultaneously estimating paid and incurred development, where each influences the other. The paid losses are part of the incurred losses, so the separate effects are from the paid and unpaid triangles, shown in Tables 19 and 20.

First the impacts on the reserve (7059.47) from the average of the paid and incurred chain ladder reserves is calculated, where the paids at the last lag are increased by the incurred-to-paid ratio at that lag. Tables 21 and 22 show the impacts of the paid and unpaid triangles, and Tables 23 and 24 show the GDFs.

Table 19: Quarg-Mack Paid Increments

	L0	**L1**	**L2**	**L3**	**L4**	**L5**	**L6**
AY0	576	1228	166	54	50	28	29
AY1	866	1082	214	70	52	64	
AY2	1412	2346	494	164	78		
AY3	2286	3006	432	126			
AY4	1868	1910	870				
AY5	1442	2568					
AY6	2044						

Table 20: Quarg-Mack Unpaid

	L0	**L1**	**L2**	**L3**	**L4**	**L5**	**L6**
AY0	402	300	164	120	100	80	43
AY1	978	604	304	248	224	106	
AY2	1492	596	446	184	150		
AY3	1216	666	346	292			
AY4	944	1104	204				
AY5	1200	396					
AY6	2978						

Table 21: Average Reserve Impact of Paid

	L0	L1	L2	L3	L4	L5	L6
A0	-0.68	-.02	0.32	0.86	2.32	5.95	13.99
A1	-0.45	0.20	0.54	1.08	2.54	6.17	
A2	-0.41	0.24	0.58	1.12	2.59		
A3	-0.36	0.30	0.64	1.18			
A4	-0.32	0.33	0.67				
A5	-0.20	0.46					
A6	1.37						

Table 22: Average Reserve Impact Unpaid

L0	L1	L2	L3	L4	L5	L6
-0.29	-0.15	-0.26	-0.72	-1.76	-4.01	14.99
-0.29	-0.15	-0.26	-0.72	-1.76	3.57	
-0.29	-0.15	-0.26	-0.72	1.77		
-0.29	-0.15	-0.26	1.08			
-0.29	-0.15	0.82				
-0.29	0.68					
0.84						

Table 23: Average Reserve GDF of Paid

	0	L1	L2	L3	L4	L5	L6
A0	1	0.068	0.109	0.140	0.233	0.476	1
A1	1	0.102	0.117	0.153	0.257	0.524	
A2	1	0.167	0.227	0.301	0.509		
A3	1	0.271	0.319	0.406			
A4	1	0.221	0.228				
A5	1	0.171					
A6	1						

Table 24: Average Reserve GDF Unpaid

	0	L1	L2	L3	L4	L5	L6
A0	1	0.067	0.106	0.139	0.232	0.464	1
A1	1	0.126	0.129	0.160	0.269	0.536	
A2	1	0.198	0.219	0.306	0.499		
A3	1	0.239	0.300	0.395			
A4	1	0.192	0.246				
A5	1	0.180					
A6	1						

The impacts of the lower left are not great, mostly because the development factors are fairly low in this example. The impacts on the upper right of both paid and unpaid losses are quite high, however. The unpaid losses not on the last diagonal have a negative impact, because they lower subsequent incurred development factors, but do not have factors applied to them. The GDFs are similar to CL in general.

The model in Venter [15] used generalized regression for both the paid and unpaid triangles, where regressors could be from either triangle or from the cumulative paid and incurred triangles. Except for the first couple of columns, the previous unpaid losses provided reasonable explanations of both the current paid increment and the current remaining unpaid. The paid and unpaid at lags 3 and on were just multiples of the previous unpaid, with a single factor for each. That is, expected paids were 33.1%, and unpaids 72.3%, of the previous unpaid. Since these sum to more than 1, there is a slight upward drift in the incurred. The lag 2 expected paid was 68.5% of the lag 1 unpaid. The best fit to the lag 2 expected unpaid was 9.1% of the lag 1 cumulative paid. For lag 1 paid, 78.1% of the lag 0 incurred was a reasonable fit. Lag 1 unpaid was more complicated, with the best fit being a regression, with constant, on lag 0 and lag 1 paids. There were also diagonal effects in both models. The residuals were best fit with a Weibull distribution. Tables 25 – 28 show the fits.

Table 25: Weibull Model Impact of Paid

	L0	L1	L2	L3	L4	L5	L6
A0	0.09	-0.18	-1.58	4.38	0.38	7.67	5.45
A1	0.04	0.26	0.59	1.90	2.75	2.32	
A2	-.37	0.33	0.42	0.57	-0.28		
A3	-.13	0.17	0.67	1.26			
A4	-.02	0.20	0.31				
A5	-.94	0.70					
A6	1.25						

Table 26: Weibull Model Impact Unpaid

L0	L1	L2	L3	L4	L5	L6
0.06	0.67	-1.02	-1.45	-1.82	0.51	4.14
-0.17	-0.44	-1.80	-0.73	0.52	2.56	
-0.20	-0.16	0.47	-1.17	3.63		
-0.09	-0.32	-1.17	2.51			
-0.10	-0.34	1.89				
-0.32	1.47					
0.65						

Table 27: Weibull Model GDF of Paid

	0	L1	L2	L3	L4	L5	L6
A0	1	0.938	0.725	0.235	0.268	0.125	.143
A1	1	0.451	0.057	0.052	0.066	0.065	
A2	1	0.192	0.347	0.377	0.290		
A3	1	0.137	0.250	0.145			
A4	1	0.094	0.277				
A5	1	0.269					
A6	1						

Table 28: Weibull Model GDF Unpaid

	0	L1	L2	L3	L4	L5	L6
A0	1	0.824	0.172	-0.058	0.015	.072	.054
A1	1	0.357	0.700	-0.044	0.115	.052	
A2	1	0.152	0.465	-0.173	0.113		
A3	1	0.050	0.136	0.123			
A4	1	0.507	0.089				
A5	1	0.687					
A6	1						

The two highest impacts for the average of paid and incurred are 14 and 15. For the Weibull they are 7.7 and 5.5. The average has two other points with impacts above 5, whereas the Weibull has none. Below 5 the impacts are roughly comparable. Since the Weibull has variance proportional to the mean squared, small observations have lower variance, and so a stronger pull on the model and higher impacts. In total, excluding the first column, the GDFs sum to 9.9, but including the diagonals (see Venter [15] for details) there are 12 parameters plus two Weibull shape parameters. The form of the model apparently does not allow the parameters to be fully expressed. The Weibull model still has more high impacts than would be desirable, but it is a clear improvement over the average of the paid and incurred. The reserve is quite a bit lower for the better-fitting Weibull model as well: 6255 vs. 7059.

7 CONCLUSION

Robust analysis has been introduced as an additional testing method for loss development models. It is able to identify points that have a large influence on the reserve, and so whose random components would also have a large influence. Through three examples, customized models were found to be more robust than standard models like CL and ODP, and in two of the examples, even better models were found as a response to the robust analysis.

REFERENCES

[1] Tampubolon, Dumaria R. 2008. "Uncertainties in the estimation of outstanding claims liability in general insurance," PhD thesis, Macquarie University, Sydney, Australia.

[2] Ye, Jianming. 1998. "On measuring and correcting the effects of data mining and model selection," *Journal of the American Statistical Association*, March 1998, 93(441): 120-131.

[3] Tukey, John W. 1960. "A survey of sampling from contaminated distributions," in I. Olkin (Ed.), *Contributions to Probability and Statistics*, (pp. 448-485) (Palo Alto: Stanford University Press).

[4] Dell'Aquila, Rosario, Elvezio Ronchetti, and Fabio Troiani. 2003. "Robust GMM analysis of models for the short rate process," *Journal of Empirical Finance* 10:373-397.

[5] Hampel, Frank R. 1968. "Contributions to the Theory of Robust Estimation." PhD Thesis, University of California, Berkeley.

[6] Bilodeau, Martin. 2001. Discussions of Papers Already Published. *North American Actuarial Journal* 5(3): 23–28. (online version included in Brazauskas, Vytaras, and Robert Serfling, 2000)

[7] Brazauskas, Vytaras, and Robert Serfling. 2000a. "Robust and Efficient Estimation of the Tail Index of a Single-Parameter Pareto Distribution," *North American Actuarial Journal* 4(4):12–27.

[8] Finkelstein, Mark, Howard G. Tucker, and Jerry Alan Veeh. 2006. "Pareto Tail Index Estimation Revisited," *North American Actuarial Journal* 10(1):12–27.

[9] Brazauskas, Vytaras and Robert Serfling. 2000b. "Robust Estimation of Tail Parameters for Two-Parameter Pareto and Exponential Models via Generalized Quantile Statistics," *Extremes* 3(3) Sept: 231-249

[10] Gather, Ursula and Verena Schultze, 1999. "Robust Estimation of Scale of an Exponential Distribution," *Statistica Neerlandica* 53(3):327-341

[11] Serfling, Robert 2002. "Efficient and Robust Fitting of Lognormal Distributions," *North American Actuarial Journal* 6(4):95-116.

[12] Venter, Gary G., 2007a. "Refining Reserve Runoff Ranges," CAS *E-Forum*, Summer 2007.

[13] Venter, Gary G., 2007b. "Generalized Linear Models beyond the Exponential Family with Loss Reserve Applications," CAS *E-Forum*, Summer 2007.

[14] Taylor, Greg C. and Frank R. Ashe, 1983. "Second Moments of Estimates of Outstanding Claims," *Journal of Econometrics* 23:37-61.

[15] Venter, Gary G., 2008. "Distribution and Value of Reserves Using Paid and Incurred Triangles," CAS Reserves Call Papers, Casualty Actuarial Society *E-Forum*, Fall 2008, http://www.casact.org/pubs/forum/ (to be posted August 2008).

[16] Quarg, Gerhard and Thomas Mack, 2004. "Munich chain-ladder—A reserving method that reduces the gap between IBNR projections based on paid losses and IBNR projections based on incurred losses," *Blätter DGVFM* Heft 4, 597–630.

Biographies of the Authors

Gary G. Venter is managing director at Guy Carpenter, LLC. He has an undergraduate degree in philosophy and mathematics from the University of California and an MS in mathematics from Stanford University. He has previously worked at Fireman's Fund, Prudential Reinsurance, NCCI, Workers Compensation Reinsurance Bureau and Sedgwick Re, some of which still exist in one form or another. At Guy Carpenter, Gary develops risk management and risk modeling methodology for application to insurance companies. He also teaches a graduate course in loss modeling at Columbia University.

+1.917.937.3277 **gary.g.venter@guycarp.com**

Dumaria R. Tampubolon is a lecturer at the Faculty of Mathematics and Natural Sciences at Institut Teknologi Bandung (Bandung Institute of Technology), Bandung, Indonesia. She has a "Sarjana S1" (similar to Honors) degree in mathematics, majoring in statistics, from Institut Teknologi Bandung; an M.Sc degree in statistics from Monash University, Melbourne, Australia; and a PhD degree (2008) in actuarial studies from Macquarie University, Sydney, Australia. Her research interests are in the area of general insurance.

+62.22.250.2545 ext.101 **dumaria@math.itb.ac.id**, **drtampubolon@gmail.com**

A Model to Test for and Accommodate Reserving Cycles

Thomas S. Wright, MA, CStat, FIA

Abstract

In recent years several commentators have noted evidence for a "reserving cycle" linked to the underwriting cycle. It seems that in many classes of non-life insurance, when premium rates are relatively low, claim development patterns tend to be longer-tailed than when premium rates are high. If this is the case, then traditional reserving methods based on an assumption that the development pattern is the same for all origin years will tend to overstate reserves for periods where premium rates were high, and understate reserves for periods where premium rates were low. The present paper reviews the evidence for a reserving cycle and discusses possible causes. A mathematical model is then proposed that accommodates the main possible causes. The purpose of this model is three-fold: (a) to test for the existence of reserving cycle effects, (b) to help identify the causes, and (c) to produce improved reserve estimates. An example analysis is presented using the proposed model. The evidence for the existence of reserving cycles is now sufficiently strong that, in the author's opinion, it is important for reserving actuaries to be aware of the possibility of cyclical effects, to investigate evidence for such effects in any reserving exercise, and (where there is strong evidence) to adjust reserve estimates accordingly. The model proposed in the present paper can be implemented in Excel and will often be a useful tool for these purposes.

Keywords. Reserving cycle, underwriting cycle, development patterns, curve fitting, least squares, premium rate indices.

1. INTRODUCTION

1.1 Research Context

1.1.1 Bob Conger's presentation at GIRO 2002

The idea of a "reserving cycle" was first given prominence by Bob Conger (then CAS President) in his keynote presentation to the 2002 GIRO Convention in the UK. For all classes of US property/casualty insurance combined, and for workers compensation alone, he showed the ratio of initial estimated ultimates (at end of the first development year) to the latest estimated ultimates (at end of 2001) for each of the previous 20 accident years. When plotted against time, this ratio appeared, in both cases, to show a cyclical pattern of under- and over-reserving. This cycle appeared to be in phase with the underwriting cycle over the nearly two complete cycles of the years 1980 to 2001. Initial reserve estimates were consistently too low in both of the "soft market" periods when premium rates were relatively low (the mid 1980s and the late 1990s), and were consistently too high in the intervening "hard markets" (when premium rates were relatively high).

The most obvious explanation is that when setting reserves soon after writing the business, insurers tend to under-estimate the magnitude of the underwriting cycle. History shows that at the lowest point of the underwriting cycle, insurers often write business at loss-making rates. But presumably they don't do this deliberately: it is only human to hope and believe that the business they recently wrote will ultimately prove profitable, and to set the initial reserves accordingly. At the other extreme of the underwriting cycle, when premium rates are buoyant, even relatively cautious (high) initial reserves may show a fairly good profit. Management might privately believe that they could reasonably set the initial reserves lower and show an even higher profit. But with a choice between declaring a very high profit now, with the possibility that this will deteriorate, and declaring a more moderate (but still healthy) profit now, with the expectation that this will allow further good news to be released as the claims run-off, it is easy to see the attraction of the latter.

If this were the whole explanation for the reserving cycle, then the actuarial profession could rest easy. If we as actuaries provide objective, unbiased estimates for the reserves, and senior management chooses to depart from these estimates for reasons such as those described above, then that is their responsibility not ours.

However, we need to be sure that actuarial reserve estimates are as good as they can be. Could it be that actuarial reserving methods are partly to blame for the reserving cycle?

1.1.2 Working party report at GIRO 2003

Bob Conger's presentation at GIRO 2002 prompted the formation of a working party tasked with investigating the existence and possible causes of a reserving cycle in the UK. This working party was chaired by Nick Line, and presented its report [3] at the 2003 GIRO convention. The working party concluded that:

(a) A reserving cycle did also exist in the UK.

(b) Standard actuarial reserving methods are probably a contributory cause of the reserving cycle.

(c) There was some (inconclusive) evidence that development patterns vary with the underwriting cycle, tending to be longer-tailed when premium rates are low.

(d) There was clear evidence that Lloyd's premium rate indices had tended to understate the true magnitude of the underwriting cycle.

Conclusion (a) is based on UK industry reserves (from regulatory returns) over the period 1985 to 2001 inclusive. This is a shorter period than that considered by Bob Conger, and covers little more than one complete underwriting cycle (including the soft market of the mid to late 1980s and the next soft market of the late 1990s). The working party looked at the non-life insurance market as a whole, and some individual classes, and concluded that the reserving cycle exists for several major classes (motor, property and liability) and that the cycles for these classes are in phase with one another.

The reserves analyzed by the working party in support of conclusion (a) were obtained from regulatory returns, and as such show booked reserves as opposed to actuarial estimates. Conclusion (b) was based on an investigation of the extent to which standard actuarial reserving methods (chain ladder (CL) and Bornheutter-Ferguson (BF)) produce cyclical under- and over-reserving if applied mechanistically. This was investigated by applying these methods to the run-off data from regulatory returns, and comparing early estimates to actual ultimates. For some (but not all) classes of insurance, the results showed a clear cyclical pattern closely following that observed in the regulatory reserves. This was clearest for long-tail liability classes.

The working party then tried to explain why these standard actuarial methods tend to give a cyclical pattern of under- and over-reserving. They postulated two main causes: the points labeled (c) and (d) above. The apparent variation in development patterns with the underwriting cycle (point (c)) violates the basic assumption of the CL and BF methods: that the development pattern is the same for all origin periods. The working party found some evidence of more rapid paid development for origin years in the "hard" part of the underwriting cycle. The paid chain ladder method would clearly tend to overstate reserves at the top of the cycle and understate at the bottom of the cycle if this is the case.

The tendency for premium-rate indices to understate the amplitude of the underwriting cycle (point (d) above) exacerbates the reserving error produced by the BF method: if the softness of a soft market is understated, then the prior expected loss ratio will also be understated, leading to an initial under-estimation of the ultimate.

[Note that if both (c) and (d) apply, their effects on paid BF reserves will be in the same direction, rather than offsetting each other. The paid BF reserve is (1-F)*(prior ultimate), where F is the expected proportion of ultimate development obtained by the chain ladder method. The CL method gives a value for F that is an average for all origin years in the run-off array. If development is quicker than average when premium rates are high (point (c)), then this will be lower than the expected proportion developed for accident years where premium rates are high, so the factor (1-F) will be too high for these years. Assuming the other factor of the BF reserve (the prior ultimate) is calculated in the usual way (as premium multiplied by average ULR divided by premium index) this factor will be too high if the premium index is too low when premium rates are high (point (d)).

Conversely, both factors of the paid BF reserve will be too low for origin years where premium rates are low.]

The 2003 GIRO working party also suggested some possible reasons why the run-off pattern might depend on the level of premiums (that is, possible explanations for point (c)). They came up with the following possible causes of longer paid development patterns in soft markets. (Note that the working party did not look for direct evidence that any of these actually occur: these points were merely suggested as possible causes.)

- When premium rates are low, insurers might be more reluctant to pay claims, leading to more protracted negotiations and longer payment delays.
- In soft markets, insurers might compete by including additional cover and relaxing terms and conditions (as well as by reducing premiums). This might result in more disputes over coverage, tending to lengthen development patterns. (Presumably tightening of terms and conditions in hard markets might equally lead to disputes, but the difference is that in this case, disputes are less likely to delay payments.)
- If upper policy limits are increased in soft markets, this would also tend to lengthen development patterns. (On the other hand, if insurers reduce deductibles in soft markets, this would tend to shorten development patterns because deductibles would be exhausted sooner.)
- If more multi-year policies are written in soft-markets, these would tend to lengthen development patterns of under-writing year cohorts (but this should not affect accident year run-off patterns).

The above four points relate to paid development. In addition, the working party noted that incurred development patterns would be longer-tailed in soft markets if insurers adopt more optimistic case-reserving practices when premium rates are low.

1.1.3 General Insurance Reserving Issues Taskforce (GRIT)

At the beginning of 2004, the UK actuarial profession created the General Insurance Reserving Issues Taskforce (GRIT). Of five specific issues given in the terms of reference for GRIT, one was "to consider the actions which the profession should take in relation to the observations made in the Reserving Cycle Working Party paper presented at GIRO 2003." GRIT produced its final report (after a consultation process within the UK profession) in March 2006 [2]. Section 7 of the GRIT report, entitled "Improving our Methods," is mostly concerned with reserving cycles. This topic is also mentioned in Sections 1.1.6, 1.1.7, 1.8.2-1.8.5, 2.6.4, 9.2 and 9.9 of the GRIT report.

GRIT carried out basically the same analysis as the 2003 working party, but using Lloyd's data where the previous working party had used insurance company data (from regulatory returns). Like the working party, GRIT applied the CL and BF methods mechanically to historical run-off triangles for different classes of business, and compared early forecasts produced by these methods to actual outcomes. Like the working party, GRIT concluded that these methods do produce a cyclical pattern of under and over reserving, and that this pattern is in phase with the underwriting cycle.

As a possible way forward, GRIT suggested (Section 7.5 of [2]) fitting cumulative Weibull distribution curves to cumulative paid development data, and allowing the scale parameter to vary cyclically. The equation for the cumulative development pattern proposed in [2] is:

$$\text{Claims}(t) = A * [\, 1 - \exp\{-(b/t)^c\}] \tag{1}$$

Here, t is development time, A is ultimate, b is a scale parameter, and c is a shape parameter. [Note that c has to be negative in order for this to be a valid cumulative development pattern: if c is positive, then Claims(t) tends to zero as t tends to infinity. The Weibull curve is usually specified using t/b instead of b/t so that the shape parameter c takes positive values: we then have Claims(t) tending to the ultimate A as t tends to infinity.]

1.1.4 Other prior research on the reserving cycle

In the UK, GRIT was replaced (following publication of its final report in 2006) by the General Insurance Reserving Oversight Committee (GI ROC). GI ROC initiated four working parties that would report to future GIRO conventions. One of these is the working party on "Implications of the underwriting and reserving cycles for reserving." By the time of GIRO 2007 this working party had not made significant progress.

Perhaps surprisingly, considering that it was Bob Conger (then president of the CAS) who first highlighted this issue, there seems to be no published research in this area by US actuaries. (At least, a search of the CAS Web Site yields nothing new.)

1.2 Objectives of the Present Paper

The present paper develops the idea (suggested in the 2006 GRIT report [2]) of fitting curves to cumulative development data in a way that allows for the possibility of cyclical variation of development patterns.

The GRIT paper suggested using a cumulative Weibull distribution function (see Equation 1) for this purpose. In the present paper, Weibull, Burr, and Inverse Burr distribution functions are used.

Any distribution function is by definition an increasing function. In practice, cumulative incurred run-off patterns often do not increase at all stages of development, so cumulative probability distribution functions would not provide a good fit. The present paper develops a family of curves that does have the flexibility to accommodate typical cumulative incurred development patterns. This family of curves is derived by modeling both reporting and payment delays using cumulative probability distribution functions. This produces two linked families of cumulative curves: one for paid, the other for incurred. By fitting these simultaneously to paid and incurred run-off data, a single ultimate is estimated for each origin year from all available data. This avoids the common problem of having one ultimate estimated from paid data and another ultimate estimated from incurred data, then having to combine the two somehow.

Parameters of the fitted curves are linked to a premium rating index so that both paid and incurred run-off curves are allowed to vary with the underwriting cycle. This is done in a way that allows for the possibility that the premium rating index might understate the true amplitude of the underwriting cycle (as found to be the case in [3]).

The paper is not concerned with directly looking for evidence of each the possible causes of cyclically varying run-off patterns discussed in Section 1.1.2. Instead, the mathematical model is developed in such a way that it will accommodate these possible causes if they exist. The model also accommodates other possible factors, such as variation in reporting delay with the underwriting cycle. It will not always be possible to distinguish the true cause using the results of fitting the proposed model.

2. CYCLICAL CURVE-FITTING METHOD

2.1 Principles of curve-fitting to claims development data

In Section 2, a model is introduced that can be used to test for the presence of cyclical development patterns, to distinguish some of the main cyclical effects, and to estimate ultimates in the presence of these effects. Initially it is assumed that the only data available are the usual aggregate cumulative paid and incurred run-off arrays, and a premium rate index. Later (Section 2.4) the use of premium or other exposure information is also considered.

The method is basically a curve-fitting method as suggested in the GRIT report [2]. When curve-fitting is used for reserving, it is common to assume that the run-off pattern is the same for all origin years. If this is not the case, origin years can sometimes be grouped so that it is approximately the case in each group. The GRIT paper suggested classifying origin years into two categories according to their position in the underwriting cycle (hard or soft). Instead of doing this, the method introduced in the present paper uses a premium rate index to allow continuous graduation between hard and soft market run-off curves.

Reserving methods have previously been developed that allow run-off patterns to gradually change across origin years: for example, the method described in Wright [4]. That method allows for trend changes in development patterns but not for cyclical changes. It is also quite complex because it is a full stochastic method which gives predictive standard errors as well as best estimates. A limitation of that method is that it requires mainly positive increments in the run-off data, so it often cannot be applied directly to incurred data without first adjusting the data in some way.

What we need now is a method that can be applied to both paid and incurred data, preferably making use of premium development data too, and which allows for cyclical changes in run-off patterns. The top priority is to develop such a method that gives good point estimates in the presence of underwriting and reserving cycles. A lower priority is rigorous assessment of standard errors: this is not considered in the present paper.

If the run-off pattern is not assumed to be the same for all origin years, then the model necessarily has more parameters than where the run-off pattern is assumed to be constant. It is a well-established statistical principle that as the number of estimated parameters increases, their reliability (when estimated from a given volume of data) generally decreases. Therefore, it is advisable to use as much relevant data as possible when estimating parameters. For this reason, the proposed method fits run-off curves to both paid and incurred data simultaneously. This also avoids the problem (met with most other reserving methods) of having one set of reserve estimates obtained from paid data and a different set of estimates obtained from incurred data, then having to combine into a single set of final estimates. In order to allow fitting to incurred data as well as paid, the family of run-off curves must allow for negative increments as these often exist in incurred data.

2.2 Model for a single origin year

2.2.1 Cumulative paid and incurred development curves

In this sub-section we consider the run-off of a single origin year. The model is generalized for multiple origin years in later sub-sections.

$F_P(t)$ is used to denote a cumulative paid run-off pattern (where t is continuous development time). This is a function that starts at 0 when $t = 0$, and increases to 1 as t tends to infinity. For paid data, although there may be occasional decreases due to salvage and subrogation, the underlying pattern is assumed to be strictly increasing. $F_P(t)$ is therefore a cumulative distribution function: its derivate $f_P(t) = dF_P(t)/dt$, can be regarded as the probability density function for the delay to payment of each dollar that is ultimately paid.

For modeling incurred run-off patterns, we need to consider reporting delays. We use $F_R(t)$ to denote the cumulative distribution function of reporting delays (in respect of each dollar that is ultimately paid). Since every claim must be reported before it is paid, we should have $F_R(t) \geq F_P(t)$ at all development times t. Exhibit 1 shows typical curves $F_R(t)$ and $F_P(t)$ for a single origin year. (The curves in Exhibit 1 are Weibull distributions with mean values of 1 year and 3.6 years respectively.)

Note that $F_R(t)$ is not the cumulative incurred development pattern: it is the distribution function of reporting delays in respect of each dollar that is ultimately paid. For example, consider an accident year with an ultimate paid amount of $100,000. Suppose the first claim is reported mid-way through accident year zero (at time $t = 0.5$), and that this claim ultimately settles for $1,000. Since this is 1% of the total ultimate for the accident year, $F_R(t)$ increases from 0 to 0.01 at $t = 0.5$. The incurred development pattern will usually differ from this. For example, suppose that when this first claim is reported the initial case reserve is set at $2,000. Since this is 2% of the total ultimate for the accident year, the incurred development pattern increases from 0 to 0.02 at $t = 0.5$. (Of course, none of these development patterns is known with certainty until the accident year concerned is fully developed.)

To model incurred development, we assume that when a claim is reported a case reserve is set up, and the amount of the case reserve (on average, in the period between reporting and eventual payment) is b-dollars for each dollar that is ultimately paid. If case reserves are set conservatively (perhaps more likely during hard markets) we will have $b > 1$. In soft markets, case reserves are more likely to be set optimistically so b may take lower values, and we might have $b < 1$.

Under the above assumptions, for each dollar of ultimate, the expected cumulative amount paid by development time t is $F_P(t)$ and the expected amount outstanding at time t is $b.\{F_R(t) - F_P(t)\}$. So if we use $F_I(t)$ to denote the expected cumulative incurred run-off pattern then (from the definition of incurred as paid plus outstanding) we have:

$$\begin{aligned} F_I(t) &= F_P(t) + b.\{ F_R(t) - F_P(t)\} \\ &= b.\ F_R(t) + (1-b).\ F_P(t). \end{aligned} \tag{2}$$

This last equation can be interpreted by noting that incurred increases by the amount b when the claim is reported (and the case reserve set up), then increases by the amount (1- b) (which is usually negative) when the claim is paid.

Because of the possibility that $b > 1$, the function $F_I(t)$ is not in general a probability distribution function because it is not strictly increasing. Both $F_R(t)$ and $F_P(t)$ are strictly increasing (from 0 to 1) but if $b > 1$, $F_I(t)$ will show the usual incurred run-off shape: increasing rapidly then decreasing towards ultimate. This is illustrated in Exhibit 2, which shows typical run-off patterns for the case $b = 1.5$ (that is, case reserves are on average 50% higher than what is ultimately paid in respect of the reported claims).

Suppose we have cumulative paid and cumulative incurred run-off data. If we assume some parametric family of curves for $F_P(t)$ and $F_R(t)$, Equation 2 then implies a parametric family for $F_I(t)$. The parameters can be estimated by fitting the curve $F_P(t)$ to the cumulative paid data, and the curve $F_I(t)$ to the incurred data. Note that b is one of the parameters that will be estimated from the data. Some suitable parametric distributions for $F_P(t)$ and $F_R(t)$ are considered in the next sub-section.

The bias factor b need not necessarily be assumed to take a constant value across development time (within each single origin year). It seems likely that the accuracy of case reserves might sometimes improve with time. This possibility can be allowed for by using a model of the form:

$$b_t = \exp\{\beta_0 + \beta_2.\max(0, t_0\text{-}t)\}/ \tag{2a}$$

Instead of a single constant parameter b, this form of model has three parameters β_0, β_2 and t_0. (A further parameter, β_1, is introduced in Section 2.3.2.) The exponentiation ensures that the bias factor b_t is always positive. The expression max(0, t_0-t) allows b_t to change between development times t=0 and t=t_0. At later development times, max(0, t_0-t) is zero so the bias factor b_t settles at $\exp\{\beta_0\}$.

2.2.2 Suitable parametric distribution families

When selecting a family of curves to fit to paid claims development data, in principle any analytic family of probability distribution functions could be tried: for example Log-Normal, Pareto, Gamma, Weibull, etc. However, in this paper we restrict attention to distribution families that have the following properties:

- The cumulative distribution function is mathematically simple. This is desirable so that curve-fitting can be carried out quickly and easily in a spreadsheet.
- Ability to accommodate a wide range of values for the ratio of mode to mean. In particular, the distribution family should include distributions with mode equal to zero as well as distributions with mode greater than zero. [Recall that the mode of a distribution is the point where the density function takes its maximum value, or equivalently, where the slope of the cumulative distribution function is greatest.] This is desirable so that the same family of distributions can reasonably be used for reporting delays and for payment delays (which is merely convenient, not strictly necessary). In some classes of insurance, reporting delays tend to be very short for the majority of claims so that the mode is close to zero. For payment delays, the mode is invariably greater than zero.

The Log-Normal distribution (for example) is not used in this paper because it does not satisfy either of these criteria. The Log-Normal cumulative distribution function can be calculated from the Normal distribution function (which is available in popular spreadsheet software) but it is relatively complex and slow to calculate compared to some simpler distribution functions. The mode and mean of a Log-Normal distribution (using the usual μ and σ parameterization) are respectively $\exp(\mu)$ and $\exp(\mu + \sigma^2/2)$, so the ratio of mode to mean is $\exp(-\sigma^2/2)$. Although this can take any value between zero and one, a mode of zero is not possible.

Based on the above criteria, three distribution families have been selected for use in the present paper to model development patterns. Other distribution families could be used within the framework developed here, and some might prove to be more suitable than the selected three. These three distribution families have been chosen for convenience, on the basis that we have to start somewhere, and because they are probably as good as any for illustrating the principles of the proposed method.

The three distribution families used in this paper are the Weibull, the Burr, and the Inverse Burr. Table 1 gives the cumulative distribution function $F(t)$ and the mean and the mode of these distributions. (The penultimate column gives conditions for the mode to be greater than zero, and the final column gives the formula for the mode when it is greater than zero.) $\Gamma(.)$ denotes the Gamma function, which can be evaluated in Excel® as EXP(GAMMALN(x)).

Table 1: Formulas for analytic delay distributions

	$F(t)$	Mean	Mode>0 if	Mode (if > 0)
Weibull	$1 - \exp\{-(t/s)^c\}$	$s.\Gamma(1+1/c)$	$c > 1$	$s.(1-1/c)^{1/c}$
Burr	$1 - 1/\{1+(t/s)^c\}^a$	$s.\Gamma(1+1/c).\Gamma(a-1/c)\ /\ \Gamma(a)$	c > 1	$s.\{(c-1).(ac+1)\}^{1/c}$
Inv Burr	$1/\{1+(s/t)^c\}^a$	$s.\Gamma(a+1/c).\Gamma(1-1/c)\ /\ \Gamma(a)$	$ac > 1$	$s.\{(ac-1).(c+1)\}^{1/c}$

In all three cases, $F(t)$ increases monotonically from 0 when t=0, towards 1 as t tends to infinity. The parameter s is a scale parameter; the parameters a and c are shape parameters. The Weibull family has just one shape parameter; the Burr and Inverse Burr families each have two shape parameters. The additional shape parameter means that the Burr and Inverse Burr families are much larger and more flexible then the Weibull family. The Burr and Inverse Burr families have some well-known sub-families. The Pareto is the sub-family of the Burr family obtained by setting c to 1. The Inverse Pareto is the sub-family of the Inverse Burr obtained by setting c to 1. The Log-Logistic is the sub-family of both Burr and Inverse Burr families obtained by setting a to 1. Each of these sub-families has one shape parameter, so in that sense, is as large as the Weibull family.

Although we use all three of these distribution families in Section 3 of this paper, in the remainder of Section 2 we use the Weibull distribution for both payment delays and reporting delays. The Weibull is used because it is a relatively simple distribution, and it serves to illustrate the principles of the proposed modeling method. No implication is intended that the Weibull is superior to any other distribution family for this purpose. As already noted, any analytic distribution family could be used within the framework developed in Section 2.2.1, and using the same principles as are illustrated below using the Weibull distribution. There is also no reason in principle why two different distribution families should not be used; one to model reporting delays $F_R(t)$, and another to model payment delays $F_P(t)$, (provided $F_R(t) \geq F_P(t)$ for all values of t).

2.2.3 Model for single origin year based on Weibull distributions

Here the Weibull distribution is used to model both reporting and payment delays. Subscripts R and P are used to distinguish parameters of the reporting and payment delay distributions. The symbol ^ is used to denote exponentiation, that is: (t/s)^$c = (t/s)^c$.

For the reporting and payment delay distributions we have:

$$F_R(t) = 1 - \exp\{-(t/s_R)^\wedge c_R\} \tag{3}$$
$$F_P(t) = 1 - \exp\{-(t/s_P)^\wedge c_P\} \tag{4}$$

So from Equation 2, the incurred development pattern is:

$$F_I(t) = 1 - b.\exp\{-(t/s_R)\hat{}c_R\} - (1-b).\exp\{-(t/s_P)\hat{}c_P\} \quad (5)$$

The above equations for $F_P(t)$ and $F_I(t)$ specify a model for the paid and incurred run-off patterns for each dollar of ultimate. If U denotes the ultimate paid (which is what we aim to estimate), then the expected amounts paid and incurred by development time t are respectively $U.F_P(t)$ and $U.F_I(t)$.

From Equations 3 and 4, the requirement $F_R(t) \geq F_P(t)$ (for all values of t) is equivalent to $(t/s_R)\hat{}c_R \geq (t/s_P)\hat{}c_P$, which in turn is equivalent to $t\hat{}(c_R - c_P) \geq (s_R\hat{}c_R)/(s_P\hat{}c_P)$. If c_R is not equal to c_P, then the left side of this inequality takes all values between zero and infinity as t varies between zero and infinity. So the only way this inequality can be true for all positive values of t is by having c_R equal to c_P (so the left side is equal to 1 for all t) and s_R less than s_P (so the right side is less than 1). However, in practice, it is of little consequence if $F_R(t)$ is less than $F_P(t)$ for high values of t (that is, where both $F_R(t)$ and $F_P(t)$ are very close to 1). So we will not insist on the constraint $c_R = c_P$. Instead, it is proposed to check that the fitted curves are reasonable by viewing them graphically. This is illustrated by Exhibit 1, in which both curves are Weibull distributions, with parameters $s_R = 1$, $c_R = 1$, $s_P = 4$, $c_P = 3$. These parameters give $F_R(t) \geq F_P(t)$ for $t \leq 8$, but $F_R(t) < F_P(t)$ for $t > 8$. Since both curves reach 99.97% development at $t = 8$, it is of no practical consequence that $F_R(t) < F_P(t)$ for $t > 8$.

2.3 Model for multiple origin years

2.3.1 Variation of parameters across origin years

In Section 2.2 we developed a model for paid and incurred development patterns of a single origin year. If the Weibull distribution is used for both reporting and payment delays (as in 2.2.3), then the model has six parameters for each origin year: U, b, s_R, c_R, s_P, c_P. This is clearly too many parameters to attempt to estimate separately for each origin year. For the latest origin year we usually have only two data values: one paid, one incurred (although there may be more if sub-annual development periods are used).

The total number of parameters needs to be reduced. To achieve this, we could try assuming initially that some of the parameters take a single constant value across all origin years. For example, we might assume that s_R, c_R and c_P take the same values for all origin years, so that only the parameters U, b and s_P vary across origin years. Setting s_R and c_R to values that are constant across all origin years is appropriate if reporting delays have the same distribution $F_R(t)$ across all origin years. For some datasets this might turn out to be a reasonable assumption. (We discuss later how this assumption can be tested. Applications of the model presented in this paper to actual datasets have shown evidence that reporting delays do sometimes vary with the underwriting cycle: possible causes are discussed in Section 4.1.3.)

By allowing b to vary across origin years, we allow for the possibility that case reserves are set up more or less conservatively at different points in the underwriting cycle. This possibility is suggested in the existing literature discussed in Section 1. By allowing the scale parameter s_P of the payment delay distribution to vary, we allow for variation in the speed of claim settlement. Previous research has found evidence that such variation does occur with the underwriting cycle (see Section 1). Clearly, the ultimate U must also be allowed to take a different value for each origin year as this is what we aim to estimate.

In the remainder of this paper, parameters that are allowed to vary across origin years have a subscript j to label the origin year. So in the Weibull model, if parameters U, b and s_P are allowed to vary, these are denoted U_j, b_j, s_{Pj}.

2.3.2 Allowing for cyclical development patterns

To allow for underwriting cycle effects, parameters of the run-off curves that are not held constant across all origin years can be linked to a known premium rate index. Interpretation of model parameters is simplified if the premium rate index (denoted Q_j for origin year j) is scaled so the mean value across all origin years is 1. We can then use equations of the form:

$$b_j = \exp\{\beta_0 + \beta_1.(Q_j\text{-}1)\} \tag{6}$$

$$s_{Pj} = \exp\{\sigma_0 + \sigma_1.(Q_j\text{-}1)\} \tag{7}$$

Here, β_0, β_1, σ_0 and σ_1 are parameters (to be estimated from the run-off data) that are assumed to take the same values for all origin years. The subtraction of 1 from the premium rate index further simplifies interpretation of the parameters: for example, $\exp\{\beta_0\}$ represents the value of b for an average year in which $Q_j = 1$. The exponentiation ensures that parameters b_j and s_{Pj} are always positive (which is necessary to produce valid development curves). Note that this form of model for b_j and s_{Pj} allows for the possibility that the known index Q_j may understate the true amplitude of the underwriting cycle. Previous research (see Section 1) suggests that this is often the case with premium rate indices. If this is the case, the parameters β_1 and σ_1 estimated from the run-off data will simply take higher values than they would if Q_j correctly reflected the amplitude of the reserving cycle. The number of parameters to be estimated from the paid and incurred run-off data is now reasonable. We have:

- Seven parameters each assumed to take a single constant value across all origin years (s_R, c_R, c_P, β_0, β_1, σ_0 and σ_1), plus
- One parameter taking a different value for each origin year (the ultimate U_j).

If there are more than eight development periods, the basic chain ladder model has more parameters than this.

If the possibility that bias factors change with development time is allowed for as described in Section 2.2.1 (Equation (2a)) then Equation (6) becomes:

$$b_t = \exp\{\beta_0 + \beta_1.(Q_j\text{-}1) + \beta_2.\max(0, t_0\text{-}t)\} \tag{6a}$$

2.3.3 Estimation of parameters by least squares

The parameters of the paid and incurred development curves can be determined by the method of least-squares. The following notation is used in this section:

- $F_{Pj}(t)$ denotes the cumulative paid development curve for origin year j. Using the Weibull model, this is given by Equation 4, with scale parameter s_P replaced by s_{Pj} from Equation 7.
- $F_{Ij}(t)$ denotes the cumulative incurred development curve for origin year j. Using the Weibull model, this is given by Equation 5, with scale parameter s_P replaced by s_{Pj} from Equation 7, and the case-reserve redundancy-factor b replaced by b_j from Equation 6 (or Equation 6a).
- P_{jd} denotes the actual cumulative paid for origin year j and development period d.
- I_{jd} denotes the actual cumulative incurred for origin year j and development period d.

The residual sum of squares is defined as the sum of squared differences between actual and expected values. This can be calculated separately for paid and incurred:

$$\begin{aligned} \text{RSS}_P &= \Sigma\ \{P_{jd} - U_j.F_{Pj}(t)\}^2 \\ \text{RSS}_I &= \Sigma\ \{I_{jd} - U_j.F_{Ij}(t)\}^2 \end{aligned} \tag{8}$$

Summation is over all origin years j and development periods d in the run-off arrays. In the case of annual development data, d denotes the development year. We use the convention that the origin year itself is development year 0, so d takes the values 0, 1, 2, etc. In the fitted curves ($F_{Pj}(t)$ and $F_{Ij}(t)$), t denotes continuous development time. Ideally this would be the exact elapsed time from the date of loss occurrence. However, since claims in a particular origin year cohort do not usually all have exactly the same date of loss occurrence, t is set to an approximate average delay from the date of loss occurrence until the end of the corresponding development period d. Table 2 gives appropriate values of t for each development year d, for both accident year and underwriting year cohorts. These approximations are based on assumptions that accidents occur uniformly in time and policies incept uniformly. Further details are given in Appendix B.

Similar approximations can be used for sub-annual development periods. Approximations such as these tend to be relatively crude for early development periods: this is discussed further in Section 2.5.3.

Table 2: Approximate mean delay in each development year

Development year (d)	0	1	2	3	4	5+
Accident year mean delay (t)	0.5	1.5	2.5	3.5	4.5	d+0.5
Underwriting year mean delay (t)	0.333	1	2	3	4	d

Given values for the parameters of the development curves (in the case of the Weibull model: s_R, c_R, c_P, β_0, β_1, σ_0 and σ_1) and a value for the ultimate U_j of each origin year, the "expected" values $U_j.F_{Pj}(t)$ and $U_j.F_{Ij}(t)$ can be calculated corresponding to each cell (j,d) of the run-off array. From these, the residual sums of squares RSS_P and RSS_I can be calculated (Equation 8). The least squares estimation method is to search for the values of the parameters (s_R, c_R, c_P, β_0, β_1, σ_0, σ_1 and U_j for each origin year) that minimize the residual sums of squares. Note that ultimates U_j can be treated as parameters of the model and estimated by least squares along with the other parameters. However for early origin years, the ultimate may already be known with some precision. If, for a particular origin year all reported claims have been settled and further claims are considered unlikely, then there is no need to estimate the ultimate U_j by least squares, and a better model will usually be obtained by using the known value of this quantity. Usually this applies only for the earlier origin years: for origin years that are not fully developed the ultimate is estimated by least squares.

2.3.4 Combining paid and incurred by weighted least squares

It is clearly possible (provided the number of data-points exceeds the number of parameters) to carry out least squares estimation separately for paid and incurred. However, the paid and incurred models have parameters in common. (In the case of the Weibull model of Section 2.3.2, the following parameters feature in both the paid and incurred models: c_p, σ_0, σ_1 and the ultimate U_j of each origin year.) If least squares estimation is carried out separately for paid and incurred, the paid data will yield one set of estimates for these parameters, and the incurred data will yield another set of values for the same parameters. This can be avoided by carrying out the least squares procedure just once based on the total residual sum of squares $\mathrm{RSS}_P + \mathrm{RSS}_I$.

This raises the question of relative weighting between paid and incurred: is it correct to give RSS_P and RSS_I equal weight by just adding them? An alternative would be to find the parameter values that minimize $\mathrm{RSS}_P + w.\mathrm{RSS}_I$ where w is a predefined weighting factor. To increase the influence of the incurred data relative to the paid data, we would choose a value for w that is greater than 1, and to give more influence to the paid data we would choose a value less than 1.

One way to justify a relative weighting on theoretical grounds would be to develop a full stochastic model that treats each value P_{jd} and I_{jd} as a random variable and gives an expression for the variance of each one. The basic theoretical justification for the least squares method is two-fold:

- The Gauss-Markov theorem states that, in linear models, weighted least squares estimates have the smallest variance of all linear unbiased estimates.
- Quasi-likelihood theory shows that weighted least squares estimates are asymptotically unbiased and efficient (that is, have minimum possible variance) even in non-linear models.

In both Gauss-Markov and quasi-likelihood theory, the weights that give optimal least squares estimates are inversely proportional to the variances of the corresponding random variables. In addition, if some of the random variables are correlated, then the residual sum of squares that is minimized should include cross terms with weights depending on the covariance between the corresponding variables.

In the present application, it is clear that the observations (P_{jd} and I_{jd}) are not all mutually independent. Since incurred is paid plus outstanding, any reasonable stochastic model would indicate a positive covariance between the values P_{jd}, I_{jd} with the same values of j and d. Furthermore, since these are cumulative values, it is likely that there is serial correlation between successive values of P_{jd} as the development period d increases in each fixed origin year j. For this reason, a full stochastic model would indicate that the optimum "residual sum of squares" to be minimized should include cross terms such as $\{P_{jd} - U_j.F_{Pj}(t)\}.\{I_{jd} - U_j.F_{Ij}(t)\}$ as well as pure squared terms such as $\{P_{jd} - U_j.F_{Pj}(t)\}^2$. A full stochastic model would also indicate the optimum relative weighting of every term in the residual sum of squares.

Development of a full stochastic model is not attempted in this paper because it would be mathematically complex and probably contentious (as it would require many assumptions about the nature of the stochastic variation in paid and incurred run-off data). Instead, the aim is to produce a method that is simple enough to be widely useful if applied intelligently. To this end, it is proposed to ignore the clear correlation that will exist between the observed values of P_{jd} and I_{jd} by including no cross-terms in the residual sum of squares. We also take no account of differing variances among the P_{jd} (and among the I_{jd}) by giving every term equal weight in RSS_P (and in RSS_I). The remaining question is: should we take this cavalier approach one step further by giving equal weight to both RSS_P and RSS_I ?

2.3.5 Empirical determination of relative weighting of paid and incurred

At this point, it is proposed to allow for the possibility that one of the two datasets (either paid or incurred) may appear to be more reliable than the other. The theory (Gauss-Markov and quasi-likelihood) suggests that the two terms (RSS_P and RSS_I) should be weighted in inverse proportion to the mean variance of paid and incurred data-points. That is, instead of minimizing RSS_P+RSS_I we should minimize $(RSS_P/\sigma^2_P)+(RSS_I/\sigma^2_I)$, where σ^2_P and σ^2_I are typical variances of individual paid and incurred observations. This is equivalent to minimizing the following total weighted sum of squares:

$$\text{Weighted sum of squares} = RSS_P + w_I.RSS_I \text{ where } w_I = \sigma^2_P/\sigma^2_I. \tag{9}$$

Instead of using a stochastic model to determine the relative magnitudes of σ^2_P and σ^2_I, a purely empirical approach can be used in which their relative magnitudes are estimated from the residuals. Standard theory suggests the variance of a paid observation be estimated as:

$$\sigma^2_P = RSS_P/(n_P - p_P) \tag{10}$$

where n_P is the number of paid observations, and p_P is the number of parameters estimated from these observations. If variances are estimated in this way we will have $(RSS_P/\sigma^2_P) = (n_P - p_P)$ and $(RSS_I/\sigma^2_I) = (n_I - p_I)$, so the total weighted sum of squares $(RSS_P/\sigma^2_P)+(RSS_I/\sigma^2_I)$ will be $(n_P + n_I) - (p_P + p_I)$. An iterative fitting procedure is necessary to achieve this. It is also necessary to divide the total parameter count into the two components p_P and p_I. Each parameter that features in the fitted curves of both paid and incurred (for example, the ultimates U_j) makes a fractional contribution to both p_P and p_I. What is believed to be a reasonable pragmatic approach is proposed for this purpose. (This is discussed further in Section 2.5.1, but no rigorous theoretical justification is claimed.)

For example, suppose we have annual paid and incurred run-off arrays for 10 origin years, so that $n_P = n_I = 55$. Suppose we are using the Weibull model described in Section 2.3.2. The paid development curves depend on 13 parameters: c_P, σ_0, σ_1 and U_1, $U_2 \ldots U_{10}$. The incurred development curves depend on 17 parameters: the same 13 as in the paid model, plus s_R, c_R, β_0 and β_1. If we count a parameter that features in both paid and incurred models as half a parameter in each model, we have: $p_P = 6.5$ and $p_I = 10.5$, which gives the correct total number of distinct parameters: $p_P + p_I = 17$. (A more refined method of counting parameters is discussed in Section 2.5.1.)

On the first iteration, we estimate the parameters by minimizing RSS_P+RSS_I (that is, using equal weights initially). Suppose this produces parameter values that give $RSS_P = 100$ and $RSS_I = 300$, so the total minimized residual sum of squares is 400. Initial estimates of the mean variances for paid and incurred are then: $\sigma^2_P = 100/(55–6.5) = 2.06$ and $\sigma^2_I = 300/(55–10.5) = 6.74$ (If we had reason to believe that σ^2_P and σ^2_I were equal, we would estimate the value as 400/(110–17) = 4.30.) σ^2_I being so much higher than σ^2_P indicates that the model does not fit the incurred data as closely as it fits the paid data. The incurred data are therefore less reliable than the paid for the purpose of projecting run-off patterns, and so should be given less weight than paid in fitting the model. So for the next iteration, instead of minimizing RSS_P+RSS_I, we minimize (RSS_P/2.06)+(RSS_I/6.74). Multiplying by 2.06, we see that this is equivalent to minimizing $RSS_P + 0.31 * RSS_I$. Minimizing this might result in $RSS_P = 95$ (that is, a closer fit to the paid data than on the first iteration) and $RSS_I = 350$ (a poorer fit to the incurred data than on the first iteration). These figures give the following revised estimates of variances: $\sigma^2_P = 95/(55–6.5) = 1.96$ and $\sigma^2_I = 305/(55–10.5) = 7.87$. So on the third iteration, we minimize $RSS_P + 0.25 * RSS_I$. Continuing in this way, convergence usually occurs after a few iterations.

If the model is set up in Excel®, the Excel solver can be used to search for the parameter values that minimize the required weighted sum of squares in each iteration. Note however, that since the fitted curves are non-linear functions of the parameters, solver does not guarantee to find the global minimum. It is advisable to try several sets of starting values if there is any doubt about the solution found by the Excel solver.

2.4 Use of premium and exposure data

2.4.1 Estimated ultimate for latest origin year

In the model described so far, no use has been made of premium or other exposure data. For the latest origin year ($j = J$ say), the ultimate U_J is estimated purely by fitting curves to development patterns, and assuming that changes in the parameters of these curves from one origin year to the next are linked to changes in the underwriting cycle (through equations such as 6 and 7). The latest origin year has one free parameter of its own (the ultimate U_J), and (assuming annual paid and incurred development data) two data-values: the actual paid and incurred amounts at the end of the zeroth development year P_{J0} and I_{J0}.

This latest origin year contributes just two terms to the weighted sum of squares:

$$\{P_{J0} - U_J.F_{PJ}(t)\}^2 + \mathrm{w}.\{I_{J0} - U_J.F_{IJ}(t)\}^2 \tag{11}$$

Here, $t = ½$ for accident year cohorts, or 1/3 for underwriting year cohorts: see Table 2.

Since U_J does not appear in any other terms of the total sum of squares, it can be adjusted to minimize the sum of the above two terms. Elementary calculus shows that this gives:

$$U_J = \{P_{J0}.F_{PJ}(t) + w.I_{J0}.F_{IJ}(t)\} / \{F_{PJ}(t)^2 + w.F_{IJ}(t)^2\} \tag{12}$$

If we write U_{PJ} and U_{IJ} for the ultimates projected from the latest paid or incurred separately (that is $U_{PJ} = P_{J0} / F_{PJ}(t)$ and $U_{IJ} = I_{J0} / F_{IJ}(t)$), then we have:

$$U_J = \{U_{PJ}.F_{PJ}(t)^2 + w.U_{IJ}.F_{IJ}(t)^2\} / \{F_{PJ}(t)^2 + w.F_{IJ}(t)^2\} \tag{13}$$

This shows that U_J is a weighted average of U_{PJ} and U_{IJ} with weights $F_{PJ}(t)^2$ and $w.F_{IJ}(t)^2$.

For example, suppose paid and incurred amounts for the latest origin year are P_{J0} = \$200 and I_{J0} = \$1000, and suppose the fitted development curves imply that these figures are respectively 20% and 110% of ultimate, that is: $F_{PJ}(t) = 0.2$ and $F_{IJ}(t) = 1.1$. Then projecting paid and incurred separately to ultimate gives the estimates: U_{PJ} = \$200/0.2 = \$1000 and U_{IJ} = \$1000 / 1.1 = \$909. Suppose further that the analysis described in Section 2.3.3 indicates that incurred sums of squares should receive a weight of 0.2 relative to paid (that is, $w = 0.2$), then we have : $F_{PJ}(t)^2 = 0.04$ and $w.F_{IJ}(t)^2 = 0.24$. Equation 13 then gives a combined estimated ultimate: U_J = (0.04 * \$1000 + 0.24 * \$909) / 0.28 = \$922.

Note that in practice, the model can be set up in Excel, and least-squares estimation carried out using the Excel solver. It is not necessary to evaluate the formulas given above: the estimated ultimates are parameters of the model that are found by the Excel solver.

2.4.2 Model for ultimate in terms of premium or other exposure information

From Equation 12, it is clear that the estimated ultimate for the final origin year will be sensitive to the values of the two observations P_{J0} and I_{J0}. The sensitivity of the estimate to these values can be reduced (hence the reliability of the estimated ultimate increased) if total premium or some other measure of exposure can be obtained. A measure of exposure other than premium is more valuable than premium, because to make use of premium, we also have to use the estimated premium rate index for the latest year (Q_J) and this is already used in $F_{PJ}(t)$ and $F_{IJ}(t)$ (through equations such as 6 and 7). Premium and exposure data will clearly also be useful for other origin years, but it is for the latest few origin years that this additional information is most valuable.

First we consider using a measure of exposure other than premium. This might be, for example, gross tonnage in a marine account, or total payroll in workers compensation. The exposure for origin year j is denoted X_j. If there are no cycles or trends in the ultimate loss per unit of exposure, then we have:

$$U_j = r.X_j + \text{random error.} \tag{14}$$

Here, the parameter *r* represents the mean ultimate loss per unit of exposure.

However, it could be that there is a trend in ultimate loss per unit of exposure. We should at least expect an inflationary trend if the exposure measure is not in dollars. In this case, we could try a model of the form $r_j = \exp(\varrho_0 + \varrho_1.j)$ (where ϱ_0 and ϱ_1 are parameters assumed to take constant values across all origin years). We can also allow for the possibility that the ultimate loss per unit of exposure varies with the underwriting cycle:

$$r_j = \exp(\varrho_0 + \varrho_1.j + \varrho_2.Q_j). \tag{15}$$

Using this model, Equation 14 becomes:

$$U_j = X_j.\exp(\varrho_0 + \varrho_1.j + \varrho_2.Q_j) + \text{random error.} \tag{16}$$

In the event that the only measure of exposure available is premium (denoted Prem_j) then underwriting cycle effects need to be removed from this by using $X_j = \text{Prem}_j / Q_j$ in the above. The possibility that the premium rate index understates the true amplitude of the underwriting cycle is accommodated (approximately) by the inclusion of Q_j in the exponential factor: in this case the parameter ϱ_2 will be lower than it would be if Q_j correctly reflected the amplitude.

Another possibility is to remove the exponentiation from Equation 16 (so any inflationary trend is approximated as linear) to give:

$$U_j = \text{Prem}_j.(\varrho_2 + \varrho_1.j/Q_j + \varrho_0/Q_j) + \text{random error.} \tag{17}$$

If premium takes several years to develop to ultimate (as is often the case in London market business because of profit-sharing, reinstatement premiums, retrospective experience rating, end-of-term exposure adjustments, etc.), then Prem_j could be obtained by applying a simple projection method (such as chain ladder) to the premium development array.

In all the above equations the "random error" term reflects real variation in loss experience from one origin year to another. As a first approximation, it is probably reasonable most of the time to assume that the variance of this is proportional to the expected ultimate U_j and to approximate this as being proportional to Prem_j/Q_j. Proportionality of variance to expected value implies that the coefficient of variation is inversely proportional to the square-root of the expected ultimate, reflecting the diversification benefit of large portfolios.

2.4.3 Use of exposure information in curve-fitting

To make use of the premium (or other exposure) information in curve-fitting by least-squares, we need to add a further term to the sum of squares that is minimized (see Section 2.3.3). From Equation 17, and the above assumption on the approximate variance of the random error term, the additional sum of squares in respect of exposure is given by:

$$\mathrm{RSS}_X = \mathrm{Prem}_0.\Sigma_j \{U_j - \mathrm{Prem}_j.(\varrho_2 + \varrho_1.j/Q_j + \varrho_0/Q_j) \}^2 .Q_j / \mathrm{Prem}_j. \tag{18}$$

Here, summation is over all origin years j. Prem_0 represents the mean value of Prem_j across all origin years (or some other suitable value, e.g., $\mathrm{Prem}_0 = \mathrm{Prem}_J$). This factor is included to offset the factor Q_j/Prem_j applied to each term in the sum: it ensures that RSS_X is in units of "dollars-squared" and of the same order of magnitude as RSS_P and RSS_I. The total sum of squares to be minimized becomes $\mathrm{RSS}_P + w_I.\mathrm{RSS}_I + w_X.\mathrm{RSS}_X$, where RSS_P and RSS_I are given by Equations 8 and 9, and appropriate values for the weights w_I and w_X can be determined iteratively using the principles described in Section 2.3.3.

For example, consider again the case of annual paid and incurred run-off arrays for 10 origin years, so that $n_P = n_I = 55$. We now have ten additional pieces of information: the estimated ultimate premiums Prem_j, denoted by $n_X = 10$. Once again using the Weibull model described in Section 2.3.2: the paid development curves depend on 13 parameters (c_P, σ_0, σ_1 and U_1, $U_2 \ldots U_{10}$), the incurred development curves depend on 17 parameters (the same 13 as in the paid model plus s_R, c_R, β_0 and β_1) and the exposure model depends on 13 parameters (the ten ultimates and the three rho-parameters of Equation 17). Parameters that feature in more than one of the three components can be counted in proportion to the number of data-points in each component. For example, as the parameter U_1 is determined using all 120 data-points, it is counted as 55/120 of a parameter in the paid model, 55/120 of a parameter in the incurred model, and 10/120 of a parameter in the exposure model. On this basis we have: $p_P = 10 * 55/120 + 3 * 55/110 = 6.08$, $p_I = 10 * 55/120 + 3 * 55/110 + 4 = 10.08$, and $p_X = 10 * 10 / 120 + 3 = 3.83$, which gives the correct total number of distinct parameters: $p_P + p_I + p_X = 20$. (A more refined method of counting parameters is proposed and discussed in Section 2.5.1.)

On the first iteration, we estimate the parameters by minimizing $\mathrm{RSS}_P+\mathrm{RSS}_I+\mathrm{RSS}_X$ (that is, using equal weights initially). Suppose this produces parameter values that give $\mathrm{RSS}_P = 102$ and $\mathrm{RSS}_I = 308$ and $\mathrm{RSS}_X = 50$, so the total minimized residual sum of squares is 460. (Note that RSS_P and RSS_I are necessarily higher than they were when RSS_X was not considered.) Initial estimates of the variances are then: $\sigma^2_P = 102/(55–6.08) = 2.09$ and $\sigma^2_I = 308/(55–10.08) = 6.86$ and $\sigma^2_x = 50 / (10 – 3.83) = 8.1$.

For the second iteration, we minimize $(\mathrm{RSS}_P/2.09)+(\mathrm{RSS}_I/6.86)+(\mathrm{RSS}_X/8.1)$, which is equivalent to minimizing: $RSS_P + 0.304 * RSS_I + 0.258 * RSS_X$. The values of RSS_P, RSS_I and RSS_X given by minimizing this weighted sum are then used to calculate revised estimates of σ^2_P, σ^2_I and σ^2_X, and the weights for the third iteration calculated using $w_I = \sigma^2_P/\sigma^2_I$ and $w_X = \sigma^2_P/\sigma^2_X$. Convergence usually occurs after a few iterations.

2.4.4 Effect of exposure information on projected ultimate for latest origin year

In Section 2.4.1 we considered the estimated ultimate for the latest origin year (U_J) obtained using just two pieces of information for that origin year: P_J and I_J. We now consider the estimate of U_J obtained by, in addition, using the premium data as described in 2.4.3.

This latest origin year now contributes three terms to the weighted sum of squares:

$$\{P_{J0} - U_J.F_{PJ}(t)\}^2 + w_I.\{I_{J0} - U_J.F_{IJ}(t)\}^2 + w_X.\{U_J - \text{Prem}_J.R_J\}^2\ .Q_J.\text{Prem}_0/\text{Prem}_J \qquad (19)$$

(Here R_J denotes ($\varrho_2 + \varrho_1.J/Q_J + \varrho_0/Q_J$), which can be regarded as the expected ultimate loss ratio for the latest origin year.)

Since U_J does not appear in any other terms of the total sum of squares, it can be adjusted to minimize the sum of the above three terms. It is easily proved that this gives:

$$U_J = \{P_{J0}.F_{PJ}(t) + w_I.I_{J0}.F_{IJ}(t) + w_X.R_J.Q_J.\text{Prem}_0\} / \{F_{PJ}(t)^2 + w_I.F_{IJ}(t)^2 + w_X.Q_J.\text{Prem}_0/\text{Prem}_J\}. \qquad (20)$$

If we write U_{PJ}, U_{IJ} and U_{XJ} for ultimates estimated respectively from paid, incurred and premium data separately (that is $U_{PJ} = P_{J0} / F_{PJ}(t)$, $U_{IJ} = I_{J0} / F_{IJ}(t)$ and $U_{XJ} = \text{Prem}_J.R_J$), then we have:

$$U_J = \{U_{PJ}.F_{PJ}(t)^2 + w_I.U_{IJ}.F_{IJ}(t)^2 + w_X.U_{XJ}.Q_J.\text{Prem}_0/\text{Prem}_J\} \\ /\{F_{PJ}(t)^2 + w_I.F_{IJ}(t)^2 + w_X.Q_J.\text{Prem}_0/\text{Prem}_J\}. \qquad (21)$$

This shows that U_J is now a weighted average of U_{PJ}, U_{IJ} and U_{xJ}.

For example, suppose paid and incurred amounts for the latest origin year are P_{J0} = \$200 and I_{J0} = \$1000, and suppose the fitted development curves imply that these figures are respectively 20% and 110% of ultimate, that is: $F_{PJ}(t) = 0.2$ and $F_{IJ}(t) = 1.1$. Then projecting paid and incurred separately to ultimate gives the estimates: U_{PJ} = \$200/0.2 = \$1000 and U_{IJ} = \$1000 / 1.1 = \$909. In addition, suppose we have Prem_J = \$1100 and, from the fitted loss-ratio model, R_J = 104%, so that U_{XJ} = \$1144. Suppose further that the analysis described in the previous section converges to $w_I = 0.29$ and $w_X = 0.24$. If the latest origin year is believed to be at the mid-point of the underwriting cycle (so $Q_J = 1$) and we have used the normalizing factor $\text{Prem}_0 = \text{Prem}_J$, then we have: $F_{PJ}(t)^2 = 0.04$, $w_I.F_{IJ}(t)^2 = 0.35$ and $w_X.Q_J.\text{Prem}_0/\text{Prem}_J = 0.24$. The above formula then gives a final estimate: U_J = (0.04 * \$1000 + 0.35 * \$909 + 0.24 * \$1144) / 0.63 = \$1004.

For earlier origin years, the influence of the exposure information will be lower because the number of terms in the sum of squares relating to paid and incurred development data is higher for earlier years, while the number of terms relating to exposure data remains at one for each origin year.

2.5 Parameter Counts and Significance Tests

2.5.1 Parameter counts

In the examples of Sections 2.3.3 and 2.4.3, the total parameter count was apportioned between the different sub-sets of data (paid, incurred and exposure data) in proportion to the number of data-points. The example of Section 2.3.3 has an equal volume of paid and incurred data (55 paid and 55 incurred observations) and no exposure data, so each parameter that features in both paid and incurred models was counted as half a parameter in each of these models. The example of Section 2.4.3 has 10 exposure observations in addition to the 55 paid and 55 incurred observations. Each parameter that contributes to all three parts of the model was counted as 10/120 of a parameter in the exposure part, and 55/120 in each of the paid and incurred parts.

The rationale for splitting parameter counts in this way is that it approximately reflects the relative influence of each type of data in determining the value of the parameter. This can be further refined by taking account of the relative weights used in the total weighted sum of squares that is minimized. For example, if we have just 55 paid and 55 incurred observations (no exposure data), and the incurred data is given a weight of 0.5 relative to the paid data (that is, $w_I = 0.5$), then each incurred observation has (on average) only half the influence of each paid observation in determining parameter values. On this basis, a parameter whose value is determined from both paid and incurred data would be counted as 2/3 determined from paid data and 1/3 from incurred data.

In general using this method, each parameter contributes to p_P, p_I, and p_X in proportion to n_P, $w_I.n_I$, and $w_X.n_X$. Each parameter must have a total count of one, so for a parameter estimated from all three data sources, the contributions to p_P, p_I, and p_X are respectively:

$$n_P \,/\, (n_P + w_I.n_I + w_X.n_X), \quad w_I.n_I \,/\, (n_P + w_I.n_I + w_X.n_X), \quad w_X.n_X \,/\, (n_P + w_I.n_I + w_X.n_X).$$

For a parameter that is not estimated from all three sources, the corresponding term(s) must be omitted from the denominator of these expressions so that the total is always one for each parameter. No rigorous theoretical justification is claimed for this method of counting parameters: it is proposed as a reasonable pragmatic approach.

The method for counting parameters used in Sections 2.3.3 and 2.4.3 is as above but with the weights w_I and w_X omitted. Including these weights in the parameter counts slightly complicates the fitting process because the relative weights (estimated from the residual sums of squares as described in Sections 2.3.3 and 2.4.3) change at each iteration, so the parameter counts will also change at each iteration.

To illustrate this, consider again the example of Section 2.4.3. For the first iteration, we use $w_I = w_X = 1$. Therefore the parameter counts are initially as in Section 2.4.3: $p_P = 6.08$, $p_I = 10.08$, $p_X = 3.83$ (giving a total parameter count of 20). Estimates of variances obtained from the first iteration are then (exactly as in Section 2.4.3): $\sigma^2_P = 102/(55–6.08) = 2.09$ and $\sigma^2_I = 308/(55–10.08) = 6.86$ and $\sigma^2_x = 50/(10–3.83) = 8.1$.

For the second iteration, the relative weights become $w_I = 2.09 / 6.86 = 0.304$, and $w_X = 2.09 / 8.1 = 0.258$, so we minimize $\text{RSS}_P + 0.304 * \text{RSS}_I + 0.258 * \text{RSS}_X$ (which again, is exactly as in Section 2.4.3). However, unlike in Section 2.4.3, the parameter counts used in estimating variances from the results of the second iteration should now factor in the weights used in the second iteration. Parameter counts that factor in these weights are:

$p_P = 3 * \{55 / (55 + 0.304 * 55)\} + 10 * \{55 / (55 + 0.304 * 55 + 0.258 * 10)\}$,

$p_I = 0.304 * p_P + 4$,

$p_X = 10 * \{0.258 * 10 / (55 + 0.304 * 55 + 0.258 * 10)\} + 3$.

These evaluate to $p_P = 9.70$, $p_I = 6.95$, $p_X = 3.35$. (As a check, we see that the total of these parameter counts is still 20.) Estimates of variances from the second iteration are then: $\sigma^2_P = \text{RSS}_P/(55 - 9.70)$, $\sigma^2_I = \text{RSS}_I/(55 - 6.95)$ and $\sigma^2_x = \text{RSS}_X / (10 - 3.35)$. These give weights w_I and w_X for the third iteration, and hence parameter counts used in estimating variances from the results of the third iteration.

2.5.2 Statistical significance tests

When an additional parameter is introduced into a model, the minimized residual sum of squares is inevitably smaller than it was before the new parameter was introduced. (This is because the model without the additional parameter is equivalent to a model in which the additional parameter is set to zero. When the new parameter is introduced, it is no longer constrained to take the value zero. When non-zero values are allowed in minimizing the residual sum of squares, it is extremely unlikely that the minimum will occur at exactly the value zero.) So by introducing an increasing number of parameters, the quality of fit (as measured by the residual sum of squares) can be progressively improved until the number of parameters is equal to the number of observations: when this occurs a perfect fit is possible and the residual sum of squares becomes zero.

Clearly we should try to avoid over-fitting: that is, we should try to avoid including parameters that reflect only the random variation of the particular dataset rather than genuine underlying effects. The purpose of statistical significance tests is to avoid over-fitting: parameters are included in a model only if they are statistically significant.

In least squares estimation, an appropriate significance test is based on the size of the decrease in the minimized residual sum of squares (RSS) when a new parameter is introduced. If the minimized RSS reduces only slightly, then the new parameter may not be statistically significant. To judge whether a decrease in the minimized RSS is statistically significant, it should be compared to the mean RSS per "degree of freedom."

To illustrate, consider again the example of Section 2.3.3 based on 55 paid observations and 55 incurred observations. Suppose that initially we fit a model with 17 parameters and that we give equal weight to incurred and paid data so parameters are estimated by minimizing RSS_P + RSS_I. Suppose (as in Section 2.3.3) that the minimized value is 400.0. Now suppose that an additional parameter is introduced into the model (this might, for example, be the parameter β_I of Equation 6), and that the minimized RSS with this new parameter included is 397.0. Is the new parameter statistically significant?

To answer this, we note that the new parameter caused a decrease of 3.0 in the minimized RSS. If this is judged to be a large decrease, then we conclude that it is unlikely to have been caused purely by chance and therefore that the new parameter is statistically significant. To judge whether the decrease of 3.0 is large (statistically significant) or small (insignificant) we need to compare it to something else. At first sight, it might seem that a suitable quantity to compare this decrease to is the mean-squared residual. There are 110 residuals in total (equal to the number of data-points) so the mean-squared residual is 397/110 (= 3.61) when the new parameter is included, and 400/110 (=3.64) when it is not included. Clearly the mean-squared residual decreases as the number of parameters increases. To allow for this, the denominator used in calculating the mean needs to be adjusted for the number of parameters in the model. Instead of dividing by the number of observations, we should divide by the number of "degrees of freedom", which is defined as the number of observations less the number of fitted parameters. The number of parameters was 17 before then new parameter was introduced and 18 afterwards, so the numbers of degrees of freedom are respectively 93 and 92. The mean-squared residual per degree of freedom is therefore 400/93 (= 4.30) before the new parameter is introduced and 397/92 (=4.32) after. To judge whether the new parameter is statistically significant, we should compare the change in the RSS of 3.0 to the mean value 4.32. Since the change in the RSS is less than the mean RSS per degree of freedom, the new parameter is not statistically significant.

Note that the change in the RSS is compared to the mean obtained from the more general model (4.32 in this example) not the mean obtained from the model excluding the parameter (4.30 in this example). This is because, if a parameter is statistically significant, the mean RSS from the model with the parameter excluded would be wrongly inflated by the exclusion.

If the decrease in the RSS obtained by introducing an additional parameter is less than the mean RSS per degree of freedom after the parameter has been introduced (as in the above example), then the additional parameter is not statistically significant. However, a change in the RSS that exceeds the mean RSS per degree of freedom is not necessarily conclusive evidence of statistical significance. Clearly, the greater the ratio of change in RSS to RSS per degree of freedom, the greater the statistical significance of the new parameter. (In the above example, this ratio is 3.0/4.32 = 0.69.) This ratio is known as the "*F*-ratio" and, to help judge its statistical significance, it can be compared to a theoretical *F*-distribution. In idealized circumstances, the theoretical *F*-distribution is the probability distribution of an *F*-ratio under the hypothesis that the additional parameter is equal to zero. Although this is not exactly the case in practice, the theoretical *F*-distribution remains a useful tool in judging the statistical significance of *F*-ratios. If an *F*-ratio is in the extreme right tail of the theoretical *F*-distribution, this is evidence against the hypothesis that the true value of the parameter is zero. In other words, the parameter is statistically significant. In our example, the appropriate theoretical *F*-distribution is that with 1 and 92 degrees of freedom (1 because one additional parameter has been introduced, 92 because after introducing the additional parameter, the RSS has 92 degrees of freedom, that is, 110 - 18). The 95^{th} percentile of the theoretical F distribution with 1 and 92 degrees of freedom is 3.95. This means that if the true value of the new parameter is zero, there is only a 5% chance that the *F*-ratio would be as high as 3.95. So an *F*-ratio in excess of 3.95 is strong evidence that the true value of the parameter is non-zero. An *F*-ratio above the 95^{th} percentile is usually judged to be statistically significant. A value above the 90^{th} percentile (2.77 in this example) would also usually be regarded as statistically significant, but with a lower degree of confidence. Any *F*-ratio greater than 1 provides some evidence that the parameter is in fact non-zero, but clearly the lower the *F*-ratio, the weaker the evidence.

In carrying out *F*-tests on weighted sums of squares, it is important to ensure that the weights are the same in both the numerator and denominator of the *F*-ratio. It would be wrong to compare a change in $\text{RSS}_P + w_I.\text{RSS}_I$ to a mean value of $\text{RSS}_P + w_I'.\text{RSS}_I$ unless w_I' is equal to w_I. If the weights are not equal, then we are not comparing like with like and the F-ratio is meaningless.

In the present paper, a model is fitted iteratively (with the number of parameters fixed) until the relative weight w_I converges (as described in 2.3.3). Additional parameters should then be introduced, initially with no change in w_I in order to carry out a valid F-test. If the F-test shows the additional parameters to be statistically significant, further iterations can then be carried out with the additional parameters included until w_I converges to a new value.

2.5.3 Extra parameters for first few development years

When fitting theoretical run-off curves to discrete aggregate development data, it often happens that the fit is relatively poor for the first one or two development years. This occurs because the first development year contains a mixture of actual delays from the accident date to the end of the development year (depending on the distribution of accident occurrence dates in the accident year). In the case of underwriting year cohorts, the situation is further complicated by the range of possible policy inception dates within the underwriting year. These effects are approximately taken into account by using the average values of t given in Table 2 (Section 2.3.3). These approximations are often poor for the first one or two development years. The accuracy of these approximations increases in later development years because the variation in accident dates within the first year is proportionately a smaller part of the total delay.

This phenomenon was observed in Wright (1989) where it was accommodated by introducing some additional parameters for the first few development periods. The same refinement can easily be introduced in the present model. It is usually only the first two development years that are significantly affected. To allow for these, up to four additional parameters are required: two for paid development and two for incurred development.

In the model of Section 2.3.2 (Equations 8 and 9) the paid and incurred expected values are modeled respectively as $U_j.F_{Pj}(t)$ and $U_j.F_{Ij}(t)$. Here, j is the origin year, t is the average development delay (given by the approximations in Table 2), $F_{Pj}(t)$ and $F_{Ij}(t)$ are the run-off curves given by Equations 4, 5, 6, and 7, and U_j is the ultimate for origin year j.

We now introduce additional parameters θ_{P0}, θ_{P1}, θ_{I0}, θ_{I1}. The subscripts 0 and 1 indicate that these parameters apply to development years 0 and 1. For these two development years, the expected cumulative paid values are modeled as $\exp(\theta_{P0}).U_j.F_{Pj}(t)$ (where t = 0.5 for accident years, 0.333 for underwriting years) and $\exp(\theta_{P1}).U_j.F_{Pj}(t)$ (where t = 1.5 for accident years, 1.0 for underwriting years), and similarly for cumulative incurred. Note that with this form of model:

- Each θ-parameter may take any real value (positive or negative).
- The value zero for a θ-parameter corresponds to the case where no adjustment is necessary (the factor exp(θ) is then one so can be omitted).
- The θ-parameters can be determined in the same was as any other parameter of the model: by least squares estimation.
- Statistical hypothesis tests can be carried out (as described in 2.5.2) to determine whether or not these additional parameters are necessary.

3. EXAMPLE ANALYSIS

3.1 Data

To illustrate the methods described in Section 2, they are applied to the development data given in Appendix A. This is based on actual data, covering underwriting years 1993 to 2006. The class of business and other details are not given here to preserve confidentiality. The numbers of paid and incurred data-points are $n_P = n_I = 105$, giving a total of 210. The premium rate index (given in Section A.1.4. of the appendix) is an estimate obtained by applying conventional projection methods to the triangles to find estimated ultimate premiums and claims. The premium rate index was then calculated as the ratio of estimated ultimate premiums to estimated ultimate claims. This was adjusted by a constant factor so the mean value of the index Q_j over the 14 underwriting years is one. (The reliability of this method of calculating a premium rate index is discussed in Section 4.2.)

3.2 Weibull model

3.2.1 Constant development pattern

First we fit the Weibull model with all parameters fixed at constant values across all origin years. In other words, we assume initially that the run-off pattern is the same for all origin years with no dependence on the underwriting cycle. This model has a total of 19 parameters: s_P, c_P, s_R, c_R, b, and $U_1 \ldots U_{14}$. First estimates of the parameters are given by minimizing the un-weighted total residual sum of squares $\mathrm{RSS}_P + \mathrm{RSS}_I$. (This is Equation 9 in the case $w_I = 1$, where RSS_P and RSS_I are given by Equation 8.) Allocating the 19 parameters between paid and incurred data as described in Section

2.5.1 gives $p_P = 8$, $p_I = 11$.

Results are shown in the second column (iteration number 1) of Table 3. The residual sums of squares are in millions. Initial estimates of typical paid and incurred variances from Equation 10 are: $\sigma^2{}_P = 276.7/97 = 2.85$ and $\sigma^2{}_I = 351.3/94 = 3.74$. The fact that $\sigma^2{}_I$ is higher than $\sigma^2{}_P$ indicates that the Weibull curve does not fit as closely to the incurred data as to the paid data. This gives a weight for the second iteration of $w_I = 2.85/3.74 = 0.763$. Parameter counts (using the method described in 2.5.1) are then:

$p_P = 16 * 105 / (105 + 0.763 * 105) = 9.07$ and $p_I = 3 + 0.763 * p_P = 9.93$.

Results of minimizing the weighted residual sum of squares are given in the corresponding column (iteration number 2) of Table 3. These results give new estimates: $\sigma^2{}_P = 263.4/95.93 = 2.75$, $\sigma^2{}_I = 366.6/95.07 = 3.86$ hence $w_I = 2.75 / 3.86 = 0.712$ for the third iteration. Continuing in this way, convergence occurs in five iterations. Using the formula for the mean of a Weibull distribution (see Section 2.2.2), the final values of the Weibull parameters imply a mean reporting delay of 1.8 years and a mean payment delay of 2.7 years. The final column of Table 3 shows the ratio of the ultimates from the converged Weibull model to the basic chain ladder ultimates obtained from just the incurred data.

Table 3: Weibull curves with constant parameters

Iteration	1	2	3	4	5	U_j as % of ICL
w_I	1	0.763	0.712	0.699	0.700	
p_P	8	9.07	9.35	9.42	9.41	
p_I	11	9.93	9.65	9.58	9.59	
RSS_P	276.7	263.4	260.2	260.2	260.2	
RSS_I	351.3	366.6	370.9	370.9	370.9	
$RSS_P+w_I.RSS_I$	628.0	543.2	524.3	519.6	520.0	
s_P	3.00	3.01	3.02	3.02	3.02	
c_P	1.41	1.40	1.40	1.40	1.40	
s_R	2.01	1.93	1.91	1.91	1.91	
c_R	1.24	1.24	1.24	1.24	1.24	
b	0.98	0.89	0.87	0.87	0.87	
U_1	13,913	13,927	13,930	13,930	13,930	102.7%
U_2	19,130	19,165	19,174	19,174	19,174	101.2%
U_3	11,200	11,217	11,221	11,221	11,221	99.6%
U_4	10,995	10,980	10,976	10,976	10,976	101.8%
U_5	12,982	12,960	12,954	12,954	12,954	93.8%
U_6	26,159	25,838	25,755	25,755	25,755	89.2%
U_7	68,255	68,256	68,257	68,257	68,257	94.8%
U_8	142,745	143,104	143,197	143,197	143,197	98.1%
U_9	128,173	128,703	128,841	128,841	128,841	101.6%
U_{10}	65,742	65,949	66,003	66,003	66,003	92.4%
U_{11}	4,445	4,377	4,359	4,359	4,359	92.0%
U_{12}	4,440	4,458	4,463	4,463	4,463	103.1%
U_{13}	7,422	7,180	7,112	7,112	7,112	86.5%
U_{14}	24,784	24,252	24,093	24,093	24,093	81.2%
ΣUj	540,385	540,365	540,334	540,334	540,334	96.4%

3.2.2 Varying bias factor

Next some of the parameters are allowed to vary with the underwriting cycle using the model of Equations 6 and 7. First we allow just the parameter b to vary, so instead of a single parameter b, we now have two parameters β_0 and β_1 (see Equation 6). The additional parameter relates to incurred data only so p_I increases by 1 giving: $p_P = 9.41$, $p_I = 10.59$.

Table 4: Weibull curves with varying *b*-parameter

Iteration	1	2
w_I	0.700	0.697
p_P	9.41	9.43
p_I	10.59	10.57
RSS_P	260.3	260.3
RSS_I	368.7	368.7
$RSS_P+w_I.RSS_I$	518.5	517.4
s_P	3.02	3.02
c_P	1.40	1.40
s_R	1.90	1.90
c_R	1.24	1.24
β_0	-0.053	-0.053
β_1	0.191	0.191
ΣU_j	537,411	537,411

The second column (iteration 1) of Table 4 shows least squares results obtained using the same value w_I = 0.700 as used in the model with b constant. The additional parameter β_1 causes the weighted RSS to fall from 520.0 to 518.5. An approximate test of statistical significance of the additional parameter is the F-test described in Section 2.5.2. This is based on the ratio of the decrease in the weighted RSS per additional parameter (which is 1.5 in this case, as there is only one additional parameter) to the mean RSS per degree of freedom in the model with 20 parameters, which is 518.5 / (210 – 20) = 2.7. If the additional parameter (β_1) is actually zero, then this ratio has approximately an F-distribution with 1 and 190 degrees of freedom, so a value in the extreme right-tail of the F-distribution would be evidence against the hypothesis that β_1 is zero. In this case, the ratio (1.5/2.7) is less than 0.5, which is not an extreme value compared to an F-distribution, so the parameter β_1 is not statistically significant. Nevertheless, β_1 being positive (0.191) is weak evidence that case estimates are strengthened in harder markets.

3.2.2 Varying paid development time-scale parameter

Since the F-test indicates no strong evidence that b varies with the underwriting cycle, we next try a model in which b is constant, but the paid-development scale parameter (s_P) is allowed to vary as in Equation 7. We start (iteration 1) with w_I = 0.700 as in Table 3, so the additional parameter contributes 1/1.700 to p_P and 0.700/1.700 to p_I, to give p_P = 10.00, p_I = 10.00. The weighted RSS becomes 510.3, which is a decrease of 9.7. This gives the *F*-ratio 9.7/(510.3/190) = 3.6, which is high, indicating that the additional parameter is statistically significant this time. σ_1 being less than zero (-0.199) indicates that s_{Pj} decreases as the premium index Q_j increases, that is, payment delays tend to be shorter in harder markets. Three further iterations are necessary for convergence as shown in Table 5.

Table 5: Weibull curves with varying payment delay parameter

Iteration	1	2	3	4
w_I	0.700	0.665	0.656	0.653
p_P	10.00	10.21	10.27	10.28
p_I	10.00	9.79	9.73	9.72
RSS_P	248.5	246.2	245.6	245.6
RSS_I	373.8	377.2	378.1	378.1
$RSS_P+w_I.RSS_I$	510.3	497.0	493.5	492.6
σ_0	1.015	1.015	1.015	1.015
σ_1	-0.199	-0.201	-0.201	-0.201
c_P	1.40	1.39	1.39	1.39
s_R	1.63	1.63	1.62	1.62
c_R	1.26	1.26	1.26	1.26
b	0.66	0.65	0.65	0.65

Next we try a model in which both b and s_P are allowed to vary (as in Equations 6 and 7). Although it seemed that β_1 was not statistically significant when s_P was held constant, it is possible that when both b and s_P are allowed to vary, both are statistically significant. With the weight w_I fixed at 0.653, the additional parameter (β_1) causes a decrease in the weighted RSS from 492.6 to 485.7, a decrease of 6.9. The *F*-ratio is 6.9 / (485.7/189) = 2.7, which is close to the 89th percentile of the corresponding *F*-distribution. Since this is not an extremely high percentile, the statistical significance of the parameter is not clear. In this case, we continue with this model for now, and re-test the significance of the β_1 parameter at a later stage. Note that the final column of Table 6 (ratio of cyclical ultimates to incurred CL ultimates) now shows a fairly clear cyclical pattern.

Table 6: Weibull curves with varying payment delay and *b* parameters

Iteration	1	2	U_j as % of incurred CL
w_I	0.653	0.655	
p_P		10.27	
p_I		10.73	
RSS_P	243.8	243.9	
RSS_I	370.4	370.2	
$RSS_P + w_I.RSS_I$	485.7	486.4	
σ_0	0.986	0.986	
σ_1	-0.251	-0.251	
c_P	1.40	1.40	
s_R	1.70	1.70	
c_R	1.25	1.25	
β_0	-0.098	-0.097	
β_1	0.513	0.513	
U_1	13,598	13,598	100.2%
U_2	19,057	19,056	100.6%
U_3	11,056	11,056	98.2%
U_4	10,792	10,793	100.1%
U_5	12,804	12,804	92.7%
U_6	25,661	25,664	88.9%
U_7	68,454	68,455	95.1%
U_8	145,666	145,662	99.8%
U_9	128,080	128,076	101.0%
U_{10}	63,611	63,610	89.1%
U_{11}	3,343	3,344	70.6%
U_{12}	3,223	3,224	74.5%
U_{13}	5,815	5,817	70.7%
U_{14}	18,283	18,284	61.6%
ΣU_j	529,443	529,444	94.5%

Table 7 shows the implied variation of s_P and b across the origin years: these values are calculated from Equations 6 and 7 using the parameter estimates given in Table 6. Comparing the hard market of 2003-2004 to the soft market of 1998-2001 these results imply:

- Approximately a one-third reduction in payment delays in the hard market (s_P decreases from about three in the soft market to about two in the hard market).
- More than doubling of case estimates in the hard market (b increased from 0.7 in the soft

market to about 1.7 in the hard market conditions). However, the statistical significance of the β_1 parameter was questionable so that this apparent cyclical effect might in fact be caused by random variation: results presented in the next sub-section suggest that this is in fact the case.

Table 7: Weibull curves with varying payment delay and *b* parameters

Origin year (j)	Q_j - 1	s_P	b
1993	0.17	2.57	0.99
1994	-0.32	2.90	0.77
1995	-0.18	2.81	0.83
1996	-0.19	2.81	0.82
1997	-0.32	2.90	0.77
1998	-0.47	3.01	0.71
1999	-0.49	3.03	0.70
2000	-0.60	3.11	0.67
2001	-0.42	2.98	0.73
2002	-0.25	2.86	0.80
2003	1.14	2.02	1.63
2004	1.34	1.92	1.81
2005	0.18	2.56	1.00
2006	0.41	2.42	1.12

3.2.2 Use of Premium data

We could now test whether there is any evidence that reporting delays also vary with the cycle by using a model like Equation 7, but for the scale parameter of reporting delay s_R. However, before doing this, we test the effect of using premium data as described in Section 2.4.3. The number of ultimate premium data-points is 14 so the total number of data-points increases to 224. First we try just one additional parameter, ϱ_0 (that is, we use Equation 17 with parameters ϱ_1 and ϱ_2 set to zero). Convergence occurs in four iterations as shown in Table 8. The value 1.066 (or 106.6%) for ϱ_0 represents the mean ultimate loss ratio on the assumption that, after on-leveling premiums using the premium rate index Q_j, the mean ultimate loss ratio is the same for all origin years.

Table 8: Weibull curves with varying payment delay and *b* parameters, using premium data

Iteration	1	2	3	4
w_I	0.655	0.701	0.726	0.737
w_X	1	1.606	1.884	2.005
	9.64	9.08	8.82	8.71
	10.32	10.36	10.40	10.42
	2.04	2.57	2.78	2.87
RSS_P	255.6	263.0	266.4	266.4
RSS_I	362.3	357.5	355.6	355.6
RSS_X	19.9	16.6	15.5	15.5
$\text{RSS}_P + w_I.\text{RSS}_I + \text{w}_x.\text{RSS}_x$	512.9	540.2	553.7	559.6
σ_0	1.029	1.038	1.041	1.041
σ_1	-0.155	-0.132	-0.125	-0.124
c_P	1.40	1.41	1.41	1.41
s_R	1.85	1.90	1.92	1.92
c_R	1.24	1.24	1.24	1.24
β_0	-0.048	-0.016	0.000	0.000
β_1	0.310	0.261	0.246	0.246
ϱ_0	1.065	1.066	1.066	1.066
ΣU_j	535,711	535,997	536,064	536,064

Since the statistical significance of the β_1 parameter was unclear when the model was calibrated using just the paid and incurred claims data, we next test the significance of this parameter when the premium data is also used in calibration. If the β_1 parameter is set to zero and least squares estimation carried out using the same weights as above (w_I = 0.737, w_X = 2.005), the effect is to increase the minimized RSS from 559.6 to 561.3. This increase is not statistically significant (F-ratio = 1.7 / (559.6 / (224 -22)) = 0.61), implying there is no clear evidence that β_1 is non-zero. Table 9 shows results for the model in which payment scale parameter s_P varies with the underwriting cycle, but the parameter *b* is the same across all underwriting years.

Including parameters ϱ_1 and ϱ_2 (with weights w_I = 0.751 and w_X = 2.294 as in Table 9) causes the weighted RSS to reduce from 570.4 to 569.7, which is clearly not a statistically significant reduction. Including a parameter that allows the reporting delay to vary with the underwriting cycle reduces the weighted RSS from 570.4 to 568.7, which again is not statistically significant (F-ratio = (570.4 – 568.7) / (568.7 / 202) = 0.60).

Table 9: Weibull curves with varying payment delay parameter only, using premium data

Iteration	1	2	3	4	U_j as % of ICL
w_I	0.737	0.743	0.749	0.751	
w_X	2.005	2.199	2.268	2.295	
p_P	8.71	8.60	8.54	8.52	
p_I	9.42	9.39	9.40	9.40	
p_X	2.87	3.02	3.06	3.08	
RSS_P	268.6	270.3	271.0	271.3	
RSS_I	358.8	358.0	357.6	357.4	
RSS_X	14.1	13.6	13.4	13.3	
$RSS_P+w_I.RSS_I + w_x.RSS_x$	561.3	566.2	569.2	570.4	
σ_0	1.054	1.056	1.056	1.056	
σ_1	-0.106	-0.102	-0.101	-0.101	
c_P	1.40	1.40	1.40	1.40	
s_R	1.79	1.80	1.81	1.81	
c_R	1.24	1.24	1.24	1.24	
b	0.77	0.78	0.78	0.78	
ϱ_0	1.070	1.070	1.070	1.070	
U_1	13,734	13,730	13,729	13,728	101.2%
U_2	19,064	19,058	19,056	19,055	100.5%
U_3	11,128	11,126	11,125	11,125	98.8%
U_4	10,823	10,818	10,816	10,816	100.4%
U_5	13,109	13,122	13,127	13,129	95.0%
U_6	25,984	26,002	26,013	26,017	90.2%
U_7	68,561	68,568	68,572	68,573	95.3%
U_8	143,808	143,746	143,723	143,713	98.5%
U_9	128,143	128,080	128,056	128,046	101.0%
U_{10}	66,568	66,659	66,687	66,698	93.4%
U_{11}	4,513	4,520	4,523	4,524	95.5%
U_{12}	3,839	3,839	3,839	3,839	88.7%
U_{13}	7,969	7,971	7,971	7,972	96.9%
U_{14}	20,748	20,747	20,747	20,747	69.9%
ΣU_j	537,993	537,987	537,984	537,982	96.0%

The results now show a smaller (and more plausible) amount of variation in mean payment delays with the underwriting cycle: compare Table 10 to Table 7. Table 10 also shows the implied mean payment delay in years (from the formula given in Table 1). The mean reporting delay is 1.63 years (the same for all origin years).

Table 10: Weibull curves with varying payment delay parameter only, using premium data

Origin year (*j*)	Q_j - 1	s_P	mean (years)
1993	0.17	2.83	2.57
1994	-0.32	2.97	2.70
1995	-0.18	2.93	2.67
1996	-0.19	2.93	2.67
1997	-0.32	2.97	2.70
1998	-0.47	3.01	2.75
1999	-0.49	3.02	2.75
2000	-0.60	3.05	2.78
2001	-0.42	3.00	2.73
2002	-0.25	2.95	2.69
2003	1.14	2.56	2.34
2004	1.34	2.51	2.29
2005	0.18	2.82	2.57
2006	0.41	2.76	2.51

3.3 Burr model

Using Burr distributions for both paid and reporting delays gives higher residual sums of squares than using the Weibull model, indicating that Burr curves provide a poorer fit to this particular dataset.

3.4 Inverse Burr model

To compare the quality of fit of the Inverse Burr and Weibull models, we fit an Inverse Burr model using the same values of the weights as in Table 8: w_I = 0.737 and w_X = 2.005. The minimized weighted RSS is 538.7 (see Table 11) which is 20.9 lower than obtained using the Weibull model (Table 8). The Inverse Burr model has two additional parameters (there are two shape parameters instead of one for both reporting and payment delays), so the decrease is 10.5 for each additional parameter. Comparing this to the RSS per degree of freedom (538.0 / 200 = 2.7) the decrease appears to be statistically significant. (Note that a formal *F*-test is not strictly valid here because the two models are not nested.) We conclude that the Inverse Burr model fits this particular dataset better than the Weibull model. Convergence occurs in six iterations as shown in Table 11 (intermediate results are not shown for all six iterations).

Table 11: Inverse Burr curves with varying payment delay and *b* parameters, using premium data

Iteration	1	6
w_I	0.737	0.544
w_X	2.005	1.520
p_P	9.29	10.60
p_I	11.85	10.77
p_X	2.87	2.62
RSS_P	230.0	215.8
RSS_I	379.1	395.8
RSS_X	15.1	17.1
$RSS_P+w_I.RSS_I + w_x.RSS_x$	538.7	457.3
σ_0	1.406	1.405
σ_1	-0.141	-0.148
a_P	0.27	0.26
c_P	4.00	4.12
s_R	2.76	2.82
a_R	0.31	0.29
c_R	3.29	3.44
β_0	-0.035	-0.011
β_1	0.237	0.257
ϱ_0	1.063	1.056

If the β_1 parameter is set to zero and least squares estimation carried out using the same weights as above (w_I = 0.544, w_X = 1.520), the minimized RSS increases from 457.3 to 459.0. This increase is not statistically significant (F-ratio = 1.7 / (457.3 / 200) = 0.74), implying (as for the Weibull model) that there is no clear evidence that β_1 is non-zero. After convergence, the final parameter values imply (using the formula for the mean of an Inverse Burr distribution from Section 2.2.2) that the mean reporting delay is 1.72 years. The mean payment delay varies with the underwriting cycle as shown in Table 13.

Including parameters ϱ_1 and ϱ_2 (with w_I and w_X as in Table 12) causes the weighted RSS to reduce from 461.2 to 460.3, which is clearly not a statistically significant reduction. Including a parameter that allows the reporting delay to vary with the underwriting cycle reduces the weighted RSS from 461.2 to 459.3, which again is not statistically significant (F-ratio = 1.9 / (459.3 / 200) = 0.83). Including additional parameters as described in Section 2.5.3 shows that these are statistically significant for this dataset, but for reasons of space, further results are not given here.

Table 13: Inverse Burr curves with varying payment delay parameter only, using premium data

Origin year (j)	Q_j - 1	s_P	mean (years)
1993	0.17	4.03	2.51
1994	-0.32	4.29	2.67
1995	-0.18	4.22	2.62
1996	-0.19	4.22	2.63
1997	-0.32	4.29	2.67
1998	-0.47	4.37	2.72
1999	-0.49	4.39	2.73
2000	-0.60	4.45	2.77
2001	-0.42	4.35	2.70
2002	-0.25	4.26	2.65
2003	1.14	3.56	2.22
2004	1.34	3.47	2.16
2005	0.18	4.02	2.50
2006	0.41	3.91	2.43

4. CONCLUSIONS

4.1 Commonly seen cycle dependencies

4.1.1 Variation of payment delays with the underwriting cycle

The example analysis of Section 3 shows evidence of payment delays lengthening in soft market origin years. The model described in this paper has been applied to several actual datasets for different classes of business and the finding that payment delays are longer in soft markets occurs consistently. This concurs with the findings of previous research described in Section 1 of the present paper. Possible causes are listed in Section 1.1.2.

4.1.2 Variation of case reserve strength with the underwriting cycle

Although there is no clear evidence that the case estimate bias factor varies with the underwriting cycle in the example analysis of Section 3, application of the model to other datasets has in many cases shown clear evidence that case reserves are set at higher levels in origin years with higher premium rates.

4.1.3 Variation of reporting delays with the underwriting cycle

Applications of the model to other datasets have shown, in some cases, evidence that reporting delays tend to be shorter in hard markets. This is not something that has been explicitly suggested in previous research. Some possible reasons why reporting delays might be extended in soft markets are discussed here. First we should note that what we actually measure as "reporting delay" is the time between the accident date and a case reserve being created on the insurer's claim administration system. This is the sum of two main components: (a) the true reporting delay (between loss occurrence and time when the insured, or broker, reports the loss to the insurer), and (*b*) the time between the claim being reported to the insurer and the initial case reserve being created. It is possible that the second component becomes longer in soft markets. Indeed there has been a recent high profile case in the UK in which senior insurance company executives were jailed for concealing reported claims. Having acknowledged that increased delays in this second component are possible, we now focus on reasons why the true reporting delay might increase in soft markets. There are several possibilities:

- In soft markets, insureds (and/or brokers) might be aware that they got a good deal on their insurance, and be concerned that on renewal the premium is likely to increase. For this reason, they might deliberately delay reporting valid claims until after renewal negotiations have been completed. This would border on fraudulent behavior by the insureds, but nevertheless is clearly possible.
- If cover is extended by relaxing terms and conditions in a soft market, insureds might genuinely fail to realize initially that they can claim for certain types of loss.
- If periods of cover have been extended beyond the usual one year in soft markets, it is possible that this is not correctly allowed for when compiling the aggregate run-off arrays. For example, when all policies run for one year, it would be correct to assume that if the accident date does not fall in the year the policy was written, then the loss should be allocated to the following accident year. However, if this method of allocating losses to accident years is continued when some policies run for more than one year, then development delays will appear (wrongly) to be extended.

4.2 Accuracy of premium rate index

In the example analysis, we used the reciprocal of estimated ULRs instead of a premium index.

Clearly this is not ideal. The ultimate ULR varies with claims experience, not just premium rate variation. A high ULR is not necessarily indicative of a soft market: it might occur simply because of unusually high loss experience. Although the results appear to show longer development tails in softer markets, could it be that in fact all we are seeing is longer development patterns when losses are exceptionally high? The results could partly reflect this, but the fact that the estimated ULRs do broadly follow a cyclical pattern (rather than just random variation) suggests that most of the variation in ULRs reflects variation in premium rates with the underwriting cycle.

Where the models show significant cyclical variation in run-off patterns, the estimated reserve for the latest year will clearly be sensitive to the value of the premium rate index for that year, and this is the most difficult year to get an accurate fix on. A worthwhile area for future research would be to predict premium rate variation of the underwriting cycle. For example, we might expect premium rates next year to be related to measurable quantities such as the number of new insurance company start-ups this year, or the amount of new capital in the insurance industry. If the underwriting cycle can be predicted from such quantities (even if only one year ahead), then the accuracy of reserves could be improved by using these quantities directly in the reserving model instead of (or as well as) the estimated premium rate index Q_j. There is a substantial literature on the underwriting cycle and its causes: this could point to suitable alternative variables to use instead of Q_j.

5. REFERENCES

[1] Conger, R. "Actuaries: Deeper? Broader? Or Uncertain?" Presented at the 2002GIRO Convention.

[2] General Insurance Reserving Issues Taskforce (chaired by A.R. Jones), "A Change Agenda for Reserving," *British Actuarial Journal* 2006, Vol. 12, No. 3:435-599.

[3] Line, N. et al., "The Cycle Survival Kit: An Investigation into the Reserving Cycle and Other Issues," GIRO Working Party Report 2003, www.actuaries.org.uk/files/pdf/proceedings/giro2003/Line.pdf.

[4] Wright, T. S., "A Stochastic Method for Reserving in General Insurance," *Journal of the Institute of Actuaries* 1990, Vol. 117:677-731 l

APPENDIX A – DATA FOR EXAMPLE

A.1.1 Paid claims

1993	1,051	3,093	6,251	9,490	12,341	12,738	13,246	13,267	13,319	13,346	13,412	13,441	13,443	13,551
1994	1,102	6,112	10,284	14,047	15,690	16,505	17,539	18,071	18,457	18,729	18,676	18,679	18,917	
1995	824	2,467	3,938	7,311	8,538	9,537	10,849	11,016	11,078	11,111	11,117	11,132		
1996	726	2,876	4,817	7,279	8,783	9,694	9,782	10,006	10,283	10,508	10,548			
1997	443	2,293	5,113	7,381	10,016	10,684	11,102	12,453	12,868	13,084				
1998	383	5,169	12,181	16,846	17,852	19,817	20,974	21,331	22,193					
1999	3,705	17,901	27,656	40,224	51,593	61,108	62,227	63,858						
2000	6,839	30,329	55,327	86,931	112,799	127,208	133,792							
2001	5,048	29,082	56,406	83,434	101,096	111,152								
2002	2,936	10,127	22,969	42,236	54,467									
2003	73	517	1,593	2,328										
2004	127	747	2,012											
2005	115	711												
2006	559													

A.1.2 Incurred claims

1993	2,609	6,730	9,659	12,084	12,801	13,458	13,569	13,552	13,414	13,461	13,471	13,481	13,584	13,569
1994	5,556	10,517	13,594	15,635	16,988	17,436	18,065	18,204	18,571	18,799	18,747	18,751	18,973	
1995	1,461	5,481	7,258	8,731	9,739	10,478	11,006	11,056	11,088	11,121	11,121	11,161		
1996	1,697	5,772	7,496	9,178	10,080	10,846	11,011	11,127	10,700	10,670	10,668			
1997	1,474	4,189	7,621	10,210	11,752	12,657	13,011	13,314	13,640	13,684				
1998	1,664	12,524	22,283	24,423	24,959	25,788	26,771	26,844	28,454					
1999	4,633	28,059	45,391	51,533	62,706	64,962	67,509	69,657						
2000	13,853	49,104	82,185	117,950	129,088	137,329	138,833							
2001	10,311	47,971	80,236	103,794	113,943	117,873								
2002	4,602	18,402	36,267	53,627	63,363									
2003	231	1,659	3,323	3,757										
2004	582	1,690	2,634											
2005	543	3,024												
2006	2,752													

A.1.3 Premium

1993	9,206	12,421	14,247	14,591	14,401	14,572	14,628	14,634	14,634	14,634	14,616	14,616	14,617	14,618
1994	4,966	9,290	11,293	11,593	11,491	11,812	11,967	11,967	11,967	11,968	11,968	11,968	11,913	
1995	4,023	8,167	9,164	8,358	8,259	8,281	8,285	8,286	8,286	8,286	8,288	8,419		
1996	3,513	8,283	8,208	8,015	8,029	8,028	8,044	8,041	8,041	8,046	8,045			
1997	2,956	6,897	8,586	8,677	8,609	8,646	8,656	8,674	8,661	8,668				
1998	3,935	8,984	13,778	13,757	13,366	13,381	13,378	13,381	13,281					
1999	14,593	29,079	32,777	33,364	33,268	33,307	33,266	33,316						
2000	16,428	47,209	52,678	53,624	53,673	53,541	53,499							
2001	28,913	63,532	65,714	66,977	66,763	67,086								
2002	22,951	46,264	48,791	48,835	48,814									
2003	5,563	8,918	9,394	9,143										
2004	3,889	6,630	8,301											
2005	4,893	7,919												
2006	11,643													

A.1.4 Premium rate index

		Basic Chain Ladder Ultimates		
Year	Q	Paid	Incurred	Premium
1993	1.170	13,551	13,569	14,618
1994	0.682	19,069	18,952	11,914
1995	0.816	11,305	11,261	8,402
1996	0.814	10,724	10,777	8,059
1997	0.684	13,317	13,813	8,680
1998	0.533	22,852	28,859	13,303
1999	0.507	67,323	71,977	33,313
2000	0.402	145,197	145,963	53,535
2001	0.581	126,142	126,812	67,199
2002	0.745	69,064	71,432	49,075
2003	2.136	3,682	4,737	9,153
2004	2.341	4,893	4,328	8,380
2005	1.184	3,257	8,225	8,847
2006	1.405	12,207	29,675	27,248

APPENDIX B – MEAN DELAYS BY DEVELOPMENT YEAR

This appendix derives the approximations given in Table 2 (Section 2.3.3) which is reproduced below for convenience. The values in the table are approximate mean delays (in years) from loss occurrence to end of development year. For accident year cohorts, development year 0 is the year in which the loss occurs. For underwriting year cohorts, development year 0 is the year of policy inception (that is, the year in which the cover provided by a policy commences).

Table 2: Approximate mean delay in each development year

Development year (*d*)	0	1	2	3	4	5+
Accident year mean delay (*t*)	0.5	1.5	2.5	3.5	4.5	d+0.5
Underwriting year mean delay (*t*)	0.333	1	2	3	4	d

The figures in Table 2 for accident year cohorts follow immediately from an assumption that losses occur uniformly throughout the accident year. The mean delay until the end of development year zero (which is the accident year itself) is then obviously half a year. The other values in the table are equally obvious for accident years.

For underwriting years it is assumed that:

(a) Policies incept uniformly throughout the year.

(b) Policies are in force for one year.

(c) Accidents occur uniformly throughout the year of cover provided by each policy.

For development year 1 we aim to find the mean delay between the accident date and the end of development year 1. Development year 1 is the year following the underwriting year. By assumptions (a) and (b) policies expire uniformly throughout development year 1, and all covered losses will have occurred by end of development year 1. By assumptions (b) and (c) the mean accident date on a policy is half a year after policy inception. By assumption (a) the mean point of policy inception is midway through the underwriting year. Therefore, over all policies, the mean accident date is one year after the start of the underwriting year, that is, at the end of development year zero. So the mean delay since accident occurrence at the end of development year 1 is $t = 1$. Clearly, for all later development years ($d > 1$) the mean delay to end of development year d is $t = d$.

For development year 0 the situation is more complex because only half of all covered losses are expected to have occurred by the end of development year 0 (because, by assumption (a), the mean policy inception date is 0.5 years before the end of development year zero). Because of this, the expected proportion of ultimate U that will be paid by end of development year zero is approximately 0.5 * $F_P(t)$ (instead of $F_P(t)$ for other development years) where t is the mean delay between accident date and end of development year 0 for the 50% accidents that occur before the end of development year 0.

Instead of explicitly including the factor 0.5 in the model for underwriting year cohorts, a factor is estimated by least squares as described in Section 2.5.3.

To find the mean delay t for the 50% of accidents expected to occur before the end of development year 0, we use s to denote the inception date of a policy: s = 0 corresponds to an inception date at the start of the underwriting year, and s = 1 corresponds to an inception date at the end of the underwriting year.

Since all policies are in force immediately before the end of the underwriting year, a delay $t = 0$ is possible on all policies regardless of the value of s. At the other extreme, a delay $t = 1$ is possible only on policies incepting at the start of the underwriting year (that is, on policies with $s = 0$). In general, for t between 0 and 1, a delay t is possible only on policies with $s < (1-t)$. So the mean delay is the weighted average of all values of t from 0 and 1, with weights proportional to $(1-t)$. That is:

Mean(t) = $\int t.(1-t).dt \,/ \int (1-t).dt$ where both integrals are from $t = 0$ to $t = 1$.

Evaluating these integrals gives: Mean(t) = (1/6)/(1/2) = 1/3.

Abbreviations and notations

The table below gives an alphabetical list of all abbreviations and notation used in the paper. Items marked * in the second column are items of data. All other quantities are calculated from the data items. The final column shows subscripts that are sometimes applied to the symbol given in the first column:

- P/R means a subscript is used to distinguish parameters relating to payment and reporting delays.
- I/P/X means a subscript is used to distinguish quantities relating to incurred, paid, and exposure data.
- *j* means this subscript is sometimes applied to distinguish values relating to different origin years.

Symbol	**Data**	**Represents**	**Subscripts**
a		shape parameter of cumulative development curve	P/R and j
a_0, a_1		parameters linking a to Q_j (as in Equations 6 and 7)	P/R
b		mean case reserve bias factor	j
β_0, β_1		parameters linking b to Q_j (see Equation 6)	
β_2		parameter linking b to development time (see Equation 6a)	
BF		abbreviation for Bornheutter-Fergusson	
c		shape parameter of cumulative development curve	P/R and j
γ_0, γ_1		parameters linking c to Q_j (as in Equations 6 and 7)	P/R
CL		abbreviation for chain ladder	
d	*	development period in run-off array: $d = 0, 1, 2\ldots$	
$F_I(t)$		cumulative incurred run-off curve	j
$F_P(t)$		cumulative paid run-off curve	j
$F_R(t)$		cumulative distribution of reporting delays	j
θ_0, θ_1		adjustments to cumulative development in years 0 and 1 (Section 2.5.3)	I/P
$\Gamma(.)$		the Gamma function of mathematics	
I_{jd}	*	cumulative incurred development data	
j	*	origin year: $j = 1, 2, \ldots J$	
J	*	number of origin years in run-off array	
n_I	*	number of observations in incurred run-off array (I_{jd})	
n_P	*	number of observations in paid run-off array (P_{jd})	
n_X	*	number of origin years with known exposure (X_j or $Prem_j$)	
p_I		number of parameters estimated from incurred data (I_{jd})	
p_P		number of parameters estimated from paid data (P_{jd})	
p_X		number of parameters estimated from exposure data (X_j or $Prem_j$)	
P_{jd}	*	cumulative paid development data	
$Prem_j$	*	ultimate premium for origin year j	
Q_j	*	premium rate index for origin year j	
r		expected ultimate loss per unit of exposure	j
R		expected ultimate loss ratio (i.e., ultimate loss per unit of premium)	j
ϱ_0, ϱ_1, ϱ_2		parameters in model for r and R (see Equation 14)	
RSS		residual sum of squares (that is, sum of squared residuals)	$I/P/$X
s		scale parameter of cumulative development curve	P/R and j
σ_0, σ_1		parameters linking s to Q_j (see Equation 7)	P/R
σ_I^2		typical variance of a cumulative incurred observation (I_{jd})	
σ_P^2		typical variance of a cumulative paid observation (P_{jd})	
σ_X^2		ratio of variance to mean for an ultimate loss amount (U_j)	
t		continuous development time	
U_j		ultimate cumulative loss for origin year j	
ULR		abbreviation for ultimate loss ratio	
w_I		weight of incurred RSS relative to paid RSS in least squares estimation	
w_X		weight of exposure RSS relative to paid RSS in least squares estimation	
X_j	*	exposure for origin year j	

Biography of the Author

Thomas Wright is a consulting actuary who has worked mainly in non-life insurance since 1988, and with Deloitte & Touche LLP in London since 2001. He is a Fellow of the Institute of Actuaries, a Fellow of the Royal Statistical Society, and a Chartered Statistician. He is the author of several papers and articles on stochastic claims reserving and he was one of the pioneers of generalized linear modeling in personal lines rate-making. In 1991 he co-authored "Statistical Motor Rating: Making Effective Use of Your Data," which was published by the UK Institute of Actuaries and has since become a standard reference work. He was joint winner of the CAS/COTOR challenge in 2005 and is currently an active member of the Working Party on Reserve Uncertainty of the Institute of Actuaries General Insurance Reserving Oversight Committee (GIROC).

thwright@deloitte.co.uk

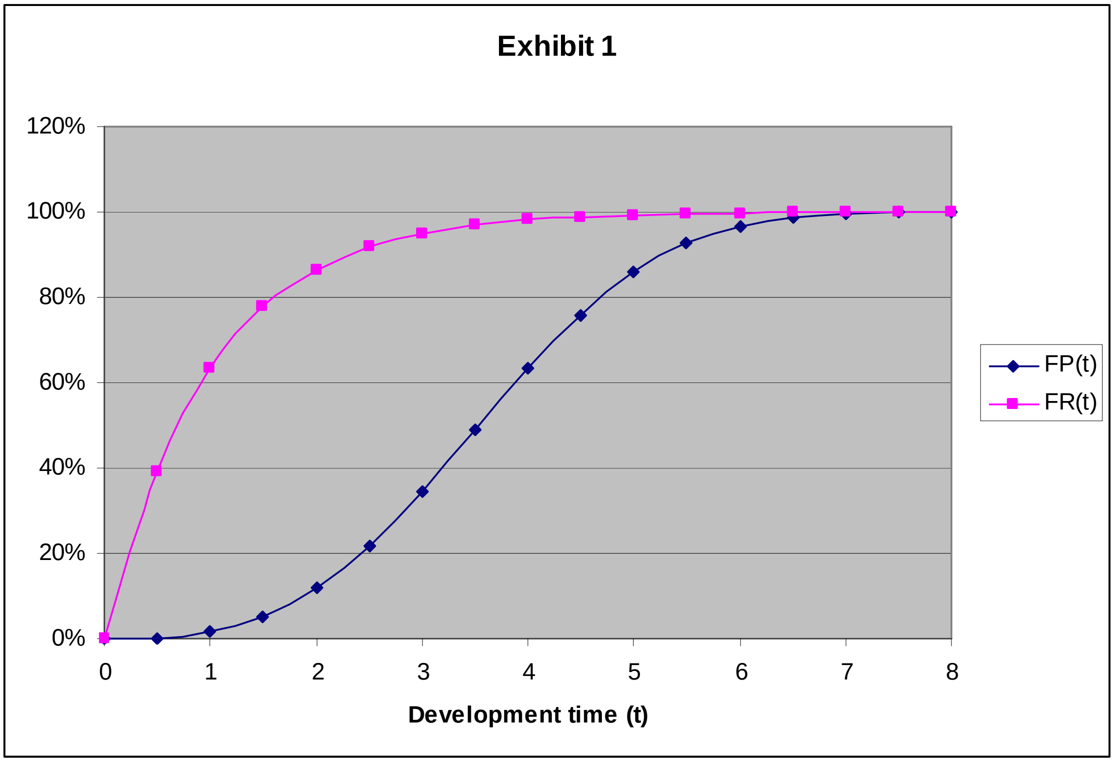


Exhibit 1
120%
100%
80%
60%
40%
20%
0%
0
1
2
3
4
5
6
7
8
Development time (t)
FP(t)
FR(t)

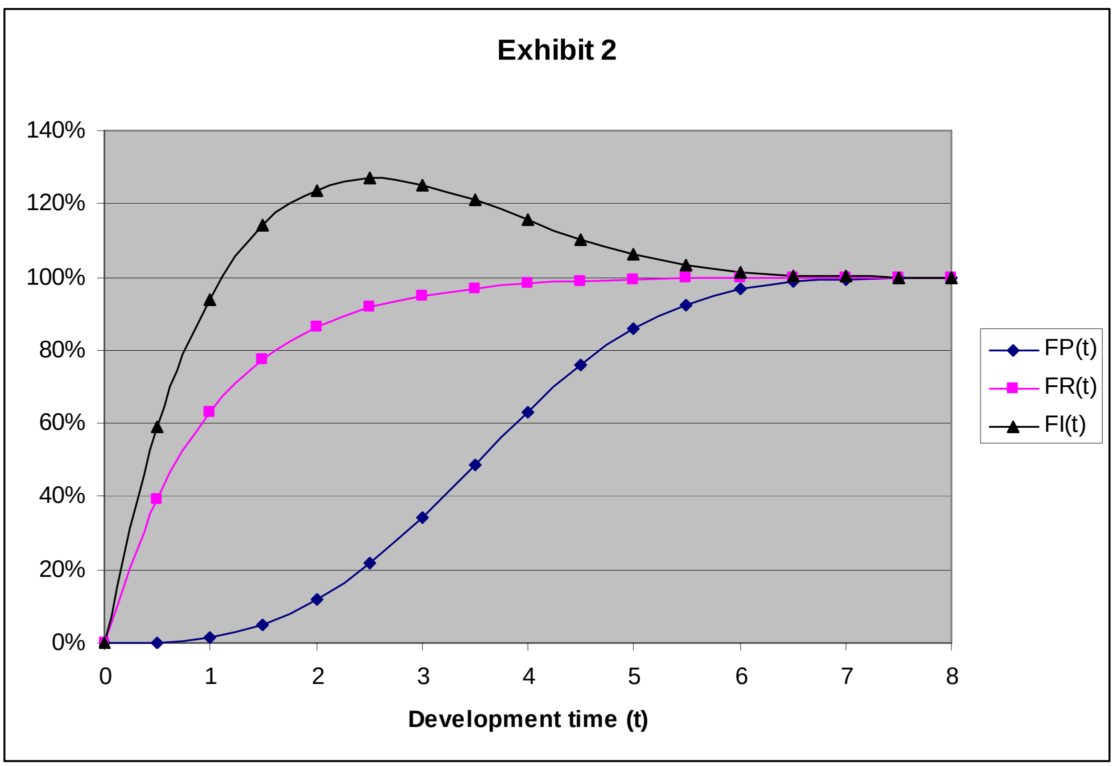


Exhibit 2
140%
120%
100%
80%
60%
40%
20%
0%
0
1
2
3
4
5
6
7
8
Development time (t)
FP(t)
FR(t)
FI(t)

Grouping Loss Distributions by Tail Behavior Part I: Discrete Families

Dan Corro
National Council on Compensation Insurance
Spring 2008

Abstract: *This three part paper addresses the task of modelling the right hand tail of a severity distribution. In Part I the excess ratio function is used to define a discrete sequence of loss distributions with related moments and similar tail behavior. Part II extends this to continuous one-parameter families and provides some examples. Part III provides the main result: that under some reasonable conditions, each such family has a limiting distribution which is exponential. The paper then exploits this to 1) group loss distributions based on tail behavior and 2) promote the choice of (mixed) exponentials to model tail behavior.*

1 Background

Even a large claim database may not suffice to give an accurate picture of the (far) right hand tail of the severity distribution of the expected losses. Consider the approach in which a distribution is built from empirical data for the more common loss amounts but is then truncated and a theoretical distribution spliced on to model the tail, where there are few actual observations. This approach has considerable appeal because most of the "bumps" of the expected loss severity are (or are believed to be) at lower loss amounts where the behavior is revealed in the observed losses. Conceptually, the expected tail should not be subject to such bumps, but rather reflect a stable pattern (i.e., if there are still bumps, you have not gone far enough to enter the "tail"). A direct measure of "bumpiness" is the presence of local modal values, or points where the derivative of the density function changes sign. "Higher order bumps" are where higher order derivatives change sign. The mathematical concept of monotonality captures this. Ideally, the tail behavior should be less bumpy, i.e., more monotone, than the overall severity distribution. The task arises, then, given a severity distribution to find a related distribution with similar behavior but one that is more monotone. Here we describe such a related distribution, what we call the "coderived" distribution. The process of constructing the coderived distribution can be repeated, and this paper is especially interested in describing the resulting sequence of distributions. These sequences emerge as canonically related families of loss distributions. This suggests an organizational scheme for continuous loss distributions and provides an alternative to the more conventional organization of loss distributions into "families" according to the arithmetic form of the density function.

The ratio of losses in excess of a given loss limit x to total losses defines a function $R(x)$ that formally resembles a survival function. The loss distribution defined by that survival function is the "coderived" distribution. Conceptually, the coderived distribution provides a "preview" into the tail. The coderived distribution is shown to exhibit (right hand) tail behavior and moments that are very closely related to those of the original loss distribution. However, the coderived distribution has a simpler, more "monotone", shape than the original, in a sense defined in the paper. There is no information lost, as the coderived distribution completely determines the original. Repeating this process of "coderiving" loss distributions yields (Part I) a discrete sequence of loss distributions that are observed (Part II) to fall within a continuous one-parameter collection of loss distributions. Such collections all have tails of the same ultimate settlement rate, again as defined later. We then (Part III) consider a simple approach to ordering loss distributions according to the "thickness" of their tails. Finally, we use these concepts to relate thickness with monotonality and ultimate settlement rate. A key finding is that the asymptotic behavior of the hazard rate function provides a natural bridge between these two perspectives. Another finding reveals a unique "fixed point" role played by the exponential class of loss distributions. Assuming a tail behavior that is sufficiently "simple", we show that the (mixed) exponential distribution has properties that favor it as a choice to fit the tail of the distribution.

2 Notation and Terminology

In this paper we consider "smooth loss distribution functions" or SLDFns, by which we mean:

Definition 1 *A function* $F\colon [0,\infty) \to [0,1]$ *is a* ***loss distribution function,*** *or* ***LDFn,*** *provided that*

- $F(0) = 0$
- $\displaystyle\lim_{x\to\infty} F(x) = 1$
- F *is nondecreasing.*

We often use the standard symbol "⇒" as an abbreviation for "implies" and more generally:

$\Rightarrow$	implies
$\Leftrightarrow$	if and only if
$\Rightarrow\Leftarrow$	contradiction.

Definition 2 *The **minimum loss of** F denoted $\alpha_F \in \mathbb{R}$ is uniquely determined by*

$$x < \alpha_F \Rightarrow F(x) = 0 \text{ and } x > \alpha_F \Rightarrow F(x) > 0.$$

Definition 3 *The **maximum loss of** F denoted $\omega_F \in \mathbb{R} \cup \{\infty\}$ is uniquely determined by*

$$x < \omega_F \Rightarrow F(x) < 1 \text{ and } x > \omega_F \Rightarrow F(x) = 1.$$

Definition 4 *F is a **smooth loss distribution function**, or SLDFn, provided F is infinitely differentiable on (α_F, ω_F), continuous on $[0, \infty)$ and the limit*

$$\lim_{x \to \omega_F} -\frac{d(\ln(1-F))}{dx} \in \mathbb{R} \cup \{\infty\}.$$

Notation 5 *For any SLDFn F, we denote the corresponding density [PDF] as f and have $f(x) = \frac{dF(x)}{dx}$ for $x \in (\alpha_F, \omega_F)$ and $f(x) = 0$ otherwise. We occasionally denote the corresponding expectation of a real valued function g defined on (α_F, ω_F) as*

$$E[g(X)] = \int_{\alpha_F}^{\omega_F} g(x) f(x) dx = \int_0^\infty g(y) f(y) dy.$$

The survival function of F is denoted $S = 1 - F$ and the mean as

$$\mu = \int_0^\infty y dF(y) = \int_0^\infty y f(y) dy = E[X].$$

We say F has finite mean provided $\mu < \infty$. For any $c \in \mathbb{R}$, we set

$$\mu^{(c)} = \int_0^\infty y^c f(y) dy = E[X^c].$$

So $\mu^{(0)} = 1$ and $\mu^{(1)} = \mu$ and we call $\mu^{(c)}$ the c-th moment of F. Provided $0 < \mu < \infty$, the excess ratio function of F is given by

$$R(x) = \frac{\int_x^\infty (y - x) f(y) dy}{\mu}$$

and we denote by $\widehat{S}$ the function

$$\widehat{S}(x) = \frac{\int_x^\infty y f(y) dy}{\mu}$$

for $x \geq 0$. *We denote the hazard rate function by*

$$\lambda(x) = \frac{f(x)}{S(x)}$$

for $x \in (0, \omega_F)$. *We let*

$$L(t) = \int_0^\infty e^{-ty} f(y)dy = \int_{\alpha_F}^{\omega_F} e^{-ty} f(y)dy = E\left[e^{-tX}\right]$$

denote the Laplace transform of F *and* $M(t) = L(-t)$ *the moment generating function. When* F *has finite mean we denote the standard deviation as*

$$\sigma = \sqrt{\int_0^\infty (y-\mu)^2 f(y)dy} = \sqrt{\mu^{(2)} - \mu^2}$$

and the coefficient of variation as $CV = \frac{\sigma}{\mu}$.*We use subscripts on* f_F, E_F, S_F, μ_F, $\mu_F^{(c)}$, R_F, $\widehat{S}_F$, σ_F, CV_F, λ_F, L_F, *and* M_F *when necessary to indicate dependence on* F.

Note that for any SLDFn F, the requirement that $\lim\limits_{x\to\infty} F(x) = 1$ *forces* $f(a) > 0$ for some $a > 0$ and so $\mu^{(c)} > 0$ for every $c \in \mathbb{R}$.

Definition 6 *For any SLDFn* F*, the* ***ultimate settlement rate*** *is*

$$\tau_F = \lim_{x \to \omega_F} \lambda_F(x).$$

Note that for any SLDFn F we have for all $x \in (0, \omega_F)$ that $S(x) > 0$ and by the chain rule

$$-\frac{d\left(\ln\left(1-F\right)\right)}{dx} = -\frac{d(\ln S(x))}{dx} = -\frac{1}{S(x)}\frac{dS(x)}{dx} = \frac{1}{S(x)}\frac{dF(x)}{dx} = \frac{f(x)}{S(x)} = \lambda(x)$$

and so, by our definition of SLDFn, τ_F is well defined.

Example 7 *The function*

$$F(x) = \left\{ \begin{array}{cc} 1 - e^{\frac{x(x-2)}{(x-1)^2}} & 0 \leq x \leq 1 \\ 1 & 1 \leq x \end{array} \right\}$$

is an SLDFn that is infinitely differentiable on $(0, \infty)$ *with* $\omega_F = 1$ *and* $\tau_F = \infty$.

We begin by noting that SLDFns are determined by their hazard rate functions:

Proposition 8 *For any SLDFn* F*:*

$$S_F(x) = e^{-\int_0^x \lambda_F(t)dt} \text{ for every } x \in [0, \omega_F).$$

Proof. We have noted that for any $z \in (0, \omega_F)$

$$\frac{d(\ln S(x))}{dx} = -\lambda(x)$$

holds for all $x \in (0, z)$. We see that $\lambda(x)$ is integrable on $[0, z]$. But then $S(x)$ and $T(x) = e^{-\int_0^x \lambda(t)dt}$ are two continuous functions with the same logarithmic derivative on $(0, z)$. It follows that

$$\begin{aligned} 0 &= \frac{d(\ln T(x) - \ln S(x))}{dx} = \frac{d(\ln \frac{T(x)}{S(x)})}{dx} \\ &\Rightarrow \ln \frac{T(x)}{S(x)} = c \text{ is constant on } (0, z) \\ &\Rightarrow \frac{T(x)}{S(x)} = e^c \text{ is constant on } (0, z) \\ &\Rightarrow T(x) = e^c S(x) \text{ for every } x \in (0, z). \end{aligned}$$

But then

$$\begin{aligned} S(0) &= 1 = e^0 = e^{-\int_0^0 \lambda(t)dt} = T(0) \Rightarrow e^c = 1 \\ &\Rightarrow S(x) = T(x) = e^{-\int_0^x \lambda(t)dt} \text{ for every } x \in [0, z). \end{aligned}$$

Since $z \in [0, \omega_F)$ was arbitrary,

$$\begin{aligned} S_F(x) &= e^{-\int_0^x \lambda_F(t)dt} \\ \text{for every } x &\in \bigcup_{z \in [0,\omega_F)} [0, z) = [0, \omega_F) \end{aligned}$$

as required. ■

We will have occasion to consider the case when the hazard rate function is increasing or decreasing. This can often be readily determined, as in:

Proposition 9 *For any SLDFn F with λ_F differentiable on $(\alpha_F, \omega_F) = (0, \infty)$:*

$$\frac{d\lambda_F}{dx} = \lambda_F^2 + \frac{\frac{df_F}{dx}}{S_F} = \lambda_F \left(\lambda_F + \frac{d \ln f_F}{dx} \right).$$

Proof. From the definition of $\lambda = \lambda_F$

$$\begin{aligned} \frac{d\lambda}{dx} &= \frac{d}{dx}\left(\frac{f}{S}\right) = \frac{S\frac{df}{dx} - f\frac{dS}{dx}}{S^2} \\ &= \frac{S\frac{df}{dx} - f(-f)}{S^2} = \frac{S\frac{df}{dx}}{S^2} + \frac{f^2}{S^2} \\ &= \left(\frac{f}{S}\right)^2 + \frac{\frac{df}{dx}}{S} = \lambda^2 + \frac{\frac{df}{dx}}{S} \\ &= \lambda^2 + \frac{\frac{df}{dx}f}{fS} = \lambda^2 + \lambda\frac{\frac{df}{dx}}{f} \\ &= \lambda^2 + \lambda\frac{d\ln f}{dx} \end{aligned}$$

as required. ■

The following proposition expresses the excess ratio function in terms of S and $\widehat{S}$.

Proposition 10 *For any SLDFn F with $\mu_F < \infty$,*

$$R_F(x) = \widehat{S_F}(x) - \frac{xS_F(x)}{\mu_F}, \text{ for every } x \geq 0.$$

Proof. From the definition of $R(x)$ we have

$$\begin{aligned}
R(x) &= \frac{1}{\mu}\int_x^\infty (y-x)f(y)dy \\
&= \frac{1}{\mu}\left[\int_x^\infty yf(y)dy - x\int_x^\infty f(y)dy\right] \\
&= \frac{1}{\mu}\left[\int_x^\infty yf(y)dy - xS(x)\right] \\
&= \widehat{S}(x) - \frac{xS(x)}{\mu}.
\end{aligned}$$

as required. ■

Proposition 11 *For any SLDFn F and $a, b, c \in \mathbb{R}$ with $a \geq b \geq 0$ and $\mu_F^{(c)} < \infty$, and further provided either $c \geq 0$ or $a > b$, we have (with the convention that $0^0 = 1$):*

$$c\int_a^\infty (y-b)^{c-1} S_F(y)dy = \int_a^\infty (y-b)^c f_F(y)dy - (a-b)^c S_F(a).$$

Proof. The case $c = 0$ reduces to the identity

$$0 = \int_a^\infty f(y)dy - (1)\,S(a) = S(a) - S(a).$$

So assume $c \neq 0$. The result follows from integration by parts

$$u = S(y) \qquad v = (y-b)^c$$

$$\begin{aligned}
c\int_a^\infty (y-b)^{c-1} S(y)dy &= \int_a^\infty S(y)\left(c\,(y-b)^{c-1}\right)dy \\
&= \int_a^\infty udv = uv]_a^\infty - \int_a^\infty vdu \\
&= (y-b)^c\, S(y)]_a^\infty - \int_a^\infty (y-b)^c\,(-f(y))\,dy
\end{aligned}$$

$$= \left(\lim_{y\to\infty}(y-b)^c\, S(y)\right) - (a-b)^c\, S(a) + \int_a^\infty (y-b)^c\, f(y)dy$$

Now clearly

$$c < 0 \Rightarrow \lim_{y\to\infty} (y-b)^c S(y) \le \lim_{y\to\infty} S(y) = 0$$

and for $y > b+1$ and $c > 0$

$$\begin{aligned}(y-b)^c S(y) &\le y^c S(y) = y^c \int_y^\infty f(x)dx \\ &= \int_y^\infty y^c f(x)dx \le \int_y^\infty x^c f(x)dx\end{aligned}$$

it follows that

$$\begin{aligned}0 &\le \lim_{y\to\infty} (y-b)^c S(y) \\ &\le \lim_{y\to\infty} \int_y^\infty x^c f(x)dx = 0 \text{ since } \int_0^\infty x^c f(x)dx = E[X^c] < \infty \\ &\Rightarrow 0 = \lim_{y\to\infty} (y-b)^c S(y)\end{aligned}$$

and we conclude that

$$c\int_a^\infty (y-b)^{c-1} S(y)dy = -(a-b)^c S(a) + \int_a^\infty (y-b)^c f(y)dy$$

and the result follows. ∎

Corollary 12 *If either $a > b$ or $c > 0$, then:*

$$\int_a^\infty (y-b)^{c-1} S_F(y)dy < \infty \Leftrightarrow \int_a^\infty (y-b)^c f_F(y)dy < \infty.$$

Proof. Clear since under the conditions we must have $(a-b)^c S(a) < \infty$. ∎

Corollary 13 *For any SLDFn F and $c \in \mathbb{R}$ with $\mu_F^{(c)} < \infty$:*

$$\mu_F^{(c)} = \left\{ \begin{array}{cc} \lim\limits_{a\to 0, a>0} \left(a^c S_F(a) + c\int_a^\infty y^{c-1} S_F(y)dy\right) & c < 0 \\ 1 & c = 0 \\ c\int_0^\infty x^{c-1} S_F(x)dx & c > 0 \end{array} \right\}.$$

Proof. Suppose first that $c < 0$. Letting $a > b = 0$ in Proposition 11

$$\begin{aligned}\mu^{(c)} &= \int_0^\infty y^c f(y)dy \\ &= \lim_{a\to 0, a>0} \int_a^\infty y^c f(y)dy \\ &= \lim_{a\to 0, a>0} \left(a^c S(a) + c\int_a^\infty y^{c-1} S(y)dy\right)\end{aligned}$$

as asserted. The result is apparent for $c = 0$. For $c > 0$ the result follows by letting $b = 0$ and $a > 0$ go to 0 in Proposition 11

$$\begin{aligned}
\mu^{(c)} &= \int_0^\infty y^c f(y)dy = \lim_{a\to 0}\int_a^\infty y^c f(y)dy = \lim_{a\to 0}\left(a^c S(a) + c\int_a^\infty y^{c-1}S(y)dy\right) \\
&= \lim_{a\to 0} a^c S(a) + \lim_{a\to 0}\left(c\int_a^\infty y^{c-1}S(y)dy\right) \\
&= \lim_{a\to 0} a^c + c\lim_{a\to 0}\int_a^\infty y^{c-1}S(y)dy \\
&= \left(\lim_{a\to 0} a\right)^c + c\int_0^\infty y^{c-1}S(y)dy \\
&= c\int_0^\infty y^{c-1}S(y)dy
\end{aligned}$$

as asserted. ■

The existence of $\mu_F^{(c)}$ for large positive c is typically discussed in terms of the existence of $\mu_F^{(n)}$ for large $n \in \mathbb{N} = \{1, 2, \ldots\}$ and with $\mu_F^{(n)}$ termed a higher moment. And it is often noted that the existence of higher moments is suggestive of a thin right hand tail. We will see how to make that mathematically precise below. The above corollary suggests that the existence $\mu_F^{(c)}$ for negative c is more subtle and we will see later that this relates with the analytic character of the distribution function, more specifically its degree of monotonality (alternating sign of higher order derivatives).

To any SLDFn F we will associate other SLDFns whose moments are closely related to those of F. The simplest case comes from the observation that the function $\widehat{S}(x) = \frac{\int_x^\infty yf(y)dy}{\mu}$ resembles a survival function.

Definition 14 *For any SLDFn F we set $\widehat{F} = 1 - \widehat{S_F}$.*

Proposition 15 *For any SLDFn F with finite mean, $\widehat{F}$ is an SLDFn with*

$$\begin{aligned}
f_{\widehat{F}}(x) &= \frac{xf(x)}{\mu},\ \alpha_{\widehat{F}} = \alpha_F,\ \omega_{\widehat{F}} = \omega_F, \\
\tau_{\widehat{F}} &= \tau_F - \frac{1}{\omega_F} \text{ for finite } \omega_F \\
\tau_{\widehat{F}} &= \tau_F \text{ for } \omega_F = \infty \\
\text{and } \mu_{\widehat{F}}^{(c)} &= \frac{\mu_F^{(c+1)}}{\mu_F} \text{ for every } c \in \mathbb{R}.
\end{aligned}$$

Proof. We have

$$\begin{aligned}
\widehat{F}(0) &= 1 - \widehat{S}(0) = 1 - \frac{\int_0^\infty yf(y)dy}{\mu} = 1 - \frac{\mu}{\mu} = 1 - 1 = 0 \\
\text{and } \frac{d\widehat{F}}{dx} &= \frac{d\left(1 - \widehat{S_F}\right)}{dx} = -\frac{d\widehat{S_F}}{dx} = \frac{-1}{\mu}\frac{d\int_x^\infty yf(y)dy}{dx} = \frac{xf(x)}{\mu} \geq 0
\end{aligned}$$

which clearly implies that $\widehat{F}$ is infinitely differentiable *on* (α_F, ω_F) and continuous and nondecreasing on $[0, \infty)$. Also

$$\infty > \mu = \int_0^\infty yf(y)dy = \lim_{x\to\infty} \int_0^x yf(y)dy + \int_x^\infty yf(y)dy$$

$$= \lim_{x\to\infty} \int_0^x yf(y)dy + \mathit{lim}_{x\to\infty} \int_x^\infty yf(y)dy = \mu + \lim_{x\to\infty} \int_x^\infty yf(y)dy$$

$$\Rightarrow \lim_{x\to\infty} \int_x^\infty yf(y)dy = 0$$

whence

$$\begin{aligned} \lim_{x\to\infty} \widehat{F}(x) &= 1 - \lim_{x\to\infty} \widehat{S}(x) = 1 - \frac{1}{\mu} \lim_{x\to\infty} \int_x^\infty yf(y)dy \\ &= 1 - \frac{0}{\mu} = 1 \end{aligned}$$

and we see that $\widehat{F}$ is an SLDFn. It is clear that $\widehat{F}$ has PDF

$$f_{\widehat{F}}(x) = \frac{d\widehat{F}}{dx} = -\frac{d\widehat{S}}{dx} = -\frac{d}{dx}\left(\frac{\int_x^\infty yf(y)dy}{\mu}\right) = \frac{xf(x)}{\mu}$$

We will make frequent use of the observation that F being an SLDFn implies that the PDF $f = f_F$ is continuous on $(0, \omega_F) \cup (\omega_F, \infty)$. In particular, we have

$$x < \omega_F \Rightarrow F(x) < 1$$

$$\Rightarrow \int_x^\infty f(y)dy = S(x) > 0$$

$$\Rightarrow \text{there exists some } z > x, \epsilon > 0 \text{ such that } \{|w - z| < \epsilon \Rightarrow f(w) > 0\}$$

$$\Rightarrow \widehat{S}(x) = \frac{\int_x^\infty yf(y)dy}{\mu} > 0$$

$$\Rightarrow \widehat{F}(x) < 1$$

moreover

$$x > \omega_F \Rightarrow F(x) = 1$$

$$\Rightarrow \int_x^\infty f(y)dy = S(x) = 0$$

$$\Rightarrow f(w) = 0 \text{ for every } w > x$$

$$\Rightarrow \widehat{S}(x) = \frac{\int_x^\infty yf(y)dy}{\mu} = 0$$

$$\Rightarrow \widehat{F}(x) = 1$$

which establishes $\omega_{\widehat{F}} = \omega_F$. Similarly

$$x < \alpha_F \Rightarrow f(x) = 0$$
$$\Rightarrow \widehat{F}(x) = \frac{\int_0^x yf(y)dy}{\mu} = \frac{\int_0^x y(0)dy}{\mu} = 0$$

moreover

$$x > \alpha_F \Rightarrow \int_0^x f(y)dy = F(x) > 0$$

$\Rightarrow$ there exist $z, \epsilon \in \mathbb{R}$ such that $0 < z < x, \epsilon > 0$ such that $\{|w - z| < \epsilon \Rightarrow f(w) > 0\}$

$$\Rightarrow \widehat{F}(x) = \frac{\int_0^x yf(y)dy}{\mu} > 0$$

which establishes $\alpha_{\widehat{F}} = \alpha_F$. Alternatively, since $f_{\widehat{F}}(x) > 0 \Leftrightarrow f(x) > 0$ it is clear that $\alpha_{\widehat{F}} = \alpha_F$ and $\omega_{\widehat{F}} = \omega_F$. Fist assume that ω_F is finite, then by l'H$\widehat{\text{o}}$pital:

$$\begin{aligned}
\tau_{\widehat{F}} &= \lim_{x\to\omega_F} \lambda_{\widehat{F}}(x) = \lim_{x\to\omega_F} \frac{xf(x)}{\mu\widehat{S}(x)} = \lim_{x\to\omega_F} \frac{x\frac{df}{dx} + f(x)}{\mu\frac{d\widehat{S}}{dx}} \\
&= -\lim_{x\to\omega_F} \frac{x\frac{df}{dx} + f(x)}{\mu f_{\widehat{F}}(x)} = -\lim_{x\to\omega_F} \frac{x\frac{df}{dx} + f(x)}{xf(x)} \\
&= -\lim_{x\to\omega_F} \left(\frac{\frac{df}{dx}}{f(x)} + \frac{1}{x}\right) = -\lim_{x\to\omega_F} \left(\frac{\frac{df}{dx}}{f(x)}\right) - \frac{1}{\omega_F} \\
&= -\lim_{x\to\omega_F} \left(\frac{f(x)}{-S(x)}\right) - \frac{1}{\omega_F} = \lim_{x\to\omega_F} (\lambda_F(x)) - \frac{1}{\omega_F} \\
&= \tau_F - \frac{1}{\omega_F}
\end{aligned}$$

as required. The same argument shows that $\tau_{\widehat{F}} = \tau_F$ for $\omega_F = \infty$. Finally

$$\mu_{\widehat{F}}^{(c)} = \int_0^\infty y^c f_{\widehat{F}}(y)dy = \int_0^\infty y^c \left(\frac{yf(y)}{\mu}\right) dy = \frac{1}{\mu}\int_0^\infty y^{c+1} f(y)dy = \frac{\mu_F^{(c+1)}}{\mu_F}$$

completing the proof. ■

Remark 16 *The distribution of $\widehat{F}$ is sometimes referred to as the time-biased distribution. It has application to sampling theory when the probability of selection increases with time of exposure or attained age.*

It is easy to generalize the time-biased distribution:

Definition 17 *For any SLDFn F and $c \in \mathbb{R}$ with $\mu_F^{(c)} < \infty$, we denote by $\widehat{F}^{[c]}$ the SLDFn with PDF*

$$f_{\widehat{F}^{[c]}}(x) = \frac{x^c f(x)}{\mu^{(c)}}.$$

10

Proposition 18 *For any SLDFn F and $c, d \in \mathbb{R}$ with $\mu_F^{(c)} < \infty$:*

$$\omega_{\widehat{F}^{[c]}} = \omega_F \quad \textit{and } \mu^{(d)}_{\widehat{F}^{[c]}} = \frac{\mu_F^{(c+d)}}{\mu_F^{(c)}}.$$

Proof. As before we see that

$$x < \omega_F \Rightarrow F(x) < 1$$

$$\begin{aligned}
&\Rightarrow \int_x^\infty f(y)dy = S(x) > 0 \\
&\Rightarrow \text{there exist } z > x, \epsilon > 0 \text{ such that } \{|w - z| < \epsilon \Rightarrow f(w) > 0\} \\
&\Rightarrow S_{\widehat{F}^{[c]}}(x) = \frac{\int_x^\infty y^c f(y)dy}{\mu^{(c)}} > 0 \\
&\Rightarrow \widehat{F}^{[c]}(x) < 1
\end{aligned}$$

and we have

$$x > \omega_F \Rightarrow F(x) = 1$$

$$\begin{aligned}
&\Rightarrow \int_x^\infty f(y)dy = S(x) = 0 \\
&\Rightarrow f(w) = 0 \text{ for every } w > x \\
&\Rightarrow S_{\widehat{F}^{[c]}}(x) = \frac{\int_x^\infty y^c f(y)dy}{\mu^{(c)}} = 0 \\
&\Rightarrow \widehat{F}^{[c]}(x) = 1
\end{aligned}$$

whence $\omega_{\widehat{F}^{[c]}} = \omega_F$ and also

$$\begin{aligned}
\mu^{(d)}_{\widehat{F}^{[c]}} &= \int_0^\infty y^d f_{\widehat{F}^{[c]}}(y)dy = \int_0^\infty y^d \left(\frac{y^c f(y)}{\mu^{(c)}}\right) dy \\
&= \frac{\mu^{(c+d)}}{\mu^{(c)}} \int_0^\infty \frac{y^{c+d} f(y)}{\mu^{(c+d)}} dy = \frac{\mu_F^{(c+d)}}{\mu_F^{(c)}}
\end{aligned}$$

as asserted. ■

Analogous to this construction (actually "dual" in a sense to be made precise below), we observe that the mean of any SLDFn F with finite mean can be expressed in terms of its survival function as $\mu = \int_0^\infty S(x)dx$. Therefore the function $\widetilde{f}(x) = \frac{S(x)}{\mu}$ is the PDF of another related SLDFn, which we denote as $\widetilde{F}$.

Definition 19 *For any SLDFn F with finite mean, the* ***coderived distribution*** *of F, which we denote by $\widetilde{F}$, is the distribution function with PDF*

$$\widetilde{f}(x) = f_{\widetilde{F}}(x) = \frac{S(x)}{\mu}.$$

Remark 20 *Observe that*

$$\frac{d\widetilde{f}(x)}{dx} = \frac{1}{\mu}\frac{dS(x)}{dx} = \frac{-f(x)}{\mu}$$
$$\Rightarrow \quad f(x) = -\mu\frac{d\widetilde{f}(x)}{dx}$$

and the PDF of the SLDFn F is obtained by differentiation, or "derived", from that of $\widetilde{F}$. Back in the days of category theory, mathematicians liked to assign the "co-" prefix when reversing arrows. So $\widetilde{F}$ is "coderived" from F, which prompts the name assigned to $\widetilde{F}$.

Klugman [5] relates the right hand tail behavior of the original distribution with that of the coderived distribution, which he terms the "equilibrium distribution". In particular, he considers the asymptotic behavior of the hazard rate functions of the two distributions. We will pursue that somewhat further in this paper. We begin with the observation that the excess ratio is the survival function of the coderived distribution:

Proposition 21 *If F is an SLDFn with finite mean, survival function S and excess ratio function R, then:*

$$\alpha_{\widetilde{F}} = 0,\ \omega_{\widetilde{F}} = \omega_F \quad \text{and } R(x) = \frac{\int_x^\infty S(y)dy}{\int_0^\infty S(y)dy} = \int_x^\infty \widetilde{f}(y)dy = S_{\widetilde{F}}(x),\ \text{for } x \geq 0.$$

Proof. Let F have PDF $f = f_F$, since $\widetilde{f}(0) = \frac{1}{\mu} > 0$ is continuous at 0, clearly $\alpha_{\widetilde{F}} = 0$. We also have

$$x < \omega_F \Rightarrow F(x) < 1$$

$$\Rightarrow \quad \mu\widetilde{f}(x) = \int_x^\infty f(y)dy = S(x) > 0$$
$$\Rightarrow \quad \widetilde{S}(x) = \int_x^\infty \widetilde{f}(y)dy > 0$$
$$\Rightarrow \quad \widetilde{F}(x) < 1$$

and moreover

$$x > \omega_F \Rightarrow F(x) = 1$$

$$\Rightarrow \quad \int_x^\infty f(y)dy = S(x) = 0$$
$$\Rightarrow \quad \widetilde{f}(w) = \frac{S(w)}{\mu} = 0 \text{ for every } w > x$$
$$\Rightarrow \quad \widetilde{S}(x) = \int_x^\infty \widetilde{f}(w)dw = 0$$
$$\Rightarrow \quad \widetilde{F}(x) = 1$$

which establishes $\omega_{\widetilde{F}} = \omega_F$. Now from Proposition 11 we have

$$\begin{aligned}\int_x^\infty S(y)dy &= \int_x^\infty yf(y)dy - xS(x) \\ &= -x\int_x^\infty f(y)dy + \int_x^\infty yf(y)dy \\ &= \int_x^\infty (y-x)f(y)dy,\end{aligned}$$

Thus

$$R(x) = \frac{\int_x^\infty (y-x)f(y)dy}{\mu} = \frac{\int_x^\infty S(y)dy}{\int_0^\infty S(y)dy} = \int_x^\infty \frac{S(y)}{\mu}dy = \int_x^\infty \widetilde{f}(y)dy.$$

as required. ■

Corollary 22 *Under the assumptions of the Proposition:*

$$\frac{dR}{dx}(x) = \frac{-S(x)}{\mu} = -\widetilde{f}(x), \text{ for every } x \geq 0.$$

Proof. By the Fundamental Theorem of Calculus

$$\frac{dR}{dx} = \frac{d}{dx}\left(\frac{\int_x^\infty S(y)dy}{\mu}\right) = \frac{-S(x)}{\mu} = -\widetilde{f}(x).$$

as required. ■

Let F be an SLDFn with finite mean. Observe that $\widetilde{F}$ is again an SLDFn and so provided $\mu_{\widetilde{F}} < \infty$ we can repeat the process to get $\widetilde{\widetilde{F}}$. More precisely, we can recursively construct the sequence of LDFns

$$\begin{aligned}\widetilde{F}^{[0]} &= F \\ \widetilde{F}^{[1]} &= \widetilde{F} \\ \widetilde{F}^{[n]} &= \widetilde{\widetilde{F}^{[n-1]}} \text{ for } n = 2,3,4,\ldots\text{provided } \mu_{\widetilde{F}^{[n-1]}} < \infty.\end{aligned}$$

and refer to $\widetilde{F}^{[n]}$ as the **n-th forward coderived LDFn** of F. It is clear that $\omega_{\widetilde{F}^{[n]}} = \omega_F$ for $n = 2,3,4,\ldots$provided $\mu_{\widetilde{F}^{[n-1]}} < \infty$.

We will soon see (Proposition 27) that quite generally the existence of an n-th forward coderived LDFn is equivalent to having a finite n-th moment

$$\widetilde{F}^{[n]} \text{ exists } \Leftrightarrow \mu^{(n)} < \infty.$$

The PDF of the coderived loss distribution is continuous and nonincreasing and so a mode of any such coderived distribution is at $x = 0$ where its PDF takes its maximum value of $\frac{1}{\mu}$. Conversely, if F is an SLDFn with nonincreasing PDF f, then it is easy to verify that $G(x) = \frac{f(0)-f(x)}{f(0)}$ is an SLDFn with coderived distribution $\widetilde{G} = F$. It is also worth noting that because the survival curve completely determines the distribution, the coderived distribution completely determines the original distribution. And indeed for any n, the n-th forward coderived LDFn, should there be one, completely determines the original LDFn. We conclude this section with a rather general observation on the existence of moments.

Proposition 23 *If F is an SLDFn with finite mean, then there exist unique $a, c \in \mathbb{R} \cup \{\infty\}$ such that:*

$$(0,1) \subseteq (a,c) = \left\{ b \in \mathbb{R} - \{a,c\} \, |\mu_F^{(b)} < \infty \right\}.$$

Proof. Set $A = \left\{ b \in \mathbb{R} | \mu_F^{(b)} < \infty \right\}$. We claim that A is a connected subset of $\mathbb{R}$. To see this, note that

$$\begin{aligned} a, c \quad &\in \quad A \Rightarrow \int_0^\infty y^a f(y)dy, \int_0^\infty y^c f(y)dy < \infty \\ &\Rightarrow \quad \int_0^1 y^a f(y)dy, \int_1^\infty y^a f(y)dy, \int_0^1 y^c f(y)dy, \int_1^\infty y^c f(y)dy < \infty. \end{aligned}$$

So suppose $a < b < c$ with $a, c \in A$, then we have

$$\begin{aligned} 0 \quad &< \quad y < 1 \Rightarrow y^a > y^b > y^c \Rightarrow \infty > \int_0^1 y^a f(y)dy > \int_0^1 y^b f(y)dy \\ 1 \quad &< \quad y \Rightarrow y^a < y^b < y^c \Rightarrow \int_1^\infty y^b f(y)dy < \int_1^\infty y^c f(y)dy < \infty \\ &\Rightarrow \quad \mu^{(b)} = \int_0^\infty y^b f(y)dy = \int_0^1 y^b f(y)dy + \int_1^\infty y^b f(y)dy < \infty \\ &\Rightarrow \quad b \in A. \end{aligned}$$

And it follows that A is connected, as claimed. Now since F has finite mean, we clearly have

$$\begin{aligned} \mu^{(0)} \quad &= \quad \int_0^1 y^0 f(y)dy = \int_0^1 1 f(y)dy = 1 < \infty \Rightarrow 0 \in A \\ \mu^{(1)} \quad &= \quad \int_0^1 y^1 f(y)dy = \int_0^1 y f(y)dy = \mu < \infty \Rightarrow 1 \in A \end{aligned}$$

and since the connected subsets of $\mathbb{R}$ are exactly the intervals, the result follows. ■

Example 24 *For the usual families of loss distributions (beta, Pareto, burr, Weibull, gamma,...) the set $\left\{ c \in \mathbb{R} | \mu^{(c)} < \infty \right\}$ is an* ***open*** *interval. The following example, provided by Derek Schaff, shows that is not always the case for* *loss distributions. Define*

$$\begin{aligned} g(x) &= \left\{ \begin{array}{cc} 0 & x \leq 2 \\ \left(\frac{1}{x\ln x}\right)^2 & x > 2 \end{array} \right\} \\ &\Rightarrow \int_0^\infty y^c g(y)dy = \int_2^\infty y^c g(y)dy \\ &= \int_{\ln 2}^\infty \frac{e^{(c-1)u}}{u^2} du \left\{ \begin{array}{cc} < \infty & c \leq 1 \\ \infty & c > 1 \end{array} \right\} \end{aligned}$$

where we used the change of variable $u = \ln y$, noting that for $c \leq 1$ the integral is dominated by the convergent integral $\int_0^\infty \frac{du}{u^2}$ while for $c > 1$ l'Hôpital shows that the integrand does not even go to 0 as $u \to \infty$. We see that

$$f_F(x) = \frac{g(x)}{\int_0^\infty g(y)dy} \Rightarrow \{c \in \mathbb{R} | E[X^c] < \infty\} = (-\infty, 1].$$

3 Moments and the Coderived Distribution

The discussion leading to the definition of a coderived loss distribution together with Proposition 10 gives the first two Items of:

Proposition 25 *For any SLDFn F with finite mean μ and all $x \geq 0$:*

1. $f_{\widetilde{F}}(x) = \frac{S_F(x)}{\mu}$
2. $S_{\widetilde{F}}(x) = R_F(x) = \widehat{S}_F(x) - x f_{\widetilde{F}}(x)$
3. $S_F(x) > 0 \Rightarrow \lambda_{\widetilde{F}}(x) = \left(\mu \frac{\widehat{S}_F(x)}{S_F(x)} - x\right)^{-1} > 0$
4. $S_F(x) > 0 \Rightarrow \lambda_{\widetilde{F}}(x) = \frac{\int_0^\infty f(x+z)dz}{\int_0^\infty z f(x+z)dz}$.

Proof. Items 1 and 2 have been established. For Item 3, we have seen that

$$S_F(x) > 0$$

$$\Rightarrow S_{\widetilde{F}}(x) > 0 \text{ and } f_{\widetilde{F}}(x) = \frac{S_F(x)}{\mu} > 0$$

and so by Items 1 and 2

$$\begin{aligned} 0 &< \lambda_{\widetilde{F}}(x) = \frac{f_{\widetilde{F}}(x)}{S_{\widetilde{F}}(x)} = \left(\frac{S_{\widetilde{F}}(x)}{f_{\widetilde{F}}(x)}\right)^{-1} = \left(\frac{\widehat{S}_F(x) - x f_{\widetilde{F}}(x)}{f_{\widetilde{F}}(x)}\right)^{-1} \\ &= \left(\frac{\widehat{S}_F(x)}{f_{\widetilde{F}}(x)} - x\right)^{-1} = \left(\frac{\widehat{S}_F(x)}{\frac{S_F(x)}{\mu}} - x\right)^{-1} = \left(\mu \frac{\widehat{S}_F(x)}{S_F(x)} - x\right)^{-1} \end{aligned}$$

as required. And then Item 4 follows from definitions and the change of variable $z = y - x$

$$\lambda_{\widetilde{F}}(x) = \left(\frac{\mu \widehat{S}_F(x)}{S_F(x)} - x \right)^{-1} = \left(\frac{\int_x^\infty y f(y) dy}{\int_x^\infty f(y) dy} - x \right)^{-1}$$

$$= \left(\frac{\int_x^\infty y f(y) dy - x \int_x^\infty f(y) dy}{\int_x^\infty f(y) dy} \right)^{-1} = \frac{\int_x^\infty f(y) dy}{\int_x^\infty y f(y) dy - \int_x^\infty x f(y) dy}$$

$$= \frac{\int_x^\infty f(y) dy}{\int_x^\infty (y - x) f(y) dy} = \frac{\int_0^\infty f(x+z) dz}{\int_0^\infty z f(x+z) dz}$$

completing the proof. ■

As was noted, the coderived distribution determines the original:

Proposition 26 *For any two SLDFns with finite means F and G*

$$F = G \quad \Leftrightarrow \quad \widetilde{F} = \widetilde{G}.$$

Proof. Trivially, $F = G \Rightarrow \widetilde{F} = \widetilde{G}$. Conversely

$$\widetilde{F} = \widetilde{G} \Rightarrow \frac{S_F(x)}{\mu_F} = f_{\widetilde{F}}(x) = f_{\widetilde{G}}(x) = \frac{S_G(x)}{\mu_G}$$

and letting $x = 0$ we have

$$\begin{aligned} \frac{1}{\mu_F} &= \frac{S_F(0)}{\mu_F} = \frac{S_G(0)}{\mu_G} = \frac{1}{\mu_G} \\ &\Rightarrow \mu_F = \mu_G \\ &\Rightarrow 1 - F = S_F = S_G = 1 - G \\ &\Rightarrow F = G \end{aligned}$$

as asserted. ■

The moments of coderived distributions are readily obtained from those of the original distribution:

Proposition 27 *If F is an SLDFn and $n \in \mathbb{N}$, then:*

$$\mu_F^{(n)} < \infty \Rightarrow \mu_{\widetilde{F}}^{(k)} = \frac{\mu_F^{(k+1)}}{(k+1)\,\mu_F} < \infty \text{ for } k = 0, 1, 2, ..., n-1.$$

Proof. For $k = 0$ we have $\mu_{\widetilde{F}}^{(0)} = 1 = \frac{\mu}{\mu} = \frac{\mu_F^{(1)}}{(1)\mu_F}$. More generally, from Proposition 23 we know that

$$\mu_F^{(n)} < \infty \Rightarrow \mu_F^{(k)} < \infty \text{ for } k = 0, 1, 2, ..., n$$

and from Corollary 13

$$\mu_{\widetilde{F}}^{(k)} = \int_0^\infty x^k f_{\widetilde{F}}(x)dx = \int_0^\infty x^k \frac{S(x)}{\mu}dx = \frac{1}{(k+1)\,\mu}\int_0^\infty x^{k+1} f(x)dx = \frac{\mu_F^{(k+1)}}{(k+1)\,\mu_F}$$

as required. ■

We see that

$$F \text{ has } n \text{ finite moments } \mu_F^{(k)},\ k = 1,2,3,...,n$$

$$\Leftrightarrow \widetilde{F} \text{ has } n-1 \text{ finite moments } \mu_{\widetilde{F}}^{(k)},\ k = 1,2,3,...,n-1.$$

Taking the coderived distribution can remove the existence of a higher moment.

The ultimate settlement rate τ_F is a useful measure of the tail behavior of a loss distribution. Our first significant result is that the tail behavior of the coderived loss variables has τ_F in common with the original, i.e., τ_F is a $\sim$invariant:

Proposition 28 *If F is an SLDFn with $\mu_F^{(n)} < \infty$, then:*

$$\tau_F = \tau_{\widetilde{F}^{[k]}}, 0 \leq k \leq n.$$

Proof. Note that by Proposition 25 and Corollary 22

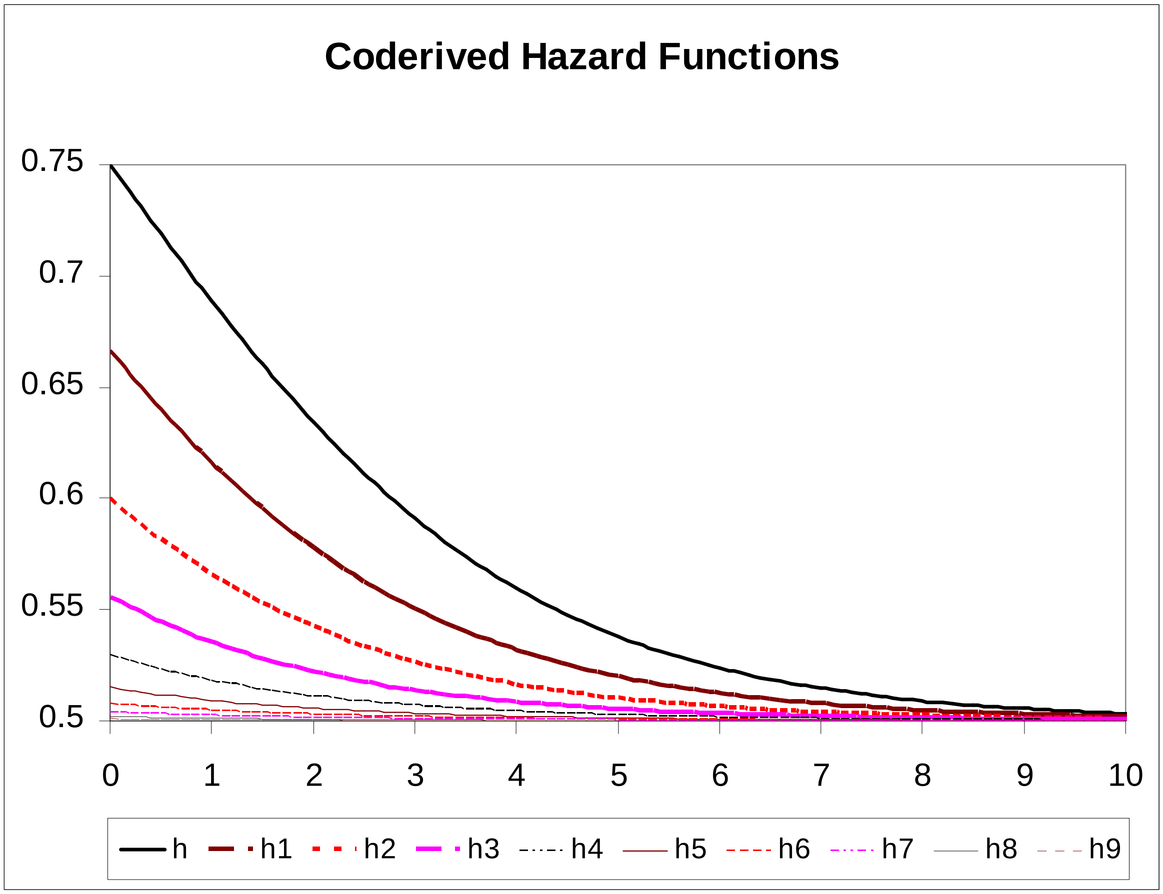
$$\begin{aligned}
\tau_{\widetilde{F}} &= \lim_{x\to\omega_F} \lambda_{\widetilde{F}}(x) \\
&= \lim_{x\to\omega_F} \frac{f_{\widetilde{F}}(x)}{S_{\widetilde{F}}(x)} \\
&= \frac{1}{\mu}\lim_{x\to\omega_F} \frac{S(x)}{R(x)}
\end{aligned}$$

and since $\lim_{x\to\omega_F} S(x) = 0 = \lim_{x\to\omega_F} R(x)$ we may invoke l'H$\hat{\text{o}}$pital

$$\begin{aligned}
\tau_{\widetilde{F}} &= \frac{1}{\mu}\lim_{x\to\omega_F} \frac{\frac{dS}{dx}}{\frac{dR}{dx}} \\
&= \frac{1}{\mu}\lim_{x\to\omega_F} \frac{-f(x)}{\frac{-S(x)}{\mu}} \\
&= \frac{\mu}{\mu}\lim_{x\to\omega_F} \frac{f(x)}{S(x)} \\
&= \lim_{x\to\omega_F} \lambda_F(x) = \tau_F
\end{aligned}$$

and since $\mu_F^{(n)} < \infty \Rightarrow \mu_F^{(k)} < \infty$, $1 \leq k \leq n$, the result for $n \geq 2$ follows by repeated application

$$\tau_F = \tau_{\widetilde{F}} = \tau_{\widetilde{F}^{[1]}} = \tau_{\widetilde{F}^{[2]}} = \cdots = \tau_{\widetilde{F}^{[n]}}$$



completing the proof. ■

Perhaps the easiest way to understand coderived distributions is to look at their hazard rate functions. In the chart below

$$h = \lambda_F \quad \text{with } \tau_F = \frac{1}{2} \quad \text{and } hx = \lambda_{\widetilde{F}[x]}, 1 \leq x \leq 10.$$

The chart illustrates how the higher coderived distributions "anticipate the tail", converging faster to the constant τ_F.

Another way to see that the coderived variable shares tail behavior is to compare the survival curve of the coderived variable with that of conditional survival excess of a particular loss amount c. More precisely, we make the:

Definition 29 *Let F be an SLDFn and c be a positive constant such that $F(c) < 1$. The* ***over c residual loss variable****, denoted $F^{>c}$ is the SLDFn*

18

determined from

$$\begin{aligned} F^{>c}(x) &= 1 - \frac{S_F(x+c)}{S_F(c)} \textit{ for every } x \geq 0 \\ \iff & S_{F^{>c}} = \frac{S_F(x+c)}{S_F(c)} \textit{ for every } x \geq 0. \end{aligned}$$

The following shows that the tail behavior of a residual variable is also akin to that of the original loss variable and that there is a simple relationship between this residual and the coderived variables:

Proposition 30 *If F is an SLDFn and c is a positive constant such that $S_F(c) > 0$,then:*

1. $\omega_{F^{>c}} = \omega_F - c$
2. $f_{F^{>c}}(x) = \frac{f_F(x+c)}{S_F(c)}$ for every $x \geq 0$
3. $d \geq 0$ such that $S_F(c+d) > 0 \Rightarrow (F^{>c})^{>d} = F^{>c+d}$
4. $\lambda_{F^{>c}}(x) = \lambda_F(x+c)$ for every $x \geq 0$
5. $\tau_{F^{>c}} = \tau_F$
6. $\widetilde{F}^{>c} = \widetilde{F^{>c}}$
7. $\mu_F < \infty \Rightarrow \mu_{F^{>c}} = \mu_F \frac{S_{\widetilde{F}}(c)}{S_F(c)} = \frac{S_{\widetilde{F}}(c)}{f_{\widetilde{F}}(c)} = \frac{1}{\lambda_{\widetilde{F}}(c)}$
8. $\left(\widetilde{F}^{[n]}\right)^{>c} = \widetilde{F^{>c}}^{[n]}$ for every $n \in \mathbb{N}$.

Proof. Note that $S_F(c) > 0 \Rightarrow F(c) < 1 \Rightarrow c \leq \omega_F \Rightarrow \omega_F - c \geq 0$. Item 1 is obvious

$$\begin{aligned} x < \omega_F - c \Rightarrow x + c < \omega_F &\Rightarrow F(x+c) < 1 \\ &\Rightarrow S(x+c) > 0 \\ &\Rightarrow F^{>c}(x) = 1 - \frac{S_F(x+c)}{S_F(c)} < 1 \end{aligned}$$

and

$$\begin{aligned} x > \omega_F - c \Rightarrow x + c > \omega_F &\Rightarrow F(x+c) = 1 \\ &\Rightarrow S(x+c) = 0 \\ &\Rightarrow F^{>c}(x) = 1 - \frac{S_F(x+c)}{S_F(c)} = 1. \end{aligned}$$

Item 2 follows from the chain rule

$$\begin{aligned} f_{F^{>c}}(x) &= \frac{d}{dx}\left(F^{>c}(x)\right) = \frac{d}{dx}\left(\frac{-S_F(x+c)}{S_F(c)}\right) \\ &= \frac{-1}{S_F(c)}\frac{d}{dx}\left(S_F(x+c)\right) = \frac{-1}{S_F(c)}\frac{d\left(S_F(x+c)\right)}{d\left(x+c\right)}\frac{d\left(x+c\right)}{dx} \\ &= \frac{-1}{S_F(c)}\left(-f_F(x+c)\right) = \frac{f_F(x+c)}{S_F(c)}. \end{aligned}$$

For Item 3

$$\begin{aligned} S_{(F^{>c})^{>d}}(x) &= \frac{S_{F^{>c}}(x+d)}{S_{F^{>c}}(d)} = \frac{\frac{S_F((x+d)+c)}{S_F(c)}}{\frac{S_F(d+c)}{S_F(c)}} = \frac{S_F(x+(c+d))}{S_F(c+d)} = S_{F^{>c+d}}(x) \\ &\Rightarrow \left(F^{>c}\right)^{>d} = F^{>c+d} \end{aligned}$$

Item 4 is immediate from Item 2

$$\lambda_{F^{>c}}(x) = \frac{f_{F^{>c}}(x)}{S_{F^{>c}}(x)} = \frac{\frac{f_F(x+c)}{S_F(c)}}{\frac{S_F(x+c)}{S_F(c)}} = \frac{f_F(x+c)}{S_F(x+c)} = \lambda_F(x+c)$$

and Item 5 is immediate from Item 4

$$\tau_{F^{>c}} = \lim_{x\to\infty}\lambda_{F^{>c}}(x) = \lim_{x\to\infty}\lambda_F(x+c) = \lim_{x\to\infty}\lambda_F(x) = \tau_F.$$

Observe next that letting $G = F^{>c}$ we have the PDF

$$f_{\widetilde{G}}(x) = \frac{S_G(x)}{\mu_G} = \frac{S_{F^{>c}}(x)}{\mu_G} = \frac{\frac{S_F(x+c)}{S_F(c)}}{\mu_G} = \frac{S_F(x+c)}{\mu_G S_F(c)}$$

while by Item 2 we also have the PDF

$$f_{\widetilde{F}^{>c}}(x) = \frac{f_{\widetilde{F}}(x+c)}{S_{\widetilde{F}}(c)} = \frac{\frac{S_F(x+c)}{\mu_F}}{S_{\widetilde{F}}(c)} = \frac{S_F(x+c)}{\mu_F S_{\widetilde{F}}(c)}$$

which implies that the two PDFs are proportional, whence equal

$$\begin{aligned} \frac{S_F(x+c)}{\mu_G S_F(c)} &= f_{\widetilde{G}}(x) = f_{\widetilde{F}^{>c}}(x) = \frac{S_F(x+c)}{\mu_F S_{\widetilde{F}}(c)} \\ &\Rightarrow \widetilde{F^{>c}} = \widetilde{G} = \widetilde{F}^{>c} \end{aligned}$$

which proves Item 6. For Item 7 just note that from the above equation with $x = 0$

$$\begin{aligned} \frac{S_F(0+c)}{\mu_G S_F(c)} &= \frac{S_F(0+c)}{\mu_F S_{\widetilde{F}}(c)} \\ &\Rightarrow \mu_{F^{>c}} = \mu_G = \mu_F\frac{S_{\widetilde{F}}(c)}{S_F(c)} = \frac{S_{\widetilde{F}}(c)}{\frac{S_F(c)}{\mu_F}} \\ &= \frac{S_{\widetilde{F}}(c)}{f_{\widetilde{F}}(c)} = \frac{1}{\frac{f_{\widetilde{F}}(c)}{S_{\widetilde{F}}(c)}} = \frac{1}{\lambda_{\widetilde{F}}(c)}. \end{aligned}$$

20

Finally, Item 8 is a straightforward induction on n using Item 6; indeed, case $n = 1$ is Item 6, and then

$$\left(\widetilde{F}^{[n+1]}\right)^{>c} = \left(\widetilde{\widetilde{F}^{[n]}}\right)^{>c} = \widetilde{\left(\widetilde{F}^{[n]}\right)^{>c}} = \widetilde{\widetilde{F^{>c}}^{[n]}} = \widetilde{F^{>c}}^{[n+1]}$$

which completes the induction and the proof. ∎

This provides a perspective on the coderived survival curve of an SLDFn, inasmuch as the coderived survival is to the original survival probability in the same proportion as the mean residual life is to the overall mean lifetime

$$\frac{S_{\widetilde{F}}(c)}{S_F(c)} = \frac{\mu_{F^{>c}}}{\mu_F}.$$

And the hazard rate function for the coderived distribution is the reciprocal of the mean residual life

$$\lambda_{\widetilde{F}}(c) = \frac{1}{\mu_{F^{>c}}}.$$

This perspective leads to a relationship between λ_F and $\lambda_{\widetilde{F}}$:

Proposition 31 *If F is an SLDFn with finite mean, then whenever λ_F is increasing (nondecreasing, decreasing, nonincreasing) on $(0, \omega_F) = \left(0, \omega_{\widetilde{F}}\right)$, then so too is $\lambda_{\widetilde{F}}$.*

Proof. Suppose λ is increasing and fix any $z > 0$, then for $y + z < \omega_F$

$$\frac{d}{dy}\left(\frac{S(y+z)}{S(y)}\right) = \frac{S(y)\frac{d}{dy}(S(y+z)) - S(y+z)\frac{d}{dy}(S(y))}{S(y)^2}$$

$$\begin{aligned}
&= \frac{S(y)(-f(y+z)) + S(y+z)f(y)}{S(y)^2} \\
&= \frac{S(y+z)f(y) - S(y)(f(y+z))}{S(y)^2} \\
&= \frac{S(y+z)}{S(y)}\lambda(y) - \frac{S(y+z)}{S(y)}\frac{f(y+z)}{S(y+z)} \\
&= \frac{S(y+z)}{S(y)}\lambda(y) - \frac{S(y+z)}{S(y)}\lambda(y+z) \\
&= \frac{S(y+z)}{S(y)}(\lambda(y) - \lambda(y+z)) < 0.
\end{aligned}$$

And so $\frac{S(y+z)}{S(y)}$ is a a decreasing function of y. It follows that

$$x < y \Rightarrow \frac{S(x+z)}{S(x)} > \frac{S(y+z)}{S(y)}$$

21

$$\Rightarrow \quad \mu_{F>x} = \int_0^\infty S_{F>x}(z)\,dz = \int_0^\infty \frac{S(x+z)}{S(x)}dz > \int_0^\infty \frac{S(y+z)}{S(y)}dz = \int_0^\infty S_{F>y}(z)\,dz = \mu_{F>y}$$

$$\Rightarrow \quad \lambda_{\widetilde{F}}(x) = \frac{1}{\mu_{F>x}} < \frac{1}{\mu_{F>y}} = \lambda_{\widetilde{F}}(y)$$

and $\lambda_{\widetilde{F}}$ is also increasing, as required. The case of λ nondecreasing follows similarly, simply by changing strict inequalities to inequalities. The cases of λ decreasing and nonincreasing follow by reversing inequalities. ■

Proposition 32 *If F is an SLDFn with finite mean and c ia a positive constant such that $\lambda_F(c) > 0$,then:*

$$\begin{aligned} \lambda_F \textit{ nondecreasing} &\Rightarrow \lambda_{\widetilde{F}}(c) \geq \lambda_F(c) \\ \lambda_F \textit{ increasing} &\Rightarrow \lambda_{\widetilde{F}}(c) > \lambda_F(c) \\ \lambda_F \textit{ nonincreasing} &\Rightarrow \lambda_{\widetilde{F}}(c) \leq \lambda_F(c) \\ \lambda_F \textit{ decreasing} &\Rightarrow \lambda_{\widetilde{F}}(c) < \lambda_F(c). \end{aligned}$$

Proof. Suppose λ is nondecreasing. For any $z > 0$

$$\begin{aligned} S(c+z) &= e^{-\int_0^{c+z}\lambda(t)dt} \\ \frac{S(c+z)}{S(c)} &= e^{\int_0^{c}\lambda(t)dt - \int_0^{c+z}\lambda(t)dt} = e^{-\int_c^{c+z}\lambda(t)dt}. \end{aligned}$$

And since λ is nondecreasing

$$t \in (c, c+z) \Rightarrow \lambda(t) \geq \lambda(c)$$

$$\begin{aligned} &\Rightarrow \int_c^{c+z}\lambda(t)dt \geq \int_c^{c+z}\lambda(c)dt = \lambda(c)\,z \\ &\Rightarrow -\int_c^{c+z}\lambda(t)dt \leq -\lambda(c)\,z \\ &\Rightarrow \frac{S(c+z)}{S(c)} = e^{-\int_c^{c+z}\lambda(t)dt} \leq e^{-\lambda(c)z}. \end{aligned}$$

And we have

$$\begin{aligned} 0 \quad &< \quad \frac{1}{\lambda_{\widetilde{F}}(c)} = \int_0^\infty \frac{S(c+z)}{S(c)}dz \leq \int_0^\infty e^{-\lambda(c)z}dz = \left[\frac{e^{-\lambda(c)z}}{-\lambda(c)}\right]_0^\infty = \frac{1}{\lambda(c)} \\ &\Rightarrow \lambda_{\widetilde{F}}(c) \geq \lambda(c) \end{aligned}$$

22

as required. The case of λ increasing follows by making the inequalities strict. The case of λ nonincreasing/decreasing follows similarly, reversing inequalities. ■

Proposition 33 *If F is an SLDFn with finite mean and $\tau_F > 0$, then $\mu_F^{(n)} < \infty$ for every $n \in \mathbb{N}$.*

Proof. We first show that $\tau_F > 0 \Rightarrow \mu_{\widetilde{F}} < \infty$. Observe that $f_{\widetilde{F}}(x) = \frac{S(x)}{\mu} > 0$ for every $x < \omega_F = \omega_{\widetilde{F}}$. By Proposition 28

$$\begin{aligned} \lim_{x\to\omega_F} \frac{f_{\widetilde{F}}(x)}{S_{\widetilde{F}}(x)} &= \lim_{x\to\omega_F} \lambda_{\widetilde{F}}(x) = \tau_{\widetilde{F}} = \tau_F > 0 \\ &\Rightarrow \quad 0 < \lim_{x\to\omega_F} \frac{S_{\widetilde{F}}(x)}{f_{\widetilde{F}}(x)} = \frac{1}{\tau_F} < \infty. \end{aligned}$$

This entails that there exist constants M and $b > 0$ such that

$$\begin{aligned} \frac{S_{\widetilde{F}}(x)}{f_{\widetilde{F}}(x)} &\leq b \text{ for every } x \in (M, \omega_F) \\ &\Rightarrow \quad S_{\widetilde{F}}(x) \leq b f_{\widetilde{F}}(x) \text{ for every } x \in (M, \omega_F). \end{aligned}$$

Whence

$$\begin{aligned} \mu_{\widetilde{F}} &= \int_0^\infty S_{\widetilde{F}}(x)dx \\ &= \int_0^M S_{\widetilde{F}}(x)dx + \int_M^{\omega_F} S_{\widetilde{F}}(x)dx \\ &\leq \int_0^M 1dx + \int_M^\infty b f_{\widetilde{F}}(x)dx \\ &\leq M + b < \infty \end{aligned}$$

as claimed. But then, again by Proposition 28, we must have that $\mu_{\widetilde{F}[n]} < \infty$ for every $n \in \mathbb{N}$. The proof is completed by induction on n, the case $n = 1$ being clear. So assume the result holds for $k \leq n$. Note that $\tau_{\widetilde{F}} = \tau_F > 0$.Then we have, by induction applied to $\widetilde{F}$ and Proposition 27

$$\mu_F^{(n+1)} = (n+1)\, \mu_{\widetilde{F}}^{(n)} \mu_F < \infty$$

completing the induction and the proof. ■

Remark 34 *The lognormal density shows that the converse does not hold.*

Proposition 35 *If F is an SLDFn with $0 < \tau_F < \infty$, then $\lim_{c\to\omega_F} \mu_{F^{>c}} = \frac{1}{\tau_F}$.*

Proof. We have from l'Hôpital

$$\lim_{c\to\omega_F} \mu_{F^{>c}} = \lim_{c\to\omega_F} \int_0^\infty S_{F^{>c}}(x)dx = \lim_{c\to\omega_F} \int_0^\infty \frac{S_F(x+c)}{S_F(c)} dx$$

$$
\begin{aligned}
&= \lim_{c\to\omega_F} \frac{\int_0^\infty S_F(x+c)dx}{S_F(c)} = \lim_{c\to\omega_F} \frac{\int_c^\infty S_F(x)dx}{S_F(c)} \\
&= \lim_{c\to\omega_F} \frac{\frac{d}{dc}\int_c^\infty S_F(x)dx}{\frac{d}{dc}S_F(c)} = \lim_{c\to\omega_F} \frac{-S_F(c)}{-f_F(c)} \\
&= \lim_{c\to\omega_F} \frac{1}{\lambda_F(c)} = \frac{1}{\lim\limits_{c\to\omega_F}\lambda_F(c)} = \frac{1}{\tau_F}
\end{aligned}
$$

as claimed. ■

Proposition 36 *If F is an SLDFn with finite mean and $0 < \tau_F < \infty$, then:*

$$\lim_{x\to\omega_F} \frac{S_{\widetilde{F}}(x)}{S_F(x)} = \frac{1}{\mu_F \tau_F}.$$

Proof. From the above propositions

$$\lim_{x\to\omega_F} \frac{S_{\widetilde{F}}(x)}{S_F(x)} = \lim_{x\to\omega_F} \frac{\mu_{F>x}}{\mu_F} = \frac{\lim\limits_{x\to\omega_F}\mu_{F>x}}{\mu_F} = \frac{\frac{1}{\tau_F}}{\mu_F} = \frac{1}{\mu_F\tau_F}$$

as claimed. ■

Proposition 37 *If F is an SLDFn with finite mean and $0 < \tau_F < \infty$, then:*

$$\lim_{x,c\to\omega_F} \frac{S_{\widetilde{F>c}}(x)}{S_{F>c}(x)} = 1.$$

Proof. From the above

$$\lim_{c,x\to\omega_F} \frac{S_{\widetilde{F>c}}(x)}{S_{F>c}(x)} = \lim_{c\to\omega_F}\left(\lim_{x\to\omega_F} \frac{S_{\widetilde{F>c}}(x)}{S_{F>c}(x)}\right)$$

$$
\begin{aligned}
&= \lim_{c\to\omega_F} \frac{1}{\mu_{F>c}\tau_{F>c}} = \lim_{c\to\omega_F} \frac{1}{\mu_{F>c}\tau_F} \\
&= \frac{1}{\tau_F}\lim_{c\to\omega_F} \frac{1}{\mu_{F>c}} = \frac{1}{\tau_F}\frac{1}{\lim\limits_{c\to\omega_F}\mu_{F>c}} \\
&= \frac{1}{\tau_F}\frac{1}{\frac{1}{\tau_F}} = \frac{1}{1} = 1
\end{aligned}
$$

as claimed. ■

Proposition 38 *For any SLDFn F such that $\mu_F^{(n)} < \infty$ for every $n \in \mathbb{N}$:*

$$L_{\widetilde{F}}(t) = \frac{1-L_F(t)}{\mu t} \quad \text{for } t > 0$$

and if F has a moment generating function, then so does $\widetilde{F}$ with

$$M_{\widetilde{F}}(t) = \frac{M_F(t)-1}{\mu t} \quad \text{for } t > 0.$$

Proof. We have, from Proposition 27, for every $t > 0$

$$
\begin{aligned}
L_{\widetilde{F}}(t) &= \sum_{k=0}^{\infty} \frac{(-1)^k \mu_{\widetilde{F}}^{(k)} t^k}{k!} \\
&= \sum_{k=0}^{\infty} \frac{(-1)^k \left(\frac{\mu_F^{(k+1)}}{(k+1)\mu} \right) t^k}{k!} \\
&= \frac{1}{\mu} \sum_{k=0}^{\infty} \frac{(-1)^k \mu_F^{(k+1)} t^k}{(k+1)!} \\
&= \frac{1}{\mu t} \sum_{k=0}^{\infty} \frac{(-1)^k \mu_F^{(k+1)} t^{k+1}}{(k+1)!}
\end{aligned}
$$

$$
\begin{aligned}
&\Rightarrow \quad -\mu t L_{\widetilde{F}}(t) = \sum_{k=0}^{\infty} \frac{(-1)^{k+1} \mu_F^{(k+1)} t^{k+1}}{(k+1)!} \\
&= \quad \sum_{j=1}^{\infty} \frac{(-1)^j \mu_F^{(j)} t^j}{j!} = L_F(t) - 1 \\
&\Rightarrow \quad L_{\widetilde{F}}(t) = \frac{1 - L_F(t)}{\mu t}.
\end{aligned}
$$

And so if $M_F(t)$ exists, it follows that

$$
M_{\widetilde{F}}(t) = L_{\widetilde{F}}(-t) = \frac{1 - L_F(-t)}{-\mu t} = \frac{L_F(-t) - 1}{\mu t} = \frac{M_F(t) - 1}{\mu t}
$$

as required. ■

A straightforward integration by parts, however, provides the stronger result:

Proposition 39 *For any SLDFn F:*

$$
L_{\widetilde{F}}(t) = \frac{1 - L_F(t)}{\mu t} \qquad \text{for } t > 0.
$$

Proof. Fix $t > 0$, we have

$$\begin{aligned}
L_{\widetilde{F}}(t) &= \int_0^\infty e^{-tx} f_{\widetilde{F}}(x)dx \\
&= \int_0^\infty e^{-tx} \frac{S(x)}{\mu} dx \\
&= \frac{1}{\mu} \int_0^\infty udv \text{ where } u = S(x) \text{ and } v = -\frac{e^{-tx}}{t} \\
&= \frac{1}{\mu} \left([uv]_0^\infty - \int_0^\infty vdu \right) \\
&= \frac{1}{\mu} \left(\left[-\frac{e^{-tx} S(x)}{t} \right]_{x=0}^{x\to\infty} - \int_0^\infty \left(-\frac{e^{-tx}}{t} \right) (-f(x))\, dx \right) \\
&= \frac{1}{\mu t} \left(1 - \int_0^\infty e^{-tx} f(x) dx \right) \\
&= \frac{1}{\mu t} (1 - L_F(t))
\end{aligned}$$

as required. ■

Another relationship between the moments of the original and the coderived distributions is:

Proposition 40 *For any SLDFn F*

$$\mu_F^{(n)} < \infty \Rightarrow \mu_{\widetilde{F}[k]} = \frac{\mu_F^{(k+1)}}{(k+1)\,\mu_F^{(k)}} \text{ for } k = 0, 1, 2, ..., n-1.$$

Proof. Note that by Proposition 23

$$\mu_F^{(n)} < \infty \Rightarrow \mu_F^{(k)} < \infty \text{ for } k = 0, 1, 2, ..., n$$

so our assumption is inductive. For $k = 0$ the assertion is

$$\mu_F = \mu_{\widetilde{F}[0]} = \frac{\mu_F^{(1)}}{(1)\,\mu_F^{(0)}} = \frac{\mu_F}{1 \cdot 1}$$

which is vacuous. For $k = 1$ the assertion is just

$$\mu_{\widetilde{F}} = \mu_{\widetilde{F}[1]} = \frac{\mu_F^{(2)}}{2\mu_F^{(1)}}$$

which holds by Proposition 27. We proceed by induction, invoking the case $n-1$ for $G = \widetilde{F}$, which is indeed an SLDFn with $\mu_{\widetilde{F}}^{(n-1)} < \infty$. Invoking Proposition

27 twice more

$$\begin{aligned}
\mu_{\widetilde{F}^{[k]}} &= \mu_{\widetilde{G}^{[k-1]}} \\
&= \frac{\mu_G^{(k)}}{k\mu_G^{(k-1)}} \\
&= \frac{\mu_{\widetilde{F}}^{(k)}}{k\mu_{\widetilde{F}}^{(k-1)}} \\
&= \left(\frac{\mu_F^{(k+1)}}{(k+1)\,\mu_F}\right) \div \left(k\left(\frac{\mu_F^{(k)}}{k\mu_F}\right)\right) \\
&= \frac{\mu_F^{(k+1)}}{(k+1)\,\mu_F^{(k)}}
\end{aligned}$$

completing the induction and the proof. ■

Corollary 41 *For any SLDFn F and positive constant c with $S_F(c) > 0$:*

$$\mu_F^{(n)} < \infty \Rightarrow \frac{S_{\widetilde{F}^{[k]}}(c)}{S_F(c)} = \frac{\mu_{F>c}^{(k)}}{\mu_F^{(k)}} \text{ for } k = 0, 1, 2, ..., n.$$

Proof. The proof is by induction. Case $k = 1$ follows from Item 7 of Proposition 30. Combining Items 7 and 8 of that same Proposition, together with Proposition 40 and the induction hypothesis

$$\begin{aligned}
\frac{S_{\widetilde{F}^{[k+1]}}(c)}{S_F(c)} &= \frac{S_{\widetilde{\widetilde{F}^{[k]}}}(c)}{S_{\widetilde{F}^{[k]}}(c)} \frac{S_{\widetilde{F}^{[k]}}(c)}{S_F(c)} = \frac{\mu_{\left(\widetilde{F}^{[k]}\right)^{>c}}}{\mu_{\widetilde{F}^{[k]}}} \frac{\mu_{F>c}^{(k)}}{\mu_F^{(k)}} \\
&= \frac{\mu_{\widetilde{F>c}^{[k]}}}{\mu_{\widetilde{F}^{[k]}}} \frac{\mu_{F>c}^{(k)}}{\mu_F^{(k)}} = \frac{\frac{\mu_{F>c}^{(k+1)}}{k+1}}{\frac{\mu_F^{(k+1)}}{k+1}} = \frac{\mu_{F>c}^{(k+1)}}{\mu_F^{(k+1)}}
\end{aligned}$$

completing the induction and the proof. ■

The following result will come in handy later when we relate the concept of coderived distribution with ultimate settlement rates and tail "thickness".

Proposition 42 *If F is an SLDFn and $n \in \mathbb{N}$ with $\mu_F^{(n)} < \infty$, then:*

$$\lim_{x \to \omega_F} \frac{f_F(x)}{f_{\widetilde{F}^{[m]}}(x)} = \frac{\tau_F^m \mu_F^{(m)}}{m!} \text{ for } m = 0, 1, 2, ..., n.$$

Proof. Note that

$$x < \omega_F = \omega_{\widetilde{F}^{[m-1]}} \Rightarrow f_{\widetilde{F}^{[m]}}(x) = \frac{S_{\widetilde{F}^{[m-1]}}(x)}{\mu_{\widetilde{F}^{[m-1]}}} > 0$$

assures that we are not dividing by 0. For $m = 0, 1$ we have

$$\lim_{x\to\omega_F} \frac{f_F(x)}{f_{\widetilde{F}^{[0]}}(x)} = \lim_{x\to\omega_F} \frac{f_F(x)}{f_F(x)} = \lim_{x\to\omega_F} 1 = 1 = \mu_F^{(0)} (\tau_F)^0$$

$$\lim_{x\to\omega_F} \frac{f_F(x)}{f_{\widetilde{F}^{[1]}}(x)} = \lim_{x\to\omega_F} \frac{f_F(x)}{f_{\widetilde{F}}(x)} = \lim_{x\to\omega_F} \frac{f_F(x)}{\frac{S_F(x)}{\mu_F}} = \mu_F \lim_{x\to\omega_F} \frac{f_F(x)}{S_F(x)} = \mu_F \lim_{x\to\omega_F} \lambda_F(x) = \mu_F^{(1)} \tau_F$$

and the formula holds for $m = 0, 1$. Proceed by induction on m noting that Proposition 23 assures that our hypothesis is inductive. Invoking the case $m = 1$ and the induction hypothesis

$$\begin{aligned}
\lim_{x\to\omega_F} \frac{f_F(x)}{f_{\widetilde{F}^{[m+1]}}(x)} &= \lim_{x\to\omega_F} \frac{f_F(x)}{f_{\widetilde{F}^{[m]}}(x)} \frac{f_{\widetilde{F}^{[m]}}(x)}{f_{\widetilde{\widetilde{F}^{[m]}}}(x)} \\
&= \lim_{x\to\omega_F} \frac{f_F(x)}{f_{\widetilde{F}^{[m]}}(x)} \lim_{x\to\omega_F} \frac{f_{\widetilde{F}^{[m]}}(x)}{f_{\widetilde{\widetilde{F}^{[m]}}}(x)} \\
&= \frac{\tau_F^m \mu_F^{(m)}}{m!} \mu_{\widetilde{F}^{[m]}} \tau_{\widetilde{F}^{[m]}}
\end{aligned}$$

And then by Propositions 40 and 28

$$\lim_{x\to\omega_F} \frac{f_F(x)}{f_{\widetilde{F}^{(m+1)}}(x)} = \frac{\tau_F^m \mu_F^{(m)}}{m!} \frac{\mu_F^{(m+1)}}{(m+1)\,\mu_F^{(m)}} \tau_F = \frac{\tau_F^{m+1} \mu_F^{(m+1)}}{(m+1)!}$$

completing the induction and the proof. ∎

Corollary 43 *If F is a non-vanishing SLDFn with $0 < \tau_F < \infty$, then:*

$$\text{for every } m, n \in \mathbb{N}, \lim_{x\to\infty} \frac{f_{\widetilde{F}^{[n]}}(x)}{f_{\widetilde{F}^{[m]}}(x)} = \frac{\tau_F^{m-n} n! \mu_F^{(m)}}{m! \mu_F^{(n)}}.$$

Proof. By Proposition 33 $\mu_F^{(k)} < \infty$ for every $k \in \mathbb{N}$ and so the proposition gives

$$\begin{aligned}
\lim_{x\to\infty} \frac{f_{\widetilde{F}^{[n]}}(x)}{f_{\widetilde{F}^{[m]}}(x)} &= \lim_{x\to\infty} \frac{f_{\widetilde{F}^{[n]}}(x)}{f_F(x)} \lim_{x\to\infty} \frac{f_F(x)}{f_{\widetilde{F}^{[m]}}(x)} \\
&= \frac{n!}{\tau_F^n \mu_F^{(n)}} \frac{\tau_F^m \mu_F^{(m)}}{m!} = \frac{\tau_F^{m-n} n! \mu_F^{(m)}}{m! \mu_F^{(n)}}
\end{aligned}$$

as asserted. ∎

Proposition 42 suggests that the series of higher coderived SLDFns $F, \widetilde{F}, \widetilde{F}^{[2]}, \widetilde{F}^{[3]}, \ldots$ share a similar right hand tail behavior and that it may sometimes be viable to approximate the right hand tail of an SLDFn F with that of a higher coderived loss distribution $\widetilde{F}^{[m]}$, adjusted by the scalar $\frac{\tau_F^m \mu_F^{(m)}}{m!}$. Generalizing Proposition 40 yet one step further, we see that all the moments of all coderived distributions are readily obtained from those of F:

28

Proposition 44 *If F is a non-vanishing SLDFn such that for every $n \in \mathbb{N}$ we have $\mu_F^{(n)} < \infty$, then:*

$$\binom{j+k}{k} \mu^{(k)}_{\widetilde{F}^{[j]}} = \frac{\mu_F^{(j+k)}}{\mu_F^{(j)}} \quad \textit{for every } j,k \in \mathbb{N} \cup \{0\}.$$

Proof. The case $j = 0$ is just $\binom{k}{k}\mu_F^{(k)} = \mu_F^{(k)} = \frac{\mu_F^{(k)}}{\mu_F^{(0)}}$ which is certainly true for all integers $k \geq 0$. For the case $j = 1$, Proposition 27 gives

$$\binom{k+1}{k} \mu^{(k)}_{\widetilde{F}^{[1]}} = (k+1)\, \mu^{(k)}_{\widetilde{F}} = (k+1) \left(\frac{\mu_F^{(k+1)}}{(k+1)\, \mu_F} \right) = \frac{\mu_F^{(k+1)}}{\mu_F^{(1)}}$$

and so the result again holds for all integers $k \geq 0$. The proof is by induction on j. Let $G = \widetilde{F}^{[j-1]}$. By Proposition 27

$$\begin{aligned} \mu^{(k)}_{\widetilde{F}^{[j]}} &= \mu^{(k)}_{\widetilde{G}} \\ &= \frac{\mu_G^{(k+1)}}{(k+1)\, \mu_G} \\ &= \frac{\mu^{(k+1)}_{\widetilde{F}^{[j-1]}}}{(k+1)\, \mu_{\widetilde{F}^{[j-1]}}}. \end{aligned}$$

Invoking induction on the numerator and Proposition 40 on the denominator, we have

$$\begin{aligned} \mu^{(k)}_{\widetilde{F}^{[j]}} &= \frac{\frac{\mu_F^{(j-1+k+1)}}{\mu_F^{(j-1)}}}{\binom{j-1+k+1}{k+1} (k+1) \left(\frac{\mu_F^{(j-1+1)}}{(j-1+1)\mu_F^{(j-1)}} \right)} \\ &= \frac{\frac{\mu_F^{(j+k)}}{\mu_F^{(j-1)}}}{\binom{j+k}{k+1} (k+1) \left(\frac{\mu_F^{(j)}}{j\mu_F^{(j-1)}} \right)} \\ &= \frac{(k+1)!(j+k-(k+1))!j\mu_F^{(j+k)}}{(j+k)!\,(k+1)\,\mu_F^{(j)}} \\ &= \frac{(k+1)!\,(j-1)!j\mu_F^{(j+k)}}{(j+k)!(k+1)\mu_F^{(j)}} \\ &= \frac{k!j!\mu_F^{(j+k)}}{(j+k)!\mu_F^{(j)}} = \binom{j+k}{k}^{-1} \frac{\mu_F^{(j+k)}}{\mu_F^{(j)}} \end{aligned}$$

completing the induction and the proof. ■

We next present a series of straightforward results on coderived loss distributions.

Proposition 45 *If F is an SLDFn such that for every $n \in \mathbb{N}$ we have $\mu_F^{(n)} < \infty$, then*

$$\text{for every } n \in \mathbb{N}, \text{ and } t > 0, \quad L_{\widetilde{F}^{[n]}}(t) = \frac{n!}{\mu_F^{(n)}} \sum_{k=0}^{\infty} \frac{(-1)^k \mu_F^{(n+k)} t^k}{(n+k)!}.$$

And if F has a moment generating function, then so does $\widetilde{F}^{[n]}$ with

$$M_{\widetilde{F}^{[n]}}(t) = \frac{n!}{\mu_F^{(n)}} \sum_{k=0}^{\infty} \frac{\mu_F^{(n+k)} t^k}{(n+k)!}.$$

Proof. We need only verify the assertion for the Laplace transform, since that clearly implies the formula for the moment generating function. For $n = 1$ the assertion becomes

$$L_{\widetilde{F}}(t) = \frac{1}{\mu} \sum_{k=0}^{\infty} \frac{(-1)^k \mu_F^{(k+1)} t^k}{(k+1)!} = \frac{-1}{\mu t} \sum_{k=0}^{\infty} \frac{(-1)^{k+1} \mu_F^{(k+1)} t^{k+1}}{(k+1)!}$$

$$= \frac{-1}{\mu t} \sum_{j=1}^{\infty} \frac{(-1)^j \mu_F^{(j)} t^j}{j!} = \frac{-1}{\mu t} \left(\sum_{j=0}^{\infty} \frac{(-1)^j \mu_F^{(j)} t^j}{j!} - 1 \right)$$

$$= \frac{-1}{\mu t} \left(L_F(t) - 1 \right) = \frac{1 - L_F(t)}{\mu t}$$

which is known to hold by Proposition 38. The proof is by induction on n. For $n > 1$ we again have by Proposition 38, induction, and Proposition 40

$$L_{\widetilde{F}^{[n]}}(t) = \frac{1 - L_{\widetilde{F}^{[n-1]}}(t)}{\mu_{\widetilde{F}^{[n-1]}} t}$$

$$= \frac{1}{\mu_{\widetilde{F}^{[n-1]}} t} \left(1 - \frac{(n-1)!}{\mu_F^{(n-1)}} \sum_{k=0}^{\infty} \frac{(-1)^k \mu_F^{(n+k-1)} t^k}{(n+k-1)!} \right)$$

$$= \frac{1}{\frac{\mu_F^{(n-1+1)}}{(n-1+1)\mu_F^{(n-1)}} t} \left(-\frac{(n-1)!}{\mu_F^{(n-1)}} \sum_{k=1}^{\infty} \frac{(-1)^k \mu_F^{(n+k-1)} t^k}{(n+k-1)!} \right)$$

$$= \frac{1}{\frac{\mu_F^{(n)}}{n\mu_F^{(n-1)}} t} \left(-\frac{(n-1)!}{\mu_F^{(n-1)}} \sum_{k=1}^{\infty} \frac{(-1)^k \mu_F^{(n+k-1)} t^k}{(n+k-1)!} \right)$$

$$= \frac{n\mu_F^{(n-1)}}{\mu_F^{(n)}} \left(\frac{(n-1)!}{\mu_F^{(n-1)}} \sum_{k=1}^{\infty} \frac{(-1)^{k-1} \mu_F^{(n+(k-1))} t^{k-1}}{(n+(k-1))!} \right)$$

$$= \frac{n!}{\mu_F^{(n)}} \left(\sum_{j=0}^{\infty} \frac{(-1)^j \mu_F^{(n+j)} t^j}{(n+j)!} \right)$$

completing the induction and the proof. ■

Proposition 46 *For any SLDFn F and for every $n \in \mathbb{N}$:*

1. $\mu_F^{(n)} < \infty \Rightarrow f_{\widetilde{F}^{[n]}}(x) = \frac{n \int_x^\infty (y-x)^{n-1} f_F(y) dy}{\mu_F^{(n)}}$
2. $\mu_F^{(n)} < \infty \Rightarrow S_{\widetilde{F}^{[n]}}(x) = \frac{\int_x^\infty (y-x)^{n} f_F(y) dy}{\mu_F^{(n)}}$
3. $\mu_F^{(n)} < \infty \Rightarrow \lambda_{\widetilde{F}^{[n]}}(x) = \frac{n \int_x^\infty (y-x)^{n-1} f_F(y) dy}{\int_x^\infty (y-x)^{n} f_F(y) dy}$ for every $x < \omega_F$
4. $m, n \in \mathbb{N}, \mu_F^{(n)} < \infty, 0 \leq m \leq n \Rightarrow \frac{d^m S_{\widetilde{F}^{[n]}}}{dx^m} = \frac{(-1)^m n! \mu_F^{(n-m)} S_{\widetilde{F}^{[n-m]}}}{(n-m)! \mu_F^{(n)}}$
5. $(CV_F)^2 = 2\left(\frac{\mu_{\widetilde{F}}}{\mu_F}\right) - 1$
6. $CV_F = 1 \Leftrightarrow \mu_F = \mu_{\widetilde{F}}$
7. $CV_F < 1 \Leftrightarrow \mu_F > \mu_{\widetilde{F}}$
8. $CV_F > 1 \Leftrightarrow \mu_F < \mu_{\widetilde{F}}$

Proof. The proof of Item 1 is by induction on n. For $n = 1$ the assertion reduces to

$$f_{\widetilde{F}}(x) = \frac{\int_x^\infty (y-x)^0 f_F(y) dy}{\mu_F} = \frac{\int_x^\infty f_F(y) dy}{\mu} = \frac{S_F(x)}{\mu}$$

and for $n = 2$ the assertion is

$$f_{\widetilde{F}^{[2]}}(x) = \frac{2 \int_x^\infty (y-x) f_F(y) dy}{\mu_F^{(2)}} = \frac{\frac{1}{\mu} \int_x^\infty (y-x) f_F(y) dy}{\frac{\mu_F^{(2)}}{2\mu}} = \frac{R_F(x)}{\frac{\mu_F^{(2)}}{2\mu}} = \frac{S_{\widetilde{F}}(x)}{\mu_{\widetilde{F}}}$$

and both hold by Proposition 25. Then we have, for $n > 1$, from Propositions 25 and 40, and induction

$$\begin{aligned} f_{\widetilde{F}^{[n]}}(x) &= \frac{S_{\widetilde{F}^{[n-1]}}(x)}{\mu_{\widetilde{F}^{[n-1]}}} \\ &= \frac{\int_x^\infty f_{\widetilde{F}^{[n-1]}}(z) dz}{\frac{\mu_F^{(n)}}{n \mu_F^{(n-1)}}} \end{aligned}$$

$$\begin{aligned} &= \frac{n \mu_F^{(n-1)}}{\mu_F^{(n)}} \int_x^\infty \frac{(n-1) \int_z^\infty (y-z)^{n-2} f_F(y) dy}{\mu_F^{(n-1)}} dz \\ &= \frac{n(n-1)}{\mu_F^{(n)}} \int_x^\infty \int_z^\infty (y-z)^{n-2} f_F(y) dy dz \end{aligned}$$

$$
\begin{aligned}
&= \frac{n(n-1)}{\mu_F^{(n)}} \int_x^\infty \int_x^y (y-z)^{n-2} f_F(y) dz dy \\
&= \frac{n(n-1)}{\mu_F^{(n)}} \int_x^\infty f_F(y) \int_x^y (y-z)^{n-2} dz dy
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{n(n-1)}{\mu_F^{(n)}} \int_x^\infty f_F(y) \left[-\frac{(y-z)^{n-1}}{n-1} \right]_{z=x}^{z=y} dy \\
&= \frac{n(n-1)}{\mu_F^{(n)}} \int_x^\infty f_F(y) \left(\frac{(y-x)^{n-1}}{n-1} \right) dy \\
&= \frac{n \int_x^\infty (y-x)^{n-1} f_F(y) dy}{\mu_F^{(n)}}
\end{aligned}
$$

completing the proof of Item 1. Item 2 follows from Item 1. Indeed, for $n = 1$ the assertion reduces to

$$
S_{\widetilde{F}}(x) = \frac{\int_x^\infty (y-x) f_F(y) dy}{\mu} = R_F(x)
$$

which holds by Proposition 25. Then we have from Item 1

$$
S_{\widetilde{F}^{[n]}}(x) = \int_x^\infty f_{\widetilde{F}^{[n]}}(z) dz
$$

$$
\begin{aligned}
&= \int_x^\infty \left(\frac{n \int_z^\infty (y-z)^{n-1} f_F(y) dy}{\mu_F^{(n)}} \right) dz \\
&= \frac{n}{\mu_F^{(n)}} \int_x^\infty \int_z^\infty (y-z)^{n-1} f_F(y) dy dz \\
&= \frac{n}{\mu_F^{(n)}} \int_x^\infty \int_x^y (y-z)^{n-1} f_F(y) dz dy \\
&= \frac{n}{\mu_F^{(n)}} \int_x^\infty f_F(y) \int_x^y (y-z)^{n-1} dz dy \\
&= \frac{n}{\mu_F^{(n)}} \int_x^\infty f_F(y) \left[-\frac{(y-z)^n}{n} \right]_{z=x}^{z=y} dz dy \\
&= \frac{1}{\mu_F^{(n)}} \int_x^\infty f_F(y) (y-x)^n dy
\end{aligned}
$$

completing the proof of Item 2. Item 3 follows from Items 1 and 2 and the fact that $\omega_{\widetilde{F}^{[n]}} = \omega_F$. Item 4 follows from Item 2 by differentiating under the integral

$$
\frac{d^m S_{\widetilde{F}^{[n]}}}{dx^m}(x) = \frac{d^m \left(\frac{\int_x^\infty (y-x)^n f_F(y) dy}{\mu_F^{(n)}} \right)}{dx^m}
$$

32

$$\begin{aligned}
&= \frac{1}{\mu_F^{(n)}} \left(\int_x^\infty \frac{d^m}{dx^m} \left((y-x)^n f_F(y) \right) dy \right) \\
&= \frac{1}{\mu_F^{(n)}} \left(\int_x^\infty \frac{(-1)^m n!}{(n-m)!} (y-x)^{n-m} f_F(y) dy \right)
\end{aligned}$$

$$\begin{aligned}
&= \frac{(-1)^m n!}{(n-m)!\mu_F^{(n)}} \left(\int_x^\infty (y-x)^{n-m} f_F(y) dy \right) \\
&= \frac{(-1)^m n! \mu_F^{(n-m)}}{(n-m)!\mu_F^{(n)}} \left(\frac{\int_x^\infty (y-x)^{n-m} f_F(y) dy}{\mu_F^{(n-m)}} \right) \\
&= \frac{(-1)^m n! \mu_F^{(n-m)} S_{\widetilde{F}^{[n-m]}}(x)}{(n-m)!\mu_F^{(n)}}
\end{aligned}$$

which establishes Item 4. Note that

$$\begin{aligned}
(CV_F)^2 &= \frac{\mu_F^{(2)} - \mu_F^2}{\mu_F^2} = \frac{\mu_F^{(2)}}{\mu_F^2} - 1 \\
&= \frac{2}{\mu_F} \left(\frac{\mu_F^{(2)}}{2\mu_F} \right) - 1 = \frac{2}{\mu_F} \left(\mu_{\widetilde{F}} \right) - 1
\end{aligned}$$

which establishes Item 5. Since Items 6, 7 and 8 follow immediately from Item 5, this completes the proof. ■

A simple but useful observation is that taking the coderived distribution commutes with a change of scale:

Definition 47 *Let F be a SLDFn and $a > 0$ be any positive constant; the SLDFn F_a is defined as*

$$F_a(x) = F(ax).$$

Proposition 48 *For every $a, c > 0$ and for every $n \in \mathbb{N} \cup \{0\}$:*

1. $\omega_{F_a} = a\omega_F$
2. $S_{F_a}(x) = S_F(ax)$
3. $f_{F_a}(x) = a f_F(ax)$
4. $\lambda_{F_a}(x) = a\lambda_F(ax)$
5. $\mu_{F_a}^{(c)} = \frac{\mu_F^{(c)}}{a^c}$
6. $\widehat{(F_a)}^{[c]} = \left(\widehat{F}^{[c]} \right)_a$
7. $\widetilde{(F_a)}^{[n]} = \left(\widetilde{F}^{[n]} \right)_a$

8. $S(ac) > 0 \Rightarrow (F_a)^{>c} = (F^{>ac})_a$

Proof. Items 1 and 2 are obvious, Item 3 follows from the chain rule

$$f_{F_a}(x) = \frac{dF_a}{dx} = \left(\frac{dF}{dz}|_{z=ax}\right)\frac{dz}{dx} = (f_F(z)|_{z=ax})\, a = af_F(ax).$$

and Item 4 is then an immediate consequence

$$\lambda_{F_a}(x) = \frac{f_{F_a}(x)}{S_{F_a}(x)} = \frac{af_F(ax)}{S_F(ax)} = a\lambda_F(ax).$$

For Item 5, use Item 3 and the change of variable $z = ay$

$$\begin{aligned}
\mu_{F_a}^{(c)} &= \int_0^\infty y^c f_{F_a}(y)dy = \int_0^\infty y^c af_F(ay)dy \\
&= \int_0^\infty \left(\frac{z}{a}\right)^c f_F(z)dz = \frac{\int_0^\infty z^c f_F(z)dz}{a^c} \\
&= \frac{\mu_F^{(c)}}{a^c}.
\end{aligned}$$

Item 6 now follows from

$$\begin{aligned}
f_{\widehat{(F_a)}^{[c]}}(x) &= \frac{x^c f_{F_a}(x)}{\mu_{F_a}^{(c)}} = \frac{x^c af_F(ax)}{\frac{\mu_F^{(c)}}{a^c}} \\
&= a\left(\frac{(ax)^c f_F(ax)}{\mu_F^{(c)}}\right) = af_{\widehat{F}^{[c]}}(ax) \\
&= f_{\left(\widehat{F}^{[c]}\right)_a}(x).
\end{aligned}$$

For item 7, note first that this holds vacuously for $n = 0$

$$\widetilde{(F_a)}^{[0]} = F_a = \left(\widetilde{F}^{[0]}\right)_a$$

and for $n = 1$

$$\begin{aligned}
f_{\widetilde{F_a}}(x) &= \frac{S_{F_a}(x)}{\mu_{F_a}} = \frac{S_F(ax)}{\frac{\mu_F}{a}} = af_{\widetilde{F}}(ax) = f_{\widetilde{F}_a}(x) \\
&\Rightarrow \widetilde{(F_a)}^{[1]} = \widetilde{F_a} = \widetilde{F}_a = \left(\widetilde{F}^{[1]}\right)_a
\end{aligned}$$

and now Item 7 follows by induction

$$\begin{aligned}
\widetilde{(F_a)}^{[n+1]} &= \widetilde{\widetilde{(F_a)}^{[n]}} = \widetilde{\left(\widetilde{F}^{[n]}\right)_a} = \left(\widetilde{\widetilde{F}^{[n]}}\right)_a \\
&= \left(\widetilde{F}^{[n+1]}\right)_a
\end{aligned}$$

34

Finally, from Item 2 we have

$$\begin{aligned} S_{(F_a)^{>c}}(x) &= \frac{S_{F_a}(x+c)}{S_{F_a}(c)} = \frac{S_F(a(x+c))}{S_F(ac)} \\ &= \frac{S_F(ax+ac)}{S_F(ac)} = S_{F^{>ac}}(ax) = S_{(F^{>ac})_a}(x) \end{aligned}$$

completing the proof. ■

Since the coderived distribution relates with the excess ratio, the following result for mixed distributions is no surprise:

Proposition 49 *Given $m \in \mathbb{N}$, SLDFns $F_1, ..., F_m$ all with finite means, and any real weights $w_i \geq 0$ with $1 = \sum_{i=1}^m w_i$, for the weighted mixture SLDFn $F = \sum_{i=1}^m w_i F_i$ with PDF $f_F = \sum_{i=1}^m w_i f_{F_i}$, we have:*

$$\begin{aligned} f_{\widetilde{F}} &= \sum_{i=1}^m u_i f_{\widetilde{F_i}} \text{ and } \widetilde{F} = \sum_{i=1}^m u_i \widetilde{F}_i, \\ \text{where } u_i &= \frac{w_i \mu_{F_i}}{\mu_F} \text{ and } 1 = \sum_{i=1}^m u_i. \end{aligned}$$

Proof. This is again a straightforward verification, from Proposition 25

$$\begin{aligned} f_{\widetilde{F}}(x) &= \frac{S_F(x)}{\mu_F} = \frac{\sum_{i=1}^m w_i S_{F_i}(x)}{\mu_F} \\ &= \frac{\sum_{i=1}^m w_i \mu_{F_i} \left(\frac{S_{F_i}(x)}{\mu_{F_i}}\right)}{\mu_F} = \sum_{i=1}^m u_i f_{\widetilde{F_i}}(x) \\ &\Rightarrow S_{\widetilde{F}}(x) = \sum_{i=1}^m u_i S_{\widetilde{F_i}}(x) \end{aligned}$$

and since clearly

$$\begin{aligned} \mu_F &= \sum_{i=1}^m w_i \mu_{F_i} \\ &\Rightarrow \sum_{i=1}^m u_i = \sum_{i=1}^m \frac{w_i \mu_{F_i}}{\mu_F} = \frac{\sum_{i=1}^m w_i \mu_{F_i}}{\mu_F} = \frac{\mu_F}{\mu_F} = 1 \end{aligned}$$

the result follows from

$$\begin{aligned} \widetilde{F}(x) &= 1 - S_{\widetilde{F}}(x) \\ &= 1 - \sum_{i=1}^m u_i S_{\widetilde{F_i}}(x) = \sum_{i=1}^m u_i - \sum_{i=1}^m u_i S_{\widetilde{F_i}}(x) \\ &= \sum_{i=1}^m u_i \left(1 - S_{\widetilde{F_i}}(x)\right) = \sum_{i=1}^m u_i \widetilde{F}_i(x). \end{aligned}$$

■

So while taking the coderived distribution does not "commute" with constructing a mixture, the coderived distribution of a mixture is nevertheless a mixture of the coderived distributions, but one in which the frequency weights of the original mix are replaced with loss weights for the coderived mix. We will find that this simple observation can prove surprisingly instructive. We will also require the:

Corollary 50 *With the notation and assumptions of Proposition 49*

$$\mu_{F_i} \leq \mu_{\widetilde{F_i}} \;\; 1 \leq i \leq m \Rightarrow \mu_F \leq \mu_{\widetilde{F}} \; .$$

Proof. Proceed by induction on m. Withou tloss of generality we may order so that:

$$\mu_{F_1} \leq \mu_{F_2} \leq \mu_{F_3} \leq ... \leq \mu_{F_m}.$$

The case $m = 1$ is clear. Let G be the mixture of $F_2, ..., F_m$ in which:

$$F_i \text{ has weight } \frac{w_i}{\sum_{i=2}^m w_i}.$$

Then G has PDF

$$f_G = \frac{\sum_{i=2}^m w_i f_{F_i}}{\sum_{i=2}^m w_i} = \frac{\sum_{i=2}^m w_i f_{F_i}}{1 - w_1}$$

and we have:

$$\begin{aligned} \mu_{F_1} &\leq \mu_{F_2} \leq \mu_G \leq \mu_{F_m} \\ &\Rightarrow \mu_G \geq \mu_{F_1}. \end{aligned}$$

Then by induction $\mu_{\widetilde{G}} \geq \mu_G$ and so

$$\begin{aligned} \mu_F &= w_1 \mu_{F_1} + (1 - w_1) \mu_G \\ \mu_{\widetilde{F}} &= \frac{w_1 \mu_{F_1} \mu_{\widetilde{F_1}} + (1 - w_1) \mu_G \mu_{\widetilde{G}}}{\mu_F} \\ &\geq \frac{w_1 \mu_{F_1}^2 + (1 - w_1) \mu_G^2}{\mu_F} \\ &= \left(\frac{w_1 \mu_{F_1}}{\mu_F} \right) \mu_{F_1} + \left(\frac{(1 - w_1) \mu_G}{\mu_F} \right) \mu_G \\ &= (w_1 - \epsilon) \mu_{F_1} + (1 - w_1 + \epsilon) \mu_G \end{aligned}$$

in which

$$\frac{w_1 \mu_{F_1}}{\mu_F} = w_1 - \epsilon$$

36

and we find that

$$\begin{aligned}
\epsilon &= w_1 - \frac{w_1 \mu_{F_1}}{\mu_F} = w_1 \left(1 - \frac{\mu_{F_1}}{\mu_F}\right) \\
&= w_1 \left(\frac{\mu_F - \mu_{F_1}}{\mu_F}\right) = w_1 \left(\frac{w_1 \mu_{F_1} + (1 - w_1)\,\mu_G - \mu_{F_1}}{\mu_F}\right) \\
&= w_1 \left(\frac{(1 - w_1)\left(\mu_G - \mu_{F_1}\right)}{\mu_F}\right) \geq 0
\end{aligned}$$

$$\Rightarrow \epsilon \geq 0.$$

We see that

$$\begin{aligned}
\mu_{\widetilde{F}} &\geq (w_1 - \epsilon)\,\mu_{F_1} + (1 - w_1 + \epsilon)\,\mu_G \\
&= w_1 \mu_{F_1} + (1 - w_1)\,\mu_G + \epsilon\left(\mu_G - \mu_{F_1}\right) \\
&= \mu_F + \epsilon\left(\mu_G - \mu_{F_1}\right) \\
&\geq \mu_F
\end{aligned}$$

completing the proof ■

Corollary 51 *With the notation and assumptions of Proposition 49*

$$CV_{F_i} \geq 1, \quad 1 \leq i \leq m \Rightarrow CV_F \geq 1.$$

Proof. Clear from Corollary 50 and Proposition 46. ■

We next show how the coderived distributions of an SLDFn F "make up a part of the tail" of F. We begin with

Lemma 52 *For any two SLDFns F and G with $w \in [0,1]$ and $c \geq 0$ with $S_F(c) S_G(c) > 0$:*

$$\begin{aligned}
(wF + (1 - w)\,G)^{>c} &= vF^{>c} + (1 - v)\,G^{>c} \\
\textit{where } v &= \frac{w S_F(c)}{w S_F(c) + (1 - w)\,S_G(c)}.
\end{aligned}$$

Proof. We have

$$1 - (wF + (1 - w)\,G)^{>c}(x) = \frac{S_{wF + (1-w)G}(x + c)}{S_{wF + (1-w)G}(c)}$$

$$\begin{aligned}
&= \frac{w S_F(x + c) + (1 - w)\,S_G(x + c)}{w S_F(c) + (1 - w)\,S_G(c)} \\
&= \frac{w S_F(c) \frac{S_F(x+c)}{S_F(c)} + (1 - w)\,S_G(c) \frac{S_G(x+c)}{S_G(c)}}{w S_F(c) + (1 - w)\,S_G(c)} \\
&= v \frac{S_F(x + c)}{S_F(c)} + (1 - v) \frac{S_G(x + c)}{S_G(c)} \\
&= v S_{F^{>c}}(x) + (1 - v) S_{G^{>c}}(x)
\end{aligned}$$

37

and it follows that

$$(wF + (1-w)\,G)^{>c} = 1 - \left(1 - (wF + (1-w)\,G)^{>c}\right)$$

$$\begin{aligned}
&= 1 - (vS_{F^{>c}} + (1-v)S_{G^{>c}}) \\
&= v - vS_{F^{>c}} + (1-v) - (1-v)S_{G^{>c}} \\
&= v\,(1 - S_{F^{>c}}) + (1-v)\,(1 - S_{G^{>c}}) \\
&= vF^{>c} + (1-v)\,G^{>c}
\end{aligned}$$

as asserted. ■

Lemma 53 *For any SLDFn F with finite mean and $0 < \tau_F$, there exists $c \geq 0$ and $w \in (0,1)$ and SLDFn G with $\omega_G = \omega_{F^{>c}}$ and*

$$\begin{aligned}
F^{>c} &= w\left(\widetilde{F}\right)^{>c} + (1-w)\,G \\
&= w\widetilde{F^{>c}} + (1-w)\,G.
\end{aligned}$$

Proof. We have

$$0 < \tau_F = \lim_{x \to \omega_F} \lambda_F(x)$$

which implies that

$$\text{there exist } c, \epsilon \text{ with } \epsilon > 0, 0 \leq c < \omega_F$$

$$\text{and } \{\lambda_{F^{>c}}(x) = \lambda_F(x+c) | x \in (0, \omega_F - c)\} \subset (\epsilon, \infty).$$

Let $w = Min\left(\frac{1}{2}, \epsilon\mu_{F^{>c}}\right)$. Then we have $\mu_{F^{>c}} > 0$ and

$$\frac{w}{\mu_{F^{>c}}} \leq \epsilon$$

$$\Rightarrow \frac{w}{\mu_{F^{>c}}} < \lambda_{F^{>c}}(x) \text{ for every } x \in (0, \omega_F - c)$$

$$\Rightarrow \frac{w}{\mu_{F^{>c}}} < \frac{f_{F^{>c}}(x)}{S_{F^{>c}}(x)} \text{ for every } x \in (0, \omega_F - c)$$

$$\Rightarrow wf_{\widetilde{F^{>c}}}(x) = \frac{wS_{F^{>c}}(x)}{\mu_{F^{>c}}} < f_{F^{>c}}(x) \text{ for every } x \in (0, \omega_F - c)$$

Make the definition

$$g(x) = \frac{f_{F^{>c}}(x) - wf_{\widetilde{F^{>c}}}(x)}{1-w} > 0 \ \text{ for } x \in (0, \omega_F - c)\,,$$

then

$$f_{F^{>c}}(x) = wf_{\widetilde{F^{>c}}}(x) + (1-w)\,g(x) \text{ for every } x \in (0, \omega_F - c)$$

whence g is a C^∞ PDF on $(0, \omega_F - c)$ and the result follows by setting $G(x) = \int\limits_0^x g(z)dz$. ■

Proposition 54 *If $n \in \mathbb{N}$ and F is an SLDFn with $\mu_F^{(n)} < \infty$ and $0 < \tau_F$, then there exist $c \geq 0$, $w \in (0,1)$, and an SLDFn G with*

$$\omega_G = \omega_{F^{>c}} \quad \textit{and}$$

$$F^{>c} = w\left(\widetilde{F}^{[n]}\right)^{>c} + (1-w)\,G = w\widetilde{F^{>c}}^{[n]} + (1-w)\,G.$$

Proof. The proof is by induction on n, the case $n = 1$ being covered by the second lemma. By induction there exists $c_1 \geq 0$ and $w_1 \in (0,1)$ and SLDFn G_1 so that

$$F^{>c_1} = w_1\left(\widetilde{F}^{[n]}\right)^{>c_1} + (1-w_1)\,G_1.$$

Again by the second lemma there exists $c_2 \geq 0$ and $w_2 \in (0,1)$ and SLDFn G_2 so that

$$\left(\widetilde{F}^{[n]}\right)^{>c_2} = w_2\left(\widetilde{\widetilde{F}^{[n]}}\right)^{>c_2} + (1-w_2)\,G_2 = w_2\left(\widetilde{F}^{[n+1]}\right)^{>c_2} + (1-w_2)\,G_2.$$

It now follows from the first lemma that there exist $u, v \in (0,1)$

$$\begin{aligned}
F^{>c_1+c_2} &= \left(F^{>c_1}\right)^{>c_2} = \left(w_1\left(\widetilde{F}^{[n]}\right)^{>c_1} + (1-w_1)\,G_1\right)^{>c_2} = \\
&= u\left(\left(\widetilde{F}^{[n]}\right)^{>c_1}\right)^{>c_2} + (1-u)\,G_1^{>c_2} \\
&= u\left(\widetilde{F}^{[n]}\right)^{>c_1+c_2} + (1-u)\,G_1^{>c_2} \\
&= u\left(\left(\widetilde{F}^{[n]}\right)^{>c_2}\right)^{>c_1} + (1-u)\,G_1^{>c_2}
\end{aligned}$$

$$\begin{aligned}
&= u\left(w_2\left(\widetilde{F}^{[n+1]}\right)^{>c_2} + (1-w_2)\,G_2\right)^{>c_1} + (1-u)\,G_1^{>c_2} \\
&= u\left(v\left(\widetilde{F}^{[n+1]}\right)^{>c_1+c_2} + (1-v)\,G_2^{>c_1}\right) + (1-u)\,G_1^{>c_2} \\
&= uv\left(\widetilde{F}^{[n+1]}\right)^{>c_1+c_2} + (1-uv)G_3
\end{aligned}$$

and setting $w = uv \in (0,1)$ and $c = c_1 + c_2$ completes the induction and the proof. ■

We have seen, Proposition 26, that the coderived distribution determines the original. So it makes sense to ask, given an SLDFn F, is there an SLDFn G (necessarily uniquely determined with finite mean) such that $\widetilde{G} = F$. This prompts:

Definition 55 *Let G be an SLDFn with finite mean and $F = \widetilde{G}$. The SLDFn G is called the **backward coderived loss distribution function** of F. We set,* *recursively*

$$\begin{aligned}\widetilde{F}^{[-1]} &= G \\ \widetilde{F}^{[-n]} &= \widetilde{\widetilde{F}^{[-n+1]}}^{[-1]} \Leftrightarrow \widetilde{\widetilde{F}^{[-n]}} = \widetilde{F}^{[-n+1]} \textit{for } n = 2, 3, 4, \ldots.\end{aligned}$$

If the LDFn $\widetilde{F}^{[-n]}$ exists for some integer $n > 0$, then $\widetilde{F}^{[-n]}$ is called the ***n-th backward coderived loss distribution*** *of F.*

Quite generally, for any loss distribution F with differentiable PDF $f(x)$ such that $\frac{df}{dx} \leq 0$, we could define the **backward coderived loss distribution** to be the distribution with survival function equal to $T(y) = \frac{f(y)}{f(0)}$. Suppose F and G are loss variables with $G = \widetilde{F}^{[-1]}$the backward coderived distribution of F. Of course, the mean of G is

$$\mu_G = \int_0^\infty S_G(y)dy = \int_0^\infty T(y)dy = \frac{\int_0^\infty f_F(y)dy}{f_F(0)} = \frac{1}{f_F(0)}$$

and for the PDF of $\widetilde{G}$ we have, as one would expect

$$\begin{aligned}f_{\widetilde{G}}(y) &= \frac{S_G(y)}{\mu_G} = \frac{\frac{f_F(y)}{f_F(0)}}{\frac{1}{f_F(0)}} = f_F(y) \\ &\Rightarrow \widetilde{G} = F.\end{aligned}$$

References

[1] Abramowitz, M. and Stegun, A.., *Handbook of Mathematical Functions*, Dover, 1972.

[2] Engl, G., Corro, D., "The 2004 NCCI Excess Loss Factors," *CAS* Forum, Fall 2006, pp. 513–571.

[3] Feller, W., *An Introduction to Probability Theory and Its Applications,* Volume II, Wiley, 1966.

[4] Keatinge, C., "Modeling Losses with the Mixed Exponential Distribution," *PCAS* LXXXVI, 1999, pp. 654–698.

[5] Klugman, S., Panjer, H., Willmot, G., *Loss Models: From Data to Decisions,* 2nd edition, Wiley, 2004

Grouping Loss Distributions by Tail Behavior Part II: Continuous Families

Dan Corro
National Council on Compensation Insurance
Spring 2008

Abstract: *This three part paper addresses the task of modelling the right hand tail of a severity distribution. In Part I the excess ratio function is used to define a discrete sequence of loss distributions with related moments and similar tail behavior. Part II extends this to continuous one-parameter families and provides some examples. Part III provides the main result: that under some reasonable conditions, each such family has a limiting distribution which is exponential. The paper then exploits this to 1) group loss distributions based on tail behavior and 2) promote the choice of (mixed) exponentials to model tail behavior.*

Remark 56 *This is Part II of a three part paper. We assume familiarity with Part I and continue our numbering from Part I.*

4 Continuous Families of Distributions

While we introduced taking the coderived distribution as a discrete process, we use Proposition 46 to generalize our definitions:

Definition 57 *For any SLDFn F and positive $c \in \mathbb{R}$ with $\mu_F^{(c)} < \infty$, the* ***c-th coderived loss distribution function*** *of F is the LDFn G with survival function*

$$S_G(x) = \frac{\int_x^\infty (y-x)^c f_F(y)dy}{\mu_F^{(c)}}$$

which we denote as $G = \widetilde{F}^{[c]}$. For $c < 0$, the ***c-th coderived loss distribution function*** *of F is the LDFn G, if such exists, satisfying*

$$\widetilde{G}^{[-c]} = F.$$

The set $\widetilde{F}^{[\mathbb{R}]} = \left\{ \widetilde{F}^{[c]} | c \in \mathbb{R} \text{ such that } \mu_F^{(c)} < \infty \right\}$ is called the ***coderived orbit*** *of the loss distribution function F.*

41

Remark 58 *It follows from this calculation or from Proposition 46 that this agrees with the earlier definition of coderived loss distribution when $c \in \mathbb{Z}$. Indeed, under either definition we trivially have, for any loss SLDFn F,*

$$\begin{aligned}\widetilde{F}^{[0]} &= F \\ \widetilde{F}^{[1]} &= \widetilde{F}\end{aligned}$$

and

$$\widetilde{\widetilde{F}^{[c]}} = \widetilde{F}^{[c+1]} \text{ for every } c \in \mathbb{R}.$$

For any SLDFn F and $c > 0$, the c-th coderived SLDFn $\widetilde{F}^{[c]}$ exists $\Leftrightarrow \mu_F^{(c)} < \infty$. This is consistent with the original construction $S_{\widetilde{F}}(x) = \frac{S_F(x)}{\mu_F}$. Consequently we chose to use the formulation

$$S_{\widetilde{F}^{[c]}}(x) = \frac{\int_x^\infty (y-x)^c f_F(y)dy}{\mu_F^{(c)}}$$

in the definition. For $c < 0$ it is sometimes useful to try the following formula

$$S_{\widetilde{F}^{[c]}}(x) = \lim_{M\to\infty} \frac{\int_x^M (y-x)^c f_F(y)dy}{\int_0^M y^c f_F(y)dy}.$$

For example in the case that F is a mixture of exponentials, $\mu_F^{(c)} < \infty$ only for $c > -1$, but $\widetilde{F}^{[c]}$ exists for every $c \in \mathbb{R}$ and in that special case the latter formula works for every $c \in \mathbb{R}$.

Proposition 59 *For any SLDFn F and positive constants $a, c \in \mathbb{R}$ with $\mu_F^{(c)} < \infty$:*

1. $S_{\widetilde{F}^{[c]}}(x) = \frac{\int_0^\infty z^c f_F(x+z)dz}{\mu_F^{(c)}} = \frac{c\int_x^\infty (y-x)^{c-1} S_F(y)dy}{\mu_F^{(c)}} = \frac{c\int_0^\infty z^{c-1} S_F(x+z)dz}{\mu_F^{(c)}}$
2. $f_{\widetilde{F}^{[c]}}(x) = \frac{c\int_x^\infty (y-x)^{c-1} f_F(y)dy}{\mu_F^{(c)}} = \frac{c\int_0^\infty z^{c-1} f_F(x+z)dz}{\mu_F^{(c)}}$
3. $\lambda_{\widetilde{F}^{[c]}}(x) = \frac{\int_x^\infty (y-x)^{c-1} f_F(y)dy}{\int_x^\infty (y-x)^{c-1} S_F(y)dy} = \frac{\int_0^\infty z^{c-1} f_F(x+z)dz}{\int_0^\infty z^{c-1} S_F(x+z)dz}$

Proof. The substitution $z \mapsto x - y$ will be used routinely to change the lower limit of integration between $y = x$ and $z = 0$. By Proposition 11 we have

$$\int_x^\infty (y-x)^c f_F(y)dy = c\int_x^\infty (y-x)^{c-1} S_F(y)dy$$

and the rest is straightforward calculation. For Item 1

$$\begin{aligned} S_{\widetilde{F}^{[c]}}(x) &= \frac{\int_x^\infty (y-x)^c f_F(y)dy}{\mu_F^{(c)}} \\ &= \frac{c\int_x^\infty (y-x)^{c-1} S_F(y)dy}{\mu_F^{(c)}}. \end{aligned}$$

For Item 2 we differentiate under the integral

$$\begin{aligned}
f_{\widetilde{F}^{[c]}}(x) &= -\frac{dS_{\widetilde{F}^{[c]}}}{dx}(x) \\
&= -\frac{d}{dx}\left(\frac{\int_x^\infty (y-x)^c f_F(y)dy}{\mu_F^{(c)}}\right) \\
&= -\left(\frac{\int_x^\infty \frac{d}{dx}\left((y-x)^c\right) f_F(y)dy}{\mu_F^{(c)}}\right) \\
&= \frac{c\int_x^\infty (y-x)^{c-1} f_F(y)dy}{\mu_F^{(c)}}
\end{aligned}$$

and Item 3 follows from Items 1 and 2. ■

Letting $\mathbf{B}$ denote the beta function; we will make use of the following results from calculus

$$\begin{aligned}
\int_a^b (x-a)^p (b-x)^q \, dx &= (b-a)^{p+q+1}\mathbf{B}(p+1,q+1) \text{ where } p>-1, q>-1 \text{ and } b>a \\
\Gamma(c)\Gamma(1-c) &= \frac{\pi}{\sin c\pi} \text{ where } 0<c<1.
\end{aligned}$$

Proposition 60 *If F is a nonvanishing SLDFn and $c \in (0,1)$, then:*

$$\mu_F^{(c)}\mu_{\widetilde{F}^{[c]}}^{(-c)} = \frac{c\pi}{\sin c\pi} \text{ and } S_F(x) = \frac{\int_x^\infty (y-x)^{-c} f_{\widetilde{F}^{[c]}}(y)dy}{\mu_{\widetilde{F}^{[c]}}^{(-c)}}.$$

Proof. Let $G = \widetilde{F}^{[c]}$,so that $\widetilde{G}^{[-c]} = F$. We have

$$\frac{\int_x^\infty (y-x)^{-c} f_G(y)dy}{\mu_G^{(-c)}} = \frac{\int_x^\infty (y-x)^{-c} f_{\widetilde{F}^{[c]}}(y)dy}{\mu_G^{(-c)}}$$

$$\begin{aligned}
&= \frac{\int_x^\infty (y-x)^{-c}\left(c\int_y^\infty (z-y)^{c-1} f_F(z)dz\right)dy}{\mu_F^{(c)}\mu_G^{(-c)}} \\
&= \frac{c}{\mu_F^{(c)}\mu_G^{(-c)}}\int_x^\infty\int_y^\infty (y-x)^{-c}(z-y)^{c-1} f_F(z)dzdy \\
&= \frac{c}{\mu_F^{(c)}\mu_G^{(-c)}}\int_x^\infty\int_x^z (y-x)^{-c}(z-y)^{c-1} f_F(z)dydz
\end{aligned}$$

Letting $\mathbf{B}$ denote the beta function and noting that $c \in (0,1) \Rightarrow -c > -1$ and $c-1 > -1$

$$\frac{\int_x^\infty (y-x)^{-c} f_G(y)dy}{\mu_G^{(-c)}} = \frac{c}{\mu_F^{(c)}\mu_G^{(-c)}}\int_x^\infty f_F(z)\left(\int_x^z (y-x)^{-c}(z-y)^{c-1}\,dy\right)dz$$

$$
\begin{aligned}
&= \frac{c}{\mu_F^{(c)}\mu_G^{(-c)}} \int_x^\infty f_F(z)\left((z-x)^{-c+(c-1)+1}\mathbf{B}(-c+1,(c-1)+1)\right)dz \\
&= \frac{c\mathbf{B}(1-c,c)}{\mu_F^{(c)}\mu_G^{(-c)}} \int_x^\infty f_F(z)\left((z-x)^0\right)dz \\
&= \frac{c\mathbf{B}(1-c,c)}{\mu_F^{(c)}\mu_G^{(-c)}} \int_x^\infty f_F(z)dz \\
&= \frac{c\frac{\pi}{\sin c\pi}}{\mu_F^{(c)}\mu_G^{(-c)}} \int_x^\infty f_F(z)dz \\
&= \frac{c\pi S_F(x)}{\mu_F^{(c)}\mu_G^{(-c)}\sin c\pi}
\end{aligned}
$$

and letting $x=0$ in the equality, it follows that

$$
\begin{aligned}
1 &= \frac{\int_0^\infty y^{-c} f_G(y)dy}{\mu_G^{(-c)}} = \frac{c\pi S_F(0)}{\mu_F^{(c)}\mu_G^{(-c)}\sin c\pi} = \frac{c\pi}{\mu_F^{(c)}\mu_G^{(-c)}\sin c\pi} \\
&\Rightarrow \mu_F^{(c)}\mu_{\widetilde{F}^{[c]}}^{(-c)} = \mu_F^{(c)}\mu_G^{(-c)} = \frac{c\pi}{\sin c\pi}
\end{aligned}
$$

and further that for every $x \in [0,\infty)$

$$
S_F(x) = S_{\widetilde{G}^{[-c]}}(x) = \frac{\int_x^\infty (y-x)^{-c} f_G(y)dy}{\mu_G^{(-c)}} = \frac{\int_x^\infty (y-x)^{-c} f_{\widetilde{F}^{[c]}}(y)dy}{\mu_{\widetilde{F}^{[c]}}^{(-c)}}
$$

as required. ■

The following result generalizes Proposition 44 and shows that with the exception of instances when the coderived distribution fails to exist, the additive group of reals acts on the set of SLDFns under this definition. This vindicates our use of the term "orbit" and gives credence to the view that this is the "correct" way to extend the definition of coderived variable from discrete to continuous.

Proposition 61 *For any SLDFn F and positive constants $c,d \in \mathbb{R}$ with $\mu_F^{(c+d)} < \infty$, letting $\mathbf{B}$ denote the beta function:*

1. $\mu_{\widetilde{F}^{[c]}}^{(d)} = \frac{(c+d+1)\mathbf{B}(d+1,c+1)\mu_F^{(c+d)}}{\mu_F^{(c)}}$
2. $\widetilde{\widetilde{F}^{[c]}}^{[d]} = \widetilde{F}^{[c+d]}$
3. $\mu_{\widetilde{F}^{[c]}} = \frac{\mu_F^{(c+1)}}{(c+1)\mu_F^{(c)}}$
4. $\left(CV_{\widetilde{F}^{[c]}}\right)^2 = \frac{2(c+1)\mu_F^{(c)}\mu_F^{(c+2)}}{(c+2)\left(\mu_F^{(c+1)}\right)^2} - 1$

44

Proof. Note first that by definition

$$\mu_F^{(c+d)} S_{\widetilde{F}^{[c+d]}}(x) = \int_x^\infty (y-x)^{c+d} f_F(y)dy.$$

On the other hand, we have

$$\mu_{\widetilde{F}^{[c]}}^{(d)} S_{\widetilde{\widetilde{F}^{[c]}}^{[d]}}(x) = \int_x^\infty (y-x)^d f_{\widetilde{F}^{[c]}}(y)dy$$

$$\begin{aligned}
&= \int_x^\infty (y-x)^d \left(\frac{c\int_y^\infty (z-y)^{c-1} f_F(z)dz}{\mu_F^{(c)}} \right) dy \\
&= \frac{c}{\mu_F^{(c)}} \int_x^\infty \int_y^\infty (y-x)^d (z-y)^{c-1} f_F(z)dzdy \\
&= \frac{c}{\mu_F^{(c)}} \int_x^\infty \int_x^z (y-x)^d (z-y)^{c-1} f_F(z)dydz
\end{aligned}$$

$$\begin{aligned}
&= \frac{c}{\mu_F^{(c)}} \int_x^\infty f_F(z) \left(\int_x^z (y-x)^d (z-y)^{c-1} dy \right) dz \\
&= \frac{c}{\mu_F^{(c)}} \int_x^\infty f_F(z) \left((z-x)^{c+d} \mathbf{B}(d+1,c) \right) dz \\
&= \frac{c\mathbf{B}(d+1,c)}{\mu_F^{(c)}} \int_x^\infty (z-x)^{c+d} f_F(z)dz
\end{aligned}$$

$$\begin{aligned}
&= \frac{\Gamma(d+1)c\Gamma(c)}{\mu_F^{(c)}\Gamma(c+d+1)} \int_x^\infty (z-x)^{c+d} f_F(z)dz \\
&= \frac{\Gamma(d+1)\Gamma(c+1)}{\mu_F^{(c)}\Gamma(c+d+1)} \int_x^\infty (z-x)^{c+d} f_F(z)dz
\end{aligned}$$

$$\begin{aligned}
&= \frac{(c+d+1)\Gamma(d+1)\Gamma(c+1)}{\mu_F^{(c)}(c+d+1)\Gamma(c+d+1)} \int_x^\infty (z-x)^{c+d} f_F(z)dz \\
&= \frac{(c+d+1)\Gamma(d+1)\Gamma(c+1)}{\mu_F^{(c)}\Gamma(c+1+d+1)} \int_x^\infty (z-x)^{c+d} f_F(z)dz \\
&= \frac{(c+d+1)\mathbf{B}(d+1,c+1)}{\mu_F^{(c)}} \int_x^\infty (z-x)^{c+d} f_F(z)dz.
\end{aligned}$$

Letting $x=0$ we have

$$\begin{aligned}
\mu_{\widetilde{F}^{[c]}}^{(d)} &= \mu_{\widetilde{F}^{[c]}}^{(d)} S_{\widetilde{\widetilde{F}^{[c]}}^{[d]}}(0) = \frac{(c+d+1)\mathbf{B}(d+1,c+1)}{\mu_F^{(c)}} \int_0^\infty z^{c+d} f_F(z)dz \\
&= \frac{(c+d+1)\mathbf{B}(d+1,c+1)}{\mu_F^{(c)}} \mu_F^{(c+d)}
\end{aligned}$$

45

which proves Item1. For Item 2 the above equations imply

$$\begin{aligned}
\mu^{(d)}_{\widetilde{F}^{[c]}} S_{\widetilde{\widetilde{F}^{[c]}}^{[d]}}(x) &= \frac{(c+d+1)\mathbf{B}(d+1,c+1)}{\mu^{(c)}_F} \int_x^\infty (z-x)^{c+d} f_F(z)dz \\
&= \frac{(c+d+1)\mathbf{B}(d+1,c+1)}{\mu^{(c)}_F} \left(\mu^{(c+d)}_F S_{\widetilde{F}^{[c+d]}}(x)\right)
\end{aligned}$$

which by Item 1 gives

$$\begin{aligned}
S_{\widetilde{\widetilde{F}^{[c]}}^{[d]}}(x) &= \frac{(c+d+1)\mathbf{B}(d+1,c+1)\mu^{(c+d)}_F S_{\widetilde{F}^{[c+d]}}(x)}{\mu^{(d)}_{\widetilde{F}^{[c]}} \mu^{(c)}_F} = S_{\widetilde{F}^{[c+d]}}(x) \\
&\Rightarrow \widetilde{\widetilde{F}^{[c]}}^{[d]} = \widetilde{F}^{[c+d]}.
\end{aligned}$$

And since

$$\begin{aligned}
\mathbf{B}(2,c+1) &= \frac{\Gamma(2)\Gamma(c+1)}{\Gamma(c+3)} = \frac{\Gamma(c+1)}{(c+2)\Gamma(c+2)} \\
&= \frac{\Gamma(c+1)}{(c+2)\,(c+1)\,\Gamma(c+1)} = \frac{1}{(c+2)\,(c+1)}
\end{aligned}$$

we see that Item 3 is just the case $d = 1$ of Item 1

$$\begin{aligned}
\mu_{\widetilde{F}^{[c]}} &= \mu^{(1)}_{\widetilde{F}^{[c]}} = \frac{(c+2)\mathbf{B}(2,c+1)\mu^{(c+1)}_F}{\mu^{(c)}_F} \\
&= \frac{(c+2)\mu^{(c+1)}_F}{(c+2)\,(c+1)\,\mu^{(c)}_F} = \frac{\mu^{(c+1)}_F}{(c+1)\,\mu^{(c)}_F}.
\end{aligned}$$

Finally, we have by Proposition 46 and Part 3

$$\begin{aligned}
\left(CV_{\widetilde{F}^{[c]}}\right)^2 &= 2\left(\frac{\mu_{\widetilde{\widetilde{F}^{[c]}}}}{\mu_{\widetilde{F}^{[c]}}}\right) - 1 = 2\left(\frac{\mu_{\widetilde{F}^{[c+1]}}}{\mu_{\widetilde{F}^{[c]}}}\right) - 1 \\
&= 2\left(\frac{\frac{\mu^{(c+2)}_F}{(c+2)\mu^{(c+1)}_F}}{\frac{\mu^{(c+1)}_F}{(c+1)\mu^{(c)}_F}}\right) - 1 \\
&= \frac{2\,(c+1)\,\mu^{(c)}_F\,\mu^{(c+2)}_F}{(c+2)\left(\mu^{(c+1)}_F\right)^2} - 1
\end{aligned}$$

and the proof is complete. ■

Proposition 62 *For any non-vanishing loss SLDFn F and positive constant c with $\mu^{(c)}_F < \infty$:*

$$\tau_F = \tau_{\widetilde{F}^{[c]}}.$$

46

Proof. We have from l'Hôpital

$$\lim_{x\to\infty}\frac{S_F(x)}{S_{\widetilde{F}^{[c]}}(x)} = \lim_{x\to\infty}\frac{-f_F(x)}{-f_{\widetilde{F}^{[c]}}(x)} = \lim_{x\to\infty}\frac{f_F(x)}{f_{\widetilde{F}^{[c]}}(x)}$$

whence

$$\begin{aligned} 1 &= \left(\lim_{x\to\infty}\frac{S_F(x)}{S_{\widetilde{F}^{[c]}}(x)}\right)\left(\lim_{x\to\infty}\frac{f_F(x)}{f_{\widetilde{F}^{[c]}}(x)}\right)^{-1} \\ &= \left(\lim_{x\to\infty}\frac{S_F(x)}{S_{\widetilde{F}^{[c]}}(x)}\right)\left(\lim_{x\to\infty}\left(\frac{f_F(x)}{f_{\widetilde{F}^{[c]}}(x)}\right)^{-1}\right) \\ &= \left(\lim_{x\to\infty}\frac{S_F(x)}{S_{\widetilde{F}^{[c]}}(x)}\right)\left(\lim_{x\to\infty}\frac{f_{\widetilde{F}^{[c]}}(x)}{f_F(x)}\right) \end{aligned}$$

$$\begin{aligned} &= \lim_{x\to\infty}\frac{S_F(x)}{S_{\widetilde{F}^{[c]}}(x)}\frac{f_{\widetilde{F}^{[c]}}(x)}{f_F(x)} \\ &= \lim_{x\to\infty}\frac{S_F(x)}{f_F(x)}\frac{f_{\widetilde{F}^{[c]}}(x)}{S_{\widetilde{F}^{[c]}}(x)} \\ &= \lim_{x\to\infty}\frac{\lambda_{\widetilde{F}^{[c]}}(x)}{\lambda_F(x)} \end{aligned}$$

$$\Rightarrow \tau_F = \lim_{x\to\infty}\lambda_F(x) = \lim_{x\to\infty}\lambda_{\widetilde{F}^{[c]}}(x) = \tau_{\widetilde{F}^{[c]}}$$

as required. ■

The relation

$$F \sim G \quad \Leftrightarrow \quad \text{there exists } c \in \mathbb{R} \text{ such that } G = \widetilde{F}^{[c]} \Leftrightarrow G \in \widetilde{F}^{[\mathbb{R}]}$$

defines an equivalence relation on the class of SLDFns

$$F = \widetilde{F}^{(0)} \Rightarrow F \sim F$$

$$F \sim G \Rightarrow \text{there exists } c \in \mathbb{R} \text{ such that } G = \widetilde{F}^{[c]} \Rightarrow F = \widetilde{G}^{[-c]} \Rightarrow G \sim F$$

$$F \sim G, G \sim H \Rightarrow \text{there exist } c, d \in \mathbb{R} \text{ such that } G = \widetilde{F}^{[c]}, H = \widetilde{G}^{[d]}$$

$$\Rightarrow H = \widetilde{G}^{[d]} = \widetilde{\left(\widetilde{F}^{[c]}\right)}^{d} = \widetilde{F}^{[c+d]} \Rightarrow F \sim H.$$

We just observed that the real-valued mapping $F \mapsto \tau_F$ is constant on equivalence classes, i.e., orbits. In this regard we make the:

Definition 63 *For any SLDFn F and real number c set*

$$\Gamma_F(c) = \lim_{x\to\omega_F}\frac{\int_x^\infty (y-x)^{c-1} f_F(y)dy}{S_F(x)}.$$

Remark 64 *Note that for the exponential distribution $F(x) = 1 - e^{-x}$ we have*

$$\begin{aligned}
\Gamma_F(c) &= \lim_{x \to \omega_F} \frac{\int_x^\infty (y-x)^{c-1} e^{-y} dy}{e^{-x}} = \lim_{x \to \infty} \int_x^\infty (y-x)^{c-1} e^{-(y-x)} dy \\
&= \lim_{x \to \infty} \int_0^\infty z^{c-1} e^{-z} dz = \lim_{x \to \infty} \Gamma(c) = \Gamma(c)
\end{aligned}$$

which helped prompt the choice of notation.

Proposition 65 *For any SLDFn F and positive constants $a, c \in \mathbb{R}$ with $\mu_F^{(c)} < \infty$:*

1. $L_{\widetilde{F}^{[c]}}(t) = \frac{c\mu_F^{(c-1)}}{t\mu_F^{(c)}} \left(1 - L_{\widetilde{F}^{[c-1]}}(t)\right)$
2. $\mu_{\widetilde{F}^{[c-1]}} = \frac{\mu_F^{(c)}}{c\mu_F^{(c-1)}}$
3. $\tau_{F_a} = a\tau_F$
4. $\Gamma_{F_a}(c) = a^{1-c}\Gamma_F(c)$.

Proof. For Item 1

$$\begin{aligned}
L_{\widetilde{F}^{[c]}}(t) &= E_{\widetilde{F}^{[c]}}\left[e^{-tX}\right] \\
&= \int_0^\infty e^{-tx} f_{\widetilde{F}^{[c]}}(x) dx \\
&= \int_0^\infty e^{-tx} \left(\frac{c \int_x^\infty (y-x)^{c-1} f_F(y) dy}{\mu_F^{(c)}} \right) dx \\
&= \frac{c}{\mu_F^{(c)}} \int_0^\infty e^{-tx} \left(\int_x^\infty (y-x)^{c-1} f_F(y) dy \right) dx \\
&= \frac{c}{\mu_F^{(c)}} \int_0^\infty u dv
\end{aligned}$$

where

$$\begin{aligned}
u &= \int_x^\infty (y-x)^{c-1} f_F(y) dy = \mu_F^{(c-1)} S_{\widetilde{F}^{[c-1]}} \text{ and } v = -\frac{e^{-tx}}{t} \\
\text{and so } du &= -\mu_F^{(c-1)} f_{\widetilde{F}^{[c-1]}}.
\end{aligned}$$

We have

$$L_{\widetilde{F}^{[c]}}(t) = \frac{c}{\mu_F^{(c)}} \left([uv]_0^\infty - \int_0^\infty v du \right)$$

$$
\begin{aligned}
&= \frac{c}{\mu_F^{(c)}} \left(\left[-\frac{e^{-tx}}{t} \int_x^\infty (y-x)^{c-1} f_F(y) dy \right]_0^\infty - \frac{\mu_F^{(c-1)}}{t} \int_0^\infty e^{-tx} f_{\widetilde{F}^{[c-1]}}(x) dx \right) \\
&= \frac{c}{\mu_F^{(c)}} \left(\frac{1}{t} \int_0^\infty y^{c-1} f_F(y) dy - \frac{\mu_F^{(c-1)}}{t} L_{\widetilde{F}^{[c-1]}}(t) \right) \\
&= \frac{c}{\mu_F^{(c)}} \left(\frac{\mu_F^{(c-1)}}{t} - \frac{\mu_F^{(c-1)}}{t} L_{\widetilde{F}^{[c-1]}}(t) \right) \\
&= \frac{c\mu_F^{(c-1)}}{t\mu_F^{(c)}} \left(1 - L_{\widetilde{F}^{[c-1]}}(t)\right)
\end{aligned}
$$

For Item 2, invoke Item 1 and Proposition 39 applied to the LDFn $\widetilde{F}^{[c-1]}$, noting that for any LDFn G, $0 < L_G(1) < 1$

$$
\begin{aligned}
L_{\widetilde{F}}(t) &= \frac{1 - L_F(t)}{\mu t} \quad \text{for } t > 0 \\
\frac{c\mu_F^{(c-1)}}{t\mu_F^{(c)}} \left(1 - L_{\widetilde{F}^{[c-1]}}(t)\right) &= L_{\widetilde{F}^{[c]}}(t) = \frac{1}{t\mu_{\widetilde{F}^{[c-1]}}} \left(1 - L_{\widetilde{F}^{[c-1]}}(t)\right)
\end{aligned}
$$

$$
\begin{aligned}
&\Rightarrow \quad \frac{c\mu_F^{(c-1)}}{\mu_F^{(c)}} = \frac{1}{\mu_{\widetilde{F}^{[c-1]}}} \\
&\Rightarrow \quad \mu_{\widetilde{F}^{[c-1]}} = \frac{\mu_F^{(c)}}{c\mu_F^{(c-1)}}.
\end{aligned}
$$

Item 3 follows from Proposition 48

$$
\tau_{F_a} = \lim_{x \to \omega_{F_a}} \lambda_{F_a}(x) = \lim_{x \to a\omega_F} a\lambda_F(ax) = a \lim_{x \to \omega_F} \lambda_F(x) = a\tau_F.
$$

And for Item 4

$$
\Gamma_{F_a}(c) = \lim_{x \to \omega_{F_a}} \frac{\int_x^{\omega_{F_a}} (y-x)^{c-1} f_{F_a}(y) dy}{S_{F_a}(x)}
$$

$$
\begin{aligned}
&= \lim_{x \to a\omega_F} \frac{\int_x^{a\omega_F} (y-x)^{c-1} a f_F(ay)\, dy}{S_F(ax)} \\
&= \lim_{x \to a\omega_F} \frac{a^{1-c} \int_x^{a\omega_F} (ay-ax)^{c-1} f_F(ay)\, a dy}{S_F(ax)} \\
&= a^{1-c} \lim_{ax \to \omega_F} \frac{\int_{ax}^{\omega_F} (z-ax)^{c-1} f_F(z) dz}{S_F(z)} \\
&= a^{1-c} \lim_{y \to \omega_F} \frac{\int_y^{\omega_F} (z-y)^{c-1} f_F(z) dz}{S_F(z)} \\
&= a^{1-c} \Gamma_F(c)
\end{aligned}
$$

as required. ■

49

Proposition 66 *If F is an SLDFn with $0 < \tau_F < \infty$ and a any positive constant, then:*

$$F_a \in \widetilde{F}^{(\mathbb{R})} \quad \Leftrightarrow \quad a = 1.$$

Proof. The $\Leftarrow$ direction is trivial. For $\Rightarrow$

$$\begin{aligned} F_a &\in \widetilde{F}^{(\mathbb{R})} \Rightarrow \tau_F = \tau_{F_a} = a\tau_F \\ 0 &< \tau_F < \infty \Rightarrow a = 1 \end{aligned}$$

as required. ■

What really prompted the notation are Items 5 and 6 of the following:

Proposition 67 *For any non-vanishing SLDFn F with finite mean and $c \in \mathbb{R}$ with $\mu_F^{(c)} < \infty$:*

1. $\Gamma_F(1) = 1$
2. $\tau_F > 0 \Rightarrow \Gamma_F(2) = \frac{1}{\tau_F}$
3. $\Gamma_F(c) = \lim\limits_{x\to\infty} \frac{\int_0^\infty z^{c-1} f_F(z+x)dz}{S_F(x)}$
4. $\Gamma_{\widetilde{F}}(c) = \Gamma_F(c)$
5. $\tau_F \Gamma_F(c+1) = c\Gamma_F(c)$
6. $\tau_F > 0$ and $c \in \mathbb{Z} \Rightarrow \Gamma_F(c) = \tau_F^{1-c}\Gamma(c)$

Proof. We clearly have

$$\Gamma_F(1) = \lim_{x\to\omega_F} \frac{\int_x^{\omega_F} (y-x)^0 f_F(y)dy}{S_F(x)} = \lim_{x\to\omega_F} \frac{\int_x^{\omega_F} f_F(y)dy}{S_F(x)} = \lim_{x\to\omega_F} \frac{S_F(x)}{S_F(x)} = \lim_{x\to\omega_F} 1 = 1$$

verifying Item 1. When $\tau_F > 0$, we have from l'Hôpital and Proposition 22

$$\begin{aligned} \Gamma_F(2) &= \lim_{x\to\omega_F} \frac{\int_x^{\omega_F} (y-x)^1 f_F(y)dy}{S_F(x)} = \lim_{x\to\omega_F} \frac{\mu_F R_F(x)}{S_F(x)} \\ &= \mu_F \lim_{x\to\omega_F} \frac{R_F(x)}{S_F(x)} = \mu_F \lim_{x\to\omega_F} \frac{\frac{-S_F(x)}{\mu_F}}{-f_F(x)} \\ &= \lim_{x\to\omega_F} \frac{S_F(x)}{f_F(x)} = \lim_{x\to\omega_F} \frac{1}{\lambda_F(x)} = \frac{1}{\tau_F} \end{aligned}$$

proving Item 2. For Item 3, just use the change of variable $z = y - x$. For Item 4, we have, using Item 3

$$\begin{aligned} \Gamma_{\widetilde{F}}(c) &= \lim_{x\to\omega_F} \frac{\int_0^{\omega_F} z^{c-1} f_{\widetilde{F}}(z+x)dz}{S_{\widetilde{F}}(x)} \\ &= \lim_{x\to\omega_F} \frac{\int_0^{\omega_F} z^{c-1} \frac{S_F(z+x)}{\mu_F} dz}{R_F(x)} \end{aligned}$$

$$\begin{aligned}
&= \frac{1}{\mu_F}\lim_{x\to\omega_F}\frac{\int_0^{\omega_F} z^{c-1}S_F(z+x)dz}{R_F(x)} \\
&= \frac{1}{\mu_F}\lim_{x\to\omega_F}\frac{\frac{d}{dx}\int_0^{\omega_F} z^{c-1}S_F(z+x)dz}{\frac{d}{dx}R_F(x)} \\
&= \frac{1}{\mu_F}\lim_{x\to\omega_F}\frac{\int_0^{\omega_F} z^{c-1}\frac{d}{dx}\left(S_F(z+x)\right)dz}{\frac{-S_F(x)}{\mu_F}}
\end{aligned}$$

$$\begin{aligned}
&= \lim_{x\to\infty}\frac{-\int_0^{\infty} z^{c-1}f_F(z+x)\frac{d(z+x)}{dx}dz}{-S_F(x)} \\
&= \lim_{x\to\omega_F}\frac{\int_0^{\omega_F} z^{c-1}f_F(z+x)dz}{S_F(x)} \\
&= \Gamma_F(c)
\end{aligned}$$

which establishes Item 4. For Item 5 we have

$$\begin{aligned}
\tau_F(c)\Gamma_F(c+1) &= \tau_{\widetilde{F}^{[c]}}\Gamma_F(c+1) \\
&= \lim_{x\to\omega_F}\lambda_{\widetilde{F}^{[c]}}(x)\Gamma_F(c+1) \\
&= \lim_{x\to\omega_F}\frac{f_{\widetilde{F}^{[c]}}(x)}{S_{\widetilde{F}^{[c]}}(x)}\Gamma_F(c+1)
\end{aligned}$$

$$\begin{aligned}
&= \lim_{x\to\omega_F}\frac{c\int_x^{\omega_F}(y-x)^{c-1}f_F(y)dy}{\int_x^{\omega_F}(y-x)^{c}f_F(y)dy}\lim_{x\to\omega_F}\frac{\int_x^{\omega_F}(y-x)^{c}f_F(y)dy}{S_F(x)} \\
&= c\lim_{x\to\omega_F}\frac{\int_x^{\omega_F}(y-x)^{c-1}f_F(y)dy}{\int_x^{\omega_F}(y-x)^{c}f_F(y)dy}\frac{\int_x^{\omega_F}(y-x)^{c}f_F(y)dy}{S_F(x)} \\
&= c\lim_{x\to\omega_F}\frac{\int_x^{\omega_F}(y-x)^{c-1}f_F(y)dy}{S_F(x)} = c\Gamma_F(c).
\end{aligned}$$

And finally, for Item 6 note that the formula holds for $c = 1$ and $c = 2$. by Items 1 and 2. Define $\Gamma^*(c) = \tau_F^{c-1}\Gamma_F(c)$,then by Item 5

$$\begin{aligned}
\Gamma^*(c+1) &= \tau_F^{c+1-1}\Gamma_F(c+1) \\
&= \tau_F^{c}\frac{c\Gamma_F(c)}{\tau_F} = c\tau_F^{c-1}\Gamma_F(c) \\
&= c\Gamma^*(c)
\end{aligned}$$

and so Γ^* and Γ satisfy the same recurrence formula and agree on 1 and 2, whence $\Gamma^* = \Gamma$ on $\mathbb{Z}$, as required. ■

Corollary 68 *If F and G are SLDFns with finite means and $\tau_F\tau_G > 0$, then*

$$\Gamma_F(c) = \Gamma_G(c) \text{ for every } c\in\mathbb{Z} \qquad \Leftrightarrow \qquad \tau_F = \tau_G.$$

Corollary 69 *If F is an SLDFn with finite mean and $\tau_F > 0$, then*

$$\tau_F = 1 \Leftrightarrow \Gamma_F(n) = n! \text{ for every } n \in \mathbb{N}.$$

Proposition 70 *If F is a non-vanishing SLDFn, then for any $c > 1$ such that $\mu_F^{(c-1)} < \infty$, we have:*

$$\Gamma_F(c) = (c-1) \lim_{x \to \omega_F} \int_0^{\omega_F} z^{c-2} \left(\frac{S_F(x+z)}{S_F(x)} \right) dz.$$

Proof. By Proposition 11 we have

$$\int_x^{\omega_F} (y-x)^{c-1} f_F(y) dy = (c-1) \int_x^{\omega_F} (y-x)^{c-2} S_F(y) dy$$

from which we find that

$$\begin{aligned}
\Gamma_F(c) &= \lim_{x \to \omega_F} \frac{\int_x^{\omega_F} (y-x)^{c-1} f_F(y) dy}{S_F(x)} \\
&= \lim_{x \to \omega_F} \frac{(c-1) \int_x^{\omega_F} (y-x)^{c-2} S_F(y) dy}{S_F(x)} \\
&= (c-1) \lim_{x \to \omega_F} \int_x^{\omega_F} (y-x)^{c-2} \left(\frac{S_F(y)}{S_F(x)} \right) dy \\
&= (c-1) \lim_{x \to \omega_F} \int_0^{\omega_F} z^{c-2} \left(\frac{S_F(x+z)}{S_F(x)} \right) dz
\end{aligned}$$

as required. ■

Proposition 71 *If F is a non-vanishing SLDFn with $\tau_F > 0$ and is such that for every $c > 0$ we have $\mu_F^{(c)} < \infty$, then:*

$$\lim_{x \to \infty} \frac{f_{\widetilde{F}^{[c]}}(x)}{f_F(x)} = \frac{\Gamma_F(c+1)}{\mu_F^{(c)}} \quad \text{for every } c > 0.$$

Proof. We have

$$f_{\widetilde{F}^{[c]}}(x) = \frac{c \int_x^{\infty} (y-x)^{c-1} f_F(y) dy}{\mu_F^{(c)}}$$

which implies that

$$\begin{aligned}
\frac{f_{\widetilde{F}^{[c]}}(x)}{f_F(x)} &= \frac{\frac{c \int_x^{\infty} (y-x)^{c-1} f_F(y) dy}{\mu_F^{(c)}}}{f_F(x)} \\
&= \frac{\frac{c \int_x^{\infty} (y-x)^{c-1} f_F(y) dy}{S_F(x)}}{\mu_F^{(c)} \frac{f_F(x)}{S_F(x)}} \\
&= \frac{\frac{c \int_x^{\infty} (y-x)^{c-1} f_F(y) dy}{S_F(x)}}{\mu_F^{(c)} \lambda_F(x)}
\end{aligned}$$

and recalling the definition $\Gamma_F(c) = \lim\limits_{x\to\infty} \frac{\int_x^\infty (y-x)^{c-1} f_F(y)dy}{S_F(x)}$, we find from Proposition 67 that

$$\begin{aligned}
\lim_{x\to\infty} \frac{f_{\widetilde{F}^{(c)}}(x)}{f_F(x)} &= \frac{c}{\mu_F^{(c)}} \lim_{x\to\infty} \frac{\frac{\int_x^\infty (y-x)^{c-1} f_F(y)dy}{S_F(x)}}{\lambda_F(x)} \\
&= \frac{c}{\mu_F^{(c)}} \frac{\lim\limits_{x\to\infty} \frac{\int_x^\infty (y-x)^{c-1} f_F(y)dy}{S_F(x)}}{\lim\limits_{x\to\infty} \lambda_F(x)} \\
&= \frac{c\Gamma_F(c)}{\mu_F^{(c)}\tau_F} = \frac{\tau_F \Gamma_F(c+1)}{\mu_F^{(c)}\tau_F} = \frac{\Gamma_F(c+1)}{\mu_F^{(c)}}
\end{aligned}$$

as required. ■

Proposition 72 *If F is an SLDFn with finite mean, then there exist unique $a, b, c \in \mathbb{R} \cup \{\infty\}$, $a \le b \le 0, 1 \le c$ such that:*

$$\begin{aligned}
(a, c) &= \left\{ x \in \mathbb{R} - \{a, c\} \mid \textit{there exists SLDFn } G \textit{ such that } G = \widetilde{F}^{[x]} \right\} \\
(b, c) &= \left\{ x \in \mathbb{R} - \{b, c\} \mid \mu_F^{(x)} < \infty \right\}.
\end{aligned}$$

Proof. It is clear from the above that both sets are connected subsets of $\mathbb{R}$ containing $(0, 1)$ and that they share a right hand endpoint c. It is also clear from what has been shown that $a \le b$. The rest follows from Proposition 23. ■

Proposition 73 *For any SLDFns F and G with $\mu_F^{(n)}, \mu_G^{(n)} < \infty$ for every $n \in \mathbb{N}$ and $\tau_F \tau_G > 0$, letting $\mathbf{B}$ denote the beta function:*

$$G = \widetilde{F_a}^{[c]} \textit{for some positive constants } a, c \in \mathbb{R}$$

$$\Leftrightarrow$$

$$a = \frac{\tau_G}{\tau_F} \textit{ and there exists } c > 0 \textit{ such that } \tau_G^k \mu_G^{(k)} = \frac{\tau_F^k (c+k+1)\mathbf{B}(k+1, c+1)\mu_F^{(c+k)}}{\mu_F^{(c)}} \textit{ for every } k \in \mathbb{N}.$$

Proof. Suppose that $G = \widetilde{F_a}^{[c]}$ for some positive constants $a, c \in \mathbb{R}$, then

$$\begin{aligned}
G &= \left(\widetilde{F_a}\right)^{[c]} = \left(\widetilde{F}^{[c]}\right)_a \\
&\Rightarrow \tau_G = \tau_{\widetilde{F_a}^{[c]}} = \tau_{\left(\widetilde{F}^{[c]}\right)_a} = a\tau_{\widetilde{F}^{[c]}} = a\tau_F \\
&\Rightarrow a = \frac{\tau_G}{\tau_F}
\end{aligned}$$

and we have for every $k \in \mathbb{N}$

$$\mu_G^{(k)} = \mu_{\widetilde{F_a}^{[c]}}^{(k)} = \frac{(c+k+1)\mathbf{B}(k+1, c+1)\mu_{F_a}^{(c+k)}}{\mu_{F_a}^{(c)}}$$

$$\begin{aligned}
&= \frac{a^c(c+k+1)\mathbf{B}(k+1,c+1)\mu_F^{(c+k)}}{a^{c+k}\mu_F^{(c)}} \\
&= \frac{(c+k+1)\mathbf{B}(k+1,c+1)\mu_F^{(c+k)}}{a^k\mu_F^{(c)}} \\
&= \frac{(c+k+1)\mathbf{B}(k+1,c+1)\mu_F^{(c+k)}}{\left(\frac{\tau_G}{\tau_F}\right)^k \mu_F^{(c)}}
\end{aligned}$$

$$\Rightarrow \tau_G^k\mu_G^{(k)} = \frac{\tau_F^k(c+k+1)\mathbf{B}(k+1,c+1)\mu_F^{(c+k)}}{\mu_F^{(c)}} \text{ for every } k \in \mathbb{N}$$

which establishes the $\Rightarrow$ direction. Conversely, letting $a = \frac{\tau_G}{\tau_F} > 0$ those same equations imply that

$$\tau_G^k\mu_G^{(k)} = \frac{\tau_F^k(c+k+1)\mathbf{B}(k+1,c+1)\mu_F^{(c+k)}}{\mu_F^{(c)}} \text{ for every } k \in \mathbb{N}$$

$$\begin{aligned}
&\Rightarrow \quad \mu_G^{(k)} = \mu_{\widetilde{F_a}^{[c]}}^{(k)} \text{ for every } k \in \mathbb{N} \\
&\Rightarrow \quad L_G = L_{\widetilde{F_a}^{[c]}} \Rightarrow G = \widetilde{F_a}^{[c]} = \widetilde{F}_a^{[c]}
\end{aligned}$$

and the proof is complete. ■

This suggests that one way to decompose the set of all SLDFns F with $0 < \tau_F < \infty$ is into disjoint $\sim$invariant subsets of "coordinated half planes" of the form

$$(0,\infty)\widetilde{F}^{[\mathbb{R}]} = \left\{(a,c) \longleftrightarrow \widetilde{F_a}^{[c]} | a \in (0,\infty), c \in \mathbb{R}\right\}.$$

Such a plane is akin to an orbit under the affine-like action of the direct product $(0,\infty) \times \mathbb{R}$ of the multiplicative group of positive reals by the additive group of reals (subgroup of a Borel subgroup of $SL_2(\mathbb{R})$). The above Proposition provides one approach for determining when two SLDFns "lie on the same plane". Note that while there are infinitely many equations to check, mathematical induction should often apply to make this doable. Also, you may need to swap roles of F and G to deal with the possibility of $c < 0$. From knowledge of moments $\mu_F^{(c)}$ as c varies for some empirical data, the above formulas show how to pick (a,c) to match the first two moments (first solve for c to match the CV

$$\frac{2\,(c+1)\,\mu_F^{(c)}\mu_F^{(c+2)}}{(c+2)\left(\mu_F^{(c+1)}\right)^2} = \left(CV_{\widetilde{F}^{[c]}}\right)^2 + 1 = \frac{\mu_{\widetilde{F}^{[c]}}^{(2)}}{\left(\mu_{\widetilde{F}^{[c]}}^{(1)}\right)^2}$$

and then determine a as the scalar adjustment to match the mean). We will soon see how to quantify the difference in the thickness of the tail between any two elements of such a plane. We will see that for loss variables F and G in different planes, we need only be able to compare one pair of representatives from the two planes to be able to compare any two elements in the union of the two planes, including, of course, F and G.

The real-valued mapping $F \mapsto \tau_F$ defined on the set of SLDFns is constant on equivalence classes, i.e., orbits. The main result of this paper is to specify the possible structures for $\widetilde{F}^{[\mathbb{R}]}$ as they relate with the ultimate settlement rate τ_F and other metrics for the "thickness" of the tail, as that concept is defined later. This part of the paper concludes with some examples. In the next and final part we will see that the structure of $\widetilde{F}^{(n)}$ becomes more "monotone", "smooth", and "tail-like" as n increases and make mathematically precise what that statement means.

5 Examples

This section presents some simple examples.

Example 74 *Uniform density: let F be uniformly distributed on the finite interval $[a,b]$ where $0 \leq a < b$. The following are well-known and readily verified*

$$F(x) = \left\{ \begin{array}{cc} 0 & x \leq a \\ \frac{x-a}{b-a} & a \leq x \leq b \\ 1 & b \leq x \end{array} \right\}$$

$$f_F(x) = \left\{ \begin{array}{cc} 0 & x < a \\ \frac{1}{b-a} & a < x < b \\ 0 & b < x \end{array} \right\}$$

$$S_F(x) = \left\{ \begin{array}{cc} 1 & x \leq a \\ \frac{b-x}{b-a} & a \leq x \leq b \\ 0 & b \leq x \end{array} \right\}$$

$$\lambda_F(x) = \left\{ \begin{array}{cc} 0 & x < a \\ \frac{1}{b-x} & a < x < b \\ \infty & x \geq b \end{array} \right\}$$

$$\begin{aligned} \alpha_F &= a \qquad \omega_F = b \\ \mu_F &= \frac{b+a}{2} \\ M_F(t) &= \frac{e^{bt} - e^{at}}{(b-a)\,t} \qquad t > 0 \end{aligned}$$

Propositions 25, 38, and 46 lead to

$$f_{\widetilde{F}}(x) = \left\{ \begin{array}{cc} \frac{2}{a+b} & x < a \\ \frac{2(b-x)}{b^2-a^2} & a \leq x \leq b \\ 0 & x > b \end{array} \right\}$$

$$S_{\widetilde{F}}(x) = \left\{ \begin{array}{cc} \frac{a+b-2x}{a+b} & x \leq a \\ \frac{(b-x)^2}{b^2-a^2} & a \leq x \leq b \\ 0 & x \geq b \end{array} \right\}$$

$$M_{\widetilde{F}}(t) = 2\frac{e^{bt} - e^{at} - (b-a)t}{(b^2 - a^2)\,t^2} \qquad t > 0$$

Observe that

$$M_F(t) = \frac{e^{bt} - e^{at}}{(b-a)\,t} = \frac{\sum_{k=0}^{\infty} \frac{(bt)^k - (at)^k}{k!}}{(b-a)\,t}$$

$$\begin{aligned} &= \sum_{k=0}^{\infty} \frac{\left(b^k - a^k\right) t^{k-1}}{(b-a)\,k!} \\ &= \sum_{k=1}^{\infty} \frac{\left(\frac{b^{(k-1)+1} - a^{(k-1)+1}}{((k-1)+1)(b-a)}\right) t^{k-1}}{(k-1)!} \\ &= \sum_{k=0}^{\infty} \frac{\left(\frac{b^{k+1} - a^{k+1}}{(k+1)(b-a)}\right) t^k}{k!} \end{aligned}$$

$$\Rightarrow \mu_F^{(k)} = \frac{b^{k+1} - a^{k+1}}{(k+1)\,(b-a)} = \frac{\sum_{j=0}^{k} b^j a^{k-j}}{k+1}$$

And so from Propositions 27 and 40

$$\begin{aligned} \mu_{\widetilde{F}}^{(k)} &= \frac{\mu_F^{(k+1)}}{(k+1)\,\mu_F} = \frac{\frac{b^{k+2} - a^{k+2}}{(k+2)(b-a)}}{(k+1)\left(\frac{b+a}{2}\right)} \\ &= 2\frac{b^{k+2} - a^{k+2}}{(k+2)\,(k+1)\,(b^2 - a^2)} \end{aligned}$$

$$\mu_{\widetilde{F}[k]} = \frac{\mu_F^{(k+1)}}{(k+1)\,\mu_F^{(k)}} = \frac{\frac{b^{k+2} - a^{k+2}}{(k+2)(b-a)}}{(k+1)\left(\frac{b^{k+1} - a^{k+1}}{(k+1)(b-a)}\right)} = \frac{b^{k+2} - a^{k+2}}{(k+2)\,(b^{k+1} - a^{k+1})}.$$

In particular

$$\mu_{\widetilde{F}} = \frac{b^3 - a^3}{3\,(b^2 - a^2)}.$$

56

Notice that for $a > 0$

$$\begin{aligned}
\mu_{\widetilde{F}^{[k]}} &= \frac{b^{k+2} - a^{k+2}}{(k+2)\left(b^{k+1} - a^{k+1}\right)} \\
&= \frac{b\left(\frac{b}{a}\right)^{k+1} - a}{(k+2)\left(\left(\frac{b}{a}\right)^{k+1} - 1\right)} \\
&= \frac{b\left(\left(\frac{b}{a}\right)^{k+1} - 1\right) + b - a}{(k+2)\left(\left(\frac{b}{a}\right)^{k+1} - 1\right)} \\
&= \frac{b}{k+2} + \frac{b-a}{(k+2)\left(\left(\frac{b}{a}\right)^{k+1} - 1\right)} \\
&\Rightarrow \lim_{k\to\infty} \mu_{\widetilde{F}^{[k]}} = 0
\end{aligned}$$

and we see that as k increases, the LDFns $\widetilde{F}^{[k]}$ become concentrated at the value 0. On the other hand, if b is the maximum loss, then $S_{\widetilde{F}^{[k]}}(b-\epsilon) > 0$ for every $k \in \mathbb{N}$, $\epsilon > 0$. More generally we have:

Example 75 *Consider the case when the PDF has finite support and is bounded away from 0, i.e., there exist $a, b, \alpha, \beta \in \mathbb{R}$ with $0 \le a < b$ and $0 < \alpha \le \beta$ such that $f_F(x) = 0$ for $x \notin [a,b]$ and $\alpha \le f_F(x) \le \beta$ $x \in [a,b]$. In this case we have, for $c > 0$*

$$\frac{\alpha\left(b^{c+1} - a^{c+1}\right)}{c+1} = \alpha \int_a^b x^c dx \le \mu_F^{(c)} \le \beta \int_a^b x^c dx = \frac{\beta\left(b^{c+1} - a^{c+1}\right)}{c+1}$$

$$\begin{aligned}
\Rightarrow \mu_{\widetilde{F}^{[c]}} &= \frac{\mu_F^{(c+1)}}{(c+1)\,\mu_F^{(c)}} \\
&\le \frac{\frac{\beta\left(b^{c+2} - a^{c+2}\right)}{c+2}}{(c+1)\left(\frac{\alpha\left(b^{c+1} - a^{c+1}\right)}{c+1}\right)} \\
&= \frac{\beta}{\alpha\,(c+2)} \frac{b^{c+2} - a^{c+2}}{b^{c+1} - a^{c+1}} \\
&= \frac{\beta}{\alpha\,(c+2)} \frac{a+b}{a+b} \frac{b^{c+2} - a^{c+2}}{b^{c+1} - a^{c+1}} \\
&= \frac{\beta\,(a+b)}{\alpha\,(c+2)} \frac{b^{c+2} - a^{c+2}}{b^{c+2} - a^{c+2} + ab(b^c - a^c)} \\
&< \frac{\beta\,(a+b)}{\alpha\,(c+2)} \frac{b^{c+2} - a^{c+2}}{b^{c+2} - a^{c+2}} \\
&= \frac{\beta\,(a+b)}{\alpha\,(c+2)}
\end{aligned}$$

57

$$\Rightarrow \lim_{c\to\infty} \mu_{\widetilde{F}^{[c]}} = 0$$

and again, as one would expect from the uniform density example, we see that as c increases the LDFns $\widetilde{F}^{[c]}$ *become concentrated at the value* 0.

Example 76 *Exponential Distribution: Let F have an exponential density with mean* $\mu > 0$. *We have:*

$$\mu_F = \sigma_F = \mu \quad CV_F = 1 \quad \alpha_F = 0, \quad \omega_F = \infty, \quad \lambda_F = \frac{1}{\mu} = \tau_F$$

$$\mu_F^{(k)} = \mu^k k! \quad L_F(t) = \frac{1}{1+\mu t}, \; t > -\frac{1}{\mu}$$

$$f_{\widetilde{F}}(x) = \frac{S_F(x)}{\mu} = \left(\frac{1}{\mu}\right) e^{-\frac{x}{\mu}} = f_F(x)$$

$$\Rightarrow \widetilde{F} = F \Rightarrow \widetilde{F}^{[n]} = F \quad \textit{for every } n \in \mathbb{N}.$$

The converse also holds

$$\begin{aligned} \widetilde{F} &= F \\ &\Rightarrow \widetilde{(F_\mu)} = \widetilde{F}_\mu = F_\mu. \end{aligned}$$

Letting $G = F_\mu$, *we have* $G = \widetilde{G}$ *and* $\mu_G = \mu_{F_\mu} = \frac{\mu_F}{\mu} = 1$. *Define* $g(x) = S_G(-x)$ *for* $x < 0$, *then*

$$\begin{aligned} \frac{dg}{dx} &= \frac{d\left(S_G(-x)\right)}{dx} = \frac{dS_G}{d(-x)}\frac{d(-x)}{dx} = \left(-f_G\left(-x\right)\right)(-1) \\ &= f_G\left(-x\right) = f_{\widetilde{G}}\left(-x\right) = \frac{S_G(-x)}{\mu_G} = S_G(-x) = g(x) \end{aligned}$$

$$\frac{dg}{dx} = g(x), \;\; g(0) = 1 \Rightarrow g(x) = e^x$$

$$\Rightarrow S_G(-y) = g(y) = e^y$$

$$\Rightarrow S_F(x) = S_{G_{\frac{1}{\mu}}}(x) = S_G\left(\frac{x}{\mu}\right) = g\left(-\frac{x}{\mu}\right) = e^{-\frac{x}{\mu}}$$

and F has an exponential density with mean $\mu > 0$. *More generally we have for* $c > 0$

$$S_F(x) = e^{-\frac{x}{\mu}}$$

$$\Rightarrow S_{\widetilde{F}^{[c]}}(x) = \frac{\int_x^\infty (y-x)^c f_F(y)dy}{\mu_F^{(c)}}$$

58

$$\begin{aligned}
&= \frac{\int_0^\infty z^c f_F(z+x)dz}{\mu_F^{(c)}} = \frac{\int_0^\infty z^c \frac{e^{-\frac{z+x}{\mu}}}{\mu}dz}{\mu_F^{(c)}} \\
&= \frac{e^{-\frac{x}{\mu}} \int_0^\infty z^c \frac{e^{-\frac{z}{\mu}}}{\mu}dz}{\mu_F^{(c)}} = \frac{e^{-\frac{x}{\mu}} \mu_F^{(c)}}{\mu_F^{(c)}} \\
&= e^{-\frac{x}{\mu}} = S_F(x)
\end{aligned}$$

$$\Rightarrow \widetilde{F}^{[c]} = F.$$

We have established

$$S_F(x) = e^{-\frac{x}{\mu}} \Leftrightarrow \widetilde{F}^{[\mathbb{R}]} = \{F\}.$$

Note too that for any $c \geq 0$

$$F^{>c}(x) = 1 - \frac{S_F(x+c)}{S_F(c)} = 1 - \frac{e^{-\frac{x+c}{\mu}}}{e^{-\frac{c}{\mu}}} = 1 - \frac{e^{-\frac{c}{\mu}} e^{-\frac{x}{\mu}}}{e^{-\frac{c}{\mu}}} = 1 - e^{-\frac{x}{\mu}} = F(x)$$

$$\Rightarrow F^{>c} = F$$

with the converse again being true, i.e., this too characterizes the exponential distribution. Indeed for any SLDFn G :

$$G^{>c} = G \quad \textit{for every } c \geq 0$$

$$\Rightarrow \omega_G = \omega_{G^{>1}} = \omega_G - 1 \Rightarrow \omega_G = \infty$$

$$\Rightarrow 1 - S_G(x) = G(x) = G^{>c}(x) = 1 - \frac{S_G(x+c)}{S_G(c)} \quad \textit{for every } x, c \geq 0$$

$$\Rightarrow S_G(x) = \frac{S_G(x+c)}{S_G(c)} \quad \textit{for every } x, c \geq 0$$

$$\Rightarrow S_G(x+y) = S_G(x)S_G(y) \quad \textit{for every } x, y \geq 0$$

$$\Rightarrow f_G(x+y) = -\frac{dS_G(x+y)}{dy} = -\frac{dS_G(x)S_G(y)}{dy}$$

$$= S_G(x)f_G(y) + S_G(y) \cdot 0 = S_G(x)f_G(y) \quad \textit{for every } x, y \geq 0$$

$$\Rightarrow f_G(x) = S_G(x)f_G(0)$$

$$\Rightarrow 1 = \int_0^\infty f_G(x)dx = f_G(0)\int_0^\infty S_G(x)dx = f_G(0)\mu_G$$

$$f_G(0) = S_G(x)f_G(0) = \frac{1}{\mu_G}$$

$$\Rightarrow f_G(x) = S_G(x)f_G(0) = \frac{S_G(x)}{\mu_G} = f_{\widetilde{G}}(x)$$

$$\Rightarrow G = \widetilde{G} \Rightarrow G = 1 - e^{-\frac{x}{\mu_G}}.$$

59

Finally, we have as noted above

$$\begin{aligned}
\Gamma_F(c) &= \lim_{x\to\infty} \frac{\int_x^\infty (y-x)^{c-1} \frac{e^{-\frac{y}{\mu}}}{\mu} dy}{e^{-\frac{x}{\mu}}} \\
&= \lim_{x\to\infty} \int_x^\infty \mu^{1-c} \left(\frac{y-x}{\mu}\right)^{c-1} e^{-\left(\frac{y-x}{\mu}\right)} \frac{dy}{\mu} \\
&= \lim_{x\to\infty} \mu^{1-c} \int_0^\infty z^{c-1} e^{-z} dz \\
&= \lim_{x\to\infty} \mu^{1-c}\Gamma(c) = \mu^{1-c}\Gamma(c).
\end{aligned}$$

As a consequence of this example:

Proposition 77 *Suppose F is an SLDFn with $(\alpha_F, \omega_F) = (0, \infty)$, $\mu_F^{(k)} < M < \infty$ for every $k \in \mathbb{N}$, and for which $\widetilde{F}^{[\infty]} = \lim_{n\to\infty} \widetilde{F}^{[n]}$ exists as a pointwise limit function and where the convergence is uniform on $(0, \infty)$. Then for all $x > 0$:*

$$\widetilde{F}^{[\infty]}(x) = 1 - e^{-\tau_F x}.$$

Proof. Let $G = \lim_{n\to\infty} \widetilde{F}^{[n]}$, uniform convergence implies that G is an SLDFn with finite mean $\mu_G \leq M < \infty$. We have

$$\tau_G = \lim_{n\to\infty} \tau_{\widetilde{F}^{[n]}} = \lim_{n\to\infty} \tau_F = \tau_F$$

$$\text{and } \widetilde{G} = \widetilde{\lim_{n\to\infty} \widetilde{F}^{[n]}} = \lim_{n\to\infty} \widetilde{\widetilde{F}^{[n]}} = \lim_{n\to\infty} \widetilde{F}^{[n+1]} = \lim_{n\to\infty} \widetilde{F}^{[n]} = G.$$

And so by the exponential example

$$\begin{aligned}
G(x) &= 1 - e^{-\frac{x}{\mu_G}} \\
&\Rightarrow \tau_F = \tau_G = \frac{1}{\mu_G} \\
&\Rightarrow G(x) = 1 - e^{-\tau_F x}
\end{aligned}$$

as required. ■

In practice, one would expect that far enough into the tail of a distribution the hazard function λ_F would be bounded and stabilized at least to being either nonincreasing or nondecreasing. And in that event, the hazard functions of the higher coderived distributions $\widetilde{F}^{[n]}$ are squeezed to the constant τ_F. Accordingly, when $\tau_F > 0$, as n increases we would expect the $\widetilde{F}^{[n]}$ to converge to the exponential density of mean $\mu = \frac{1}{\tau_F}$. This points toward a special role for the exponential density when fitting the tail of a loss distribution. More precisely, we have:

Proposition 78 *Suppose F is an SLDFn with λ_F either nonincreasing or nondecreasing and with $0 < \tau_F < \infty$. Then for any $x > 0$:*

$$\widetilde{F}^{[\infty]}(x) = \lim_{n\to\infty} \widetilde{F}^{[n]}(x) = 1 - e^{-\tau_F x}.$$

Proof. By Proposition 33, the assumption $0 < \tau_F < \infty$ assures all moments are finite. Since λ_F is either nondecreasing or nonincreasing, Propositions 31 and 32 imply that either $\lambda \leq \widetilde{\lambda} \leq \widetilde{\widetilde{\lambda}} \leq$ ...or $\lambda \geq \widetilde{\lambda} \geq \widetilde{\widetilde{\lambda}} \geq$ .... In either event, we have pointwise convergence on $(0, \omega_F) = (0, \infty)$ of the hazard function sequence $\lambda, \widetilde{\lambda}, \widetilde{\widetilde{\lambda}}, ...$ to a $\sim$invariant hazard function. As in the proof of the previous Proposition, this entails uniform convergence of $\lambda, \widetilde{\lambda}, \widetilde{\widetilde{\lambda}}, ...$ to the constant τ_F on the interval $[x, \infty)$, which then entails that $\widetilde{F}^{[n]}(y)$ converge to $1 - e^{-\tau_F y}$ as $n \to \infty$ for any $y \geq x$, and the result follows. ∎

Similarly, recall from Proposition 37 that $\lim_{x,c \to \omega_F} \frac{S_{\widetilde{F^{>c}}}(x)}{S_{F^{>c}}(x)} = 1$ where F is any SLDFn with finite mean and $0 < \tau_F < \infty$. Then the idea is again that the "far tail" of F is captured as $G = F^{>c}$ for c large and where we have

$$\lim_{x \to \omega_G} \frac{S_{\widetilde{G}}(x)}{S_G(x)} = 1$$

$$\Rightarrow G \approx \widetilde{G} \approx 1 - e^{-\tau_G x} = 1 - e^{-\tau_{(F^{>c})} x} = 1 - e^{-\tau_F x}.$$

This suggests that quite generally, when $0 < \tau_F, \mu_F < \infty$, the exponential density of mean $\frac{1}{\tau_F}$ appears as a natural way to model the structure of the far tail of the distribution. We will see in the next section that analytic properties of exponentials, and more generally mixed exponentials, again make them a natural choice for modeling tail behavior. This strengthens the theoretical justification for the methodology used to fit tails when calculating ELFs in [2] which also derives some general formulas for splicing tails on loss distributions.

Example 79 *Pareto Density: Let $F = \Pi(\alpha, \theta)$ have the Pareto density with parameters α and θ :*

$$F(x) = \Pi(\alpha, \theta; x) = 1 - \left(\frac{\theta}{x + \theta}\right)^{\alpha} \qquad \omega_F = \infty$$

$$S_F(x) = \left(\frac{\theta}{x + \theta}\right)^{\alpha}$$

$$f_F(x) = \frac{\alpha \theta^{\alpha}}{(x + \theta)^{\alpha+1}} \qquad \lambda_F(x) = \frac{\frac{\alpha\theta^{\alpha}}{(x+\theta)^{\alpha+1}}}{\left(\frac{\theta}{x+\theta}\right)^{\alpha}} = \frac{\alpha}{x + \theta} \qquad \tau_F = 0$$

$$k \in \mathbb{N} \text{ and } k < \alpha \Rightarrow \mu_F^{(k)} = \frac{\theta^k k!}{(\alpha - 1) \cdots (\alpha - k)}$$

$$\begin{aligned} F_a(x) &= F(ax) = 1 - \left(\frac{\theta}{ax + \theta}\right)^{\alpha} \\ &= 1 - \left(\frac{\frac{\theta}{a}}{x + \frac{\theta}{a}}\right)^{\alpha} = \Pi(\alpha, \frac{\theta}{a}; x) \end{aligned}$$

$$
\begin{aligned}
F^{>c}(x) &= 1-\frac{S_F(x+c)}{S_F(c)} = 1-\frac{\left(\frac{\theta}{x+c+\theta}\right)^{\alpha}}{\left(\frac{\theta}{c+\theta}\right)^{\alpha}} \\
&= 1-\left(\frac{\theta+c}{x+(\theta+c)}\right)^{\alpha} = \Pi(\alpha,\theta+c;x)
\end{aligned}
$$

$$
\Rightarrow \Pi(\alpha,\theta)^{>c} = \Pi(\alpha,\theta+c)
$$

$$
f_{\widetilde{F}}(x) = \frac{\left(\frac{\theta}{x+\theta}\right)^{\alpha}}{\frac{\theta}{(\alpha-1)}} = \frac{(\alpha-1)\,\theta^{\alpha-1}}{(x+\theta)^{(\alpha-1)+1}} = f_{\Pi(\alpha-1,\theta)}(x)
$$

$$
\Rightarrow \widetilde{\Pi}(\alpha,\theta) = \Pi(\alpha-1,\theta)
$$

$$
\Rightarrow \text{ for every } k\in\mathbb{Z},\ \widetilde{\Pi}^{[k]}(\alpha,\theta) = \left\{\begin{array}{cc} \Pi(\alpha-k,\theta) & k<\alpha \\ \nexists & k\geq\alpha \end{array}\right\}.
$$

More generally we have for $c>0$

$$
\begin{aligned}
S_{\widetilde{F}^{[c]}}(x) &= \frac{\int_x^{\infty}(y-x)^c f_F(y)dy}{\int_0^{\infty} y^c f_F(y)dy} \\
&= \frac{\int_0^{\infty} z^c f_f(z+x)dz}{\int_0^{\infty} y^c f_F(y)dy} \\
&= \frac{\int_0^{\infty} z^c \frac{\alpha\theta^{\alpha}}{(z+x+\theta)^{\alpha+1}}dz}{\int_0^{\infty} y^c \frac{\alpha\theta^{\alpha}}{(y+\theta)^{\alpha+1}}dy}
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{\left(\frac{\theta^{\alpha}}{(x+\theta)^{\alpha}}\right)\int_0^{\infty} z^c \frac{\alpha(x+\theta)^{\alpha}}{(z+x+\theta)^{\alpha+1}}dz}{\frac{\theta^c\Gamma(c+1)\Gamma(\alpha-c)}{\Gamma(\alpha)}} \\
&= \frac{\left(\frac{\theta}{x+\theta}\right)^{\alpha}\frac{(x+\theta)^c\Gamma(c+1)\Gamma(\alpha-c)}{\Gamma(\alpha)}}{\frac{\theta^c\Gamma(c+1)\Gamma(\alpha-c)}{\Gamma(\alpha)}} = \left(\frac{\theta}{x+\theta}\right)^{\alpha-c}
\end{aligned}
$$

$$
\Rightarrow \text{ for every } c\in\mathbb{R},\ \widetilde{\Pi}^{(c)}(\alpha,\theta) = \left\{\begin{array}{cc} \Pi(\alpha-c,\theta) & c<\alpha \\ \textit{Does not exist} & c\geq\alpha \end{array}\right\}
$$

and we see that in this case the natural parametrization of the orbit $\widetilde{F}^{[\mathbb{R}]}$ *relates linearly with the* α *parameter of the usual arithmetic formula and with an orbit corresponding to a fixed value of the* θ *parameter*

$$
\begin{aligned}
\widetilde{F}^{[\mathbb{R}]} &= \left\{\widetilde{F}^{[r]} | r\in\mathbb{R} \text{ such that } \mu_F^{(r)}<\infty\right\} \\
&= \{\Pi(\alpha-r,\theta)|r\in(0,\alpha)\} \\
&= \{\Pi(s,\theta)|s>0\}.
\end{aligned}
$$

62

Note too that for $a > 0$

$$F(x) = \Pi(\alpha, \theta; x)$$

$$\begin{aligned}
\Rightarrow \quad F_a(x) = F(ax) &= 1 - \left(\frac{\theta}{ax+\theta}\right)^{\alpha} \\
&= 1 - \left(\frac{\frac{\theta}{a}}{x+\frac{\theta}{a}}\right)^{\alpha} = \Pi\left(\alpha, \frac{\theta}{a}; x\right)
\end{aligned}$$

$$\Rightarrow \Pi(\alpha,\theta)_a = \Pi\left(\alpha, \frac{\theta}{a}\right)$$

and we see that the "half plane" $(0,\infty)\widetilde{F}^{[\mathbb{R}]}$ $\sim$*invariant subset here more resembles a "quadrant" and corresponds to the Pareto density "family" of distributions*

$$(0,\infty)\widetilde{\Pi(\alpha,\theta)}^{[\mathbb{R}]} = (0,\infty)\widetilde{F}^{[\mathbb{R}]} = \{\Pi(s,t)|s>0, t>0\}.$$

Example 80 *Lognormal Density: Let* $F = \Lambda(\mu,\sigma)$ *have the Lognormal density with* $F(x) = \Phi\left(\frac{\ln x - \mu}{\mu}\right)$*. In this case:*

$$\omega_F = \infty, \tau_F = 0 \text{ and } \mu_F^{(n)} = e^{n\mu + \frac{n^2\sigma^2}{2}} < \infty \text{ for every } n \in \mathbb{N}$$

$$\Rightarrow \widetilde{F}^{[n]} \text{ exists for every } n \in \mathbb{N}.$$

We see from Proposition 40 that

$$\mu_{\widetilde{F}^{[n]}} = \frac{\mu_F^{(n+1)}}{(n+1)\,\mu_F^{(n)}} = \frac{e^{(n+1)\mu + \frac{(n+1)^2\sigma^2}{2}}}{(n+1)\,e^{n\mu+\frac{n^2\sigma^2}{2}}} = \frac{e^{\mu+\frac{(2n+1)\sigma^2}{2}}}{n+1}$$

$$\Rightarrow \lim_{n\to\infty} \mu_{\widetilde{F}^{[n]}} = 0.$$

Also, the mode of F *is* $e^{\mu-\sigma^2} > 0 \Rightarrow \nexists\ \widetilde{F}^{[-1]}$.

Perhaps the most useful example for the practical application of these ideas is:

Example 81 *Mixed Exponential Distribution: Let* F *be a mixture of exponential densities. More precisely, for some* m, $1 \le m \le \infty$, *and for any real weights* $w_i > 0$ *with* $1 = \sum_{i=1}^m w_i$ *and parameters* $\mu_i > 0$ *ordered so that* $\mu_i < \mu_{i+1}$ *and with* $\sum_{i=1}^m w_i\mu_i < \infty$*. Then consider the weighted mixture SLDFn variable* $F = \daleth(m, \langle w_i\rangle, \langle\mu_i\rangle)$

$$\begin{aligned}
\daleth(m, \langle w_i\rangle, \langle\mu_i\rangle; x) &= F(x) = 1 - \sum_{i=1}^m w_i e^{-\frac{x}{\mu_i}}. \\
S_F(x) &= \sum_{i=1}^m w_i e^{-\frac{x}{\mu_i}}.
\end{aligned}$$

Then we have, by Proposition 49

$$\begin{aligned}\mu_F &= \sum_{i=1}^{m} w_i \mu_i < \infty \\ S_{\widetilde{F}}(x) &= \frac{\sum_{i=1}^{m} w_i \mu_i e^{-\frac{x}{\mu_i}}}{\mu_F}\end{aligned}$$

and generally

$$S_{\widetilde{F}^{[n]}}(x) = \frac{\sum_{i=1}^{m} w_i \mu_i^n e^{-\frac{x}{\mu_i}}}{\sum_{i=1}^{m} w_i \mu_i^n} \text{ for every } n \in \mathbb{Z} \text{ such that } \sum_{i=1}^{m} w_i \mu_i^n < \infty.$$

Proposition 9 can be used to verify that λ_F is decreasing provided $m > 1$. Indeed

$$m > 1 \Rightarrow$$

$$\frac{1}{\mu_1} = \frac{1}{\mu_1}\frac{S_F(x)}{S_F(x)} = \frac{1}{\mu_1}\frac{\sum_{i=1}^{m} w_i e^{-\frac{x}{\mu_i}}}{S_F(x)}$$

$$= \frac{\sum_{i=1}^{m} \frac{w_i}{\mu_1} e^{-\frac{x}{\mu_i}}}{S_F(x)} > \frac{\sum_{i=1}^{m} \frac{w_i}{\mu_i} e^{-\frac{x}{\mu_i}}}{S_F(x)} = \frac{f_F(x)}{S_F(x)} = \lambda_F(x)$$

$$> \frac{\sum_{i=1}^{m} \frac{w_i}{\mu_m} e^{-\frac{x}{\mu_i}}}{S_F(x)} = \frac{1}{\mu_m}\frac{\sum_{i=1}^{m} w_i e^{-\frac{x}{\mu_i}}}{S_F(x)} = \frac{1}{\mu_m}\frac{S_F(x)}{S_F(x)} = \frac{1}{\mu_m}$$

$$\Rightarrow \frac{1}{\mu_1} > \lambda_F(x) > \frac{1}{\mu_m}$$

and similarly we find that

$$\frac{-\frac{df}{dx}}{S(x)} = \frac{\sum_{i=1}^{m} \frac{w_i}{\mu_i^2} e^{-\frac{x}{\mu_i}}}{S(x)} > \frac{\frac{1}{\mu_1}\sum_{i=1}^{m} \frac{w_i}{\mu_i} e^{-\frac{x}{\mu_i}}}{S(x)} = \frac{1}{\mu_1}\frac{f(x)}{S(x)} = \frac{1}{\mu_1}\lambda(x) > \lambda(x)^2$$

$$\Rightarrow 0 > \lambda(x)^2 + \frac{\frac{df}{dx}}{S(x)} = \frac{d\lambda}{dx}$$

$$\Rightarrow \lambda_F \text{ is decreasing}$$

as asserted. Note also that Corollary 51 implies that the $CV_F \geq 1$ for any mixed exponential. In fact

$$m > 1 \Rightarrow CV_F > 1 \Rightarrow \mu_{\widetilde{F}^{[n]}} < \mu_{\widetilde{F}^{[n+1]}} \text{ for every } n \in \mathbb{Z}.$$

When $1 < m < \infty$ we clearly have

$$\sum_{i=1}^{m} w_i \mu_i^n < \infty \text{ for every } n \in \mathbb{Z}$$

and it follows that $\widetilde{F}^{[\mathbb{R}]} \cong \mathbb{R}$ as ordered sets, with no first or last element. From Proposition 49 we see that

$$\lim_{n\to\infty} \mu_{\widetilde{F}^{[n]}} = \mu_m \text{ and } \lim_{n\to\infty} \mu_{\widetilde{F}^{[-n]}} = \mu_1$$

and we see that for mixed exponentials there are readily identified limiting distributions equal to exponential distributions

$$\begin{aligned}
\widetilde{\daleth(m, \langle w_i\rangle, \langle\mu_i\rangle)}^{[\infty]} &= \lim_{n\to\infty} \widetilde{\daleth(m, \langle w_i\rangle, \langle\mu_i\rangle)}^{[n]} = \daleth(1, 1, \mu_m) \\
\widetilde{\daleth(m, \langle w_i\rangle, \langle\mu_i\rangle)}^{[-\infty]} &= \lim_{n\to\infty} \widetilde{\daleth(m, \langle w_i\rangle, \langle\mu_i\rangle)}^{[-n]} = \daleth(1, 1, \mu_1)
\end{aligned}$$

This also illustrates what was just established more generally for the case of decreasing hazard functions.

The next example generalizes the mixed exponential and illustrates a construction that is "dual" to that of the coderived distribution:

Example 82 *Let $G = G(w)$ be a (not necessarily continuous) LDFn with PDF*

$$g(w) = \frac{dG}{dw}$$

and finite mean $\mu_G < \infty$. As above, there is the related LDFn $\widehat{G}$ with PDF

$$\widehat{g}(w) = \frac{wg(w)}{\mu_G}$$

$$\widehat{G}(w) = \int_0^w \widehat{g}(z)dz = \frac{\int_0^w zg(z)dz}{\mu_G}$$

which conforms with our earlier notation and as before we set

$$\widehat{G} = \widehat{G}^{[1]}$$

$$\widehat{G}^{[k]} = \widehat{\widehat{G}^{[k-1]}} \text{ for } k \in \mathbb{N} \text{ and } \mu_{\widehat{G}^{[k-1]}} < \infty$$

$$\mu_{\widehat{G}^{[k]}} = \frac{\mu_G^{(k+1)}}{\mu_G^{(k)}}.$$

This relates with the mixed exponential coderived distribution via a Laplace-like transformation. Define

$$L_G(x) = \int_0^\infty e^{-xw} dG = \int_0^\infty e^{-xw} g(w)dw$$

$$\Rightarrow L_G(0) = \int_0^\infty g(w)dw = 1$$

65

and the function L_G resembles a survival function of a mixed exponential distribution. We set

$$\begin{aligned}\mathcal{L}^*(G) &= 1 - L_G \\ \mathcal{L}^*(G)(x) &= 1 - \int_0^\infty e^{-xw} g(w) dw\end{aligned}$$

which associates with the LDFn G another LDFn $F = \mathcal{L}^*(G)$. Now observe that, differentiating under the integral

$$\begin{aligned}f_F(x) &= \frac{d\mathcal{L}^*(G)}{dx} = -\int_0^\infty \frac{d}{dx}\left(e^{-xw}\right) g(w) dw \\ &= -\frac{\mu_G}{\mu_G}\int_0^\infty \left(e^{-xw}\right)(-w)\, g(w) dw \\ &= \mu_G \int_0^\infty e^{-xw} \widehat{g}(w) dw \\ &= \mu_G \left(1 - \mathcal{L}^*(\widehat{G})(x)\right).\end{aligned}$$

Let $H = \mathcal{L}^*(\widehat{G})$, we have for the PDF of $\widetilde{H} = \widetilde{\mathcal{L}^*(\widehat{G})}$

$$\begin{aligned}f_{\widetilde{H}} &= \frac{S_H}{\mu_H} = \frac{1-H}{\mu_H} \\ &= \frac{\mu_G(1-H)}{\mu_G \mu_H} \\ &= \frac{\mu_G\left(1 - \mathcal{L}^*(\widehat{G})\right)}{\mu_G \mu_H} \\ &= \frac{f_F}{\mu_G \mu_H}.\end{aligned}$$

Now since both $f_{\widetilde{H}}$ and f_F are PDFs of the LDFns $\widetilde{F}$ and $\mathcal{L}(\widehat{G})$, respectively, it follows that

$$\begin{aligned}1 &= \int_0^\infty f_{\widetilde{H}}(x)dx = \int_0^\infty \frac{f_F(x)}{\mu_G \mu_H} dx = \frac{\int_0^\infty f_F(x)dx}{\mu_G \mu_H} = \frac{1}{\mu_G \mu_H} \\ &\Rightarrow \mu_G = \frac{1}{\mu_H} = \frac{1}{\mu_{\mathcal{L}^*(\widehat{G})}} \Rightarrow \mu_{\mathcal{L}^*(\widehat{G})} = \frac{1}{\mu_G} \\ &\Rightarrow f_{\widetilde{H}} = f_F \\ &\Rightarrow \mathcal{L}^*(G) = F = \widetilde{H} = \widetilde{\mathcal{L}^*(\widehat{G})}.\end{aligned}$$

This may be summarized in the commutative diagram:

$$\begin{array}{ccc} G & \xrightarrow{\mathcal{L}^*} & F \\ \hat{}\downarrow & & \uparrow\sim \\ \widehat{G} & \xrightarrow{\mathcal{L}^*} & H \end{array}$$

which illustrates that the rather trivially "derived" construction $G \to \widehat{G}$ of the time-biased distribution is "dual" under $\mathcal{L}^*$ to determining a coderived or equilibrium distribution or equivalently to determining the excess ratio curve. We have

$$\mathcal{L}^*(\widehat{G}) = \widetilde{\mathcal{L}^*\left(\widehat{\widehat{G}}\right)} = \widetilde{\mathcal{L}^*\left(\widehat{G}^{[2]}\right)}$$

$$\Rightarrow \mathcal{L}^*(G) = \widetilde{\mathcal{L}^*(\widehat{G})} = \left(\widetilde{\widetilde{\mathcal{L}^*\left(\widehat{G}^{[2]}\right)}}\right) = \widetilde{\mathcal{L}^*\left(\widehat{G}^{[2]}\right)}^{[2]}$$

and by induction

$$\mathcal{L}^*(G) = \widetilde{\mathcal{L}^*(\widehat{G}^{[1]})}^{[1]} = \widetilde{\mathcal{L}^*\left(\widehat{G}^{[2]}\right)}^{[2]} = ... = \widetilde{\mathcal{L}^*\left(\widehat{G}^{[n]}\right)}^{[n]} \quad \textit{for every } n \in \mathbb{N}$$

$$\Rightarrow \widetilde{\mathcal{L}^*(G)}^{[-n]} = \mathcal{L}^*\left(\widehat{G}^{[n]}\right) \quad \textit{for every } n \in \mathbb{Z}.$$

Example 83 *Gamma Density: Let $F = \Gamma(\alpha, \theta)$ have the Gamma density:*

$$f_F(x) = \frac{\left(\frac{x}{\theta}\right)^\alpha e^{-\frac{x}{\theta}}}{x\Gamma(\alpha)}$$

$$\mu_F^{(k)} = \theta^k (\alpha + k - 1) \cdots \alpha \textit{ for } -\alpha < k \in \mathbb{Z}.$$

We see from Proposition 40 that

$$\mu_{\widetilde{F}^{[n]}} = \frac{\mu_F^{(n+1)}}{(n+1)\,\mu_F^{(n)}} = \frac{\theta^{n+1}(\alpha+n)\cdots\alpha}{(n+1)\,\theta^n(\alpha+n-1)\cdots\alpha} = \frac{\theta(\alpha+n)}{n+1}$$

$$\Rightarrow \lim_{n\to\infty} \mu_{\widetilde{F}^{[n]}} = \theta.$$

Letting

$$G = \widetilde{F}^{[\infty]} = \lim_{n\to\infty} \widetilde{F}^{[n]}$$

we have

$$\widetilde{G} = G \quad \textit{and } \mu_G = \theta \Rightarrow G = \beth(1, \langle 1 \rangle, \langle \theta \rangle)$$

and the limiting distribution is independent of α and is recognized as exponential of mean θ. Finally, observe that

$$\mathcal{L}^*(F)(x) = 1 - \int_0^\infty e^{-xw} f_F(w)dw = 1 - \int_0^\infty e^{-xw} \frac{\left(\frac{w}{\theta}\right)^\alpha e^{-\frac{w}{\theta}}}{w\Gamma(\alpha)} dw$$

$$= 1 - \int_0^\infty \frac{w^{\alpha-1}e^{-\frac{w}{\theta}-xw}}{\theta^\alpha\Gamma(\alpha)}dw = 1 - \int_0^\infty \frac{w^{\alpha-1}e^{-w\left(x+\frac{1}{\theta}\right)}}{\theta^\alpha\Gamma(\alpha)}dw$$

$$\begin{aligned} &= 1 - \int_0^\infty \frac{\left(\frac{u}{x+\frac{1}{\theta}}\right)^{\alpha-1}e^{-u}}{\left(x+\frac{1}{\theta}\right)\theta^\alpha\Gamma(\alpha)}du \quad \textit{where} \quad u = w\left(x+\frac{1}{\theta}\right) \\ &= 1 - \frac{\left(x+\frac{1}{\theta}\right)^{-\alpha}}{\theta^\alpha}\int_0^\infty \frac{u^{\alpha-1}e^{-u}}{\Gamma(\alpha)}du \end{aligned}$$

$$= 1 - \frac{1}{\left(x+\frac{1}{\theta}\right)^\alpha\theta^\alpha} = 1 - \left(\frac{1}{\theta x+1}\right)^\alpha = 1 - \left(\frac{\frac{1}{\theta}}{x+\frac{1}{\theta}}\right)^\alpha$$

$$= \Pi(\alpha, \frac{1}{\theta}; x)$$

and we have that $\mathcal{L}^*(F) = \Pi(\alpha, \frac{1}{\theta})$. *And so for* $\alpha > 1$

$$f_{\widehat{F}}(x) = \frac{xf_F(x)}{\mu_F} = \frac{\left(\frac{x}{\theta}\right)^\alpha e^{-\frac{x}{\theta}}}{\theta\alpha\Gamma(\alpha)} = \frac{\left(\frac{x}{\theta}\right)^{\alpha+1} e^{-\frac{x}{\theta}}}{x\Gamma(\alpha+1)} = f_{\Gamma(\alpha+1,\theta)}(x)$$

$$\Rightarrow \widehat{F} = \Gamma(\alpha+1,\theta)$$

$$\Pi(\alpha,\frac{1}{\theta}) = \mathcal{L}^*(F) = \widetilde{\mathcal{L}^*(\widehat{F})} = \mathcal{L}^*(\widetilde{\Gamma(\alpha+1,\theta)}) = \Pi(\widetilde{\alpha+1,}\frac{1}{\theta})$$

as had already been observed in Example 79 above.

Example 84 *Weibull: Let* $F = W(\tau,\theta)$ *have the Weibull density:*

$$F(x) = W(\tau,\theta;x) = 1 - e^{-\left(\frac{x}{\theta}\right)^\tau} \quad \omega_F = \infty$$

$$f_F(x) = \frac{\tau\left(\frac{x}{\theta}\right)^\tau e^{-\left(\frac{x}{\theta}\right)^\tau}}{x}$$

$$\begin{aligned} \lambda_F(x) &= \frac{f_F(x)}{S_F(x)} = \frac{\tau\left(\frac{x}{\theta}\right)^\tau e^{-\left(\frac{x}{\theta}\right)^\tau}}{xe^{-\left(\frac{x}{\theta}\right)^\tau}} \\ &= \frac{\tau}{x}\left(\frac{x}{\theta}\right)^\tau = \frac{\tau x^{\tau-1}}{\theta^\tau} \end{aligned}$$

$$\tau_F = \left\{\begin{array}{cc} 0 & \tau < 1 \\ \frac{1}{\theta} & \tau = 1 \\ \infty & \tau > 1 \end{array}\right\}$$

$$\mu_F^{(k)} = \theta^k\Gamma\left(1+\frac{k}{\tau}\right),\ k > -\tau.$$

Recall the definition of the incomplete gamma function:

$$\Gamma(\alpha; x) = \int_0^x \frac{t^{\alpha-1}e^{-t}}{\Gamma(\alpha)}dt$$

and define

$$G(x) = 1 - \Gamma\left(1 + \frac{1}{\tau}; \left(\frac{x}{\theta}\right)^\tau\right) + \frac{xe^{-\left(\frac{x}{\theta}\right)^\tau}}{\theta\Gamma\left(1+\frac{1}{\tau}\right)}$$

then

$$\begin{aligned}
\frac{dG}{dx} &= -\frac{-\tau\left(\frac{x}{\theta}\right)^{\tau\left(1+\frac{1}{\tau}\right)}e^{-\left(\frac{x}{\theta}\right)^\tau}}{x\Gamma\left(1+\frac{1}{\tau}\right)} + \frac{-x\frac{\tau\left(\frac{x}{\theta}\right)^\tau e^{-\left(\frac{x}{\theta}\right)^\tau}}{x} + e^{-\left(\frac{x}{\theta}\right)^\tau}}{\theta\Gamma\left(1+\frac{1}{\tau}\right)} \\
&= \frac{\tau\left(\frac{x}{\theta}\right)^{\tau+1}e^{-\left(\frac{x}{\theta}\right)^\tau}}{x\Gamma\left(1+\frac{1}{\tau}\right)} + \frac{-\tau\left(\frac{x}{\theta}\right)^\tau e^{-\left(\frac{x}{\theta}\right)^\tau} + e^{-\left(\frac{x}{\theta}\right)^\tau}}{\theta\Gamma\left(1+\frac{1}{\tau}\right)} \\
&= \frac{\tau\left(\frac{x}{\theta}\right)^{\tau}e^{-\left(\frac{x}{\theta}\right)^\tau}}{\theta\Gamma\left(1+\frac{1}{\tau}\right)} + \frac{-\tau\left(\frac{x}{\theta}\right)^\tau e^{-\left(\frac{x}{\theta}\right)^\tau} + e^{-\left(\frac{x}{\theta}\right)^\tau}}{\theta\Gamma\left(1+\frac{1}{\tau}\right)} \\
&= \frac{\tau\left(\frac{x}{\theta}\right)^{\tau}e^{-\left(\frac{x}{\theta}\right)^\tau} - \tau\left(\frac{x}{\theta}\right)^\tau e^{-\left(\frac{x}{\theta}\right)^\tau} + e^{-\left(\frac{x}{\theta}\right)^\tau}}{\theta\Gamma\left(1+\frac{1}{\tau}\right)} \\
&= \frac{e^{-\left(\frac{x}{\theta}\right)^\tau}}{\theta\Gamma\left(1+\frac{1}{\tau}\right)} = \frac{S_F(x)}{\mu_F} = f_{\widetilde{F}}(x) \\
&\Rightarrow G = \widetilde{F} = \widetilde{W(\tau,\theta)}.
\end{aligned}$$

References

[1] Abramowitz, M. and Stegun, A.., *Handbook of Mathematical Functions*, Dover, 1972.

[2] Engl, G., Corro, D., "The 2004 NCCI Excess Loss Factors," *CAS* Forum, Fall 2006, pp. 513–571.

[3] Feller, W., *An Introduction to Probability Theory and Its Applications,* Volume II, Wiley, 1966.

[4] Keatinge, C., "Modeling Losses with the Mixed Exponential Distribution," *PCAS* LXXXVI, 1999, pp. 654–698.

[5] Klugman, S., Panjer, H., Willmot, G., *Loss Models: From Data to Decisions,* 2nd edition, Wiley, 2004

Grouping Loss Distributions by Tail Behavior Part III: Ordering Distributions

Dan Corro
National Council on Compensation Insurance
Spring 2008

Abstract: *This three part paper addresses the task of modelling the right hand tail of a severity distribution. In Part I the excess ratio function is used to define a discrete sequence of loss distributions with related moments and similar tail behavior. Part II extends this to continuous one-parameter families and provides some examples. Part III provides the main result: that under some reasonable conditions, each such family has a limiting distribution which is exponential. The paper then exploits this to 1) group loss distributions based on tail behavior and 2) promote the choice of (mixed) exponentials to model tail behavior.*

Remark 85 *This is the final part of a three part paper. We assume familiarity with Parts I and II and continue our numbering from those earlier parts.*

6 Orbits and Tail Behavior

We have seen that the orbit $\widetilde{F}^{[\mathbb{R}]}$ of an SLDFn F says something about the existence of moments and the $\sim$invariant τ_F. In this section we investigate the structural possibilities for the orbits $\widetilde{F}^{[\mathbb{R}]}$ and relate it to analytic behavior naturally associated with tail behavior. We make the following:

Definition 86 *A C^∞ function $T\colon [0,\infty) \to \mathbb{R}$ is **monotone of degree n** provided*

$$(-1)^k \frac{d^kT}{dx^k}(x) \geq 0 \text{ for } k = 0, 1, 2, ..., n \text{ and for every } x \in (0,\infty).$$

*T is **completely monotone** provided T is monotone of degree n for every $n \in \mathbb{N}$.*

Note that while the concept of monotone of degree n is peculiar to this paper, this is the standard definition of completely monotone (sometimes called totally monotone). As an immediate consequence of this definition we have:

Proposition 87 *For any SLDFn F:*

$\widetilde{F}^{[-n]}$ *exists for* $n \in \mathbb{N} \Leftrightarrow S_F$ *is monotone of degree* $n \Leftrightarrow S_{\widetilde{F}}$ *is monotone of degree* $n+1$

$$\widetilde{F}^{[-n]} \text{ exists for every } n \in \mathbb{N} \quad \Leftrightarrow \quad S_F \text{ is completely monotone.}$$

Proof. Clear from the definition of the backward coderived LDFn.

Example 88 *The survival function* $S_F(x) = \sum_{i=1}^{m} w_i e^{-\frac{x}{\mu_i}}$ *for a mixture of exponentials is completely monotone.*

■

Proposition 89 *If* $T(x) = \int_0^\infty e^{-xt} g(t) dt$ *for some integrable function* $g\colon (0,\infty) \to [0,\infty)$*, then T is completely monotone.*

Proof. This follows from differentiation under the integral

$$\frac{d^n T}{dx^n}(x) = \frac{d^n \int_0^\infty e^{-xt} g(t) dt}{dx^n} = \int_0^\infty \frac{d^n g(t) e^{-xt}}{dx^n} dt$$

$$= \int_0^\infty g(t) \frac{d^n e^{-xt}}{dx^n} dt = (-1)^n \int_0^\infty t^n g(t) e^{-xt} dt$$

$$\Rightarrow (-1)^n \frac{d^n T}{dx^n}(x) = \int_0^\infty t^n g(t) e^{-xt} dt \geq 0$$

completing the proof. ■

Remark 90 *A theorem of Bernstein establishes the converse; and we will soon make use of that theorem.*

Example 91 *Consider the survival function* $S_F(x) = e^{-\sqrt{x}}$*. In this case we have ([1], #29.3.83, p. 1026)*

$$S_F(x) = e^{-\sqrt{x}} = \int_0^\infty e^{-xt} g(t) dt \text{ where } g(t) = \frac{e^{-\frac{1}{4t}}}{2\sqrt{\pi t^3}}$$

and so S_F is completely monotone. Observe that we also have

$$f_F(x) = -\frac{dS_F}{dx}(x) = \frac{e^{-\sqrt{x}}}{2\sqrt{x}}$$

$$\lambda_F(x) = \frac{1}{2\sqrt{x}}$$

$$\Rightarrow \tau_F = 0.$$

71

Note too that the SLDFn F has all finite moments by Proposition 40, since

$$\begin{aligned}\mu_F^{(n)} &= n\int_0^\infty x^{n-1}S_F(x)dx = n\int_0^\infty x^{n-1}e^{-\sqrt{x}}dx \\ &= n\int_0^\infty u^{2n-2}e^{-u}2udu \text{ where } u=\sqrt{x}, x=u^2, dx=2udu \\ &= 2n\int_0^\infty u^{2n-1}e^{-u}du = 2n\,(2n-1)! < \infty.\end{aligned}$$

The following generalizes an earlier observation on mixed exponentials:

Proposition 92 *For any SLDFn F, if the survival function $S_F(x)$ has the form*

$$S_F(x) = \int_0^\infty e^{-xt}g(t)dt$$

for some integrable function $g\colon (0,\infty)\to[0,\infty)$, then $CV_F \geq 1$.

Proof. Note first that

$$1 = S_F(0) = \int_0^\infty g(t)dt.$$

and so by Schwartz

$$\begin{aligned}\left(\int_0^\infty \frac{g(t)}{t}dt\right)^2 &= \left(\int_0^\infty \sqrt{g(t)}\frac{\sqrt{g(t)}}{t}dt\right)^2 \\ &\leq \int_0^\infty \left(\sqrt{g(t)}\right)^2 dt \int_0^\infty \left(\frac{\sqrt{g(t)}}{t}\right)^2 dt \\ &= \int_0^\infty g(t)dt\int_0^\infty \frac{g(t)}{t^2}dt = \int_0^\infty \frac{g(t)}{t^2}dt.\end{aligned}$$

Observe next that for any fixed $t>0$, from what has been observed for the exponential distribution of parameter $\theta=\frac{1}{t}$ (example 76)

$$\begin{aligned}1 = \int_0^\infty \frac{e^{-\frac{x}{\theta}}}{\theta}dx &= t\int_0^\infty e^{-xt}dx \\ \Rightarrow \int_0^\infty e^{-xt}dx &= \frac{1}{t} \\ \theta = \int_0^\infty x\frac{e^{-\frac{x}{\theta}}}{\theta}dx &= t\int_0^\infty xe^{-xt}dx \\ \Rightarrow \int_0^\infty xe^{-xt}dx &= \frac{\theta}{t} = \frac{1}{t^2}.\end{aligned}$$

Now we compute, using Fubini

$$\begin{aligned}
\mu_F &= \int_0^\infty S_F(x)dx = \int_0^\infty \int_0^\infty e^{-xt} g(t)dtdx \\
&= \int_0^\infty \int_0^\infty e^{-xt} g(t)dxdt = \int_0^\infty g(t) \int_0^\infty e^{-xt}dxdt \\
&= \int_0^\infty \frac{g(t)}{t}dt.
\end{aligned}$$

Similarly, from Proposition 27

$$\begin{aligned}
\mu_F^{(2)} &= 2\int_0^\infty xS_F(x)dx = 2\int_0^\infty x\int_0^\infty e^{-xt}g(t)dtdx \\
&= 2\int_0^\infty g(t)\int_0^\infty xe^{-xt}dxdt = 2\int_0^\infty \frac{g(t)}{t^2}dt.
\end{aligned}$$

Now it follows that

$$\begin{gathered}
\sigma_F^2 + \mu_F^2 = \mu_F^{(2)} \\
\Rightarrow CV_F^2 + 1 = \frac{\sigma_F^2}{\mu_F^2} + 1 = \frac{\mu_F^{(2)}}{\mu_F^2} \\
= \frac{2\int_0^\infty \frac{g(t)}{t^2}dt}{\left(\int_0^\infty \frac{g(t)}{t}dt\right)^2} \geq 2 \\
\Rightarrow CV_F^2 \geq 1 \Rightarrow CV_F \geq 1
\end{gathered}$$

as required. ■

We noted in the examples that the fixed points under the coderived loss construction are exactly the exponential densities. In fact, by a theorem of Serge Bernstein, we have the following characterization of the exponential survival curve that we will find useful and that may even be of some independent interest:

Proposition 93 *For any C^∞ function $T\colon (0,\infty) \to \mathbb{R}$:*

$$\left\{\begin{array}{c}
T \textit{ is completely monotone} \\
\\
1 = \int_0^\infty T(x)dx \\
\\
\textit{There exists some } m \in \mathbb{N} \textit{ such that} \\
(-1)^m \frac{d^m T}{dx^m}(x) = T(x) \textit{ for every } x \in (0,\infty)
\end{array}\right\} \Leftrightarrow T(x) = e^{-x}.$$

Proof. $\Leftarrow$) Clear.

$\Rightarrow$) We have already noted that the case $m = 1$ holds, so assume $m > 1$. Letting $f = -\frac{dT}{dx}(x)$ we clearly have

$$1 = \int_0^\infty T(x)dx \Rightarrow \lim_{x\to\infty} T(x) = 0 \text{ and so } f(x) \geq 0 \text{ with } 1 = \int_0^\infty f(x)dx$$

and $f = f_F$ is the PDF for an SLDFn F for which $T = S_F$. As per Proposition 87, since $T = S_F$ is completely monotone we have the series of backward coderived loss variables

$$\begin{aligned}
S_{\widetilde{F}^{[-1]}}(x) &= \frac{f(x)}{f(0)} = \frac{-\frac{dT}{dx}(x)}{-\frac{dT}{dx}(0)} = \frac{\frac{dT}{dx}(x)}{\frac{dT}{dx}(0)} \\
S_{\widetilde{F}^{[-2]}}(x) &= \frac{-\frac{d}{dx}\left(S_{\widetilde{F}^{[-1]}}\right)(x)}{-\frac{d}{dx}\left(S_{\widetilde{F}^{[-1]}}\right)(0)} = \frac{\frac{d^2T}{dx^2}(x)}{\frac{d^2T}{dx^2}(0)} \\
&\vdots \\
S_{\widetilde{F}^{[-k]}}(x) &= \frac{-\frac{d}{dx}\left(S_{\widetilde{F}^{[-k+1]}}\right)(x)}{-\frac{d}{dx}\left(S_{\widetilde{F}^{[-k+1]}}\right)(0)} = \frac{\frac{d^kT}{dx^k}(x)}{\frac{d^kT}{dx^k}(0)}
\end{aligned}$$

Let $G = \widetilde{F}^{[-m]}$. We have

$$S_G = S_{\widetilde{F}^{[-m]}} = \frac{\frac{d^mT}{dx^m}}{\frac{d^mT}{dx^m}(0)} = \frac{T}{T(0)} = S_F$$

$$\Rightarrow G = F$$

$$\Rightarrow \widetilde{G}^{[k]} = \left(\widetilde{\widetilde{F}^{[-m]}}\right)^{[k]} = \widetilde{F}^{[k-m]}, \text{ for every } k \in \mathbb{Z}.$$

Now Bernstein's theorem implies that since T, $\frac{dT}{dx}$, $\frac{d^2T}{dx^2}$,...are all completely monotone, we can represent each of the $S_{\widetilde{F}^{[k]}}$ as a Laplace transform, as in Proposition 92 from which we conclude from Proposition 92 that $CV_{\widetilde{F}^{[k]}} \geq 1$ for every $k \in \mathbb{Z}$. But then by Proposition 46

$$\mu_F \leq \mu_{\widetilde{F}} \leq \mu_{\widetilde{F}^{[2]}} \leq ... \leq \mu_{\widetilde{F}^{[m]}} = \mu_F$$

$$\Rightarrow \mu_{\widetilde{F}^{[k]}} = \mu_F = \int_0^\infty T(x)dx = 1, \text{ for every } k \in \mathbb{Z}.$$

We claim that $\mu_F^{(k)} = k!$ for every $k \in \mathbb{N}$. We verify this by induction. Indeed we just observed the case $n = 1$ and by Proposition 40 and the induction hypothesis

$$1 = \mu_{\widetilde{F}^{[k]}} = \frac{\mu_F^{(k+1)}}{(k+1)\,\mu_F^{(k)}} = \frac{\mu_F^{(k+1)}}{(k+1)\,k!} = \frac{\mu_F^{(k+1)}}{(k+1)!}$$

$$\Rightarrow \mu_F^{(k+1)} = (k+1)!.$$

It only remains to observe that

$$\mu_F^{(k)} = k! \text{ for every } k \in \mathbb{N} \cup \{0\} \Rightarrow L_F(t) = \frac{1}{1+t} = L_{\beth(1,\langle 1\rangle,\langle 1\rangle)}(t)$$
$$\Rightarrow F = \beth(1, \langle 1\rangle, \langle 1\rangle)$$
$$\Rightarrow T = S_F = S_{\beth(1,\langle 1\rangle,\langle 1\rangle)} = e^{-x}$$

completing the proof. ■

Lemma 94 *If $c > 0$ is a fixed irrational number and $g\colon [0,\infty) \to \mathbb{R}$ is a continuous function satisfying*

$$g(n) \le g(n+1) \text{ for every } n \in \mathbb{N} \text{ and } g(x) = g(x+c) \text{ for every } x \in [0,\infty),$$

then g is constant, i.e., $g(x) = g(0)$ for every $x \in [0,\infty)$.

Proof. Consider the equivalence relation $\equiv$ on $[0,\infty)$ defined by

$$x \equiv y \Leftrightarrow \frac{x-y}{c} \in \mathbb{Z}.$$

Note that because $g(x) = g(x+c)$ for every $x \in [0,\infty)$, the function g is a continuous function well-defined on the equivalence classes of $[0,\infty)$. Note that

$$x \equiv x_1 \text{ and } y \equiv y_1$$
$$\Rightarrow \frac{x-x_1}{c} = z \in \mathbb{Z} \text{ and } \frac{y-y_1}{c} = w \in \mathbb{Z}$$
$$\text{but then } \frac{(x+y)-(x_1+y_1)}{c} = z + w \in \mathbb{Z}$$
$$\Rightarrow x + y \equiv x_1 + y_1$$
$$\text{and } \frac{(x-y)-(x_1-y_1)}{c} = z - w \in \mathbb{Z}$$
$$\Rightarrow x - y \equiv x_1 - y_1.$$

We claim that the sequence $A = \{\overline{n} | n \in \mathbb{N}, \overline{n} \in [0,c)\}$ of equivalence class representatives is dense in $[0,c)$. Assume given $d \in [0,c)$ and $0 < \epsilon_1 < c - d$. We have

$$\begin{aligned} \text{for every } n,m &\in \mathbb{N}, \quad n \equiv m \neq n \Rightarrow 0 \neq \frac{m-n}{c} = z \in \mathbb{Z} \Rightarrow c = \frac{m-n}{z} \in \mathbb{Q}, \text{ a contradiction } \Rightarrow\Leftarrow \\ &\Rightarrow \text{ sequence } A \text{ has distinct numbers in compact set } [0,c] \\ &\Rightarrow A \text{ has a cluster point.} \end{aligned}$$

Since there is a cluster point and the elements of A are distinct, it follows that

$$\text{there exist } m,n \in \mathbb{N} \text{ such that } m > n,\ \overline{m},\overline{n} \in [0,c) \text{ and } |\overline{m} - \overline{n}| < \frac{\epsilon_1}{4}$$

$$\Rightarrow \text{there exists } l \in \mathbb{N} \text{ such that } \sum_{k=1}^{l} (\overline{m} - \overline{n}) \in (d - \epsilon_1, d + \epsilon_1) \subset [0, c)$$

$$\text{but } \sum_{k=1}^{l} (\overline{m} - \overline{n}) = \overline{\sum_{k=1}^{l} (m - n)} = \overline{l\,(m - n)}$$

$$l\,(m - n) \in \mathbb{N} \Rightarrow \phi \neq A \cap (d - \epsilon_1, d + \epsilon_1) \cap [0, c)$$

and so A is dense in $[0, c)$ as claimed. Now since g is continuous

$$A \text{ dense in } [0, c)$$

$$\Rightarrow g(A) \text{ dense in } \{g(x)\,|x \in [0, c)\} = \{g(x)\,|x \in [0, \infty)\} = \mathrm{Im}(g).$$

Note that since g is continuous on the compact set $[0, c]$, we know that g has a maximum. But then by our assumptions, the sequence $\{g(n)|n \in \mathbb{N}\}$ is nondecreasing and bounded above. So we can set

$$\lim_{n \to \infty} g(n) = \alpha < \infty.$$

Now assume given any $\epsilon_2 > 0$, it follows that

$$\text{there exists } M \in \mathbb{N} \text{ such that } g(n) \in (\alpha - \epsilon_2, \alpha] \text{ for every } n > M$$

$$g \text{ periodic} \Rightarrow g(n) \in (\alpha - \epsilon_2, \alpha] \text{ for every } n \in \mathbb{N}$$

$$\{g(n)|n \in \mathbb{N}\} \text{ dense in } \mathrm{Im}(g) \Rightarrow \mathrm{Im}(g) \subseteq (\alpha - \epsilon_2, \alpha].$$

But since $\epsilon_2 > 0$, was arbitrary, we have

$$\{g(n)|n \in \mathbb{N}\} \subseteq \bigcap_{n \in \mathbb{N}} (\alpha - \frac{1}{n}, \alpha] = [\alpha, \alpha]$$

$$\{g(n)|n \in \mathbb{N}\} \text{ dense in } \mathrm{Im}(g) \Rightarrow \mathrm{Im}(g) \subseteq \overline{\{g(n)|n \in \mathbb{N}\}} \subseteq \overline{[\alpha, \alpha]} = [\alpha, \alpha]$$

$$\mathrm{Im}(g) \subseteq [\alpha, \alpha] \Longrightarrow g(x) = \alpha = g(0) \text{ for every } x \in [0, \infty)$$

and the proof is complete. ■

Theorem 95 *For any SLDFn F with finite mean, the following are equivalent:*

1. there exists $r \in \mathbb{R}, r > 0$ such that $F = \widetilde{F}^{[r]}$
2. $S_F(x) = e^{-\frac{x}{\mu}}$
3. $\widetilde{F}^{[\mathbb{R}]} = \{F\}$

Proof. Suppose $F = \widetilde{F}^{[r]}$ for some $r > 0$ We first claim that $\omega_F = \infty$. Indeed, if $\omega_F = b < \infty$. then clearly $\mu_F^{(k)} < \infty$ for every $k \in \mathbb{N}$ and by the intermediate value theorem

$$\begin{aligned}
\text{there exist } a_k &\in (0, b] \text{ such that } a_1 = \mu_F = \mu_F^{(1)} \\
\text{and } \mu_F^{(k+1)} &= \int_0^b x^{k+1} f_F(x) dx = \int_0^b x \left(x^k f_F(x)\right) dx \\
&= a_k \int_0^b x^k f_F(x) dx = a_k \mu_F^{(k)} \\
&\Rightarrow \quad \mu_{\widetilde{F}^{[k]}} = \frac{\mu_F^{(k+1)}}{(k+1)\,\mu_F^{(k)}} \\
&= \frac{a_k \mu_F^{(k)}}{(k+1)\,\mu_F^{(k)}} \\
&= \frac{a_k}{(k+1)} \leq \frac{b}{(k+1)} \\
&\Rightarrow \quad \lim_{k\to\infty} \mu_{\widetilde{F}^{[k]}} = 0
\end{aligned}$$

But $F = \widetilde{F}^{[r]} \Rightarrow \mu_{\widetilde{F}^{[kr]}} = \mu_F > 0$ for all $k \in \mathbb{N}$. Since the function $m(c) = \mu_{\widetilde{F}^{[c]}}$ is evidently continuous, we have

$$\begin{aligned}
0 &= \lim_{k\to\infty} \mu_{\widetilde{F}^{[k]}} = \lim_{c\to\infty} \mu_{\widetilde{F}^{[c]}} \\
&= \lim_{k\to\infty} \mu_{\widetilde{F}^{[kr]}} = \lim_{k\to\infty} \mu_F = \mu_F > 0 \quad \Rightarrow\Leftarrow
\end{aligned}$$

This contradiction implies that $\omega_F = \infty$ and $S_F \colon (0, \infty) \to \mathbb{R}$ is a C^∞ function. We now prove that $S_F(x) = e^{-\frac{x}{\mu}}$. Consider first the case $r = m \in \mathbb{N}$. Assume that $F = \widetilde{F}^{[\mathbb{R}]}$ and let $G = F_\mu$. Then by Proposition 48 and the fact that

$$\begin{aligned}
\widetilde{G}^{[m]} &= \widetilde{(F_\mu)}^{[m]} = \left(\widetilde{F}^{[m]}\right)_\mu = F_\mu = G \\
&\Rightarrow \widetilde{G}^{[-m]} = G \\
&\Rightarrow S_G = S_{\widetilde{G}^{[-m]}}.
\end{aligned}$$

Since S_G is clearly completely monotone, it follows from the same CV argument as in the proof of Proposition 93 that

$$\begin{aligned}
1 &= \mu_G = \mu_{\widetilde{G}^{[k]}} \quad \text{for every } k \in \mathbb{Z} \\
&\Rightarrow f_{\widetilde{G}^{[k]}}(0) = 1 \quad \text{for every } k \in \mathbb{Z} \\
&\Rightarrow S_G = S_{\widetilde{G}^{[-m]}} = \frac{(-1)^m \frac{d^m S_G}{dx^m}}{f_{\widetilde{G}^{[-m+1]}}(0)} = (-1)^m \frac{d^m S_G}{dx^m}.
\end{aligned}$$

and so the characterization of that lemma implies that

$$\begin{aligned} S_G(y) &= e^{-y} \\ F &= G_{\frac{1}{\mu}} \\ &\Rightarrow S_F(x) = S_{G_{\frac{1}{\mu}}}(x) = S_G\left(\frac{x}{\mu}\right) = e^{-\frac{y}{\mu}}. \end{aligned}$$

This completes the proof for $m \in \mathbb{N}$. Consider next the case $r \in \mathbb{Q}$ let $r = \frac{n}{m}$, with $m, n \in \mathbb{N}$.and $F = \widetilde{F}^{[r]}$ We claim that

$$F = \widetilde{F}^{\left[\frac{kn}{m}\right]} \text{ for every } k \in \mathbb{N}.$$

This is a simple verification by induction, for $k = 1$ this reduces to $F = \widetilde{F}^{\left[\frac{n}{m}\right]}$, which is true, and then

$$\begin{aligned} F &= \widetilde{F}^{\left[\frac{kn}{m}\right]} \\ &\Rightarrow \widetilde{F}^{\left[\frac{(k+1)n}{m}\right]} = \widetilde{F}^{\left[\frac{(k+1)n}{m}\right]} = \widetilde{F}^{\left[\frac{kn}{m}+\frac{n}{m}\right]} = \widetilde{F}^{\left[\frac{kn}{m}+r\right]} \\ &= \widetilde{\widetilde{F}^{\left[\frac{kn}{m}\right]}}^{[r]} = \widetilde{F}^{[r]} = F \end{aligned}$$

completing the induction. But then it follows that

$$F = \widetilde{F}^{\left[\frac{mn}{m}\right]} = \widetilde{F}^{[n]} \text{ for } n \in \mathbb{N} \Rightarrow S_F(x) = e^{-\frac{x}{\mu}}$$

by the case $r = m \in \mathbb{N}$, completing the proof in the rational case. Finally, consider next the case $r \in \mathbb{R}$ with r irrational. As above, the assumptions imply that we can represent each of the $S_{\widetilde{F}^{[c]}}$ as a Laplace transform, as in Proposition 92 from which we conclude that $CV_{\widetilde{F}^{[c]}} \geq 1$ for every $c \in \mathbb{R}$ We clearly have $\mu_{\widetilde{F}^{[c]}} < \infty$ for every $c \geq 0$ so we define

$$g(c) = \mu_{\widetilde{F}^{[c]}} \text{ for } c \geq 0.$$

Then g is continuous on $[0, \infty)$ and by Proposition 46

$$\begin{aligned} g(c) &= \mu_{\widetilde{F}^{[c]}} \leq \mu_{\widetilde{\widetilde{F}^{[c]}}} = \mu_{\widetilde{F}^{[c+1]}} = g(c+1) \\ g(c+r) &= \mu_{\widetilde{F}^{[r+c]}} = \mu_{\widetilde{\widetilde{F}^{[r]}}^{[c]}} = \mu_{\widetilde{F}^{[c]}} = g(c). \end{aligned}$$

And so the lemma implies that

$$\mu_{\widetilde{F}^{[c]}} = g(c) = g(0) = \mu_{\widetilde{F}^{[0]}} = \mu_F = \mu \text{ for every } c \geq 0.$$

We claim that $\mu_F^{(k)} = k!\mu^k$ for every $k \in \mathbb{N}$. We verify this by induction. Indeed the case $n = 1$ being apparent. By Proposition 40 and the induction hypothesis

$$\mu = \mu_{\widetilde{F}^{[k]}} = \frac{\mu_F^{(k+1)}}{(k+1)\,\mu_F^{(k)}} = \frac{\mu_F^{(k+1)}}{(k+1)\,k!\mu^k} = \frac{\mu_F^{(k+1)}}{(k+1)!\mu^k}$$

78

$$\Rightarrow (k+1)!\mu^{k+1} = \mu_F^{(k+1)}$$

completing the induction. It only remains to observe that

$$\mu_F^{(k)} = k!\mu^k \text{ for every } k \in \mathbb{N} \cup \{0\} \Rightarrow L_F(t) = L_{\daleth(1,\langle 1\rangle,\langle\mu\rangle)}(t)$$
$$\Rightarrow F = \daleth(1, \langle 1\rangle, \langle\mu\rangle)$$
$$\Rightarrow T(x) = S_F(x) = S_{\daleth(1,\langle 1\rangle,\langle\mu\rangle)}(x) = e^{-\frac{x}{\mu}}.$$

We have shown

$$\left(\text{there exists some } r \in \mathbb{R}, r > 0 \text{ such that } F = \widetilde{F}^{[r]}\right) \Rightarrow \left(S_F(x) = e^{-\frac{x}{\mu}}\right)$$

but clearly

$$\left(S_F(x) = e^{-\frac{x}{\mu}}\right) \Rightarrow \left(\widetilde{F}^{[\mathbb{R}]} = \{F\}\right)$$

$$\text{and} \quad \left(\widetilde{F}^{[\mathbb{R}]} = \{F\}\right) \Rightarrow \left(\text{there exists } r \in \mathbb{R}, r > 0 \text{ such that } F = \widetilde{F}^{[r]}\right)$$

and the result follows. ■

We observe that, except when $\widetilde{F}^{[\mathbb{R}]} = \{F\}$ is the singleton orbit of an exponential, once we have selected an SLDFn $G \in \widetilde{F}^{[\mathbb{R}]}$ the elements $H \in \widetilde{F}^{[\mathbb{R}]}$ are uniquely expressible in the form $H = \widetilde{G}^{[c]}$ in the sense $c = c(H)$ is uniquely determined. It is most natural to just take $G = F$. This enables us to describe the possibilities for the structure of the orbit $\widetilde{F}^{[\mathbb{R}]}$ as related to a subset of $\mathbb{R}$, an interval actually, via the bijection

$$\Omega\colon \widetilde{F}^{[\mathbb{R}]} \to \Omega\left(\widetilde{F}^{[\mathbb{R}]}\right) \subseteq \mathbb{R} \qquad \text{where } \Omega\left(\widetilde{F}^{[c]}\right) = c \in \mathbb{R}.$$

In effect, this is a canonical 1-dimensional continuous parametrization of the orbits $\widetilde{F}^{[\mathbb{R}]}$ of non-exponential SLDFns F. We summarize this observation in:

Proposition 96 *For any SLDFn $F \neq \daleth(1, \langle 1\rangle, \langle\mu_F\rangle)$ with finite mean,* $[0,1] \subseteq \Omega\left(\widetilde{F}^{[\mathbb{R}]}\right)$ *and the possibilities for* $\Omega\left(\widetilde{F}^{[\mathbb{R}]}\right)$ *and* τ_F *are:*

1. there exist $c, d \in \mathbb{R}$ with $\Omega\left(\widetilde{F}^{[\mathbb{R}]}\right) \in \{[c,,d], [c,d), (c,d], (c,d)\}, \tau_F = 0$

   $\Leftrightarrow$there exists $n \in \mathbb{N}$ such that $\mu_F^{(n)} = \infty$ and S_F is not completely monotone.

2. there exists $c \in \mathbb{R}$ with $\Omega\left(\widetilde{F}^{[\mathbb{R}]}\right) \in \{[c,\infty), (c,\infty)\}, c \leq 0, \tau_F \in [0,\infty]$

   $\Leftrightarrow \mu_F^{(n)} < \infty$ for every $n \in \mathbb{N}$ and S_F is not completely monotone.

3. there exists $d \in \mathbb{R}$ with $\Omega\left(\widetilde{F}^{[\mathbb{R}]}\right) \in \{(-\infty,d], (-\infty,d)\}, d > 0, \tau_F = 0$

   $\Leftrightarrow$there exists $n \in \mathbb{N}$ such that $\mu_F^{(n)} = \infty$ and S_F is completely monotone.

4. $\Omega\left(\widetilde{F}^{[\mathbb{R}]}\right) = (-\infty, \infty), \tau_F \in (0, \infty)$

   $\Leftrightarrow \mu_F^{(n)} < \infty$ for every $n \in \mathbb{N}$ and S_F is completely monotone.

With all possibilities actually occurring. Indeed we have:

1. Let $F(x) = \frac{x^2}{1+x^2}$. We have

$$S_F(x) = \frac{1}{1+x^2}$$

$$\mu_F = \frac{\pi}{2} < \infty$$

$$f_F(x) = \frac{2x}{(1+x^2)^2}$$

$$\mu_F^{(a)} = 2\int_0^\infty \frac{x^{a+1}}{x^4+2x^2+1}dx < \infty$$

$$\Rightarrow a+1 < 3 \Rightarrow a < 2$$

$$\Rightarrow d \leq 2$$

$$\frac{df_F}{dx} = 2\frac{1-3x^2}{(1+x^2)^3} \Rightarrow \text{there exists a mode at } \frac{1}{\sqrt{3}} > 0$$

$$\Rightarrow c > -1.$$

2. Lognormal or any loss distribution with finite support and mode > 0.

3. Pareto.

4. Mixed Exponential.

The following is also clear from the above:

Proposition 97 $\Omega\left(\widetilde{F}^{[\mathbb{R}]}\right) = (\infty, \infty) \Leftrightarrow$*there exists LDFn G with $F = \mathcal{L}^*(G)$ and $\mu_G^{(n)} < \infty$ for every $n \in \mathbb{N}$.*

7 Ordering Loss Distributions

In this section we introduce a way to order SLDFns based on differences between hazard rate functions. We then relate this with the orbit structure of the previous section.

Proposition 98 *For any SLDFns F and G with $\omega_F = \omega_G$:*

$$\lim_{x\to\omega_F}\frac{f_F(x)}{f_G(x)} = \lim_{x\to\omega_F}\frac{S_F(x)}{S_G(x)} = e^{\int_0^{\omega_F}(\lambda_G-\lambda_F)(t)dt}$$

Proof. All but the last equality is clear from l'Hôpital, but then

$$\begin{aligned}
\lim_{x\to\omega_F} \frac{S_F(x)}{S_G(x)} &= \lim_{x\to\omega_F} \frac{e^{-\int_0^x \lambda_F(t)dt}}{e^{-\int_0^x \lambda_G(t)dt}} \\
&= \lim_{x\to\omega_F} e^{-\int_0^x \lambda_F(t)dt+\int_0^x \lambda_G(t)dt} \\
&= \lim_{x\to\omega_F} e^{\int_0^x (\lambda_G-\lambda_F)(t)dt} \\
&= e^{\lim_{x\to\omega_F} \int_0^x (\lambda_G-\lambda_F)(t)dt} \\
&= e^{\int_0^{\omega_F} (\lambda_G-\lambda_F)(t)dt}
\end{aligned}$$

as required. ∎

Definition 99 *For two SLDFns F and G set*

$$\upsilon(F,G) = e^{\int_0^{Min(\omega_F,\omega_G)} (\lambda_G-\lambda_F)(t)dt}.$$

Provided $\upsilon(F,G)$ exists, define the relations ***thicker than*** *and* ***strictly thicker than*** *by*

$$\begin{aligned}
F &\succeq G \Leftrightarrow \omega_F \geq \omega_G \text{ or } (\omega_F = \omega_G \text{ and } \upsilon(F,G) \geq 1) \\
F &\succ G \Leftrightarrow \omega_F > \omega_G \text{ or } (\omega_F = \omega_G \text{ and } \upsilon(F,G) > 1).
\end{aligned}$$

Remark 100 *Note that*

$$\begin{aligned}
\omega_F &= \omega_G = \infty \text{ and } \tau_F < \tau_G \Rightarrow \int_0^{\omega_F} (\lambda_G - \lambda_F)(t)dt = \infty \Rightarrow \upsilon(F,G) = e^{\infty} = \infty > 1 \\
&\Rightarrow F \succ G.
\end{aligned}$$

Example 101 *Let $F(x) = 1 - (x+1)e^{-x}$. We have*

$$S_F(x) = (x+1)e^{-x}$$

$$f_F(x) = -\frac{dS_F}{dx}(x) = -\left((x+1)e^{-x}(-1) + e^{-x}\right) = xe^{-x}$$

$$\lambda_F(x) = \frac{f_F(x)}{S_F(x)} = \frac{xe^{-x}}{(x+1)e^{-x}} = \frac{x}{x+1} \Rightarrow \tau_F = 1$$

$$\mu_F = \int_0^\infty x f_F(x)dx = \int_0^\infty x^2 e^{-x}dx = 2$$

$$f_{\widetilde{F}}(x) = \frac{S_F(x)}{\mu_F} = \frac{(x+1)e^{-x}}{2}$$

$$\upsilon(F,\widetilde{F}) = \lim_{x\to\infty} \frac{f_F(x)}{f_{\widetilde{F}}(x)} = \lim_{x\to\infty} \frac{xe^{-x}}{\frac{(x+1)e^{-x}}{2}} = 2\lim_{x\to\infty} \frac{x}{x+1} = 2 > 1$$

$$\Rightarrow F \succ \widetilde{F}.$$

81

Example 102 *Let* $F(x) = 1 - \frac{e^{-x}}{x+1}$. *We have*

$$S_F(x) = \frac{e^{-x}}{x+1}$$

$$f_F(x) = -\frac{dS_F}{dx}(x) = -\left(\frac{(x+1)\,e^{-x}(-1) - e^{-x}}{(x+1)^2}\right) = \frac{(x+2)\,e^{-x}}{(x+1)^2}$$

$$\lambda_F(x) = \frac{f_F(x)}{S_F(x)} = \frac{\frac{(x+2)e^{-x}}{(x+1)^2}}{\frac{e^{-x}}{x+1}} = \frac{x+2}{x+1} \Rightarrow \tau_F = 1$$

$$\begin{aligned} \mu_F &= \int_0^\infty S_F(x)dx = \int_0^\infty \frac{e^{-x}}{x+1}dx = \int_1^\infty \frac{e^{-u+1}}{u}du \\ &= e\int_1^\infty \frac{e^{-u}}{u}du < e\int_1^\infty e^{-u}du = e\left[-e^{-u}\right]_1^\infty = \frac{e}{e} = 1 \end{aligned}$$

$$f_{\widetilde{F}}(x) = \frac{S_F(x)}{\mu_F} = \frac{\frac{e^{-x}}{x+1}}{\mu_F}$$

$$\upsilon(F, \widetilde{F}) = \lim_{x\to\infty} \frac{f_F(x)}{f_{\widetilde{F}}(x)} = \lim_{x\to\infty} \frac{\frac{(x+2)e^{-x}}{(x+1)^2}}{\frac{\frac{e^{-x}}{x+1}}{\mu_F}} = \mu_F \lim_{x\to\infty} \frac{x+2}{x+1} = \mu_F < 1$$

$$\Rightarrow F \prec \widetilde{F}.$$

Proposition 103 *Given SLDFns* F *and* G *with* $\omega_F = \omega_G$ *and constants* $a, b > 0$ *such that the limit* $\rho_F(\frac{a}{b}) = \lim_{x\to\omega_F} \frac{S_F(\frac{a}{b}x)}{S_F(x)}$ *exists. Then*

$$\upsilon(F_a, G_b) = \rho_F(\frac{a}{b})\upsilon(F, G)$$

Proof. We have

$$\begin{aligned} \upsilon(F_a, G_b) &= \lim_{x\to\omega_F} \frac{S_{F_a}(x)}{S_{G_b}(x)} = \lim_{x\to\omega_F} \frac{S_F(ax)}{S_G(bx)} \\ &= \lim_{x\to\omega_F} \frac{S_F\left(a(\frac{x}{b})\right)}{S_G\left(b(\frac{x}{b})\right)} = \lim_{x\to\omega_F} \frac{S_F\left((\frac{a}{b})x\right)}{S_G(x)} \\ &= \lim_{x\to\omega_F} \frac{S_F((\frac{a}{b})x)}{S_F(x)} \frac{S_F(x)}{S_G(x)} \\ &= \lim_{x\to\omega_F} \frac{S_F((\frac{a}{b})x)}{S_F(x)} \lim_{x\to\omega_F} \frac{S_F(x)}{S_G(x)} \\ &= \rho_F(\frac{a}{b})\upsilon(F, G) \end{aligned}$$

as required. ■

Recall the following from set theory:

82

Definition 104 *Given a set A with relation $\geqslant$, A is* ***partially ordered*** *under $\geqslant$ provided for every $a, b, c \in A$*

1. *(reflexive) $a \geqslant a$*
2. *(antisymmetric) $(a \geqslant b$ and $b \geqslant a) \Rightarrow a = b$*
3. *(transitive) $(a \geqslant b$ and $b \geqslant c) \Rightarrow a \geqslant c$*

It is straightforward to verify that $\succeq$ defines a partial order relation on the equivalence classes of SLDFns modulo the equivalence relation

$$F \approx G \Leftrightarrow \omega_F = \omega_G \text{ and } \upsilon(F, G) = 1.$$

As usual, the case in which the hazard function is either increasing or decreasing is especially easy:

Proposition 105 *If F is any SLDFn with finite mean and $\tau_F > 0$, then for all $m < n \in \mathbb{N}$:*

$$\begin{array}{lll} \lambda_F \text{ increasing} & \Rightarrow & \widetilde{F}^{[m]} \succ \widetilde{F}^{[n]} \\ \lambda_F \text{ decreasing} & \Rightarrow & \widetilde{F}^{[n]} \succ \widetilde{F}^{[m]}. \end{array}$$

Proof. Clear from Propositions 33 and 32. Observe that λ_F increasing or decreasing implies that $\lambda_F > 0$ on (α_F, ω_F). Now, all moments are finite, so $\widetilde{F}^{[n]}$ exists and so the assertion at least makes sense. We have

$$\begin{aligned} &\lambda_F \text{ increasing } \Rightarrow \lambda_{\widetilde{F}^{[m]}} \text{ increasing} \\ &\Rightarrow \lambda_{\widetilde{F}^{[n]}} = \lambda_{\widetilde{\widetilde{F}^{[n-1]}}} > \lambda_{\widetilde{F}^{[n-1]}} \geq \lambda_{\widetilde{F}^{[m]}} \text{ on } (\alpha_F, \omega_F) \\ &\Rightarrow \int_{\alpha_F}^{\omega_F} \left(\lambda_{\widetilde{F}^{[n]}} - \lambda_{\widetilde{F}^{[m]}}\right)(t)dt > 0 \\ &\Rightarrow \upsilon(\widetilde{F}^{[m]}, \widetilde{F}^{[n]}) = e^{\int_{\alpha_F}^{\omega_F}\left(\lambda_{\widetilde{F}^{[n]}} - \lambda_{\widetilde{F}^{[m]}}\right)(t)dt} > e^0 = 1 \\ &\Rightarrow \widetilde{F}^{[m]} \succ \widetilde{F}^{[n]} \end{aligned}$$

as asserted. The result for λ_F decreasing follows similarly, reversing inequalities. ■

Proposition 106 *For any SLDFns F and G with $\omega_F = \omega_G = \infty$*

$$\begin{aligned} \upsilon &= \upsilon(F, G) < \infty \\ &\Rightarrow \text{ for every } \epsilon > 0 \text{ there exists an } M \text{ such that } |S_F(x) - \upsilon S_G(x)| < \epsilon \text{ for every } x > M. \end{aligned}$$

Proof. Clear. Given $\epsilon > 0$

$$\omega_G = \infty \Rightarrow S_G(x) > 0 \text{ for every } x \geq 0$$
$$0 = \lim_{x\to\infty} S_G(x)$$
$$\Rightarrow \text{there exists } M_1 \text{ such that } 0 < |S_G(x)| < \sqrt{\epsilon} \text{ for every } x > M_1$$
$$\upsilon = \upsilon(F,G) = \lim_{x\to\infty} \frac{S_F(x)}{S_G(x)} < \infty$$
$$\Rightarrow \text{there exists } M_2 \text{ such that } 0 \leq \left|\frac{S_F(x)}{S_G(x)} - \upsilon\right| < \sqrt{\epsilon} \text{ for every } x > M_2.$$

Then setting $M = \max(M_1, M_2)$ we have

$$x > M$$
$$\Rightarrow |S_F(x) - \upsilon S_G(x)| = |S_G(x)| \left|\frac{S_F(x)}{S_G(x)} - \upsilon\right| < \sqrt{\epsilon}\sqrt{\epsilon} = \epsilon$$

as required. ■

Proposition 107 *For any SLDFns F and G with $\omega_F = \omega_G = \infty$ and for which* $0 \leq \tau_F, \tau_G \leq \infty$:

$$0 < \upsilon(F,G) < \infty \Rightarrow \tau_F = \tau_G.$$

Proof. Set $\lim_{x\to\infty} \frac{f_F(x)}{f_G(x)} = \lim_{x\to\infty} \frac{S_F(x)}{S_G(x)} = \upsilon$

$$\begin{aligned} 1 &= \upsilon\frac{1}{\upsilon} = \lim_{x\to\infty} \frac{f_F(x)}{f_G(x)} \lim_{x\to\infty} \frac{S_G(x)}{S_F(x)} = \lim_{x\to\infty} \frac{f_F(x)}{f_G(x)} \frac{S_G(x)}{S_F(x)} \\ &= \lim_{x\to\infty} \frac{f_F(x)}{S_F(x)} \frac{S_G(x)}{f_G(x)} = \lim_{x\to\infty} \lambda_F(x) \frac{1}{\lambda_G(x)}. \end{aligned}$$

Consider first the case $0 < \tau_F, \tau_G < \infty$

$$1 = \lim_{x\to\infty} \lambda_F(x) \frac{1}{\lim\limits_{x\to\infty} \lambda_G(x)} = \tau_F \frac{1}{\tau_G} \Rightarrow \tau_F = \tau_G.$$

We have

$$1 = \lim_{x\to\infty} \lambda_F(x) \frac{1}{\lim\limits_{x\to\infty} \lambda_G(x)} \text{ and so}$$
$$0 = \tau_F = \lim_{x\to\infty} \lambda_F(x) \Rightarrow 0 = \lim_{x\to\infty} \lambda_G(x) = \tau_G.$$

and by the same token

$$1 = \lim_{x\to\infty} \lambda_F(x) \frac{1}{\lim\limits_{x\to\infty} \lambda_G(x)} \text{ and so}$$
$$\infty = \tau_F = \lim_{x\to\infty} \lambda_F(x) \Rightarrow \infty = \lim_{x\to\infty} \lambda_G(x) = \tau_G$$

and the result follows. ■

Example 108 *The converse is false:*

$$\lambda_F(x) = 1, \lambda_G(x) = 1 + \frac{1}{x}$$
$$\Rightarrow \int_0^\infty (\lambda_G - \lambda_F)(t)dt = \int_0^\infty \frac{dt}{t} = \infty$$
$$\Rightarrow \tau_F = \tau_G = 1 \text{ with } \upsilon(F, G) = \infty.$$

Proposition 109 *For any SLDFns F and G with $\omega_F = \omega_G = \infty$ and for which $F \succeq G$:*

$$c > 0 \text{ such that } \mu_F^{(c)} < \infty \Rightarrow \mu_G^{(c)} < \infty.$$

Proof. We have

$$F \succeq G$$
$$\Rightarrow \upsilon = \upsilon(F, G) = \lim_{x\to\infty} \frac{S_F(x)}{S_G(x)} \geq 1.$$

Consider first the case $\upsilon > 1$

$$\text{there exists } M_1 \text{ such that } S_F(x) > S_G(x), \text{ for every } x > M_1$$
$$\Rightarrow \mu_G^{(c)} = c\int_0^\infty x^{c-1}S_G(x)dx$$
$$= c\int_0^{M_1} x^{c-1}S_G(x)dx + c\int_{M_1}^\infty x^{c-1}S_G(x)dx$$
$$\leq cM_1^{c-1}M_1 + c\int_{M_1}^\infty x^{c-1}S_F(x)dx$$
$$\leq cM_1^c + c\mu_F^{(c)} < \infty.$$

So now consider the case $\upsilon = 1$

$$\lim_{x\to\infty} \frac{S_G(x)}{S_F(x)} = \frac{1}{\upsilon} = 1$$

$$\Rightarrow \text{there exists } M_2 \text{ such that } S_F(x) > 0 \text{ and } \left|\frac{S_G(x)}{S_F(x)} - 1\right| < \frac{1}{2}, \text{ for every } x > M_2$$

$$\Rightarrow \frac{S_G(x)}{S_F(x)} < \frac{3}{2}, \text{ for every } x > M_2$$
$$\Rightarrow S_G(x) < \frac{3}{2}S_F(x), \text{ for every } x > M_2$$
$$\Rightarrow \mu_G^{(c)} = c\int_0^\infty x^{c-1}S_G(x)dx$$
$$= c\int_0^{M_2} x^{c-1}S_G(x)dx + c\int_{M_2}^\infty x^{c-1}S_G(x)dx$$

$$\leq cM_2^{c-1}M_2 + c\int_{M_2}^{\infty} x^{c-1}\frac{3}{2}S_F(x)dx$$
$$\leq cM_2^c + \frac{3c}{2}\int_0^{\infty} x^{c-1}S_F(x)dx \leq cM_2^c + \frac{3}{2}\mu_F^{(c)} < \infty$$

and the proof is complete. ■

Proposition 110 *For any SLDFns F and G with $\omega_F = \omega_G = \infty$:*

$$0 \leq \tau_F < \tau_G \Rightarrow F \succ G.$$

Proof. This is straightforward:

$$\tau_F < \tau_G$$
$$\Rightarrow \text{there exist } M, \epsilon > 0 \text{ such that } \lambda_F(x) < \lambda_G(x) - \epsilon \text{ for every } x > M$$
$$\Rightarrow \int_M^{\infty} (\lambda_G - \lambda_F)(t)dt \geq \int_M^{\infty} \epsilon dt = \infty.$$

But then

$$\begin{aligned}
\int_0^{\infty} (\lambda_G - \lambda_F)(t)dt &= \int_0^{M} (\lambda_G - \lambda_F)(t)dt + \int_M^{\infty} (\lambda_G - \lambda_F)(t)dt \\
&= \int_0^{M} (\lambda_G - \lambda_F)(t)dt + \infty = \infty
\end{aligned}$$
$$\Rightarrow \lim_{x\to\infty}\frac{S_F(x)}{S_G(x)} = e^{\int_0^{\infty}(\lambda_G-\lambda_F)(t)dt} = e^{\infty} = \infty > 1$$
$$\Rightarrow F \succ G$$

as asserted. ■

Given any two SLDFns F and G and assuming τ_F and τ_G are known, the comparative thickness reduces to evaluating the limit $\upsilon(F,G)$ when $\tau_F = \tau_G$. But we know that the set of SLDFns F for which τ_F is a specified constant is acted on by the additive group $\mathbb{R}$ via taking the coderived distributions (when they exist) and is thus decomposed into orbits $\widetilde{F}^{[\mathbb{R}]}$under that action. The structure of those orbits was described in the previous section and we can orient ourselves within an orbit as to the "more or less tail-like" the distribution is in the "analytic" sense that $\widetilde{F}^{[c]}$ is more tail-like than $\widetilde{F}^{[d]}$exactly when $c > d$ (here "more tail like" means higher degree of monotonality. And we have seen that one may sacrifice the existence of moments to achieve that). The next result finally draws together the two perspectives of the paper and shows how the structure of those orbits relates with "thickness":

Proposition 111 *If F and G are SLDFns with $\omega_F = \omega_G$ and $0 < \tau_F = \tau_G < \infty$, then:*

$$\textit{for every } m,n \in \mathbb{N},\ \upsilon\left(\widetilde{F}^{[m]}, \widetilde{G}^{[n]}\right) = \upsilon(F,G)\,\frac{m!\tau_F^{n-m}\mu_G^{(n)}}{n!\mu_F^{(m)}}.$$

86

Proof. From the Corollary 43

$$\lim_{x\to\infty}\frac{f_{\widetilde{F}^{[n]}}(x)}{f_{\widetilde{F}^{[m]}}(x)}=\frac{\tau_F^{m-n}n!\mu_F^{(m)}}{m!\mu_F^{(n)}}$$

whence

$$\upsilon\left(\widetilde{F}^{[m]},\widetilde{G}^{[n]}\right)=\lim_{x\to\omega_F}\frac{f_{\widetilde{F}^{[m]}}(x)}{f_{\widetilde{G}^{[n]}}(x)}=\lim_{x\to\omega_F}\frac{f_{\widetilde{F}^{[m]}}(x)}{f_F(x)}\frac{f_F(x)}{f_G(x)}\frac{f_G(x)}{f_{\widetilde{G}^{[n]}}(x)}$$

$$=\lim_{x\to\omega_F}\frac{f_{\widetilde{F}^{[m]}}(x)}{f_F(x)}\lim_{x\to\omega_F}\frac{f_F(x)}{f_G(x)}\lim_{x\to\omega_G}\frac{f_{G^{[0]}}(x)}{f_{\widetilde{G}^{[n]}}(x)}$$

$$=\left(\lim_{x\to\omega_F}\frac{f_{F^{[0]}}(x)}{f_{\widetilde{F}^{[m]}}(x)}\right)^{-1}\upsilon(F,G)\frac{\tau_G^{n-0}(0!)\mu_G^{(n)}}{n!\mu_G^{(0)}}$$

$$=\left(\frac{\tau_F^m\mu_F^{(m)}}{m!}\right)^{-1}\upsilon(F,G)\frac{\tau_G^n\mu_G^{(n)}}{n!}$$

$$=\upsilon(F,G)\frac{m!}{\tau_F^m\mu_F^{(m)}}\frac{\tau_G^n\mu_G^{(n)}}{n!}=\upsilon(F,G)\frac{m!}{n!}\frac{\tau_F^{n-m}\mu_G^{(n)}}{\mu_F^{(m)}}$$

as required. ■

Corollary 112 *For any SLDFns F and G with $\omega_F=\omega_G$ and $0<\tau_F=\tau_G<\infty$:*

$$\widetilde{F}^{[m]}\succeq\widetilde{G}^{[n]}\Leftrightarrow\upsilon(F,G)\frac{m!\tau_F^{n-m}\mu_G^{(n)}}{n!\mu_F^{(m)}}\geq1.$$

Corollary 113 *For any SLDFn F with $0<\tau_F<\infty$:*

$$\widetilde{F}^{[m]}\succeq\widetilde{F}^{[n]}\Leftrightarrow\frac{m!\tau_F^{n-m}\mu_F^{(n)}}{n!\mu_F^{(m)}}\geq1.$$

Corollary 114 *For any SLDFn F with $0<\tau_F<\infty$:*

$$\widetilde{F}^{[m]}\succeq F\Leftrightarrow m!\geq\tau_F^m\mu_F^{(m)}.$$

Corollary 115 *For any SLDFn F with $0<\tau_F<\infty$:*

there exists $k\geq0$ such that $\widetilde{F}^{[m]}\succeq F$ for every $m\geq k$.

Proof. Observe that for all $x,t>0$:

$$0<e^{-tx}<1$$

and so the integral

$$L(t)=\int_0^\infty e^{-tx}f(x)dx=\int_0^\infty\left|e^{-tx}f(x)\right|dx\leq\int_0^\infty|f(x)|\,dx=1$$

is absolutely convergent. We have for any $x > 0$:

$$e^{\tau_F x} = \sum_{k=0}^{\infty} \frac{(\tau_F x)^k}{k!} = \sum_{k=0}^{\infty} \left| \frac{(-1)^k \tau_F^k x^k}{k!} \right|$$

and the power series expansion

$$e^{-\tau_F x} = \sum_{k=0}^{\infty} \frac{(-\tau_F x)^k}{k!} = \sum_{k=0}^{\infty} \frac{(-1)^k \tau_F^k x^k}{k!}$$

is absolutely convergent and so can be integrated term by term:

$$\begin{aligned} L_F(\tau_F) &= \int_0^{\infty} e^{-\tau_F x} f(x)dx = \int_0^{\infty} \sum_{k=0}^{\infty} \left(\frac{(-1)^k \tau_F^k x^k f(x)}{k!} \right) dx \\ &= \sum_{k=0}^{\infty} \left(\left(\frac{(-1)^k \tau_F^k}{k!} \right) \int_0^{\infty} x^k f(x)dx \right) \\ &= \sum_{k=0}^{\infty} \frac{(-1)^k \mu_F^{(k)} \tau_F^k}{k!}. \end{aligned}$$

Since the terms of any convergent series must converge to 0:

there exists $k \geq 0$ such that $\dfrac{\mu_F^{(m)} \tau_F^m}{m!} = \left| \dfrac{(-1)^m \mu_F^{(m)} \tau_F^m}{m!} \right| < 1$ for every $m \geq k$.

And by the previous corollary:

$$\widetilde{F}^{[m]} \succeq F \quad \text{for every } m \geq k$$

as required. ■

Proposition 116 *If F is an SLDFn for which the orbit $\widetilde{F}^{[\mathbb{R}]}$ has a last element then:*

$$\widetilde{F}^{[m]} \succeq \widetilde{F}^{[n]} \Leftrightarrow m \geq n.$$

Proof. Let $\widetilde{F}^{[l]}$ be the last element of $\widetilde{F}^{[\mathbb{R}]}, \Omega\left(\widetilde{F}^{[\mathbb{R}]}\right) = (-\infty, l]$. Suppose first that $\widetilde{F}^{[m]} \succeq \widetilde{F}^{[n]}$, in this case, we have

$$l \geq m, l \geq n$$

$$\Rightarrow l - m = \text{highest finite moment of } \widetilde{F}^{[m]}$$

$$\Rightarrow l - n = \text{highest finite moment of } \widetilde{F}^{[n]}$$

but then from Proposition 109

$$\widetilde{F}^{[m]} \succeq \widetilde{F}^{[n]} \text{ and } \mu_{\widetilde{F}^{[m]}}^{(l-m)} < \infty$$

88

$$
\begin{gathered}
\Rightarrow \mu_{\widetilde{F}^{[n]}}^{(l-m)} < \infty \\
\Rightarrow l - m \leq l - n \\
\Rightarrow -m \leq -n \\
\Rightarrow m \geq n
\end{gathered}
$$

establishing one direction. For the converse, suppose that $m \geq n$, and by way of contradiction that $\widetilde{F}^{[m]} \succeq \widetilde{F}^{[n]}$ is false. Then since by Proposition 42 the limit $\upsilon\left(\widetilde{F}^{[m]}, \widetilde{F}^{[n]}\right)$ exists, we must have

$$
\begin{gathered}
\upsilon\left(\widetilde{F}^{[m]}, \widetilde{F}^{[n]}\right) < 1 \\
\Rightarrow \upsilon\left(\widetilde{F}^{[n]}, \widetilde{F}^{[m]}\right) = \frac{1}{\upsilon\left(\widetilde{F}^{[m]}, \widetilde{F}^{[n]}\right)} > 1 \\
\Rightarrow \widetilde{F}^{[n]} \succ \widetilde{F}^{[m]} \\
\text{above direction} \Rightarrow n \geq m \quad \text{and } n \neq m \\
\Rightarrow n > m \quad \Rightarrow\Leftarrow
\end{gathered}
$$

and this contradiction completes the proof. ■

Proposition 117 *If F is the SLDFn of a mixed exponential density, then for all $k, n \in \mathbb{N}$:*

$$\widetilde{F}^{[k]} \succeq \widetilde{F}^{[n]} \quad \Leftrightarrow \quad k \geq n.$$

Proof. Let F be a mixture of exponential densities. More precisely, for some m, $1 \leq m \leq \infty$, weights $w_i > 0$ with $1 = \sum_{i=1}^{m} w_i$ and parameters $\mu_i > 0$ ordered so that $\mu_i < \mu_{i+1}$ and with $\sum_{i=1}^{m} w_i \mu_i < \infty$ set

$$F = \beth(m, \langle w_i \rangle, \langle \mu_i \rangle)$$

with survival function

$$S_F(x) = \sum_{i=1}^{m} w_i e^{-\frac{x}{\mu_i}}.$$

Then we have, by Proposition 49

$$\mu_F = \sum_{i=1}^{m} w_i \mu_i < \infty$$

$$S_{\widetilde{F}}(x) = \frac{\sum_{i=1}^{m} w_i \mu_i e^{-\frac{x}{\mu_i}}}{\mu_F} = \sum_{i=1}^{m} u_i e^{-\frac{x}{\mu_i}} \text{ where } u_i = \frac{w_i \mu_i}{\mu_F}.$$

And it follows that

$$\frac{S_F(x)}{S_{\widetilde{F}}(x)} = \frac{\sum_{i=1}^{m} w_i e^{-\frac{x}{\mu_i}}}{\sum_{i=1}^{m} u_i e^{-\frac{x}{\mu_i}}} = \frac{e^{\frac{x}{\mu_m}} \sum_{i=1}^{m} w_i e^{-\frac{x}{\mu_i}}}{e^{\frac{x}{\mu_m}} \sum_{i=1}^{m} u_i e^{-\frac{x}{\mu_i}}}$$

89

$$= \frac{\sum_{i=1}^{m} w_i e^{\frac{x}{\mu_m} - \frac{x}{\mu_i}}}{\sum_{i=1}^{m} u_i e^{\frac{x}{\mu_m} - \frac{x}{\mu_i}}} = \frac{w_m + \sum_{i=1}^{m-1} w_i e^{\left(\frac{\mu_i - \mu_m}{\mu_i \mu_m}\right)x}}{u_m + \sum_{i=1}^{m-1} u_i e^{\left(\frac{\mu_i - \mu_m}{\mu_i \mu_m}\right)x}}$$

$$\mu_i - \mu_m < 0, 1 \le i \le m-1$$

$$\Rightarrow \lim_{x\to\infty} \frac{S_F(x)}{S_{\widetilde{F}}(x)} = \lim_{x\to\infty} \frac{w_m + \sum_{i=1}^{m-1} w_i e^{\left(\frac{\mu_i - \mu_m}{\mu_i \mu_m}\right)x}}{u_m + \sum_{i=1}^{m-1} u_i e^{\left(\frac{\mu_i - \mu_m}{\mu_i \mu_m}\right)x}}$$

$$= \frac{w_m + \lim_{x\to\infty} \sum_{i=1}^{m-1} w_i e^{\left(\frac{\mu_i - \mu_m}{\mu_i \mu_m}\right)x}}{u_m + \lim_{x\to\infty} \sum_{i=1}^{m-1} u_i e^{\left(\frac{\mu_i - \mu_m}{\mu_i \mu_m}\right)x}} = \frac{w_m + 0}{u_m + 0} = \frac{w_m}{u_m}$$

$$= \frac{w_m}{\frac{w_m \mu_m}{\mu_F}} = \frac{\mu_F}{\mu_m} < 1$$

$$\Rightarrow \widetilde{F} \succeq F$$

and then by transitivity.$\widetilde{F}^{[k-n]} \succeq F$, and the result follows by replacing F with .$\widetilde{F}^{[n]}$, which also has a mixed exponential density.

As one would expect, there are continuous analogues for many of these "momentous" observations: ■

Proposition 118 *For any SLDFns F and G with $0 < \tau_F = \tau_G < \infty$ and positive $c, d \in \mathbb{R}$:*

$$\upsilon\left(\widetilde{F}^{[d]}, \widetilde{G}^{[c]}\right) = \upsilon(F, G)\, \frac{\Gamma(d+1)\tau_F^{c-d}\mu_G^{(c)}}{\Gamma(c+1)\mu_F^{(d)}}.$$

Proof. This is clear from Proposition 71

$$\upsilon\left(\widetilde{F}^{[d]}, \widetilde{G}^{[c]}\right) = \lim_{x\to\infty} \frac{f_{\widetilde{F}^{[d]}}(x)}{f_{\widetilde{G}^{[c]}}(x)} = \lim_{x\to\infty} \frac{f_{\widetilde{F}^{[d]}}(x)}{f_F(x)} \frac{f_F(x)}{f_G(x)} \frac{f_G(x)}{f_{\widetilde{G}^{[c]}}(x)}$$

$$= \lim_{x\to\infty} \frac{f_{\widetilde{F}^{[d]}}(x)}{f_F(x)} \lim_{x\to\infty} \frac{f_F(x)}{f_G(x)} \lim_{x\to\infty} \frac{f_G(x)}{f_{\widetilde{G}^{[c]}}(x)}$$

$$= \left(\lim_{x\to\infty} \frac{f_F(x)}{f_{\widetilde{F}^{[d]}}(x)}\right)^{-1} \upsilon(F, G)\, \frac{\tau_G^c \mu_G^{(c)}}{\Gamma(c+1)} = \left(\frac{\tau_F^d \mu_F^{(d)}}{\Gamma(d+1)}\right)^{-1} \upsilon(F, G)\, \frac{\tau_Y^c \mu_G^{(c)}}{\Gamma(c+1)}$$

$$= \upsilon(F, G)\, \frac{\Gamma(d+1)}{\tau_F^d \mu_F^{(d)}} \frac{\tau_G^c \mu_G^{(c)}}{\Gamma(c+1)} = \upsilon(F, G)\, \frac{\Gamma(d+1)}{\Gamma(c+1)} \frac{\tau_F^{c-d} \mu_G^{(c)}}{\mu_F^{(d)}}$$

as required. ■

Corollary 119 *For any SLDFns F and G with $0 < \tau_F = \tau_G < \infty$ and positive $c, d \in \mathbb{R}$:*

$$\widetilde{F}^{[d]} \succeq \widetilde{G}^{[c]} \Leftrightarrow \upsilon(F, G)\, \frac{\Gamma(d+1)\tau_F^{c-d}\mu_G^{(c)}}{\Gamma(c+1)\mu_F^{(d)}} \ge 1.$$

Corollary 120 *For any SLDFn F with $0 < \tau_F < \infty$ and positive $c, d \in \mathbb{R}$:*

$$\widetilde{F}^{[d]} \succeq \widetilde{F}^{[c]} \Leftrightarrow \frac{\Gamma(d+1)\tau_F^{c-d}\mu_F^{(c)}}{\Gamma(c+1)\mu_F^{(d)}} \geq 1.$$

Corollary 121 *For any SLDFn F with $0 < \tau_F < \infty$ and positive $d \in \mathbb{R}$:*

$$\widetilde{F}^{[d]} \succeq F \Leftrightarrow \Gamma(d+1) \geq \tau_F^d \mu_F^{(d)}.$$

We conclude this section with a couple more examples.

Example 122 *Consider the Pareto density $F = \Pi(\alpha, \theta)$:*

$$S_F(x) = \left(\frac{\theta}{x+\theta}\right)^{\alpha}$$

$$\frac{\partial S}{\partial \alpha} = \left(\frac{\theta}{x+\theta}\right)^{\alpha} \ln\left(\frac{\theta}{x+\theta}\right) \leq 0$$

$$\frac{\partial S}{\partial \theta} = \alpha\theta\left(\frac{\theta}{x+\theta}\right)^{\alpha-1}\left(\frac{1}{x+\theta}\right)^2 = \alpha\theta^{\alpha}(x+\theta)^{-\alpha-1} > 0$$

which suggests that, all else equal, F gets "thicker" as θ increases or α decreases. Now let $F = \Pi(\alpha, \theta)$ $G = \Pi(\beta, \vartheta)$ be two Pareto densities, $\alpha, \beta, \theta, \vartheta \in (0, \infty)$. Then $\omega_F = \omega_G = \infty$ and $\tau_F = \tau_G = 0$ with

$$\upsilon(F, G) = e^{\int_0^{\infty}(\lambda_G - \lambda_F)(t)dt}$$

$$\int_0^{\infty}(\lambda_G - \lambda_F)(t)dt = \int_0^{\infty}\left(\frac{\beta}{t+\vartheta} - \frac{\alpha}{t+\theta}\right)dt$$

$$= \left[\beta\ln(t+\vartheta) - \alpha\ln(t+\theta)\right]_{t=0}^{t=\infty} = \left[\ln\left((t+\vartheta)^{\beta}\right) - \ln(t+\theta)^{\alpha}\right]_{t=0}^{t=\infty}$$

$$= \left[\ln\left(\frac{(t+\vartheta)^{\beta}}{(t+\theta)^{\alpha}}\right)\right]_{t=0}^{t=\infty} = \lim_{t\to\infty}\left[\ln\left(\frac{(t+\vartheta)^{\beta}}{(t+\theta)^{\alpha}}\right)\right] - \ln\left(\frac{\vartheta^{\beta}}{\theta^{\alpha}}\right)$$

$$= \ln\left(\lim_{t\to\infty}\left(\frac{(t+\vartheta)^{\beta}}{(t+\theta)^{\alpha}}\right)\right) + \ln\left(\frac{\theta^{\alpha}}{\vartheta^{\beta}}\right)$$

$$= \ln\left(\lim_{t\to\infty}\left(\frac{t+\vartheta}{t+\theta}\right)^{\alpha}(t+\vartheta)^{\beta-\alpha}\right) + \ln\left(\frac{\theta^{\alpha}}{\vartheta^{\beta}}\right)$$

$$= \ln\left(\lim_{t\to\infty}1^{\alpha}(t+\vartheta)^{\beta-\alpha}\right) + \ln\left(\frac{\theta^{\alpha}}{\vartheta^{\beta}}\right)$$

$$= \ln\left(\lim_{t\to\infty}(t+\vartheta)^{\beta-\alpha}\right) + \ln\left(\frac{\theta^{\alpha}}{\vartheta^{\beta}}\right)$$

$$= \left\{ \begin{array}{cc} +\infty & \beta > \alpha \\ \ln\left(\left(\frac{\theta}{\vartheta}\right)^{\alpha}\right) & \beta = \alpha \\ -\infty & \beta < \alpha \end{array} \right\}.$$

We see that

$$\upsilon(F,G) = e^{\int_0^{\infty}(\lambda_G - \lambda_F)(t)dt} = \left\{ \begin{array}{cc} +\infty & \beta > \alpha \\ \left(\frac{\theta}{\vartheta}\right)^{\alpha} & \beta = \alpha \\ 0 & \beta < \alpha \end{array} \right\}$$

and that

$$\Pi(\alpha,\theta) \succeq \Pi(\beta,\vartheta) \Leftrightarrow \upsilon(F,G) \geq 1$$

$$\Leftrightarrow \beta > \alpha \ \textit{or} \ (\beta = \alpha \ \textit{and} \ \theta \geq \vartheta)$$

and

$$\Pi(\alpha,\theta) \succ \Pi(\beta,\vartheta) \Leftrightarrow \upsilon(F,G) > 1$$

$$\Leftrightarrow \beta > \alpha \ \textit{or} \ (\beta = \alpha \ \textit{and} \ \theta > \vartheta)$$

which conforms to what was suggested before.

Example 123 *Let* $F = \Pi(\alpha,\theta)$ *and* $G(x) = \frac{x^2}{1+x^2}$ *:*

$$\lambda_F(x) = \frac{\alpha}{\theta + x} \Rightarrow \tau_F = 0$$

$$\lambda_G(x) = \frac{f_G(x)}{S_G(x)} = \frac{\frac{2x}{(1+x^2)^2}}{\frac{1}{1+x^2}} = \frac{2x}{1+x^2} \Rightarrow \tau_G = 0$$

$$\int_0^{\infty} (\lambda_G - \lambda_F)(t)dt = \int_0^{\infty} \left(\frac{2t}{1+t^2} - \frac{\alpha}{\theta + t}\right) dt$$

$$= \left[\ln\left(1+t^2\right) - \alpha \ln(\theta + t)\right]_{t=0}^{t=\infty} = \lim_{t\to\infty} \left[\ln\left(1+t^2\right) - \ln(\theta+t)^{\alpha}\right] + \alpha \ln(\theta)$$

$$= \lim_{t\to\infty}\left[\ln\left(\frac{1+t^2}{(\theta+t)^{\alpha}}\right)\right] + \ln(\theta^{\alpha}) = \ln\left[\lim_{t\to\infty}\left(\frac{1+t^2}{(\theta+t)^{\alpha}}\right)\right] + \ln(\theta^{\alpha})$$

$$= \ln\left[\lim_{t\to\infty}\left(\frac{2t}{\alpha(\theta+t)^{\alpha-1}}\right)\right] + \ln(\theta^{\alpha})$$

$$= \left\{ \begin{array}{cc} +\infty & \alpha \leq 1 \\ \ln\left[\lim_{t\to\infty}\left(\frac{2}{\alpha(\alpha-1)(\theta+t)^{\alpha-2}}\right)\right] + \ln(\theta^{\alpha}) & \alpha > 1 \end{array} \right\}$$

$$= \left\{ \begin{array}{cc} +\infty & \alpha \leq 1 \\ 1 + \ln(\theta^{\alpha}) & \alpha = 2 \\ \ln\left[\frac{2}{\alpha(\alpha-1)}\lim_{t\to\infty}\left((\theta+t)^{2-\alpha}\right)\right] - \ln(\theta^{\alpha}) & \alpha > 1, \alpha \neq 2 \end{array} \right\}$$

92

$$= \left\{ \begin{array}{cc} +\infty & \alpha \leq 1 \\ \ln(1) + \ln\left(\theta^2\right) & \alpha = 2 \\ \ln\left[\frac{2}{\alpha(\alpha-1)}\left(+\infty\right)\right] + \ln\left(\theta^\alpha\right) = +\infty & 1 < \alpha < 2 \\ \ln\left[\frac{2}{\alpha(\alpha-1)}\left(0\right)\right] + \ln\left(\theta^\alpha\right) = -\infty & \alpha > 2 \end{array} \right\}$$

$$= \left\{ \begin{array}{cc} +\infty & \alpha < 2 \\ \ln\left(\theta^2\right) & \alpha = 2 \\ -\infty & \alpha > 2 \end{array} \right\}$$

$$\upsilon(F,G) = e^{\int_0^\infty (\lambda_G - \lambda_F)(t)dt} = \left\{ \begin{array}{cc} +\infty & \alpha < 2 \\ \theta^2 & \alpha = 2 \\ 0 & \alpha > 2 \end{array} \right\}$$

$$\begin{aligned} \alpha &< 2 \Rightarrow \upsilon(F,G) = \infty > 1 \Rightarrow F \succ G \\ \alpha &> 2 \Rightarrow \upsilon(F,G) = 0 < 1 \Rightarrow G \succ F \end{aligned}$$

$$\begin{aligned} \alpha &= 2, \theta > 1 \Rightarrow \upsilon(F,G) > 1 \Rightarrow F \succ G \\ \alpha &= 2, \theta = 1 \Rightarrow \upsilon(F,G) = 1 \Rightarrow F \approx G \\ \alpha &= 2, \theta < 1 \Rightarrow \upsilon(F,G) < 1 \Rightarrow G \succ F. \end{aligned}$$

8 Conclusion

For a given continuous loss distribution F with finite mean, we have seen that the ratio of losses in excess of a given loss limit x to total losses defines a function $R(x)$ that formally resembles a survival function. The loss distribution defined by that survival function was defined to be the "coderived" distribution $\widetilde{F}$. This coderived distribution was shown to exhibit (right hand) tail behavior and moments that are very closely related to those of the original loss distribution (Propositions 27 and 28). Moreover, this coderived distribution has a simpler, more "monotone", structure than the original (Proposition 87). We observed that this coderived distribution completely determines the original distribution (Proposition 26). Repeating this process yields a discrete sequence of loss distributions $F, \widetilde{F}, \widetilde{\widetilde{F}}, \ldots$within a continuous, one-parameter collection of loss distributions (Remark 58). Such collections all have tails with the same ultimate settlement rate $\tau_F = \tau_{\widetilde{F}} = \tau_{\widetilde{\widetilde{F}}}$ (Proposition 28). We described a simple approach to ordering loss distributions according to the "thickness" of their tails (Definition 99) and related thickness with monotonality and ultimate settlement rate (Proposition 111). A key finding is that the asymptotic behavior of the hazard rate as captured by the ultimate settlement rate $\tau_F = \lim\limits_{x \to \omega_F} \lambda_F(x)$, provides a natural bridge between these two perspectives. We observed that if the hazard rate function is increasing or decreasing, then the sequence of coderived distributions converges to an exponential loss distribution (Proposition 78). We conclude that when modeling loss severity (where the hazard rate function is reasonably well-behaved, e.g. with only finitely many turning points, and where there is no cap), there is a uniquely determined exponential distribution with canonical properties that favor it as a choice to splice onto to the model as the right hand tail. If it is impractical to go far enough out into the tail to make the tail close to monotone (near constant hazard rate), one should consider fitting a mixed exponential. The reader is invited to consult [2] for both a discussion of tail-splicing and as a case study of this approach.

References

[1] Abramowitz, M. and Stegun, A.., *Handbook of Mathematical Functions*, Dover, 1972.

[2] Engl, G., Corro, D., "The 2004 NCCI Excess Loss Factors," *CAS* Forum, Fall 2006, pp. 513–571.

[3] Feller, W., *An Introduction to Probability Theory and Its Applications,* Volume II, Wiley, 1966.

[4] Keatinge, C., "Modeling Losses with the Mixed Exponential Distribution," *PCAS* LXXXVI, 1999, pp. 654–698.

[5] Klugman, S., Panjer, H., Willmot, G., *Loss Models: From Data to Decisions,* 2nd edition, Wiley, 2004

Modelling The Claims Development Result For Solvency Purposes

Michael Merz, Mario V. Wüthrich

Abstract

We assume that the claims liability process satisfies the distribution-free chain-ladder model assumptions. For claims reserving at time I we predict the total ultimate claim with the information available at time I and, similarly, at time $I+1$ we predict the same total ultimate claim with the (updated) information available at time $I+1$. The claims development result at time $I+1$ for accounting year $(I, I+1]$ is then defined to be the difference between these two successive predictions for the total ultimate claim. In [6, 10] we have analyzed this claims development result and we have quantified its prediction uncertainty. Here, we simplify, modify and illustrate the results obtained in [6, 10]. We emphasize that these results have direct consequences for solvency considerations and were (under the new risk-adjusted solvency regulation) already implemented in industry.

Keywords. Stochastic Claims Reserving, Chain-Ladder Method, Claims Development Result, Loss Experience, Incurred Losses Prior Accident Years, Solvency, Mean Square Error of Prediction.

1. INTRODUCTION

We consider the problem of quantifying the uncertainty associated with the development of claims reserves for prior accident years in general insurance. We assume that we are at time I and we predict the total ultimate claim at time I (with the available information up to time I), and one period later at time $I+1$ we predict the same total ultimate claim with the updated information available at time $I+1$. The difference between these two successive predictions is the so-called claims development result for accounting year $(I, I+1]$. The realization of this claims development result has a direct impact on the profit & loss (P&L) statement and on the financial strength of the insurance company. Therefore, it also needs to be studied for solvency purposes. Here, we analyze the prediction of the claims development result and the possible fluctuations around this prediction (prediction uncertainty). Basically we answer two questions that are of practical relevance:

(a) In general, one predicts the claims development result for accounting year $(I, I+1]$ in the budget statement at time I by 0. We analyze the uncertainty in this prediction. This is a *prospective view*: "how far can the realization of the claims development result deviate from 0?"

Remark: we discuss below, why the claims development result is predicted by 0.

(b) In the P&L statement at time $I+1$ one then obtains an observation for the claims development result. We analyze whether this observation is within a reasonable range around 0 or whether it is an outlier. This is a *retrospective view*. Moreover, we discuss the possible categorization of this uncertainty.

So let us start with the description of the budget statement and of the P&L statement, for an example we refer to Table 1. The budget values at Jan. 1, year I, are predicted values for the next accounting year $(I, I+1]$. The P&L statement are then the observed values at the end of this accounting year $(I, I+1]$.

Positions a) and b) correspond to the premium income and its associated claims (generated by the premium liability). Position d) corresponds to expenses such as acquisition expenses, head office expenses, etc. Position e) corresponds to the financial returns generated on the balance sheet/assets. All these positions are typically well-understood. They are predicted at Jan. 1, year I (budget values) and one has their observations at Dec. 31, year I in the P&L statement, which describes the financial closing of the insurance company for accounting year $(I, I+1]$.

Modelling The Claims Development Result For Solvency Purposes

	budget values at Jan. 1, year I	P&L statement at Dec. 31, year I
a) premiums earned	4'000'000	4'020'000
b) claims incurred current accident year	-3'200'000	-3'240'000
c) loss experience prior accident years	0	-40'000
d) underwriting and other expenses	-1'000'000	-990'000
e) investment income	600'000	610'000
income before taxes	400'000	360'000

Table 1: Income statement, in \$ 1'000

However, position c), "loss experience prior accident years", is often much less understood. It corresponds to the difference between the claims reserves at time $t = I$ and at time $t = I + 1$ adjusted for the claim payments during accounting year $(I, I+1]$ for claims with accident years prior to accounting year I. In the sequel we will denote this position by the **claims development result** (CDR). We analyze this position within the framework of the distribution-free chain-ladder (CL) method. This is described below.

Short-term vs. long-term view

In the classical claims reserving literature, one usually studies the total uncertainty in the claims development until the total ultimate claim is finally settled. For the distribution-free CL method this has first been done by Mack [7]. The study of the total uncertainty of the full run-off is a *long-term view*. This classical view in claims reserving is very important for solving solvency questions, and almost all stochastic claims reserving methods which have been proposed up to now concentrate on this long term view (see Wüthrich-Merz [9]). However, in the present work we concentrate on a second important view, the *short-term view*. The short-term view is important for a variety of reasons:

Modelling The Claims Development Result For Solvency Purposes

- If the short-term behaviour is not adequate, the company may simply not get to the "long-term", because it will be declared insolvent before it gets to the long term.

- A short-term view is relevant for management decisions, as actions need to be taken on a regular basis. Note that most actions in an insurance company are usually done on a yearly basis. These are for example financial closings, pricing of insurance products, premium adjustments, etc.

- Reflected through the annual financial statements and reports, the short-term performance of the company is of interest and importance to regulators, clients, investors, rating agencies, stock-markets, etc. Its consistency will ultimately have an impact on the financial strength and the reputation of the company in the insurance market.

Hence our goal is to study the development of the claims reserves on a yearly basis where we assume that the claims development process satisfies the assumptions of the distribution-free chain-ladder model. Our main results, Results 3.1-3.3 and 3.5 below, give an improved version of the results obtained in [6, 10]. De Felice-Moriconi [4] have implemented similar ideas referring to the random variable representing the "Year-End Obligations" of the insurer instead of the CDR. They obtained similar formulas for the prediction error and verified the numerical results with the help of the bootstrap method. They have noticed that their results for aggregated accident years always lie below the analytical ones obtained in [6]. The reason for this is that there is one redundant term in (4.25) of [6]. This is now corrected, see formula (A.4) below. Let us mention that the ideas presented in [6, 10] were already successfully implemented in practice. Prediction error estimates of Year-End Obligations in the overdispersed Poisson model have been derived by ISVAP [5] in a field study on a large sample of Italian MTPL companies. A field study in line with [6, 10] has been published by AISAM-ACME [1]. Moreover, we would also like to mention that during the writing of this paper we have learned that simultaneously similar ideas have been developed by Böhm-Glaab [2].

Modelling The Claims Development Result For Solvency Purposes

2. METHODOLOGY

2.1 Notation

We denote cumulative payments for accident year $i \in \{0,\ldots,I\}$ until development year $j \in \{0,\ldots,J\}$ by $C_{i,j}$. This means that the ultimate claim for accident year i is given by $C_{i,J}$. For simplicity, we assume that $I = J$ (note that all our results can be generalized to the case $I > J$). Then the outstanding loss liabilities for accident year $i \in \{0,\ldots,I\}$ at time $t = I$ are given by

$$R_i^I = C_{i,J} - C_{i,I-i}, \tag{2.1}$$

and at time $t = I+1$ they are given by

$$R_i^{I+1} = C_{i,J} - C_{i,I-i+1}. \tag{2.2}$$

Let

$$D_I = \{C_{i,j};\, i + j \le I \text{ and } i \le I\} \tag{2.3}$$

denote the claims data available at time $t = I$ and

$$D_{I+1} = \{C_{i,j};\, i + j \le I+1 \text{ and } i \le I\} = D_I \cup \{C_{i,I-i+1};\, i \le I\} \tag{2.4}$$

denote the claims data available one period later, at time $t = I+1$. That is, if we go one step ahead in time from I to $I+1$, we obtain new observations $\{C_{i,I-i+1};\, i \le I\}$ on the new diagonal of the claims development triangle (cf. Figure 1). More formally, this means that we get an enlargement of the σ-field generated by the observations D_I to the σ-field generated by the observations D_{I+1}, i.e.

$$\sigma(D_I) \to \sigma(D_{I+1}). \tag{2.5}$$

2.2 Distribution-free chain-ladder method

We study the claims development process and the CDR within the framework of the well-known distribution-free CL method. That is, we assume that the cumulative payments $C_{i,j}$ satisfy the assumptions of the distribution-free CL model. The distribution-free CL model has been introduced by Mack [7] and has been used by many other actuaries. It is probably the most popular claims reserving method because it is simple and it delivers, in general, very accurate results.

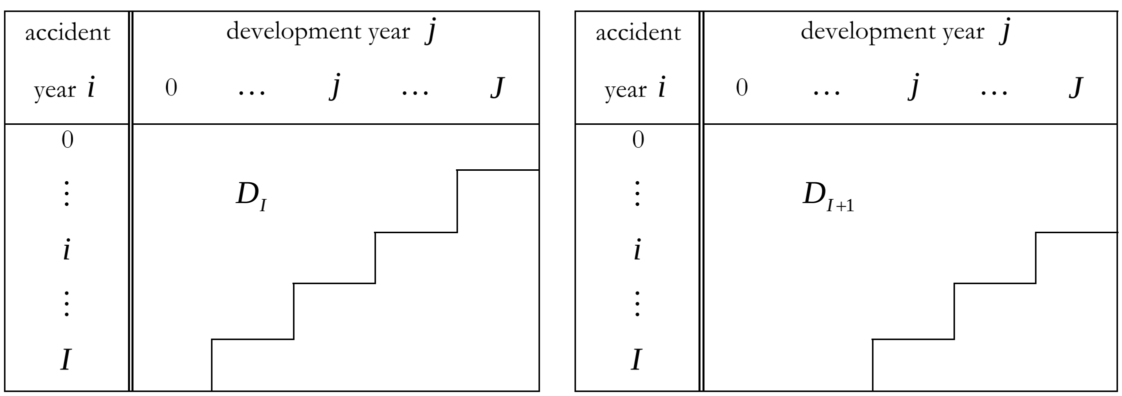


Figure 1: Loss development triangle at time $t = I$ and $t = I + 1$

Model Assumptions 2.1

- Cumulative payments $C_{i,j}$ in different accident years $i \in \{0, \ldots, I\}$ are independent.
- $(C_{i,j})_{j \geq 0}$ are Markov processes and there exist constants $f_j > 0$, $\sigma_j > 0$ such that for all $1 \leq j \leq J$ and $0 \leq i \leq I$ we have

$$E\left[C_{i,j} \,\middle|\, C_{i,j-1}\right] = f_{j-1}\, C_{i,j-1}, \tag{2.6}$$

$$Var\left(C_{i,j} \,\middle|\, C_{i,j-1}\right) = \sigma_{j-1}^2\, C_{i,j-1}. \tag{2.7}$$

□

Remarks 2.2

- In the original work of Mack [7] there were weaker assumptions for the definition of the distribution-free CL model, namely the Markov process assumption was replaced by an assumption only on the first two moments (see also Wüthrich-Merz [9]).

- The derivation of an estimate for the estimation error in [10] was done in a time-series framework. This imposes stronger model assumptions. Note also that in (2.7) we require that the cumulative claims $C_{i,j}$ are positive in order to get a meaningful variance assumption.

Model Assumptions 2.1 imply (using the tower property of conditional expectations)

$$E\left[C_{i,J} \,\middle|\, D_I\right] = C_{i,I-i} \prod_{j=I-i}^{J-1} f_j \qquad \text{and} \qquad E\left[C_{i,J} \,\middle|\, D_{I+1}\right] = C_{i,I-i+1} \prod_{j=I-i+1}^{J-1} f_j \ . \tag{2.8}$$

This means that for known CL factors f_j we are able to calculate the conditionally expected ultimate claim $C_{i,J}$ given the information D_I and D_{I+1}, respectively.

Of course, in general, the CL factors f_j are not known and need to be estimated. Within the framework of the CL method this is done as follows:

1. At time $t = I$, given information D_I, the CL factors f_j are estimated by

$$\hat{f}_j^I = \frac{\sum_{i=0}^{I-j-1} C_{i,j+1}}{S_j^I}, \qquad \text{where} \qquad S_j^I = \sum_{i=0}^{I-j-1} C_{i,j} \ . \tag{2.9}$$

2. At time $t = I+1$, given information D_{I+1}, the CL factors f_j are estimated by

Modelling The Claims Development Result For Solvency Purposes

$$\hat{f}_j^{I+1} = \frac{\sum_{i=0}^{I-j} C_{i,j+1}}{S_j^{I+1}}, \qquad \text{where} \qquad S_j^{I+1} = \sum_{i=0}^{I-j} C_{i,j} . \tag{2.10}$$

This means the CL estimates $\hat{f}_j^{I+1}$ at time $I+1$ use the increase in information about the claims development process in the new observed accounting year $(I, I+1]$ and are therefore based on the additional observation $C_{I-j,\,j+1}$.

Mack [7] proved that these are unbiased estimators for f_j and, moreover, that $\hat{f}_j^m$ and $\hat{f}_l^m$ ($m = I$ or $I+1$) are uncorrelated random variables for $j \neq l$ (see Theorem 2 in Mack [7] and Lemma 2.5 in [9]). This implies that, given $C_{i,I-i}$,

$$\hat{C}_{i,j}^I = C_{i,I-i}\ \hat{f}_{I-i}^I \cdots \hat{f}_{j-2}^I\ \hat{f}_{j-1}^I \tag{2.11}$$

is an unbiased estimator for $E\left[C_{i,j} \middle| D_I\right]$ with $j \geq I-i$ and, given $C_{i,I-i+1}$,

$$\hat{C}_{i,j}^{I+1} = C_{i,I-i+1}\ \hat{f}_{I-i+1}^{I+1} \cdots \hat{f}_{j-2}^{I+1}\ \hat{f}_{j-1}^{I+1} \tag{2.12}$$

is an unbiased estimator for $E\left[C_{i,j} \middle| D_{I+1}\right]$ with $j \geq I-i+1$.

Remarks 2.3

- The realizations of the estimators $\hat{f}_0^I, \ldots, \hat{f}_{J-1}^I$ are known at time $t = I$, but the realizations of $\hat{f}_0^{I+1}, \ldots, \hat{f}_{J-1}^{I+1}$ are unknown since the observations $C_{I,1}, \ldots, C_{I-J+1,J}$ during the accounting year $(I, I+1]$ are unknown at time I.

Modelling The Claims Development Result For Solvency Purposes

- When indices of accident and development years are such that there are no factor products in (2.11) or (2.12), an empty product is replaced by 1. For example, $\hat{C}^{I}_{i,I-i} = C_{i,I-i}$ and $\hat{C}^{I+1}_{i,I-i+1} = C_{i,I-i+1}$.

- The estimators $\hat{C}^{I+1}_{i,j}$ are based on the CL estimators at time $I+1$ and therefore use the increase in information given by the new observations in the accounting year from I to $I+1$.

2.3 Conditional mean square error of prediction

Assume that we are at time I, that is, we have information D_I and our goal is to predict the random variable $C_{i,J}$. Then, $\hat{C}^{I}_{i,J}$ given in (2.11) is a D_I-measurable predictor for $C_{i,J}$. At time I, we measure the prediction uncertainty with the so-called conditional mean square error of prediction (MSEP) which is defined by

$$msep_{C_{i,J} \mid D_I}\left(\hat{C}^{I}_{i,J}\right) = E\left[\left(C_{i,J} - \hat{C}^{I}_{i,J}\right)^2 \mid D_I\right] \tag{2.13}$$

That is, we measure the prediction uncertainty in the $L^2\left(P\left[\,\cdot \mid D_I\right]\right)$-distance. Because $\hat{C}^{I}_{i,J}$ is D_I-measurable this can easily be decoupled into process variance and estimation error:

$$msep_{C_{i,J} \mid D_I}\left(\hat{C}^{I}_{i,J}\right) = Var\left(C_{i,J} \mid D_I\right) + \left(E\left[C_{i,J} \mid D_I\right] - \hat{C}^{I}_{i,J}\right)^2. \tag{2.14}$$

This means that $\hat{C}^{I}_{i,J}$ is used as predictor for the random variable $C_{i,J}$ and as estimator for the expected value $E\left[C_{i,J} \mid D_I\right]$ at time I. Of course, if the conditional expectation $E\left[C_{i,J} \mid D_I\right]$ is known at time I (i.e. the CL factors f_j are known), it is used as predictor for $C_{i,J}$ and the estimation error term vanishes. For more information on conditional and unconditional MSEP's we refer to Chapter 3 in [9]:

2.4 Claims development result (CDR)

We ignore any prudential margin and assume that claims reserves are set equal to the expected outstanding loss liabilities conditional on the available information at time I and $I+1$, respectively. That is, in our understanding "best estimate" claims reserves correspond to conditional expectations which implies a self-financing property (see Corollary 2.6 in [8]). For known CL factors f_j the conditional expectation $E\left[C_{i,J} \mid D_I\right]$ is known and is therefore used as predictor for $C_{i,J}$ at time I. Similarly, at time $I+1$ the conditional expectation $E\left[C_{i,J} \mid D_{I+1}\right]$ is used as predictor for $C_{i,J}$. Then the **true claims development result** (true CDR) for accounting year $\left(I, I+1\right]$ is defined as follows.

Definition 2.4 (True claims development result)

The **true CDR** *for accident year* $i \in \{1,\ldots,I\}$ *in accounting year* $\left(I, I+1\right]$ *is given by*

$$\begin{aligned} CDR_i\left(I+1\right) &= E\left[R_i^I \mid D_I\right] - \left(X_{i,I-i+1} + E\left[R_i^{I+1} \mid D_{I+1}\right]\right) \\ &= E\left[C_{i,J} \mid D_I\right] - E\left[C_{i,J} \mid D_{I+1}\right], \end{aligned} \tag{2.15}$$

where $X_{i,I-i+1} = C_{i,I-i+1} - C_{i,I-i}$ *denotes the incremental payments. Furthermore, the true aggregate is given by*

$$\sum_{i=1}^{I} CDR_i\left(I+1\right). \tag{2.16}$$

Using the martingale property we see that

Modelling The Claims Development Result For Solvency Purposes

$$E\left[CDR_i(I+1) \mid D_I\right] = 0. \tag{2.17}$$

This means that for known CL factors f_j the expected true CDR (viewed from time I) is equal to zero. Therefore, for known CL factors f_j we refer to $CDR_i(I+1)$ as the **true** CDR. This also justifies the fact that in the budget values of the income statement position c) "loss experience prior accident years" is predicted by $0 (see position c) in Table 1).
The prediction uncertainty of this prediction 0 can then easily be calculated, namely,

$$msep_{CDR_i(I+1)|D_I}(0) = Var\left(CDR_i(I+1) \mid D_I\right) = E\left[C_{i,J} \mid D_I\right]^2 \frac{\sigma_{I-i}^2 / f_{I-i}^2}{C_{i,I-i}}. \tag{2.18}$$

For a proof we refer to formula (5.5) in [10] (apply recursively the model assumptions), and the aggregation of accident years can easily be done because accident years i are independent according to Model Assumptions 2.1.

Unfortunately the CL factors f_j are in general not known and therefore the true CDR is not observable. Replacing the unknown factors by their estimators, i.e., replacing the expected ultimate claims $E\left[C_{i,J} \mid D_I\right]$ and $E\left[C_{i,J} \mid D_{I+1}\right]$ with their estimates $\hat{C}_{i,J}^{I}$ and $\hat{C}_{i,J}^{I+1}$, respectively, the true CDR for accident year i $\left(1 \le i \le I\right)$ in accounting year $\left(I, I+1\right]$ is predicted/estimated in the CL method by:

Definition 2.5 (Observable claims development result)

The **observable CDR** *for accident year* $i \in \{1, \ldots, I\}$ *in accounting year* $\left(I, I+1\right]$ *is given by*

$$\widehat{CDR}_i(I+1) = \hat{R}_i^{D_I} - \left(X_{i,I-i+1} + \hat{R}_i^{D_{I+1}}\right) = \hat{C}_{i,J}^{I} - \hat{C}_{i,J}^{I+1}, \tag{2.19}$$

where $\hat{R}_i^{D_I}$ *and* $\hat{R}_i^{D_{I+1}}$ *are defined below by* (2.21) *and* (2.22), *respectively. Furthermore, the observable aggregate* CDR *is given by*

$$\sum_{i=1}^{I} C\hat{D}R_i\left(I+1\right). \tag{2.20}$$

Note that under the Model Assumptions 2.1, given $C_{i,I-i}$,

$$\hat{R}_i^{D_I} = \hat{C}_{i,J}^{I} - C_{i,I-i} \qquad \left(1 \le i \le I\right), \tag{2.21}$$

is an unbiased estimator for $E\left[R_i^I \mid D_I\right]$ and, given $C_{i,I-i+1}$,

$$\hat{R}_i^{D_{I+1}} = \hat{C}_{i,J}^{I+1} - C_{i,I-i+1} \qquad \left(1 \le i \le I\right), \tag{2.22}$$

is an unbiased estimator for $E\left[R_i^{I+1} \mid D_{I+1}\right]$.

Remarks 2.6

- We point out the (non-observable) true claims development result $CDR_i\left(I+1\right)$ is approximated by an observable claims development result $C\hat{D}R_i\left(I+1\right)$. In the next section we quantify the quality of this approximation (retrospective view).

- Moreover, the observable claims development result $C\hat{D}R_i\left(I+1\right)$ is the position that occurs in the P&L statement at Dec. 31, year I. This position is in the budget statement predicted by 0. In the next section we also measure the quality of this prediction, which determines the solvency requirements (prospective view).

- We emphasize that such a solvency consideration is only a one-year view. The remaining run-off can, for example, be treated with a cost-of-capital loading that is based on the one-year observable claims development result (this has, for example, been done in the Swiss Solvency Test).

3. MSEP OF THE CLAIMS DEVELOPMENT RESULT

Our goal is to quantify the following two quantities:

$$msep_{\widehat{CDR}_i(I+1)|D_I}(0) = E\left[\left(\widehat{CDR}_i(I+1) - 0\right)^2 \mid D_I\right], \tag{3.1}$$

$$msep_{CDR_i(I+1)|D_I}\left(\widehat{CDR}_i(I+1)\right) = E\left[\left(CDR_i(I+1) - \widehat{CDR}_i(I+1)\right)^2 \mid D_I\right]. \tag{3.2}$$

- The first conditional MSEP gives the prospective solvency point of view. It quantifies the prediction uncertainty in the budget value 0 for the observable claims development result at the end of the accounting period. In the solvency margin we need to hold risk capital for possible negative deviations of $CDR_i(I+1)$ from 0.

- The second conditional MSEP gives a retrospective point of view. It analyzes the distance between the true CDR and the observable CDR. It may, for example, answer the question whether the true CDR could also be positive (if we would know the true CL factors) when we have an observable CDR given by \$ -40'000 (see Table 1). Hence, the retrospective view separates pure randomness (process variance) from parameter estimation uncertainties.

In order to quantify the conditional MSEP's we need an estimator for the variance parameters σ_j^2. An unbiased estimate for σ_j^2 is given by (see Lemma 3.5 in [9])

$$\hat{\sigma}_j^2 = \frac{1}{I-j-1} \sum_{i=0}^{I-j-1} C_{i,j} \left(\frac{C_{i,j+1}}{C_{i,j}} - \hat{f}_j \right)^2. \tag{3.3}$$

Modelling The Claims Development Result For Solvency Purposes

3.1 Single accident year

In this section we give estimators for the two conditional MSEP's defined in (3.1)-(3.2). For their derivation we refer to the appendix. We define

$$\hat{\Delta}_{i,J}^{I} = \frac{\hat{\sigma}_{I-i}^{2} \Big/ \left(\hat{f}_{I-i}^{I}\right)^{2}}{S_{I-i}^{I}} + \sum_{j=I-i+1}^{J-1} \left(\frac{C_{I-j,j}}{S_{j}^{I+1}}\right)^{2} \frac{\hat{\sigma}_{j}^{2} \Big/ \left(\hat{f}_{j}^{I}\right)^{2}}{S_{j}^{I}}, \tag{3.4}$$

$$\hat{\Phi}_{i,J}^{I} = \sum_{j=I-i+1}^{J-1} \left(\frac{C_{I-j,j}}{S_{j}^{I+1}}\right)^{2} \frac{\hat{\sigma}_{j}^{2} \Big/ \left(\hat{f}_{j}^{I}\right)^{2}}{C_{I-j,j}}, \tag{3.5}$$

$$\hat{\Psi}_{i}^{I} = \frac{\hat{\sigma}_{I-i}^{2} \Big/ \left(\hat{f}_{I-i}^{I}\right)^{2}}{C_{i,I-i}} \tag{3.6}$$

and

$$\hat{\Gamma}_{i,J}^{I} = \hat{\Phi}_{i,J}^{I} + \hat{\Psi}_{i}^{I} \geq \hat{\Phi}_{i,J}^{I}. \tag{3.7}$$

We are now ready to give estimators for all the error terms. First of all the variance of the true CDR given in (2.18) is estimated by

$$V\hat{a}r\left(CDR_{i}(I+1)\middle| D_{I}\right) = \left(\hat{C}_{i,J}^{I}\right)^{2} \hat{\Psi}_{i}^{I}. \tag{3.8}$$

The estimator for the conditional MSEP's are then given by:

Modelling The Claims Development Result For Solvency Purposes

Result 3.1 (Conditional MSE estimator for a single accident year)

We estimate the conditional MSEP's given in (3.1)-(3.2) *by*

$$\widehat{msep}_{\widehat{CDR}_i(I+1)|D_I}(0) = \left(\widehat{C}_{i,J}^I\right)^2 \left(\widehat{\Gamma}_{i,J}^I + \widehat{\Delta}_{i,J}^I\right), \tag{3.9}$$

$$\widehat{msep}_{\widehat{CDR}_i(I+1)|D_I}\left(\widehat{CDR}_i(I+1)\right) = \left(\widehat{C}_{i,J}^I\right)^2 \left(\widehat{\Phi}_{i,J}^I + \widehat{\Delta}_{i,J}^I\right). \tag{3.10}$$

This immediately implies that we have

$$\begin{aligned}\widehat{msep}_{\widehat{CDR}_i(I+1)|D_I}(0) &= \widehat{msep}_{CDR_i(I+1)|D_I}\left(\widehat{CDR}_i(I+1)\right) + \widehat{Var}\left(CDR_i(I+1)\middle| D_I\right) \\ &\geq \widehat{msep}_{CDR_i(I+1)|D_I}\left(\widehat{CDR}_i(I+1)\right).\end{aligned} \tag{3.11}$$

Note that this is intuitively clear since the true and the observable CDR should move into the same direction according to the observations in accounting year $(I, I+1]$. However, the first line in (3.11) is slightly misleading. Note that we have derived estimators which give an equality on the first line of (3.11). However, this equality holds true only for our estimators where we neglect uncertainties in higher order terms. Note, as already mentioned, for typical real data examples higher order terms are of negligible order which means that we get an approximate equality on the first line of (3.11) (see also derivation in (A.2)). This is similar to the findings presented in Chapter 3 of [9].

3.2 Aggregation over prior accident years

When aggregating over prior accident years, one has to take into account the correlations between different accident years, since the same observations are used to estimate the CL factors and are then applied to different accident years (see also Section 3.2.4 in [9]). Based on the definition of the conditional MSEP for the true aggregate CDR around the aggregated observable CDR the following estimator is obtained:

Result 3.2 (Conditional MSEP for aggregated accident years, part I)

For aggregated accident years we obtain the following estimator

$$\widehat{msep}_{\sum_{i=1}^{I} CDR_i(I+1)\mid D_I}\left(\sum_{i=1}^{I} \widehat{CDR}_i(I+1)\right) \tag{3.12}$$
$$= \sum_{i=1}^{I} \widehat{msep}_{CDR_i(I+1)\mid D_I}\left(\widehat{CDR}_i(I+1)\right) + 2\sum_{k>i>0} \hat{C}_{i,J}^{I}\, \hat{C}_{k,J}^{I}\left(\hat{\Phi}_{i,J}^{I} + \hat{\Lambda}_{i,J}^{I}\right)$$

with

$$\hat{\Lambda}_{i,J}^{I} = \frac{C_{i,I-i}}{S_{I-i}^{I+1}} \frac{\hat{\sigma}_{I-i}^{2} \Big/ \left(\hat{f}_{I-i}^{I}\right)^2}{S_{I-i}^{I}} + \sum_{j=I-i+1}^{J-1} \left(\frac{C_{I-j,j}}{S_{j}^{I+1}}\right)^2 \frac{\hat{\sigma}_{j}^{2} \Big/ \left(\hat{f}_{j}^{I}\right)^2}{S_{j}^{I}}. \tag{3.13}$$

For the conditional MSEP of the aggregated observable CDR around 0 we need an additional definition.

$$\hat{\Xi}_{i,J}^{I} = \hat{\Phi}_{i,J}^{I} + \frac{\hat{\sigma}_{I-i}^{2} \Big/ \left(\hat{f}_{I-i}^{I}\right)^2}{S_{I-i}^{I+1}} \geq \hat{\Phi}_{i,J}^{I}. \tag{3.14}$$

Result 3.3 (Conditional MSEP for aggregated accident years, part II)

For aggregated accident years we obtain the following estimator

$$\widehat{msep}_{\sum_{i=1}^{I} \widehat{CDR}_i(I+1)\mid D_I}(0) \tag{3.15}$$
$$= \sum_{i=1}^{I} \widehat{msep}_{\widehat{CDR}_i(I+1)\mid D_I}(0) + 2\sum_{k>i>0} \hat{C}_{i,J}^{I}\, \hat{C}_{k,J}^{I}\left(\hat{\Xi}_{i,J}^{I} + \hat{\Lambda}_{i,J}^{I}\right).$$

Note that (3.15) can be rewritten as follows:

$$\begin{aligned}
&\hat{msep}_{\sum_{i=1}^{I} C\hat{D}R_i(I+1)|D_I}(0) \\
&\quad = \hat{msep}_{\sum_{i=1}^{I} CDR_i(I+1)|D_I}\left(\sum_{i=1}^{I} C\hat{D}R_i(I+1)\right) \\
&\qquad + \sum_{i=1}^{I} V\hat{a}r\left(CDR_i(I+1)\middle|D_I\right) + 2\sum_{k>i>0} \hat{C}_{i,J}^{I}\hat{C}_{k,J}^{I}(\hat{\Xi}_{i,J}^{I} - \hat{\Phi}_{i,J}^{I}) \\
&\quad \geq \hat{msep}_{\sum_{i=1}^{I} CDR_i(I+1)|D_I}\left(\sum_{i=1}^{I} C\hat{D}R_i(I+1)\right).
\end{aligned} \tag{3.16}$$

Hence, we obtain the same decoupling for aggregated accident years as for single accident years.

Remarks 3.4 (Comparison to the classical Mack [7] formula)

In Results 3.1-3.3 we have obtained a natural split into process variance and estimation error. However, this split has no longer this clear distinction as it appears. The reason is that the process variance also influences the volatility of $\hat{f}_j^{I+1}$ and hence is part of the estimation error. In other approaches one may obtain other splits, e.g. in the credibility chain ladder method (see Bühlmann et al. [3]) one obtains a different split. Therefore we modify Results 3.1.-3.3 which leads to a formula that gives interpretations in terms of the classical Mack [7] formula, see also (4.2)-(4.3) below.

Result 3.5

Modelling The Claims Development Result For Solvency Purposes

For single accident years we obtain from Result 3.1

$$\hat{msep}_{\hat{CDR}_i(I+1)|D_I}(0) = \left(\hat{C}_{i,J}^I\right)^2 \left(\hat{\Gamma}_{i,J}^I + \hat{\Delta}_{i,J}^I\right) \tag{3.17}$$

$$= \left(\hat{C}_{i,J}^I\right)^2 \left(\frac{\hat{\sigma}_{I-i}^2 / \left(\hat{f}_{I-i}^I\right)^2}{C_{i,I-i}} + \frac{\hat{\sigma}_{I-i}^2 / \left(\hat{f}_{I-i}^I\right)^2}{S_{I-i}^I} + \sum_{j=I-i+1}^{J-1} \frac{C_{I-j,j}}{S_j^{I+1}} \frac{\hat{\sigma}_j^2 / \left(\hat{f}_j^I\right)^2}{S_j^I} \right).$$

For aggregated accident years we obtain from Result 3.3

$$\hat{msep}_{\sum_{i=1}^{I} \hat{CDR}_i(I+1)|D_I}(0) = \sum_{i=1}^{I} \hat{msep}_{\hat{CDR}(I+1)|D_I}(0) \tag{3.18}$$

$$+ 2 \sum_{k>i>0} \hat{C}_{i,J}^I \hat{C}_{k,J}^I \left[\frac{\hat{\sigma}_{I-i}^2 / \left(\hat{f}_{I-i}^I\right)^2}{S_{I-i}^I} + \sum_{j=I-i+1}^{J-1} \frac{C_{I-j,j}}{S_j^{I+1}} \frac{\hat{\sigma}_j^2 / \left(\hat{f}_j^I\right)^2}{S_j^I} \right].$$

We compare this now to the classical Mack [7] formula. For single accident years the conditional MSEP of the predictor for the ultimate claim is given in Theorem 3 in Mack [7] (see also Estimator 3.12 in [9]). We see from (3.17) that the conditional MSEP of the CDR considers only the first term of the process variance of the classical Mack [7] formula $(j = I - i)$ and for the estimation error the next diagonal is fully considered $(j = I - i)$ but all remaining runoff cells $(j \geq I - i + 1)$ are scaled by $C_{i,I-i} / S_j^{I+1} \leq 1$. For aggregated accident years the conditional MSEP of the predictor for the ultimate claim is given on page 220 in Mack [7] (see also Estimator 3.16 in [9]). We see from (3.18) that the conditional MSEP of the CDR for aggregated accident years considers the estimation error for the next accounting year $(j = I - i)$ and all other accounting years $(j \geq I - i + 1)$ are scaled by $C_{i,I-i} / S_j^{I+1} \leq 1$.

Hence we have obtained a different split that allows for easy interpretations in terms of the Mack [7] formula. However, note that these interpretations only hold true for linear approximations (A.1), otherwise the picture is more involved.

4. NUMERICAL EXAMPLE AND CONCLUSIONS

For our numerical example we use the dataset given in Table 2. The table contains cumulative payments $C_{i,j}$ for accident years $i \in \{0,1,\ldots,8\}$ at time $I = 8$ and at time $I+1=9$. Hence this allows for an explicitly calculation of the observable claims development result.

	$j=0$	1	2	3	4	5	6	7	8
$i=0$	2'202'584	3'210'449	3'468'122	3'545'070	3'621'627	3'644'636	3'669'012	3'674'511	3'678'633
$i=1$	2'350'650	3'553'023	3'783'846	3'840'067	3'865'187	3'878'744	3'898'281	3'902'425	3'906'738
$i=2$	2'321'885	3'424'190	3'700'876	3'798'198	3'854'755	3'878'993	3'898'825	3'902'130	
$i=3$	2'171'487	3'165'274	3'395'841	3'466'453	3'515'703	3'548'422	3'564'470		
$i=4$	2'140'328	3'157'079	3'399'262	3'500'520	3'585'812	3'624'784			
$i=5$	2'290'664	3'338'197	3'550'332	3'641'036	3'679'909				
$i=6$	2'148'216	3'219'775	3'428'335	3'511'860					
$i=7$	2'143'728	3'158'581	3'376'375						
$i=8$	2'144'738	3'218'196							
$\hat{f}_j^I$	1.4759	1.0719	1.0232	1.0161	1.0063	1.0056	1.0013	1.0011	
$\hat{f}_j^{I+1}$	1.4786	1.0715	1.0233	1.0152	1.0072	1.0053	1.0011	1.0011	
$\hat{\sigma}_j^2$	911.43	189.82	97.81	178.75	20.64	3.23	0.36	0.04	

Table 2: Run-off triangle (cumulative payments, in \$ 1'000) for time $I=8$ and $I=9$

Modelling The Claims Development Result For Solvency Purposes

Table 2 summarizes the CL estimates $\hat{f}_j^I$ and $\hat{f}_j^{I+1}$ of the age-to-age factors f_j as well as the variance estimates $\hat{\sigma}_j^2$ for $j = 0,\ldots,7$. Since we do not have enough data to estimate σ_7^2 (recall $I = J$) we use the extrapolation given in Mack [7]:

$$\hat{\sigma}_7^2 = \min\left\{\hat{\sigma}_6^2, \hat{\sigma}_5^2, \hat{\sigma}_6^4 \big/ \hat{\sigma}_5^2\right\}. \tag{4.1}$$

Using the estimates $\hat{f}_j^I$ and $\hat{f}_j^{I+1}$ we calculate the claims reserves $\hat{R}_i^{D_I}$ for the outstanding claims liabilities R_i^I at time $t = I$ and $X_{i,I-i+1} + \hat{R}_i^{D_{I+1}}$ for $X_{i,I-i+1} + R_i^{I+1}$ at time $t = I+1$, respectively. This then gives realizations of the observable CDR for single accident years and for aggregated accident years (see Table 3). Observe that we have a negative observable aggregate CDR at time $I+1$ of about \$ -40'000 (which corresponds to position c) in the P&L statement in Table 1).

i	$\hat{R}_i^{D_I}$	$X_{i,I-i+1} + \hat{R}_i^{D_{I+1}}$	$\widehat{CDR}_i(I+1)$
0	0	0	0
1	4'378	4'313	65
2	9'348	7'649	1'698
3	28'392	24'046	4'347
4	51'444	66'494	-15'050
5	111'811	93'451	18'360
6	187'084	189'851	-2'767
7	411'864	401'134	10'731
8	1'433'505	1'490'962	-57'458
Total	2'237'826	2'277'900	-40'075

Table 3: Realization of the observable CDR at time $t = I+1$, in \$ 1'000

The question which we now have is whether the true aggregate CDR could also be positive if we had known the true CL factors f_j at time $t = I$ (retrospective view). We therefore perform the variance and MSEP analysis using the results of Section 3. Table 4 provides the estimates for single and aggregated accident years.

On the other hand we would like to know, how this observation of \$ -40'000 corresponds to the prediction uncertainty in the budget values, where we have predicted that the CDR is \$ 0 (see position c) in Table 1). This is the prospective (solvency) view.

We observe that the estimated standard deviation of the true aggregate CDR is equal to \$ 65'412, which means that it is not unlikely to have the true aggregate CDR in the range of about \$ $\pm$40'000. Moreover, we see that the square root of the estimate for the MSEP between true and observable CDR is of size \$ 33'856 (see Table 4), this means that it is likely that the true CDR has the same sign as the observable CDR which is \$ -40'000. Therefore also the knowledge of the true CL factors would probably have led to a negative claims development experience.

Moreover, note that the prediction 0 in the budget values has a prediction uncertainty relative to the observable CDR of \$ 81'080 which means that it is not unlikely to have an observable CDR of \$ -40'000. In other words the solvency capital/risk margin for the CDR should directly be related to this value of \$ 81'080.

i	$\hat{R}_i^{D_I}$	$\hat{Var}^{1/2}$	$\hat{msep}_{CDR\mid D_I}\left(\hat{CDR}\right)^{1/2}$	$\hat{msep}_{\hat{CDR}\mid D_I}(0)^{1/2}$	$msep_{Mack}^{1/2}$
0	0				
1	4'378	395	407	567	567
2	9'348	1'185	900	1'488	1'566
3	28'392	3'395	1'966	3'923	4'157
4	51'444	8'673	4'395	9'723	10'536
5	111'811	25'877	11'804	28'443	30'319
6	187'084	18'875	9'100	20'954	35'967
7	411'864	25'822	11'131	28'119	45'090
8	1'433'505	49'978	18'581	53'320	69'552
$\text{cov}^{1/2}$		0	20'754	39'746	50'361
Total	2'237'826	65'412	33'856	81'080	108'401

Table 4: Volatilities of the estimates in \$ 1'000 with:

Modelling The Claims Development Result For Solvency Purposes

$\hat{R}_i^{D_I}$	estimated reserves at time $t = I$, cf. (2.21)
$\hat{Var}^{1/2}$	estimated std. dev. of the true CDR, cf. (3.8)
$\hat{msep}_{\hat{CDR}\mid D_I}\left(\hat{CDR}\right)^{1/2}$	estimated $msep^{1/2}$ between true and observable CDR, cf. (3.10) and (3.12)
$\hat{msep}_{\hat{CDR}\mid D_I}(0)^{1/2}$	prediction std. dev. of 0 compared to $\hat{CDR}_i(I+1)$, cf. (3.9) and (3.15)
$msep_{Mack}^{1/2}$	$msep^{1/2}$ of the ultimate claim, cf. Mack [7] and (4.3)

Note that we only consider the one-year uncertainty of the claims reserves run-off. This is exactly the short term view/picture that should look fine to get to the long term. In order to treat the full run-off one can then add, for example, a cost-of-capital margin to the remaining run-off which ensures that the future solvency capital can be financed. We emphasize that it is important to add a margin which ensures the smooth run-off of the whole liabilities after the next accounting year.

Finally, these results are compared to the classical Mack formula [7] for the estimate of the conditional MSEP of the ultimate claim $C_{i,J}$ by $\hat{C}_{i,J}^I$ in the distribution-free CL model. The Mack formula [7] gives the total uncertainty of the full run-off (long term view) which estimates

$$msep_{Mack}\left(\hat{C}_{i,J}^I\right) = E\left[\left(C_{i,J} - \hat{C}_{i,J}^I\right)^2 \mid D_I\right] \tag{4.2}$$

and

$$msep_{Mack}\left(\sum_{i=1}^{I}\hat{C}_{i,J}^I\right) = E\left[\left(\sum_{i=1}^{I}C_{i,J} - \sum_{i=1}^{I}\hat{C}_{i,J}^I\right)^2 \mid D_I\right], \tag{4.3}$$

see also Estimator 3.16 in [9]. Notice that the information in the next accounting year (diagonal $I+1$) contributes substantially to the total uncertainty of the total ultimate claim over prior accident years. That is, the uncertainty in the next accounting year is \$ 81’080 and the total uncertainty is $ 108'401. Note that we have chosen a short-tailed line of business so it is clear that a lot of uncertainty is already contained in the next accounting year. Generally speaking, the portion of uncertainty which is already contained in the next accounting year is larger for short-tailed business than for long-tailed business since in long-tailed business the adverse movements in the claims reserves emerge slowly over many years. If one chooses long-tailed lines of business then the one-year risk is about 2/3 of the full run-off risk. This observation is inline with a European field study in different companies, see AISAM-ACME [1].

APPENDIX A. PROOFS AND DERIVATIONS

Assume that a_j are positive constants with $1 >> a_j$ then we have

$$\prod_{j=1}^{J}\left(1+a_j\right)-1 \approx \sum_{j=1}^{J} a_j \,, \tag{A.1}$$

where the right-hand side is a lower bound for the left-hand side. Using the above formula we will approximate all product terms from our previous work [10] by summations.

Derivation of Result 3.1. We first give the derivation of Result 3.1 for a single accident year. Note that the term $\hat{\Delta}_{i,J}^{I}$ is given in formula (3.10) of [10]. Henceforth there remains to derive the terms $\hat{\Phi}_{i,J}^{I}$ and $\hat{\Gamma}_{i,J}^{I}$.

For the term $\hat{\Phi}_{i,J}^{I}$ we obtain from formula (3.9) in [10]

$$\left[1+\frac{\hat{\sigma}_{I-i}^{2}\Big/\left(\hat{f}_{I-i}^{I}\right)^{2}}{C_{i,I-i}}\right]\left(\prod_{j=I-i+1}^{J-1}\left(1+\frac{\hat{\sigma}_{j}^{2}\Big/\left(\hat{f}_{j}^{I}\right)^{2}}{C_{I-j,j}}\left(\frac{C_{I-j,j}}{S_{j}^{I+1}}\right)^{2}\right)-1\right)$$

$$\approx \left[1+\frac{\hat{\sigma}_{I-i}^2 \Big/ \left(\hat{f}_{I-i}^{I}\right)^2}{C_{i,I-i}}\right] \sum_{j=I-i+1}^{J-1} \frac{\hat{\sigma}_j^2 \Big/ \left(\hat{f}_j^{I}\right)^2}{C_{I-j,j}} \left(\frac{C_{I-j,j}}{S_j^{I+1}}\right)^2 \tag{A.2}$$

$$\approx \sum_{j=I-i+1}^{J-1} \frac{\hat{\sigma}_j^2 \Big/ \left(\hat{f}_j^{I}\right)^2}{C_{I-j,j}} \left(\frac{C_{I-j,j}}{S_j^{I+1}}\right)^2 = \hat{\Phi}_{i,J}^{I},$$

where the approximations are accurate because $1 >> \frac{\hat{\sigma}_{I-i}^2 \Big/ \left(\hat{f}_{I-i}^{I}\right)^2}{C_{i,I-i}}$ for typical claims reserving data.

For the term $\hat{\Gamma}_{i,J}^{I}$ we obtain from (3.16) in [10]

$$\left(\left[1+\frac{\hat{\sigma}_{I-i}^2 \Big/ \left(\hat{f}_{I-i}^{I}\right)^2}{C_{i,I-i}}\right] \prod_{j=I-i+1}^{J-1} \left(1+\frac{\hat{\sigma}_j^2 \Big/ \left(\hat{f}_j^{I}\right)^2}{C_{I-j,j}} \left(\frac{C_{I-j,j}}{S_j^{I+1}}\right)^2\right)\right) - 1 \tag{A.3}$$

$$\approx \frac{\hat{\sigma}_{I-i}^2 \Big/ \left(\hat{f}_{I-i}^{I}\right)^2}{C_{i,I-i}} + \sum_{j=I-i+1}^{J-1} \frac{\hat{\sigma}_j^2 \Big/ \left(\hat{f}_j^{I}\right)^2}{C_{I-j,j}} \left(\frac{C_{I-j,j}}{S_j^{I+1}}\right)^2 = \hat{\Psi}_i^{I} + \hat{\Phi}_{i,J}^{I} = \hat{\Gamma}_{i,J}^{I}.$$

Henceforth, Result 3.1 is obtained from (3.8), (3.14) and (3.15) in [10].

□

Derivations of Results 3.2 and 3.3. We now turn to Result 3.2. All that remains to derive are the correlation terms.

We start with the derivation of $\hat{\Lambda}_{k,J}^{I}$ (this differs from the calculation in [6]). From (4.24)-(4.25) in [6] we see that for $i < k$ the cross covariance term of the estimation error

$$C_{i,I-i}^{-1}\, C_{k,I-k}^{-1}\, E\left[\widehat{CDR}_i(I+1) \middle|\, D_I\right] E\left[\widehat{CDR}_k(I+1) \middle|\, D_I\right]$$

is estimated by resampled values $\hat{f}_j$, given D_I, which implies

Modelling The Claims Development Result For Solvency Purposes

$$E\left[\left(\prod_{j=I-i}^{J-1}\hat{f}_j^I - f_{I-i}\prod_{j=I-i+1}^{J-1}\left(\frac{S_j^I}{S_j^{I+1}}\hat{f}_j^I + f_j\frac{C_{I-j,j}}{S_j^{I+1}}\right)\right)\right.$$

$$\left.\times\left(\prod_{j=I-k}^{J-1}\hat{f}_j^I - f_{I-k}\prod_{j=I-k+1}^{J-1}\left(\frac{S_j^I}{S_j^{I+1}}\hat{f}_j^I + f_j\frac{C_{I-j,j}}{S_j^{I+1}}\right)\right)\middle| D_I\right] \qquad \text{(A.4)}$$

$$=\left(\prod_{j=I-i}^{J-1}E\left[\left(\hat{f}_j^I\right)^2 \mid D_I\right] + f_{I-i}^2\prod_{j=I-i+1}^{J-1}E\left[\left(\frac{S_j^I}{S_j^{I+1}}\hat{f}_j^I + f_j\frac{C_{I-j,j}}{S_j^{I+1}}\right)^2 \mid D_I\right]\right.$$

$$- f_{I-i}^2\prod_{j=I-i+1}^{J-1}E\left[\hat{f}_j^I\left(\frac{S_j^I}{S_j^{I+1}}\hat{f}_j^I + f_j\frac{C_{I-j,j}}{S_j^{I+1}}\right)\middle| D_I\right]$$

$$\left.-\prod_{j=I-i}^{J-1}E\left[\hat{f}_j^I\left(\frac{S_j^I}{S_j^{I+1}}\hat{f}_j^I + f_j\frac{C_{I-j,j}}{S_j^{I+1}}\right)\middle| D_I\right]\right)\prod_{j=I-k}^{I-i-1}f_j\,.$$

Note that the last two lines differ from (4.25) in [6]. This last expression is now equal to (see also Section 4.1.2 in [6])

$$=\left\{\prod_{j=I-i}^{J-1}\left(\frac{\sigma_j^2}{S_j^I} + f_j^2\right) + f_{I-i}^2\prod_{j=I-i+1}^{J-1}\left(\left(\frac{S_j^I}{S_j^{I+1}}\right)^2\frac{\sigma_j^2}{S_j^I} + f_j^2\right)\right.$$

$$\left.- f_{I-i}^2\prod_{j=I-i+1}^{J-1}\left(\frac{S_j^I}{S_j^{I+1}}\frac{\sigma_j^2}{S_j^I} + f_j^2\right) - \prod_{j=I-i}^{J-1}\left(\frac{S_j^I}{S_j^{I+1}}\frac{\sigma_j^2}{S_j^I} + f_j^2\right)\right\}\prod_{j=I-k}^{I-i-1}f_j\,.$$

Next we use (A.1), so we see that the last line can be approximated by

$$\approx\left\{\sum_{j=I-i}^{J-1}\frac{\sigma_j^2/f_j^2}{S_j^I} + \sum_{j=I-i+1}^{J-1}\left(\frac{S_j^I}{S_j^{I+1}}\right)^2\frac{\sigma_j^2/f_j^2}{S_j^I}\right.$$

$$\left.-\sum_{j=I-i+1}^{J-1}\frac{S_j^I}{S_j^{I+1}}\frac{\sigma_j^2/f_j^2}{S_j^I} - \sum_{j=I-i}^{J-1}\frac{S_j^I}{S_j^{I+1}}\frac{\sigma_j^2/f_j^2}{S_j^I}\right\}\prod_{j=I-k}^{I-i-1}f_j\prod_{j=I-i}^{J-1}f_j^2$$

Modelling The Claims Development Result For Solvency Purposes

$$= \left\{ \frac{\sigma_{I-i}^2 / f_{I-i}^2}{S_{I-i}^I} - \frac{S_{I-i}^I}{S_{I-i}^{I+1}} \frac{\sigma_{I-i}^2 / f_{I-i}^2}{S_{I-i}^I} + \sum_{j=I-i+1}^{J-1} \left(1 - \frac{S_j^I}{S_j^{I+1}} \right)^2 \frac{\sigma_j^2 / f_j^2}{S_j^I} \right\} \prod_{j=I-k}^{I-i-1} f_j \prod_{j=I-i}^{J-1} f_j^2 .$$

Next we note that $1 - S_j^I / S_j^{I+1} = C_{I-j,j} / S_j^{I+1}$ hence this last term is equal to

$$= \left\{ \frac{C_{i,I-i}}{S_{I-i}^{I+1}} \frac{\sigma_{I-i}^2 / f_{I-i}^2}{S_{I-i}^I} + \sum_{j=I-i+1}^{J-1} \left(\frac{C_{I-j,j}}{S_j^{I+1}} \right)^2 \frac{\sigma_j^2 / f_j^2}{S_j^I} \right\} \prod_{j=I-k}^{I-i-1} f_j \prod_{j=I-i}^{J-1} f_j^2 .$$

Hence plugging in the estimators for f_j and σ_j^2 at time I yields the claim.

Hence there remains to calculate the second term in Result 3.2. From (3.13) in [10] we again obtain the claim, using that $1 >> \frac{\hat{\sigma}_{I-i}^2 / \left(\hat{f}_{I-i}^I \right)^2}{C_{i,I-i}}$ for typical claims reserving data.

So there remains to derive Result 3.3. The proof is completely analogous, the term containing $\hat{\Lambda}_{i,J}^I$ was obtained above. The term $\hat{\Xi}_{i,J}^I$ is obtained from (3.17) in [10] analogous to (A.3).

This completes the derivations.

□

5. REFERENCES

[1] AISAM-ACME (2007). AISAM-ACME study on non-life long tail liabilities. Reserve risk and risk margin assessment under Solvency II. October 17, 2007.

[2] Böhm, H., Glaab, H. (2006). Modellierung des Kalenderjahr-Risikos im additiven und multiplikativen Schadenreservierungsmodell. Talk presented at the German ASTIN Colloquium.

[3] Bühlmann, H., De Felice, M., Gisler, A., Moriconi, F., Wüthrich, M.V. (2008). Recursive credibility formula for chain ladder factors and the claims development result. Preprint, ETH Zurich.

[4] De Felice, M., Moriconi, F. (2006). Process error and estimation error of year-end reserve estimation in the distribution free chain-ladder model. Alef Working Paper, Rome, November 2006.

Modelling The Claims Development Result For Solvency Purposes

[5] ISVAP (2006). Reserves requirements and capital requirements in non-life insurance. An analysis of the Italian MTPL insurance market by stochastic claims reserving models. Report prepared by De Felice, M., Moriconi, F., Matarazzo, L., Cavastracci, S. and Pasqualini, S., October 2006.

[6] Merz, M., Wüthrich, M.V. (2007). Prediction error of the expected claims development result in the chain ladder method. *Bulletin of Swiss Association of Actuaries*, No. 1, 117-137.

[7] Mack, T. (1993). Distribution-free calculation of the standard error of chain ladder reserve estimates. *Astin Bulletin*, Vol. 23, No. 2, 213-225.

[8] Wüthrich, M.V., Bühlmann, H., Furrer, H. (2008). Consistent Actuarial Valuation. Springer, Berlin.

[9] Wüthrich, M.V., Merz, M. (2008). Stochastic Claims Reserving Methods in Insurance. Wiley Finance.

[10] Wüthrich, M.V., Merz, M., Lysenko, N. (2008). Uncertainty in the claims development result in the chain ladder method. Accepted for publication in *Scand. Act. J.*

Biographies of the Authors

Merz Michael[1] is Assistant Professor for Statistics, Risk and Insurance at University of Tübingen (Germany). He was awarded in 2004 with the SCOR Actuarial Prize for his doctoral thesis in risk theory.

Wüthrich Mario V.[2] is Senior Researcher and Lecturer at ETH Zurich (Switzerland) in the field actuarial and financial mathematics. He holds a PhD in Mathematics from ETH Zurich and serves on the board of the Swiss Association of Actuaries.

[1] University Tübingen, Faculty of Economics, D-72074 Tübingen, Germany.
michael.merz@uni-tuebingen.de

[2] ETH Zurich, Department of Mathematics, CH-8092 Zurich, Switzerland.
mario.wuethrich@math.ethz.ch

Principles of the Chain-Ladder "Method" Selecting and Updating Claims Development Factors

Rajesh Sahasrabuddhe, FCAS, MAAA

Abstract

There has been significant discussion recently regarding the roles of "models"[1] and "methods" in actuarial practice. I believe that much of this discussion is misguided as it is based on an imprecise and arbitrary distinction. I believe that "methods" are more appropriately considered to be a subclass of "models," rather than a wholly different class of estimation procedures. More specifically,as with "models", I believe that there **are** statistical assumptions underlying "methods."

If we accept this conclusion, then it becomes incumbent on actuaries to apply statistical theory when using methods. The most common method is the chain-ladder method. In this paper, as an example, I re-examine the process of selecting and updating claim[2] development factors under this new paradigm.

1. Methods versus Models

In the Fall 2005 CAS Forum, the CAS Working Party on Quantifying Variability in Reserve Estimates published *The Analysis and Estimation of Loss & ALAE Variability: A Summary Report*. This paper proposed the following definitions:

> *Method: A systematic procedure for estimating future payments for loss and allocated loss adjustment expense. Methods are algorithms or series of steps followed to determine an estimate; they do not involve the use of any statistical assumptions that could be used to validate reasonableness or to calculate standard error. Well known examples include the chain-ladder (development factors) method or the Bornhuetter-Ferguson method.* ***Within the context of*** **[the Working Party]** ***paper****, "methods" refer to algorithms for calculating future payment estimates, not methods for estimating model parameters. (emphasis added)*
>
> *Model: A mathematical or empirical representation of how losses and allocated loss adjustment expenses emerge and develop. The model accounts for known and inferred properties and is used to project future emergence and development. An example of a mathematical model is a formulaic representation that provides the best fit for the available historical data. Mathematical models may be parametric (see below) or non-parametric. Mathematical models are known as "closed form" representations, meaning that they are represented by mathematical formulas. An example of an empirical representation of how losses and allocated loss adjustment expenses emerge and develop is the frequency distribution produced by the set of all reserve values generated by a particular application of the chain ladder method. Empirical distributions are, by*

[1] The use of quotation marks is intended to indicate usages of the terms "models" and "methods" that the author believes to be incorrect.

[2] In this paper, we use the terms "claims" instead of "loss" in order to be consistent with Actuarial Standard of Practice (ASOP) No. 43, *Property/Casualty Unpaid Claim Estimates*.

> *construction, not in "closed form" as there is no underlying requirement that there be an underlying mathematical model.*[3]

It should be noted that these definitions were restricted to a specific context and that they were presented in a non-refereed paper. Despite this circumstance, the Actuarial Standards Board adopted these definitions in Actuarial Standard of Practice (ASOP) No. 43, *Property/Casualty Unpaid Claim Estimates*. ASOP No. 43 includes the following definitions:

> *2.5 Method—A systematic procedure for estimating the unpaid claims.*
>
> *2.6 Model—A mathematical or empirical representation of a specified phenomenon.*[4]

In addition, the ASOP document includes the following comment and response related to these definitions:

> *Section 2.5, Method and 2.6, Model*
>
> *Comment One commentator stated, "There are definite differences between 'methods' and 'models' that are much more substantial and fundamental than" what is in the proposed standard. The commentator suggested that more complete definitions be taken from the CAS Working Party paper on reserve variability.*
>
> *Response The definitions in the standard are abbreviated versions of what is in the referenced Working Party paper. The reviewers believe that further elaboration is unnecessary, although reference to various CAS publications has been added to appendix 1.* [5]

I believe that this was an unfortunate decision by the Actuarial Standards Board. These definitions appear to reinforce the notion that "methods" and "models" are actually different. The acceptance of these definitions within a binding document might also result in a *de facto* acceptance of these definitions without being subject to a refereed process.

"Methods" are defined as algorithms without statistical assumptions whereas "models" are defined as mathematical representations. The definition and cited examples imply that only an understanding of algebra and arithmetic are necessary to use "methods." In contrast, "models" appear to require more advanced statistical skills. These definitions are misguided. The definitions are also somewhat dangerous as a layperson would (rightly) question whether the training of an FCAS is required to use "methods."

For the "methods" crowd, this definition has the unfortunate result that they are not forced to statistically evaluate their estimation methodologies. After all, statistical tests cannot be performed in the absence of statistical assumptions. For the profession, this has a dangerous consequence as it devalues the skills required to perform actuarial calculations.

---

[3] CAS Working Party on Quantifying Variability in Reserve Estimates. The Analysis and Estimation of Loss & ALAE Variability: A Summary Report. Casualty Actuarial Society Forum (Fall 2005), 29-146. (Page 38)

[4] Actuarial Standards Board of American Academy of Actuaries, "Actuarial Standard of Practice No. 43, Property/Casualty Unpaid Claim Estimates (Doc. No. 106)," 2007. (Page 3)

[5] Ibid, Page 15

I believe that it is more appropriate to consider "methods" as a type or subclass of "models." Let us consider the plain-English definition of Model:

> *a simplified version of something complex used in analyzing and solving problems or making predictions*[6]

Cleary, the chain-ladder and Bornhuetter Ferguson methods, which are listed as examples of "methods," would also be considered models under this definition. Consider that the paid claims development method for estimating unpaid claim amounts may also be presented as:

$$\hat{U}_i = P_i \times \left(\prod_i \hat{f}_j\right) - 1$$

where: U = Unpaid Claims

P = Paid Claims

f_j = the estimated incremental claims development factor between j and $j + 1$

and i = the age of an accident period.

Under the definitions proposed by the Working Party and adopted by the Actuarial Standards Board, would this be considered a "method" or would it be considered a "model?" We should now see that the distinction is arbitrary.

I believe that it would have been more useful to focus on types or classes of models such as, but not limited to:

- arithmetic
- stochastic
- parametric
- deterministic
- empirical
- non-parametric

With this paradigm, we can better analyze **deterministic models** such as chain-ladder and Bornhuetter-Ferguson. An example of this analysis focused on the selection of the incremental claims development factors is presented in this paper. Other analyses, such as a review of the quality of the models themselves and correlations between development columns are beyond the scope of this paper – but they become possible under the new paradigm.

2. Review of the Properties of Statistical Estimators

We should now consider "selected incremental claims development factors" as estimators of the parameters of a model. We then consider the following properties of estimators in evaluating the quality of our claims development factors:

- Unbiasedness – An estimator ($\hat{\theta}$) is considered unbiased if its expected value is equal to the true value of the parameter (θ). That is:

  $$E[\hat{\theta}] = \theta$$

  A somewhat more relaxed constraint is that the estimator be asymptotically unbiased. That is:

[6] Encarta dictionary

$$\lim_{n \to \infty} E[\hat{\theta}] = \theta$$

- Efficiency – An estimator is considered efficient if its sampling distribution has a relatively small standard deviation.
- Consistency – An estimator is considered consistent if it is more likely to be close to its true value when the sample size is increased.
- Sufficiency – An estimator is considered sufficient if it uses all of the information in the sample.
- Robustness / Resistance – An estimator is considered resistant or robust if it is relatively unaffected by outliers.

3. Comparisons of Common Methods of Selecting Claims Development Factors

We now consider four common methods of selecting claims development factors: (i) all-year averages (weighted, or unweighted) (ii) averages of recent observations (iii) Ex hi/low averages and (iv) judgment. For purposes of this discussion, we should assume that there are no distorting influences on the data.

Table 1
Common Estimators of Claims Development Factors

	Estimator			
Property	All-Year Average	Average of Recent Observations	Ex-Hi/Low Averages	Judgment
Unbiasedness	Yes	Yes	Yes	Unknown
Efficiency	Unknown	Unknown	Unknown	Unknown
Consistency	Yes	Not Applicable (Fixed sample size)	Yes	Unknown
Sufficiency	Yes	No	No	Unknown
Robustness / Resistance	Unknown	Unknown	Yes	Probably

The conclusion that we should draw from this table is that, under current commonly used methods for estimating claims development factors, we understand very little about the quality of those factors. This situation is further exacerbated when we consider that the typical basis for selected claims development factors is "actuarial judgment" based on a review of various averages. This leads us to the unfortunate conclusion that we understand relatively little about the quality of the resulting estimates of ultimate claims.

4. Statistical Estimation Methods

We now consider two alternative statistical methods for estimating claims development: maximum likelihood and regression. We use the 12-24 month General Liability Excluding Mass Torts development experience published by the Reinsurance Association of America as our test data. The results of the estimation considering both of these methods and a comparison to the traditional techniques listed above are presented in Exhibit A.

I do not intend to imply that these are the only available statistical tools that may be used to estimate claims development factors. They are presented here as two possible examples. A discussion of the advantages and disadvantages between maximum likelihood, regression and alternative parameter estimation methodologies is beyond the scope of this paper.

Furthermore, I recognize that these estimation methods do not always have all of the desired properties listed in the prior section. For example, the maximum likelihood estimator is not always unbiased. However, what is important is that we realize where these methods fall short as compared to the (almost complete) lack of knowledge associated with traditional estimators.

Most importantly, the knowledge that we have about these estimators will allow us to update the development factors only when appropriate. That is, I believe that, too often, unpaid claim estimates are impacted by differences in judgments applied year-to-year or quarter-to-quarter. This (understandably) reduces the confidence that stakeholders have in actuarial work product.

4.1. Maximum Likelihood Estimators

The advantage of maximum likelihood estimators (MLE) is that they are: (i) asymptotically unbiased (ii) asymptotically efficient, (iii) consistent and, (iv) for large samples, the MLE is normally distributed. The principal difficulty with maximum likelihood estimation is that the procedure requires the assumption of a model form. However, this does provide a benefit in that we would then expect the MLE to be robust / resistant.

There are three steps to develop the MLE for claims development factors. First, we must determine the appropriate distribution form for the claims development factors. Then, using this distribution, we must formulate the maximum likelihood function. Finally we must determine the parameters that maximize the likelihood function.

In the attached example we assume the following distributional form:

$$(\text{claims development factor} - 1) \sim \text{LogNormal}\ (\mu, \sigma)$$

The likelihood functions and log-likelihood functions may then be, respectively, written as:

$$\text{L} = \prod f(x; \mu, \sigma)$$

$$\ln L = \sum \ln f(x; \mu, \sigma),:$$

We then can use numerical methods to solve for the parameters that maximize the likelihood or equivalently maximize the log-likelihood[7].

4.2. Regression (Least Squares Estimator)

We can also use regression techniques to estimate the claims development. For convenience, we will refer to the resulting estimator as the "regression estimator" (RE). Under the assumptions of chain ladder method that claims at a given age are proportional to the claims at the prior age, the RE will be unbiased. Heuristically, we would also expect it to be asymptotically efficient, consistent, sufficient and robust.

REs are developed by solving for the *X*-coefficient of the following regression equation:

$$Y = mX + \ \varepsilon$$

This is the equation for regression through the origin (intercept=0). *Y* and *X* are the claims at 24 and 12 months, respectively. The X coefficient, *m* = *Y* / *X*, represents the estimate of the claims development factor.

[7] In this particular example it is well known that the MLE for the μ and σ parameters of the lognormal distribution are the mean and standard deviations of the logarithms of the data.

5. Updating Claims Development Factors

A significant benefit of defining a model in terms of statistical estimators is that it provides valuable guidance in updating the model. That is, it removes the arbitrariness associated with updates to development factors determined using traditional methods.

For example, assume that with the RAA-GL data presented in the example, we had observed a development factor of X in the next period. The question then becomes: should we revise our estimator of the claims development factor? Too often, that question is answered "yes" without thought. In fact, "yes" may be the only possible answer if our claims development factor is based on a "traditional approach." The answer should be: "Only if our new observation results in an updated estimator that is statistically significantly different from the prior estimator." We can use hypothesis testing to determine whether a change in the claims development factor is warranted.

5.1. Maximum Likelihood Estimators

In the example presented, we use the Likelihood Ratio test to determine whether a development factor estimator developed using maximum likelihood should be updated. That is, we test the null hypothesis that there should be no change to the estimator. The alternative hypothesis is that, the estimator should be updated.

The Likelihood Ratio test statistic is calculated as 2 times the difference in the log-likelihood values. This test statistic has a chi-square distribution with degrees of freedom equal to the number of parameters. The log-likelihoods are calculated including the <u>new</u> data.

Furthermore, it should be noted that there is no restriction on the data used in the calculation of the test statistic. That is, even if the initial parameters are calculated using all available observations, we are free to test for whether an update is required using, for example, only the most recent five observations. Stated differently, the decision as to the data used in the estimation process is independent of the hypothesis test.

Exhibits B1 and B2 present examples where the new observation does not support and does support, respectively, a change to the claims development factor estimator.

5.2. Regression (Least Squares Estimator)

Similarly, we can use hypothesis testing to determine whether a development factor estimator developed using regression should be updated. We perform this test by calculating the predicted Y values using the following relationship:

$$\hat{Y} = mX$$

We then fit the following regression line to the predicted-Y values:

$$Y - \hat{Y} = aY + \varepsilon$$

We can then test for significance of the regression coefficient. If the regression coefficient is significant, we then reject the null hypothesis. Exhibits C1 and C2 present examples where the new observation does not support and does support, respectively, a change to the claims development factor estimator.

6. Acknowledgments

The author wishes to thank Katy Siu and Bernard Chan for their reviews of this paper. Any errors that remain herein are the responsibility of the author. As with many research topics, the concepts presented herein are a "work-in-progress." The author would welcome your comments. Please consult the CAS member directory for contact information.

RAA
General Liability Excluding Mass Torts
Selecting Claims Development Factors
Reported Incurred Claims

(1)	(2)	(3)	(4)	(5)	(6)	(7)
			(3) / (2)	ln [(4) - 1]		ln [(6)]
				Statistics for Maximum Likelihood		
Accident Year	at 12 mos. (*A*)	at 24 mos. (*B*)	Observed (*X*)	Y = ln(*X* - 1)	f(y; μ, σ)	Log-Likelihood
1989	49,997	139,166	2.7835	0.578570442	0.620243308	-0.477643445
1990	70,104	201,662	2.8766	0.629467965	0.801247311	-0.221585627
1991	79,614	208,748	2.6220	0.483660668	0.307094784	-1.180598835
1992	56,265	190,867	3.3923	0.872249604	0.850514412	-0.161913922
1993	68,133	199,866	2.9335	0.65931547	0.895211433	-0.110695351
1994	68,530	241,658	3.5263	0.9267596	0.661954402	-0.412558604
1995	69,055	253,640	3.6730	0.983206773	0.461090693	-0.774160524
1996	102,320	295,607	2.8890	0.636070974	0.823202378	-0.194553205
1997	115,360	330,745	2.8671	0.624369451	0.783933695	-0.243430835
1998	138,160	468,526	3.3912	0.871788697	0.851967055	-0.160207421
1999	151,311	565,163	3.7351	1.006171101	0.386360609	-0.950984125
2000	178,943	562,916	3.1458	0.763504918	1.050062421	0.048849611
2001	187,203	671,424	3.5866	0.950347825	0.576352116	-0.551036493
2002	183,601	692,642	3.7725	1.019763639	0.34516186	-1.063741814
2003	149,925	494,121	3.2958	0.831076094	0.963885674	-0.036782587

	Estimator Value
Traditional Estimators	
All-Year Weighted Average	3.3064
Five Year Weighted Average	3.5092
X-Hi/Low Average	3.2381
Statistical Estimators	
Maximum Likelihood	
Model Form *assumes LDF-1 is lognormally distributed*	
μ	0.7891
σ	0.175177205
$\boxed{\text{LDF}}$	$\boxed{3.2014}$
Log-Likelihood	-6.491043178
Regression through the origin	
Rgression Model	*B* = *m***A* + ε
$\boxed{\text{Coefficient } (m)}$	$\boxed{3.3743}$
Standard Error of Coefficient *SE* (*m*)	0.0896

RAA
General Liability Excluding Mass Torts
Selecting Claims Development Factors
Reported Incurred Claims

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
			(3) / (2)	ln [(4) - 1]			ln [(6)]	ln [(7)]
					Statistics for Maximum Likelihood			
Accident Year	at 12 mos. (*A*)	at 24 mos. (*B*)	Observed (*X*)	Y = ln(*X* - 1)	f(y; μ_0, σ_0)	f(y; μ_a, σ_a)	Log-Likelihood	Log-Likelihood
					H_0	H_a	H_0	H_a
1989	49,997	139,166	2.7835	0.578570442	0.609738605	0.558088576	-0.49472493	-0.583237591
1990	70,104	201,662	2.8766	0.629467965	0.805137803	0.747959337	-0.216741832	-0.290406665
1991	79,614	208,748	2.6220	0.483660668	0.285169761	0.255255323	-1.254670623	-1.365490968
1992	56,265	190,867	3.3923	0.872249604	0.873308617	0.895304603	-0.135466272	-0.11059128
1993	68,133	199,866	2.9335	0.65931547	0.908646836	0.852297506	-0.095798779	-0.159819627
1994	68,530	241,658	3.5263	0.9267596	0.6702392	0.706931176	-0.400120615	-0.346821965
1995	69,055	253,640	3.6730	0.983206773	0.456795588	0.497404281	-0.78351928	-0.698352141
1996	102,320	295,607	2.8890	0.636070974	0.829189261	0.771901938	-0.18730685	-0.25889776
1997	115,360	330,745	2.8671	0.624369451	0.786225562	0.729238497	-0.240511553	-0.315754444
1998	138,160	468,526	3.3912	0.871788697	0.874878068	0.896706774	-0.133670753	-0.109026367
1999	151,311	565,163	3.7351	1.006171101	0.378578156	0.417908939	-0.971332738	-0.87249172
2000	178,943	562,916	3.1458	0.763504918	1.086090261	1.059350837	0.082584332	0.057656303
2001	187,203	671,424	3.5866	0.950347825	0.57879619	0.618480075	-0.546804867	-0.480490303
2002	183,601	692,642	3.7725	1.019763639	0.335822123	0.373795744	-1.091173654	-0.98404577
2003	149,925	494,121	3.2958	0.831076094	0.995671493	1.000599101	-0.004337902	0.000598921
2004	**New Obseravtion**		**3.7000**	0.993251773	0.421722988	0.461945273	-0.863406607	-0.772308852
Statistical Estimators					0.000650999	0.000682683	-7.337002925	-7.289480228

<u>Maximum Likelihood</u>

Model Form *assumes LDF-1 is lognormally distributed*

	Hyothesis Testing H_0	H_a
μ	0.7891	0.8018
σ	0.169237259	0.171153501
LDF	3.2014	3.2297
Log-Likelihood	-7.337002925	-7.289480228
Change in Log_likelihood	0.047522697	
Likelihhod Ratio Test Statistic	0.095045395	
Critical Value	5.991464547	Chi-Square (2 d.f.)
	Accept H_0	

RAA
General Liability Excluding Mass Torts
Selecting Claims Development Factors
Reported Incurred Claims

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
			(3) / (2)	ln [(4) - 1]			ln [(6)]	ln [(7)]
					Statistics for Maximum Likelihood			
Accident Year	at 12 mos. (*A*)	at 24 mos. (*B*)	Observed (*X*)	Y = ln(*X* - 1)	f(y; μ_0, σ_0)	f(y; μ_a, σ_a)	Log-Likelihood	Log-Likelihood
					H_0	H_a	H_0	H_a
1989	49,997	139,166	2.7835	0.578570442	0.609738605	0.526960015	-0.49472493	-0.640630606
1990	70,104	201,662	2.8766	0.629467965	0.805137803	0.611993102	-0.216741832	-0.491034267
1991	79,614	208,748	2.6220	0.483660668	0.285169761	0.35433519	-1.254670623	-1.037511951
1992	56,265	190,867	3.3923	0.872249604	0.873308617	0.680420218	-0.135466272	-0.385044704
1993	68,133	199,866	2.9335	0.65931547	0.908646836	0.654525535	-0.095798779	-0.42384468
1994	68,530	241,658	3.5263	0.9267596	0.6702392	0.606929497	-0.400120615	-0.499342644
1995	69,055	253,640	3.6730	0.983206773	0.456795588	0.511158897	-0.78351928	-0.671074784
1996	102,320	295,607	2.8890	0.636070974	0.829189261	0.621970701	-0.18730685	-0.474862293
1997	115,360	330,745	2.8671	0.624369451	0.786225562	0.604091376	-0.240511553	-0.504029808
1998	138,160	468,526	3.3912	0.871788697	0.874878068	0.680931052	-0.133670753	-0.384294223
1999	151,311	565,163	3.7351	1.006171101	0.378578156	0.469315257	-0.971332738	-0.756480547
2000	178,943	562,916	3.1458	0.763504918	1.086090261	0.734687447	0.082584332	-0.308310113
2001	187,203	671,424	3.5866	0.950347825	0.57879619	0.568643622	-0.546804867	-0.564501365
2002	183,601	692,642	3.7725	1.019763639	0.335822123	0.444295045	-1.091173654	-0.811266421
2003	149,925	494,121	3.2958	0.831076094	0.995671493	0.717284422	-0.004337902	-0.332282833
2004	**New Obseravtion**		**5.6000**	1.526056303	3.90759E-05	0.006127827	-10.15000459	-5.094915038
Statistical Estimators					6.03201E-08	1.54664E-06	-16.62360091	-13.37942628

Maximum Likelihood

Model Form *assumes LDF-1 is lognormally distributed*

	Hyothesis Testing	
	H_0	H_a
μ	0.7891	0.8351
σ	0.169237259	0.242228661
LDF	3.2014	3.3052
Log-Likelihood	-16.62360091	-13.37942628
Change in Log_likelihood	3.244174636	
Likelihhod Ratio Test Statistic	6.488349272	
Critical Value	5.991464547	Chi-Square (2 d.f.)
	Reject H_0	

RAA
General Liability Excluding Mass Torts
Selecting Claims Development Factors
Reported Incurred Claims

(1)	(2)	(3)	(4)	(5)	(6)	(7)
			(3) / (2)	(3) / (1)	(5) * (2)	(6) - (3)
Accident Year	at 12 mos. (*A*)	at 24 mos. (*B*)	Observed (*X*)	Estiamtor H_0	Predicted *B*	Residuals
1989	49,997	139,166	2.7835	3.3743	168,703	29,537
1990	70,104	201,662	2.8766	3.3743	236,550	34,888
1991	79,614	208,748	2.6220	3.3743	268,639	59,891
1992	56,265	190,867	3.3923	3.3743	189,853	-1,014
1993	68,133	199,866	2.9335	3.3743	229,899	30,033
1994	68,530	241,658	3.5263	3.3743	231,239	-10,419
1995	69,055	253,640	3.6730	3.3743	233,010	-20,630
1996	102,320	295,607	2.8890	3.3743	345,255	49,648
1997	115,360	330,745	2.8671	3.3743	389,256	58,511
1998	138,160	468,526	3.3912	3.3743	466,189	-2,337
1999	151,311	565,163	3.7351	3.3743	510,564	-54,599
2000	178,943	562,916	3.1458	3.3743	603,802	40,886
2001	187,203	671,424	3.5866	3.3743	631,673	-39,751
2002	183,601	692,642	3.7725	3.3743	619,519	-73,123
2003	149,925	494,121	3.2958	3.3743	505,887	11,766
2004	175,000	647,500	3.7000	3.3743	590,497	-57,003

Statistical Estimators

<u>Regression through the origin</u>

	Hyothesis Testing H_0	H_a
LDF	3.3743	3.4141
Test Statistic	-0.020898502	
Standard Error	0.024561098	
t Statistic	-0.850878164	
d.f.	15	
Critical Value at 5%	2.131449536	
	Accept H_0	

RAA
General Liability Excluding Mass Torts
Selecting Claims Development Factors
Reported Incurred Claims

(1)	(2)	(3)	(4) (3) / (2)	(5) (3) / (1)	(6) (5) * (2)	(7) (6) - (3)
Accident Year	at 12 mos. (*A*)	at 24 mos. (*B*)	Observed (*X*)	Estiamtor H_0	Predicted *B*	Residuals
1989	49,997	139,166	2.7835	3.3743	168,703	29,537
1990	70,104	201,662	2.8766	3.3743	236,550	34,888
1991	79,614	208,748	2.6220	3.3743	268,639	59,891
1992	56,265	190,867	3.3923	3.3743	189,853	-1,014
1993	68,133	199,866	2.9335	3.3743	229,899	30,033
1994	68,530	241,658	3.5263	3.3743	231,239	-10,419
1995	69,055	253,640	3.6730	3.3743	233,010	-20,630
1996	102,320	295,607	2.8890	3.3743	345,255	49,648
1997	115,360	330,745	2.8671	3.3743	389,256	58,511
1998	138,160	468,526	3.3912	3.3743	466,189	-2,337
1999	151,311	565,163	3.7351	3.3743	510,564	-54,599
2000	178,943	562,916	3.1458	3.3743	603,802	40,886
2001	187,203	671,424	3.5866	3.3743	631,673	-39,751
2002	183,601	692,642	3.7725	3.3743	619,519	-73,123
2003	149,925	494,121	3.2958	3.3743	505,887	11,766
2004	175,000	980,000	5.6000	3.3743	590,497	-389,503

Statistical Estimators

Regression through the origin

	Hyothesis Testing H_0	H_a
LDF	3.3743	3.6462
Test Statistic	-0.116444321	
Standard Error	0.049658948	
t Statistic	2.34488095	
d.f.	15	
Critical Value at 5%	2.131449536	
	Reject H0	

Risk Margins in Fair Value Loss Reserves:
Required Capital for Unpaid Losses and its Cost

Michael G. Wacek, FCAS, MAAA

Abstract

There is a general consensus that, in the absence of a trading market for loss reserves, a reasonable estimate of the "fair value" of unpaid losses is the risk-free present value of an unbiased estimate of those losses plus a market-based risk margin. If the risk margin is defined as the risk-free present value of the market-clearing cost of the capital required to support the unpaid losses during the run-off period, the size of the risk margin depends on the amount of required capital. Existing literature shows how to calculate the risk margin in cases where the amount of required capital is specified exogenously. However, the European Solvency II directive defines the capital requirement as of any given date as an endogenous variable equal to the amount needed to ensure solvency over a one-year time horizon with 99.5% confidence. This paper derives and illustrates an integrated framework for quantifying the required capital, the implied cost-of-capital-based risk margin and the fair value reserve from the expected volatility, payment and other characteristics of an unpaid loss portfolio consistent with the Solvency II standard. The conceptual framework presented has application to both fair value reserving and economic capital modeling.

Keywords. Economic capital, fair value loss reserve, hindsight loss reserve estimate, risk margin, Solvency II, stochastic loss reserve modeling

1. INTRODUCTION

In a 2007 Casualty Actuarial Forum paper entitled "Consistent Measurement of Property-Casualty Risk-Based Capital Adequacy" [6], Wacek included formulas for calculation of the transfer value (or "fair value") of unbiased reserves for unpaid losses. His underlying premise was that the risk margin embedded in the fair value reserve is based on the market cost of the capital required to support the unpaid losses as they run off[1]. The risk margin formulas derived in that paper are easily applied in cases where the amount of required capital has already been explicitly specified. However, the paper provided no guidance on how to proceed in cases where required capital is defined indirectly as a function of potential loss reserve outcomes, e.g., in terms of the one-year Expected Policyholder Deficit (EPD), Value-at-Risk (VaR) or Tail Value-at-Risk (TVaR) at some specified target confidence level.

This paper partially fills the gap in [6] by presenting formulas and a procedure for determining the fair value of unpaid losses in the case where the capital requirement is based

[1] This is the Solvency II definition, which, according to PricewaterhouseCoopers, is intended to be "a market-consistent 'economic' approach to valuation of assets and liabilities." This approach is also "conceptually in line with proposals for a revised IFRS for insurance contracts." For more background on Solvency II and its implications, see PWC's 2007 paper, "Gearing up for Solvency II" [4]. The quotes included here are from that paper. IFRS refers to "International Financial Reporting Standards."

on a target VaR measure instead of on pre-specified capital-to-reserve ratios. In addition, it incorporates more realistic assumptions about interest rates, in particular, that rates can vary by maturity and over time. The focus here is on VaR, because the capital adequacy standard embedded in the European Solvency II directive is based on a Value-at-Risk measure at the 99.5% confidence level ($VaR_{99.5\%}$).

Other authors have explored the issue of risk margins in fair value reserves. In a 2004 research project partially funded by the Casualty Actuarial Society, Tillinghast actuaries Conger, Hurley and Lowe and PricewaterhouseCoopers actuaries Gutterman, Littmann, Tarrant and Thomas estimated market-based risk margins using various approaches [1] [5][2]. Conger, Hurley and Lowe estimated historical risk margins without reference to capital. Gutterman et al included a cost-of-capital method among the four approaches they modeled. In a 2006 paper Feldblum [3] discussed a cost-of-capital approach to determining risk margins. However, like Wacek [6], both Feldblum and Gutterman et al treated the amount of required capital in their cost-of-capital models as an exogenous variable. In contrast, in this paper we model required capital as an endogenous variable. We show how to use the characteristics of the unpaid loss portfolio itself to determine the amount of capital implied by the Solvency II $VaR_{99.5\%}$ standard, the risk margin based on the cost of that capital and, ultimately, the fair value of the unpaid losses.

The paper comprises four main sections including this introduction, plus two appendices. In Section 2, using the Solvency II conceptual framework, we derive the key formulas and a recursive procedure for the calculation of required capital, risk margins and fair value reserves for unpaid losses. While that section includes a number of formulas, some of which *look* daunting, the fact is that the mathematics does not go beyond basic algebra and probability concepts. In Section 3 we illustrate a detailed practical application of the formulas and procedure presented in Section 2. In Section 4 we briefly summarize the key points and implications of the paper, and identify some areas for further research. Appendix A describes how to determine forward interest rates from the standard yield curve. Appendix B presents the derivation of a formula used in Section 2. A complete list of abbreviations and notations appears after the appendices followed by a list of references.

The terminology and notation used in this paper are generally consistent with [6][3]. That paper used the term "transfer value" rather than "fair value" but the meaning of both terms

---

[2] We provide separate references to these two self-contained papers published together in one volume.
[3] Familiarity with that paper is assumed, especially with the contents of sections 1, 2 (particularly 2.2 and 2.3) and Appendix B.

with respect to loss reserves is the same. Because "fair value" has become the more popular term, we adopt that usage in this paper.

2. DETERMINING THE FAIR VALUE OF UNPAID LOSSES

Conceptually, the fair value loss reserve is intended to be the price at which the loss reserve liability could be irrevocably transferred to a third party[4]. Because loss reserves are not normally traded, it is impossible to observe market prices directly. Instead, the fair value reserve must be determined indirectly as the risk-free present value of unpaid losses plus a risk margin reflecting the market-clearing cost of the capital required to minimize the risk of insolvency over a one-year time horizon due to loss reserve inadequacy.

In formula terms the fair value $T(L_n)$[5] of unpaid losses L_n at time *n* is the sum:

$$T(L_n) = PV(L_n) + R'_n, \tag{2.1}$$

where $PV(L_n)$ is the risk-free present value sum of the future loss and R'_n is the loss reserve risk margin, both as of time *n*.

2.1 The Present Value $PV(L_n)$

The calculation of $PV(L_n)$ requires knowledge of the amounts and timing of the expected future loss payments $P_{n+1}, P_{n+2}, P_{n+3}, \ldots, P_{n+k}$[6], where *k* represents the number of future loss payments, as well as an assumption about the risk-free yield curve. If we assume a flat yield curve, i.e., that the risk-free rate is the same irrespective of the time to maturity, we can use a single rate *r* in our present value analysis. In that case, if we assume that all loss payments are made at the midpoint of each payment year, then the value of $PV(L_n)$ is given by the formula:

$$\begin{aligned} PV(L_n) &= (1+\tfrac{1}{2}r)\cdot P_{n+1}\cdot v + (1+\tfrac{1}{2}r)\cdot P_{n+2}\cdot v^2 + (1+\tfrac{1}{2}r)\cdot P_{n+3}\cdot v^3 + \ldots + (1+\tfrac{1}{2}r)\cdot P_{n+k}\cdot v^k \\ &= (1+\tfrac{1}{2}r)\cdot(P_{n+1}\cdot v + P_{n+2}\cdot v^2 + P_{n+3}\cdot v^3 + \ldots + P_{n+k}\cdot v^k), \end{aligned} \tag{2.2}$$

---

4 Our use of the term "loss" should be understood to include claim adjusting and defense costs as well as the administrative expenses associated with managing a portfolio of claims. Those costs would be assumed by a third party in the case of an irrevocable transfer.

5 We retain the notation $T(L_n)$ to represent the fair value of unpaid losses in order to remain consistent with [6], where the equivalent term "transfer value" was used instead of "fair value".

6 This definition of P_{n+i}, for $1 \le i \le k-1$, as an *expected value* as of time *n* of a future loss payment matches the one used in Appendix B of [6]. The reader should be aware that in the body of [6], P_{n+i} refers to the *actual* payment in the year ending at time *n+i*.

where $v = (1+r)^{-1}$ and $1+\frac{1}{2}r$ is the adjustment factor required to reflect our assumption that loss payments are made at the midpoint of each year. The flat yield curve assumption allows us to factor $1+\frac{1}{2}r$ out of each term.

However, because our intent is to develop a practical framework that can be applied in the real world, where interest rates typically vary by maturity, we assume that risk-free interest rates can display that kind of variation. That requires us to introduce notation that differentiates between rates for different maturities.

Let r_m represent the annual yield to maturity as of time *n* on the risk-free fixed income instrument maturing in *m* years[7]. $v_m = (1+r_m)^{-1}$ is the corresponding one-year discount factor. r_m is also known as the "spot" rate. The standard yield curve is sometimes called the "spot rate curve" to distinguish it from other curves, including the "forward rate" curve.

The forward rate $r_{f:m}$ as of time *n* is the annual yield, between time *n+f* and *n+f +m*, on risk-free fixed income instruments maturing at or after time *n+f +m*. For a discussion of how forward rates can be derived from the spot rate curve, see Appendix A.

Having introduced the necessary notation, we can now generalize Formula (2.2) to allow for risk-free rates that vary by maturity:

$$\begin{aligned} PV(L_n) = {} & (1+\tfrac{1}{2}r_{0.5:0.5})\cdot P_{n+1}\cdot v_1 + (1+\tfrac{1}{2}r_{1.5:0.5})\cdot P_{n+2}\cdot v_2^2 \\ & + (1+\tfrac{1}{2}r_{2.5:0.5})\cdot P_{n+3}\cdot v_3^3 + \ldots + (1+\tfrac{1}{2}r_{k-0.5:0.5})\cdot P_{n+k}\cdot v_k^k \end{aligned} \tag{2.3}$$

where $r_{0.5:0.5}$ is the six-month forward rate for six-month risk-free money, and, in general for integer $0 \le j \le k-1$, $r_{j+0.5:0.5}$ is the *j*-year + six-month forward rate for six-month risk-free money. The factor $1+\frac{1}{2}r_{j+0.5:0.5}$ in each term adjusts the loss payment assumption from year-end to mid-year.

Formula (2.3) accords with an insurer investment policy of buying a set of risk-free zero coupon securities at time *n* to fund the payment of losses plus interest at each year-end. In order to be in a position to meet each expected mid-year loss payment obligation, the insurer simultaneously enters into a set of forward sales of six-month risk-free zero coupon securities, whose par values match the par values of the purchased zeroes.

For example, in accordance with that policy, at time *n* the insurer would purchase a one-year risk-free zero-coupon security having par value $(1+\frac{1}{2}r_{0.5:0.5})\cdot P_{n+1}$ (and market value $(1+\frac{1}{2}r_{0.5:0.5})\cdot P_{n+1}\cdot v_1$) and at the same time enter into a six-month forward sale of a six-month risk-free zero-coupon security with par value $(1+\frac{1}{2}r_{0.5:0.5})\cdot P_{n+1}$ (and forward price of

[7] Note that *m* does not have to be an integer.

P_{n+1}). That combination of purchase and matching forward sale would guarantee proceeds of P_{n+1} at time $n+\frac{1}{2}$, which the insurer could use to make the expected loss payment due at that time[8]. The second year's loss payment would be funded by the purchase, also at time *n*, of two-year risk-free zeroes with par value $(1+\frac{1}{2}r_{1.5:0.5})\cdot P_{n+2}$ and the simultaneous eighteen-month forward sale of six-month risk-free zeroes with the same par value. The funding of the third and subsequent years' loss payments would be addressed in a similar way.

In general, the present value $PV(L_{n+i})$ of the expected unpaid loss L_{n+i} as of time *n*+*i* is given by:

$$\begin{aligned} PV(L_{n+i}) &= (1+\tfrac{1}{2}r_{i+0.5:0.5})\cdot P_{n+i+1}\cdot v_{i:1} + (1+\tfrac{1}{2}r_{i+1.5:0.5})\cdot P_{n+i+2}\cdot v_{i:2}^{2} \\ &+ (1+\tfrac{1}{2}r_{i+2.5:0.5})\cdot P_{n+i+3}\cdot v_{i:3}^{3} + \ldots + (1+\tfrac{1}{2}r_{k-0.5:0.5})\cdot P_{n+k}\cdot v_{i:k}^{k} \end{aligned} \tag{2.4}$$

for integers $0 \le i \le k-1$, where $v_{i:m}$ represents the *i*-year forward one-year discount factor as of time *n* for *m*-year risk-free money.

2.2 The Risk Margin R'_n

The risk margin R'_n is the second component of the fair value $T(L_n)$ of unpaid losses L_n at time *n*. It is the risk-free present value sum of expected future risk charges based on the market cost of the capital required at time *n* and beyond to support the unpaid losses as they run off.

Let C_n^R represent the amount of capital required at time *n* to support the unpaid losses L_n for the next year (to time *n*+1). It is expected that after the passage of a year the unpaid loss amount will be L_{n+1} and that the amount of capital required at time *n*+1 for the following year will be C_{n+1}^R. In general, based on the sequence of expected unpaid loss amounts $L_n, L_{n+1}, L_{n+2}, \cdots, L_{n+k-1}$[9] at times *n*, *n*+1, *n*+2, …, *n*+*k*-1, respectively, we can anticipate that a sequence of expected capital amounts $C_n^R, C_{n+1}^R, C_{n+1}^R, \cdots, C_{n+k-1}^R$ will be needed to support the unpaid losses as they run off.

We assume that the capital provider demands the market-clearing annualized after-tax return on equity of *roe* for each year the capital is exposed. Assuming a constant market

---

[8] An equivalent alternative, of course, would be simply to buy a six-month zero-coupon instrument. Indeed, we could have expressed Formula (2.3) in terms of discount factors corresponding to an initial maturity six months out and at annual intervals thereafter. However, for the purposes of our presentation it is helpful to arrange for all cash flows to occur at the end of each year.

[9] This definition of L_{n+i}, for $1 \le i \le k-1$, as an *expected value* as of time *n* of a future unpaid loss amount matches the one used in Appendix B of [6]. The reader should be aware that in the body of [6], L_{n+i} refers to the *actual* unpaid loss amount at time *n*+*i*.

return on equity requirement and given a market-clearing tax rate of *tax*, the annual pre-tax return requirement on the required capital is $roe_{PT} = \dfrac{roe}{1-tax}$, a portion of which will be provided by the risk-free interest earned on the capital itself. For example, if r_1 is the one-year risk-free rate as of time *n*, then the portion of the required rate of return on capital C_n^R between time *n* and time *n*+1 that must come from the underwriting assets set aside to fund unpaid losses is $roe_{PT} - r_1$. The cost of the capital required to support the unpaid losses L_n for this first year of the run-off is $(roe_{PT} - r_1) \cdot C_n^R$.

The comparable expected rate of return on C_{n+1}^R between times *n*+1 and *n*+2 is $roe_{PT} - r_{1:1}$, where $r_{1:1}$ is the one-year forward rate as of time *n* for one-year money. We use the forward rate $r_{1:1}$ in order to match the expected deployment of C_{n+1}^R at time *n*+1. The use of the forward rate mimics the effect of entering into a one-year forward contract at time *n* to invest C_{n+1}^R in a one-year zero-coupon security at time *n*+1. The cost of the capital required to support the expected remaining unpaid losses L_{n+1} for this second year of the run-off is $(roe_{PT} - r_{1:1}) \cdot C_{n+1}^R$.

The annual costs, expected as of time *n*, related to the capital to support unpaid losses over the entire the run-off period are represented by the sequence $(roe_{PT} - r_1) \cdot C_n^R$, $(roe_{PT} - r_{1:1}) \cdot C_{n+1}^R$, $(roe_{PT} - r_{2:1}) \cdot C_{n+2}^R$, …, $(roe_{PT} - r_{k-1:1}) \cdot C_{n+k-1}^R$, where $r_{i:1}$ is the *i*-year forward rate as of time *n* for one-year money and $0 \le i \le k-1$[10].

We are now in a position to express R_n' as the following present value sum:

$$\begin{aligned} R_n' &= (roe_{PT} - r_1) \cdot C_n^R \cdot v_1 + (roe_{PT} - r_{1:1}) \cdot C_{n+1}^R \cdot v_2^2 + (roe_{PT} - r_{2:1}) \cdot C_{n+2}^R \cdot v_3^3 \\ &\quad + \ldots + (roe_{PT} - r_{k-1:1}) \cdot C_{n+k-1}^R \cdot v_k^k \end{aligned} \tag{2.5}$$

where v_1, v_2, v_3,…, v_k are the one-year risk-free discount factors implied by the yields at time *n* on fixed income instruments maturing in one, two, three, …, and *k* years, respectively[11].

R_n' can also be expressed recursively in terms of the risk margin R_{n+1}' associated with the expected unpaid losses L_{n+1} at time *n*+1:

$$R_n' = v_1 \cdot ((roe_{PT} - r_1) \cdot C_n^R + R_{n+1}'), \tag{2.6}$$

where

[10] Because the zero-year forward rate as of time *n* for one-year money is the same as the "spot" rate, we can use the notation $r_{0:1}$ and r_1 interchangeably.

[11] $v_{i+1} = (1 + r_{i+1})^{-1}$ for integer $0 \le i \le k-1$.

$$R'_{n+1} = (roe_{PT} - r_{1:1}) \cdot C^R_{n+1} \cdot v_{1:1} + (roe_{PT} - r_{2:1}) \cdot C^R_{n+2} \cdot v_{1:1} \cdot v_{2:1}$$
$$+ (roe_{PT} - r_{3:1}) \cdot C^R_{n+3} \cdot v_{1:1} \cdot v_{2:1} \cdot v_{3:1} + \ldots + (roe_{PT} - r_{k-1:1}) \cdot C^R_{n+k-1} \cdot v_{1:1} \cdot v_{2:1} \cdot v_{3:1} \cdots v_{k-1:1}$$

See Appendix B for a derivation of Formula (2.6).

If the sequence of expected required capital $C^R_n, C^R_{n+1}, C^R_{n+1}, \cdots, C^R_{n+k-1}$ is known, whether from the application of prescribed capital factors or through some other means, we can use Formula (2.5) or (2.6) to first determine R'_n and then Formula (2.1) to determine $T(L_n)$[12].

In general, the risk margin R'_{n+i} associated with the expected unpaid losses L_{n+i} at time $n+i$ can be expressed in terms of the risk margin R'_{n+i+1} associated with the expected unpaid losses L_{n+i+1} one year later at time $n+i+1$:

$$R'_{n+i} = v_{i:1} \cdot ((roe_{PT} - r_{i:1}) \cdot C^R_{n+i} + R'_{n+i+1}), \tag{2.7}$$

where $0 \le i \le k-1$.

2.3 Funding Assets S_{n+1} and Funding Need t_{n+1}

Let's assume that the required capital sequence has not been directly specified, and that instead the capital requirement has been described in the form of the objective to ensure that the one-year probability of insolvency due to fair value loss reserve inadequacy is no more than $1-\alpha$. In Value-at-Risk terms that implies capital calibration at the α confidence level and a time horizon of one year[13].

Specifically, that objective establishes the required capital C^R_n at time n as the amount needed in addition to assets equal to the fair value $T(L_n)$ of the unpaid losses L_n to ensure (with a probability of α) adequate funding of those unpaid losses L_n one year out (at time n+1). The total *funding assets* available at time n+1, including accumulated interest at the risk-free rate r_1, will be the amount defined by $S_{n+1} = (T(L_n) + C^R_n) \cdot (1 + r_1)$.

The *funding need* at time n+1 will be the amount equal to the fair value of the one-year hindsight estimate of L_n. The term *one-year hindsight estimate of* L_n is a succinct way of referring to the unpaid losses remaining at time n+1 plus the losses paid during the preceding year. It can be represented at time n by the random variable $h_{n+1} = l_{n+1} + p_{n+1}$, where l_{n+1} and p_{n+1} are also random variables defined as of time n that correspond to the unpaid and paid loss components, respectively, of the hindsight estimate.

[12] We assume that all other parameters needed for Formulas (2.1) and (2.5) or (2.6) are known.
[13] Under Solvency II, α =99.5%.

Let $t_{n+1} = T(h_{n+1})$ represent the random variable, defined as of time *n*, for the fair value of the one-year hindsight estimate of L_n at time *n*+1. The fair value of the one-year hindsight estimate is the sum of the fair values of the unpaid and paid loss components[14]. The fair value of the time *n*+1 unpaid loss estimate is the sum of its present value and the associated risk margin. Because paid losses require no capital support, the fair value of the paid component as of time *n*+1 simply reflects an interest adjustment. Putting all of this together allows us to express t_{n+1} as:

$$\begin{aligned} t_{n+1} &= T(l_{n+1}) + T(p_{n+1}) \\ &= PV(l_{n+1}) + R'_{n+1}(l_{n+1}) + p_{n+1} \cdot (1 + \tfrac{1}{2} r_{0.5:0.5}) \end{aligned} \qquad (2.8)$$

where $R'_{n+1}(l_{n+1})$ is the random variable for the required risk margin associated with the unpaid loss component l_{n+1}. We can recombine the terms in (2.8) involving l_{n+1} and p_{n+1} to express t_{n+1} more succinctly as:

$$t_{n+1} = PV(h_{n+1}) + R'_{n+1}(l_{n+1}), \qquad (2.9)$$

where $PV(h_{n+1}) = PV(l_{n+1}) + p_{n+1} \cdot (1 + \frac{1}{2} r_{0.5:0.5})$ represents the random variable for the present value of the one-year hindsight estimate of L_n as of time *n*+1.

If we assume that the relationship between the risk margin R'_{n+1} associated with the expected unpaid loss L_{n+1} at time *n*+1 and the present value $PV(L_{n+1})$ of that unpaid loss, embodied in the ratio $\dfrac{R'_{n+1}}{PV(L_{n+1})}$, is representative of the general relationship between the risk margin and the present value of the associated unpaid loss at time *n*+1, then we can express $R'_{n+1}(l_{n+1})$ as follows:

$$R'_{n+1}(l_{n+1}) = PV(l_{n+1}) \cdot \frac{R'_{n+1}}{PV(L_{n+1})} \qquad (2.10)$$

and we can then rewrite Formula (2.9) as:

$$t_{n+1} = PV(h_{n+1}) + PV(l_{n+1}) \cdot \frac{R'_{n+1}}{PV(L_{n+1})} \qquad (2.11)$$

[14] See Section 2.2 or 2.3 of [6].

2.4 Solving for C_n^R

We can express the one-year solvency objective in terms of the relationship between the funding need and the funding assets at time *n*+1 in the following alternative, but equivalent, ways:

$$\Pr ob(t_{n+1} \geq S_{n+1}) \leq 1 - \alpha\,, \tag{2.12}$$

$$\int_0^{S_{n+1}} t_{n+1} dt_{n+1} \geq \alpha \tag{2.13}$$

$$VaR_\alpha(t_{n+1}) = T_{n+1}^{-1}(\alpha) \leq S_{n+1}\,, \tag{2.14}$$

where $VaR_\alpha(t_{n+1})$ refers to the Value-at-Risk with respect to t_{n+1} at the α confidence level and T_{n+1}^{-1} represents the inverse distribution function of t_{n+1}, both of which define the funding need at the α confidence level.

The value of C_n^R that satisfies the following equilibrium relationship between the funding need at the α confidence level at time *n*+1 and the available funding assets at that time represents the amount of capital required at time *n* to support unpaid losses of L_n:

$$VaR_\alpha(t_{n+1}) = (T(L_n) + C_n^R) \cdot (1 + r_1) \tag{2.15}$$

Solving for C_n^R, we arrive at:

$$C_n^R = v_1 \cdot VaR_\alpha(t_{n+1}) - T(L_n)\,, \tag{2.16}$$

where $v_1 = (1 + r_1)^{-1}$ represents the one-year risk-free discount factor as of time *n*.

Using Formula (2.6) to expand Formula (2.1) we obtain the following formula for $T(L_n)$ in terms of C_n^R and R'_{n+1}:

$$T(L_n) = PV(L_n) + v_1 \cdot ((roe_{PT} - r_1) \cdot C_n^R + R'_{n+1})\,, \tag{2.17}$$

Substituting the Formula (2.17) expression for $T(L_n)$ into (2.16) and isolating C_n^R, we obtain a revised formula for required capital at time *n*:

$$\begin{aligned} C_n^R &= \frac{v_1 \cdot VaR_\alpha(t_{n+1}) - (PV(L_n) + v_1 \cdot R'_{n+1})}{1 + v_1 \cdot (roe_{PT} - r_1)} \\ &= \frac{VaR_\alpha(t_{n+1}) - (PV(L_n) \cdot (1 + r_1) + R'_{n+1})}{1 + roe_{PT}} \end{aligned} \tag{2.18}$$

As a fair value quantity the α-quantile funding need $VaR_\alpha(t_{n+1})$ includes an embedded risk margin, which we can isolate using Formula (2.11):

$$VaR_\alpha(t_{n+1}) = VaR_\alpha(PV(h_{n+1})) + PV(L_{n+1} \mid VaR_\alpha(PV(h_{n+1}))) \cdot \frac{R'_{n+1}}{PV(L_{n+1})} \tag{2.19}$$

where $VaR_\alpha(PV(h_{n+1}))$ represents the time n+1 present value of the one-year hindsight loss estimate at the α confidence level and $PV(L_{n+1} \mid VaR_\alpha(PV(h_{n+1}))$ represents the time n+1 present value of the *unpaid loss* component of $VaR_\alpha(PV(h_{n+1}))$ [15].

We can see from Formula (2.19) that the α-quantile funding need at time n+1 contemplates funding for not only the run-off of the unpaid losses but also for the risk margin needed to cover the cost of the capital required to support the unpaid losses during the run-off period.

If we substitute the Formula (2.19) expression of $VaR_\alpha(t_{n+1})$ into Formula (2.18) and rearrange the terms in the numerator, we obtain the following formula for C_n^R:

$$C_n^R = \frac{VaR_\alpha(PV(h_{n+1})) + PV(L_{n+1} \mid VaR_\alpha(PV(h_{n+1}))) \cdot \dfrac{R'_{n+1}}{PV(L_{n+1})} - PV(L_n)\cdot(1+r_1) - R'_{n+1}}{1 + roe_{PT}}$$

$$= \frac{VaR_\alpha(PV(h_{n+1})) - PV(L_n)\cdot(1+r_1) + R'_{n+1} \cdot \left(\dfrac{PV(L_{n+1} \mid VaR_\alpha(PV(h_{n+1})))}{PV(L_{n+1})} - 1 \right)}{1 + roe_{PT}}$$

$$= \frac{F_{n+1} + f_{n+1} \cdot R'_{n+1}}{1 + roe_{PT}} \tag{2.20}$$

where

$$F_{n+1} = VaR_\alpha(PV(h_{n+1})) - PV(L_n)\cdot(1+r_1) \tag{2.21}$$

defines the additional amount needed at time n+1 to bring present value loss funding up to the α confidence level, and

$$f_{n+1} = \frac{PV(L_{n+1} \mid VaR_\alpha(PV(h_{n+1})))}{PV(L_{n+1})} - 1 \tag{2.22}$$

[15] The notation $PV(L_{n+1} \mid VaR_\alpha(PV(h_{n+1})))$ is intended to convey the idea that the random variable $PV(l_{n+1})$ collapses to a single present value unpaid loss amount when conditioned on the specific present value hindsight estimate $VaR_\alpha(PV((h_{n+1}))$ and that that present value unpaid loss amount is the one included within $VaR_\alpha(PV((h_{n+1}))$.

is the percentage by which the time n+1 present value unpaid losses embedded in the one-year hindsight estimate at the α confidence level exceed the expected time n+1 present value unpaid loss amount.

The general formula for the anticipated future capital C_{n+i}^{R} required to support the expected unpaid losses L_{n+i} at time n+i, for $0 \le i \le k-1$, is given by:

$$C_{n+i}^{R} = \frac{F_{n+i+1} + f_{n+i+1} \cdot R'_{n+i+1}}{1 + roe_{PT}} \tag{2.23}$$

where

$$F_{n+i+1} = VaR_{\alpha}(PV(h_{n+i+1})) - PV(L_{n+i}) \cdot (1 + r_{i:1}) \tag{2.24}$$

$$f_{n+i+1} = \frac{PV(L_{n+i+1} \mid VaR_{\alpha}(PV(h_{n+i+1})))}{PV(L_{n+i+1})} - 1 \tag{2.25}$$

2.5 Recursive Procedure for C_n^R and R'_n

The expected unpaid loss amount L_{n+k} at time n+k is zero. At that point and beyond, the capital requirement C_{n+k}^{R} and the risk margin R'_{n+k} also are zero. At time n+k-1, because the terms depending on R'_{n+k} drop out, Formulas (2.23) and (2.6) simplify to:

$$C_{n+k-1}^{R} = \frac{F_{n+k}}{1 + roe_{PT}} \tag{2.26}$$

and

$$R'_{n+k-1} = v_{n+k-1:1} \cdot (roe_{PT} - r_{n+k-1:1}) \cdot C_{n+k-1}^{R} \tag{2.27}$$

By working recursively backward from time n+k-1, it is possible to determine the required capital and risk charges at any time from n through n+k-1. This can be achieved by the executing the following procedure, the first four steps of which do not rely on recursive relationships:

1) Tabulate risk-free spot rates r_m for $0 \le m \le k$ and the implied forward rates for one-year maturities based on Formula (A.2)[16].

2) Calculate and tabulate $PV(L_{n+i})$ for $0 \le i \le k-1$ using Formula (2.4).

[16] We suggest U.S. Treasury rates, but we acknowledge that others may prefer a different risk-free benchmark.

3) Model $PV(h_{n+i+1}) = PV(l_{n+i+1}) + p_{n+i+1} \cdot (1+\frac{1}{2} r_{i:1})$ for $0 \le i \le k-1$ and tabulate $VaR_{\alpha}(PV(h_{n+i+1}))$ and $PV(L_{n+i+1} \mid VaR_{\alpha}(PV(h_{n+i+1}))$[17].

4) Calculate and tabulate F_{n+i+1} for $0 \le i \le k-1$ using Formula (2.24).

5) Calculate and tabulate f_{n+i+1} for $0 \le i \le k-1$ using Formula (2.25).

6) Calculate C^{R}_{n+k-1} and R'_{n+k-1} using the following recursive procedure:

   a) First calculate C^{R}_{n+k-1} using Formula (2.26) and then R'_{n+k-1} (a function of C^{R}_{n+k-1}) using (2.27).

   b) Calculate C^{R}_{n+k-2} (a function of R'_{n+k-1}) and R'_{n+k-2} (a function of R'_{n+k-1} and C^{R}_{n+k-2}), in that order, using Formulas (2.23) and (2.7), respectively, with the formula subscript *i* replaced in every case by *k*-2.

   c) Similarly, calculate C^{R}_{n+k-3} and R'_{n+k-3}, in that order, using Formulas (2.23) and (2.7), respectively, with the formula subscript *i* replaced in every case by *k*-3.

   d) Continue stepwise in this fashion by decrementing the subscript by one and calculating the values of C^{R}_{n+i} and R'_{n+i}, in that order, using Formulas (2.23) and (2.7), respectively, with the formula subscript *i* chosen to reflect the decremented subscript for the step. Repeat until C^{R}_{n} and R'_{n} have been calculated, and then stop. The required capital and the required risk margin as of time *n* have been determined.

7) Calculate and tabulate required capital ratios to unpaid losses $c_{n+i} = C^{R}_{n+i} / L_{n+i}$ for $0 \le i \le k-1$. (Optional)

8) Use Formula (2.1) to calculate the fair value of unpaid losses $T(L_n)$ as of time *n*.

3. ILLUSTRATION

In this section we present a realistic illustration of the procedure described in Section 2.5 using unpaid loss and volatility patterns based mainly on Schedule P data reported by a diversified U.S. insurer as of December 31, 2007[18].

For purposes of illustration we make the following assumptions:

1) *n*=2007 corresponds to the valuation date of December 31, 2007.

2) The unbiased unpaid loss estimate as of December 31, 2007 is $L_{2007} = \$10{,}000$.

[17] Discussion about how to perform this modeling is beyond the scope of this paper. For one approach, see Appendix C of [6]. Another alternative is to fit distributions to historical one-year hindsight loss relationships.

[18] The derivation and discussion of those patterns is beyond the scope of this paper.

3) The applicable risk-free rates are the U.S. Treasury rates as of December 31, 2007 as shown in Table 1[19].

4) The unpaid losses as of December 31, 2007 run off over ten years (k=10) as shown in Table 2.

5) The market-clearing pre-tax return on equity is a constant $roe_{PT} = 18.75\%$, based on market-clearing $roe = 15\%$ and $tax = 20\%$[20].

6) Required capital is calibrated to VaR_α with $\alpha = 99.5\%$ over a one-year time horizon.

7) Forward interest rates for six-month and one-year money maturing on the same date are equal: $r_{j+0.5:0.5} = r_{j:1}$ for $0 \le j \le k-1$[21].

We illustrate the eight steps of the procedure by constructing a series of eight corresponding tables, each of which contains the key inputs and outputs of the respective step. In addition, we provide two additional tables which illustrate the cash flows associated with the fair value reserve run-off (Table 9) and the adequacy of the required capital to ensure fair value funding at the 99.5% confidence level (Table 10).

Table 1 summarizes the risk-free interest rates used in this illustration. The spot rates comprising the U.S. Treasury yield curve as of December 31, 2007 have been tabulated in Column (2) by the number of years m to maturity. For example, the one-year spot rate was 3.34% and the spot rate for the two-year maturity was 3.05%. The m-1 year forward rates for one-year money, derived from the December 2007 spot rates using Formula (A.2) from Appendix A, appear in Column (3). For example the one-year forward rate for one-year money was 2.76%, which was calculated using Formula (A.2) with f=1 and m=1 as follows:

$$r_{1:1} = \frac{(1+r_2)^2}{1+r_1} - 1 = \frac{(1.0305)^2}{1.0334} - 1 = 2.76\%$$

The one-year forward discount factors for one-year money are shown in Column (4). They were calculated from the forward rates in Column (3) using the formula $v_{m-1:1} = \frac{1}{1+r_{m-1:1}}$, where $1 \le m \le 10$. For example, the one-year forward one-year discount factor is 97.31%, which was calculated as $v_{1:1} = \frac{1}{1+r_{1:1}} = \frac{1}{1+2.76\%}$.

[19] We assume that U.S. Treasury rates are reasonable proxies for risk-free rates. Note that some researchers dispute that notion. That debate is beyond the scope of this paper.
[20] These return on equity and tax rate assumptions are merely illustrative, but are not unrealistic.
[21] The purpose of this assumption is merely to avoid having to introduce of an additional set of forward rates.

TABLE 1 Risk-Free Interest Rate Summary As of December 31, 2007			
(1)	(2)	(3)	(4)
Number of Years	Spot Rate	*m*-1 Year Forward One-Year Rate	*m*-1 Year Forward One-Year Discount Factor
m	r_m [1]	$r_{m-1:1}$ [2]	$v_{m-1:1}$ [3]
1	3.34%	3.34%	96.77%
2	3.05%	2.76%	97.31%
3	3.07%	3.11%	96.98%
4	3.25%	3.79%	96.35%
5	3.45%	4.25%	95.92%
6	3.57%	4.17%	96.00%
7	3.70%	4.48%	95.71%
8	3.81%	4.58%	95.62%
9	3.92%	4.80%	95.42%
10	4.04%	5.13%	95.12%

1 Source: U.S. Treasury website; spot rates shown for 4-year, 8-year and 9-year maturities are interpolated values.
2 Formula (A.2) (Appendix A)
3 Formula (A.3) (Appendix A)

Table 2 shows the expected run-off pattern of the unpaid losses as of December 31, 2007. Column (2) shows initial unpaid losses of $10,000 as of December 2007 and the expected remaining unpaid losses at successive December valuation dates through 2017, at which time all losses are expected to have been paid. The expected loss payments are shown in Column (3). The first payment of $2,839 is expected to be made during 2008, and the last payment of $160 is expected to be made in 2017. Column (4) shows the present value of the expected unpaid losses as of December 31, 2007 and the expected present values of the expected unpaid losses at successive December valuation dates. The December 2007 present value $PV(L_{2007}) = \$9{,}148$ was calculated from the expected losses payments in Column (3) using Formula (2.3) and the risk-free rates tabulated in Table 1[22]. Following the expected loss payment of $2,839 during 2008, the remaining unpaid losses as of December 31, 2008 are expected to be $7,161. The expected present value $PV(L_{2008})$ as of December 2008 of that expected unpaid loss amount is $6,566, which was calculated using Formula (2.4) and forward rates derived from Table 1[23].

TABLE 2

Unpaid Loss Reserve and Expected Run-off
As of December 31, 2007

(1)	(2)	(3)	(4)
Year Ending 12/07+ i	Expected Unpaid Losses	Expected Paid Losses	Expected PV of Unpaid Losses
$n+i$ [1]	L_{n+i}	P_{n+i}	$PV(L_{n+i})$ [2]
2007	$10,000	n/a	$9,148
2008	$7,161	$2,839	$6,566
2009	$5,105	$2,055	$4,664
2010	$3,568	$1,538	$3,247
2011	$2,467	$1,100	$2,249
2012	$1,717	$750	$1,579
2013	$1,206	$511	$1,123
2014	$752	$454	$709
2015	$442	$310	$425
2016	$160	$283	$156
2017	$0	$160	$0

1 n=2007, value of $0 \le i \le k = 10$ implied by valuation year
2 Formula 2.4, using Column (3) and forward discount rates based on Table 1

[22] For the mid-year loss payment adjustment we used the simplifying assumption that the forward rates for six-month and one-year money having the same maturity date are the same: $r_{j+0.5:.05} = r_{j:1}$ for $0 \le j \le k-1$.

[23] The actual present value of unpaid losses as of December 31, 2008 can vary from the expected due to a change in interest rates and/or a change in the unpaid loss estimate.

TABLE 3

Present Value Hindsight Statistics One Year Out
99.5% Confidence Level
At Successive Annual Valuation Dates through 2017
Expected as of December 31, 2007

(1)	(2)	(3)	(4)	(5)
Year Ending 12/07+ i	Expected PV of Unpaid Losses	PV of Expected Hindsight One Year Out	PV of 99.5% Hindsight One Year Out	PV of Unpaid in 99.5% Hindsight One Year Out
$n+i$ [1]	$PV(L_{n+i})$ [2]	$PV(L_{n+i}) \cdot (1+r_{i:1})$	$VaR_{99.5\%}(PV(h_{n+i+1}))$ [3]	[4]
2007	$9,148	$9,453	$11,067	$7,840
2008	$6,566	$6,748	$8,186	$5,847
2009	$4,664	$4,809	$6,189	$4,294
2010	$3,247	$3,370	$4,857	$3,359
2011	$2,249	$2,345	$3,427	$2,395
2012	$1,579	$1,645	$2,591	$1,797
2013	$1,123	$1,173	$1,819	$1,125
2014	$709	$742	$1,257	$668
2015	$425	$445	$803	$241
2016	$156	$164	$292	$0
2017	$0	$0	$0	$0

1 n=2007, value of $0 \le i \le k = 10$ implied by valuation year
2 Table 2, Column (4)
3 From stochastic hindsight loss analysis
4 $PV(L_{n+i+1} \mid VaR_\alpha(PV(h_{n+i+1}))$ from stochastic hindsight loss analysis

Table 3 summarizes the key results needed from the modeling of the one-year hindsight loss estimate represented by the random variable $h_{2007+i+1}$ for $0 \le i \le k = 10$. The details underlying that analysis are beyond the scope of this paper, but let us assume that we know the values of $VaR_{99.5\%}(PV(h_{2007+i+1}))$ and $PV(L_{2007+i+1} \mid VaR_{99.5\%}(PV(h_{2007+i+1}))$, which we have tabulated in Columns (4) and (5) respectively. Column (3) shows the expected present value one-year hindsight estimate $PV(L_{2007+i})\cdot(1+r_{i:1})$ as of December 2007+*i*+1, which provides a useful baseline comparison for the 99.5% quantile hindsight estimate in Column (4). Column (2) shows the present value $PV(L_{2007+i})$ as of December 2007+*i* in order to provide context for the entries in Column (3).

For example, as of December 31, 2007 the present value of unpaid losses is $PV(L_{2007}) = \$9{,}148$, as shown in Column (2). Reflecting interest at a rate of $r_1 = 3.34\%$, the expected value of that \$9,148 one year out on December 31, 2008 is \$9,453. That amount, shown in Column (3), is also the present value of the expected hindsight estimate as of that date. The present value as of December 31, 2008 of the one-year hindsight estimate at the 99.5% confidence level $VaR_{99.5\%}(PV(h_{2008}))$ is shown in Column (4) as \$11,067, which is 17% higher than the baseline value of \$9,453. As the loss portfolio runs off, the gap between the 99.5% quantile present value hindsight estimate and the baseline estimate is expected to increase. For example, the expected 99.5% level present value hindsight estimate of December 2011 unpaid losses of $VaR_{99.5\%}(PV(h_{2012})) = \$3{,}427$ is 46% higher than the baseline of \$2,345. By December 2016 the gap is expected to widen further to 78% (\$292 vs. \$164). This pattern is a manifestation of the expectation of increasing one-year volatility in the unpaid loss estimates as the portfolio ages.

Column (5) shows the expected present value one year out from each valuation date of the portion of the one-year hindsight estimate that is expected to remain unpaid as of that date. For example, as of December 31, 2007 the expected December 31, 2008 present value of the unpaid portion of the one-year hindsight estimate of \$11,067 is \$7,840, expressed formally as: $PV(L_{2008} \mid VaR_{99.5\%}(PV(h_{2008}) = \$11{,}067) = \$7{,}840$.

Table 4 illustrates the calculation of $F_{2007+i+1}$ for $0 \le i \le k = 10$, which represents the additional amount needed one year out from each valuation date December 2007+*i* to bring present value loss funding up to the 99.5% confidence level. Columns (2) and (3), both taken from Table 3, represent the expected and the 99.5% confidence level present value hindsight estimates one year out, respectively. For example, as of December 31, 2007 the expected present value of the one-year hindsight estimate one year out is \$9,453. That amount, shown in Column (2), meets the present value loss funding requirement as of December 31, 2008, if the loss payments in 2008 and beyond follow the expected pattern. However, at the 99.5% quantile, the present value one-year hindsight estimate one year out is \$11,067, shown in Column (3), which implies that an additional amount of $F_{2008} = \$1{,}614$ is needed to ensure full present value loss funding one year out at the 99.5% quantile. The additional required funding amounts one year out, shown in Column (4), generally decline as the portfolio runs

TABLE 4

Additional Loss Funding Need One Year Out
99.5% Confidence Level

At Successive Annual Valuation Dates through 2017
Expected as of December 31, 2007

(1)	(2)	(3)	(4)
Year Ending 12/07+ i	PV of Expected Hindsight One Year Out	PV of 99.5% Hindsight One Year Out	Additional Loss Funding Need One Year Out at 99.5%
$n+i$ [1]	$PV(L_{n+i}) \cdot (1+r_{i:1})$ [2]	$VaR_{99.5\%}(PV(h_{n+i+1}))$ [3]	F_{n+i+1} [4]
2007	\$9,453	\$11,067	\$1,614
2008	\$6,748	\$8,186	\$1,439
2009	\$4,809	\$6,189	\$1,381
2010	\$3,370	\$4,857	\$1,487
2011	\$2,345	\$3,427	\$1,082
2012	\$1,645	\$2,591	\$946
2013	\$1,173	\$1,819	\$646
2014	\$742	\$1,257	\$515
2015	\$445	\$803	\$358
2016	\$164	\$292	\$128
2017	\$0	\$0	\$0

1 n=2007, value of $0 \le i \le k = 10$ implied by valuation year
2 Table 3, Column (3)
3 Table 3, Column (4)
4 Formula 2.24: (4)-(3)

off, reaching $F_{2017} = \$128$ as of December 31, 2016. The additional funding requirement one year out from December 31, 2017 is $F_{2018} = \$0$, because the final loss payment is expected to occur during 2017.

Table 5 illustrates the calculation of $f_{2007+i+1}$ for $0 \le i \le k-1 = 9$. $f_{2007+i+1}$ is the amount by which the present value of the unpaid loss component of the 99.5% quantile one-year hindsight estimate of the unpaid loss L_{2007+i} as of December 2007+*i*+1 exceeds the expected present value of the unpaid loss at that same valuation date, expressed as a ratio to the latter. The expected present values of unpaid losses one year out $PV(L_{2007+i+1})$ appear in Column (2). The values of $PV(L_{2007+i+1} \mid VaR_{99.5\%}(PV(h_{2007+i+1}))$, representing the present value unpaid loss components of the 99.5% quantile hindsight estimates, are shown in Column (3). Column (4) shows the values of $f_{2007+i+1}$, which are calculated from the entries in Columns (2) and (3) using Formula (2.25). For example, in the row corresponding to the December 31, 2007 valuation date, the entry for f_{2008} in Column (4) of 19.4% is the ratio of the Column (2) entry of \$7,840 to the Column (3) entry of \$6,566, less one. The value $f_{2008} = 19.4\%$ tells us that the expected present value unpaid loss amount one year out $PV(L_{2008} \mid VaR_{99.5\%}(PV(h_{2008})) = \$7{,}840$ embedded in the 99.5% quantile present value one-year hindsight estimate $VaR_{99.5\%}(PV(h_{2008})) = \$11{,}067$ of the December 31, 2007 unpaid loss $L_{2007} = 10{,}000$ is 19.4% higher than the expected present value loss $PV(L_{2008}) = \$6{,}566$ as of December 31, 2008[24]. That in turn implies a 19.4% higher risk margin requirement as of December 31, 2008 at the 99.5% confidence loss level than at the expected loss level. As of December 31, 2016 the risk margin top-up factor one year out is treated as $f_{2017} = 0\%$. Both the expected and 99.5% quantile present value hindsight estimates one year out from December 31, 2016 are zero, which implies that the risk margin $R'_{2017} = \$0$.

Table 6 summarizes the recursive calculation of C^R_{2007+i} and R'_{2007+i} for $0 \le i \le k = 10$. Columns (2) and (4) are retabulations of F_{n+i+1} and f_{n+i+1} from Tables 4 and 5, respectively. Column (3) shows the expected risk margin one year out. This is a retabulation of the risk margins shown in Column (7), shifted by one row. For example, as of December 2007 the expected risk margin *one year out* shown in Column (3) is $R'_{2008} = \$1{,}009$, which is also the amount shown in Column (7) as the expected risk margin as of December 2008. The expected risk margin one year out as of December 2016 is $R'_{2017} = \$0$, because the unpaid loss amount as of December 2017 is zero, which implies no further capital or risk margin requirement.

[24] See Table 3, Column (4) for $VaR_{99.5\%}(PV(h_{2008})) = \$11{,}067$ and Table 2, Column (2) for $L_{2007} = 10{,}000$.

TABLE 5

Growth in Risk Margin Need One Year Out
99.5% Confidence Level
At Successive Annual Valuation Dates through 2017
Expected as of December 31, 2007

(1)	(2)	(3)	(4)
Year Ending 12/07+ *i*	Expected PV of Unpaid Losses One Year Out	PV of Unpaid in 99.5% Hindsight One Year Out	Additional Risk Margin Need One Year Out at 99.5%
$n+i$ [1]	$PV(L_{n+i+1})$ [2]	[3]	f_{n+i+1} [4]
2007	$6,566	$7,840	19.4%
2008	$4,664	$5,847	25.4%
2009	$3,247	$4,294	32.2%
2010	$2,249	$3,359	49.3%
2011	$1,579	$2,395	51.7%
2012	$1,123	$1,797	60.1%
2013	$709	$1,125	58.6%
2014	$425	$668	57.2%
2015	$156	$241	55.1%
2016	$0	$0	0.0%

1 n=2007, value of $0 \le i \le k-1=9$ implied by valuation year
2 Table 3, Column (2) one row down
3 Table 3, Column (5)
4 Formula 2.25: (3)/(2)-1

TABLE 6

Required Capital and Risk Margins
Calibrated to 99.5% Confidence Level
At Successive Annual Valuation Dates through 2017
Expected as of December 31, 2007

(1)	(2)	(3)	(4)	(5)	(6)	(7)
Year Ending 12/07+ i	Additional Loss Funding Need One Year Out at 99.5%	Expected Risk Margin One Year Out	Additional Risk Margin Need One Year Out at 99.5%	Required Capital	Annual Pre-Tax Cost of Capital (Paid One Year Out)	Expected Risk Margin
$n+i$ [1]	F_{n+i+1} [2]	R'_{n+i+1} [3]	f_{n+i+1} [4]	C^{R}_{n+i} [5]	[6]	R'_{n+i} [7]
2007	$1,614	$1,009	19.4%	$1,524	$235	$1,204
2008	$1,439	$815	25.4%	$1,386	$222	$1,009
2009	$1,381	$632	32.2%	$1,334	$209	$815
2010	$1,487	$441	49.3%	$1,435	$215	$632
2011	$1,082	$309	51.7%	$1,045	$152	$441
2012	$946	$191	60.1%	$893	$130	$309
2013	$646	$114	58.6%	$601	$86	$191
2014	$515	$54	57.2%	$460	$65	$114
2015	$358	$14	55.1%	$308	$43	$54
2016	$128	$0	0.0%	$108	$15	$14
2017	$0	$0	0.0%	$0	$0	$0

1 n=2007, value of $0 \le i \le k = 10$ implied by valuation year
2 Table 4, Column (4)
3 Column (7) one row down
4 Table 5, Column (4)
5 Formula 2.23: $[(2)+(4)\times(3)]/(1+roe_{PT})$; $roe_{PT} = roe/(1-tax)$; with $roe =$ 15% and $tax = 20\%$
6 $(roe_{PT} - r_{i1}) \cdot C^{R}_{n+i}$; $r_{i:1}$ from Table 1, Column (3) with i=m-1; C^{R}_{n+i} from Column (5)
7 Formula 2.7: $v_{i:1} \times$ [(6) × (5)+(7) one row down]$/(1+roe_{PT})$; $v_{i:1}$ from Table 1, Column (4) with i=m-1

In accordance with step 6(a) of the procedure described in Section 2.5, we start with the last year-end valuation date before the expected final loss payment in 2017, which is December 31, 2016. Because $R'_{2017} = \$0$, Formula (2.23) simplifies to Formula (2.26) and the required capital C^R_{2016} at that date is a function only of F_{2017} and roe_{PT}. Given $roe_{PT} = 15\% / (1 - 20\%) = 18.75\%$ and the value of $F_{2017} = \$128$ shown in Column (2), application of Formula (2.26) results in $C^R_{2016} = \$128 / 1.1875 = \$108$, which appears in Column (5). Next, because $R'_{2017} = \$0$, Formula (2.7) simplifies to Formula (2.27), which defines R'_{2016} simply as the cost of capital $(roe_{PT} - r_{9:1}) \cdot C^R_{2016}$ payable on December 31, 2017 (tabulated in Column (6)), discounted back to December 31, 2016 at the forward rate $r_{9:1}$. Using $r_{9:1} = 5.13\%$ and $v_{9:1} = 95.12\%$ from Table 1 together with $roe_{PT} = 18.75\%$ and $C^R_{2016} = \$108$ in Formula (2.27), $R'_{2016} = 95.12\% \cdot (18.75\% - 5.13\%) \cdot \$108 = \$14$, which appears in Column (7). This completes step 6(a).

Continuing with step 6(b), we back up one year to December 31, 2015. Formulas (2.23) and (2.7) yield requirements $C^R_{2015} = (\$358 + 55.1\% \cdot \$14) / 1.1875 = \$308$ (Column (5)) and $R'_{2015} = 95.42\% \cdot [(18.75\% - 4.80\%) \cdot \$308 + \$14] = 95.42\% \cdot (\$43 + \$14) = \$54$ (Column (7)).

In step 6(c), again using Formulas (2.23) and (2.7), now with $n + i = 2014$, the implied requirements are $C^R_{2014} = (\$515 + 57.2\% \cdot \$54) / 1.1875 = \$460$, shown in Column (5), and $R'_{2014} = 95.62\% \cdot [(18.75\% - 4.58\%) \cdot \$460 + \$54] = 95.62\% \cdot (\$65 + \$54) = \$114$, shown in Column (7).

In accordance with step 6(d), we continue in this fashion to populate Table 6 by working backward one year at a time until reaching the December 31, 2007 valuation date, at which point C^R_{2007} and R'_{2007} are calculated as $C^R_{2007} = \$1{,}614 + 19.4\% \cdot \$1{,}009 / 1.1875 = \$1{,}524$ and $R'_{2007} = 96.77\% \cdot [(18.75\% - 3.34\%) \cdot \$1{,}524 + \$1{,}009] = 96.77\% \cdot (\$235 + \$1{,}009) = \$1{,}204$.

While the ultimate objective of steps 6(a-d) is to determine the risk margin R'_{2007} as of December 31, 2007, valuable byproducts of the recursive procedure summarized in Table 6 are the expected required capital C^R_{2007+i} and risk margin R'_{2007+i} at each successive December valuation date during the run-off period.

Table 7 summarizes the required capital as a ratio to the expected unpaid losses as of December 2007 and at successive December valuation dates through 2016. It shows that the unpaid loss run-off and volatility patterns used in this illustration imply a required capital ratio that starts at 15% of unpaid losses at December 2007 and can be expected to rise during the run-off period, peaking at 70% as of December 2015. We do not know whether that pattern of generally increasing required capital ratios as a run-off portfolio ages is a general phenomenon or a unique result arising from the data used in this illustration. It seems plausible that the one-year volatility of unpaid loss estimates generally increases as a loss portfolio ages, and it seems likely that, in turn, that would lead to a higher capital requirement for a loss portfolio in run-off. However, further study would be required to determine a definitive answer to that question.

TABLE 7
Ratios of Required Capital to Unpaid Loss
Calibrated to 99.5% Confidence Level
At Successive Annual Valuation Dates through 2017
Expected as of December 31, 2007

(1)	(2)	(3)	(4)
Year Ending 12/07+ i	Expected Unpaid Losses	Required Capital	Required Capital Ratio
$n+i$ [1]	L_{n+i} [2]	C^R_{n+i} [3]	c_{n+i} [4]
2007	$10,000	$1,524	15%
2008	$7,161	$1,386	19%
2009	$5,105	$1,334	26%
2010	$3,568	$1,435	40%
2011	$2,467	$1,045	42%
2012	$1,717	$893	52%
2013	$1,206	$601	50%
2014	$752	$460	61%
2015	$442	$308	70%
2016	$160	$108	68%
2017	$0	$0	n/a

1 n=2007, value of $0 \le i \le k = 10$ implied by valuation year
2 Table 2, Column (2)
3 Table 6, Column (5)
4 $c_{n+i} = L_{n+i} / C^R_{n+i}$: (3)/(2)

TABLE 8 Fair Value Reserves Capital Calibration at 99.5% Confidence Level At Successive Annual Valuation Dates through 2017 Expected as of December 31, 2007						
(1)	(2)	(3)	(4)	(5)	(6)	(7)
Year Ending 12/07+ i	Expected Unpaid Losses	Expected PV of Unpaid Losses	Expected Risk Margin	Fair Value Reserve	Risk Margin Ratio to PV of Unpaid Losses	Fair Value Reserve Ratio to Unpaid Losses
$n+i$ [1]	L_{n+i} [2]	$PV(L_{n+i})$ [3]	R'_{n+i} [4]	$T(L_{n+i})$ [5]	$\frac{R'_{n+i}}{PV(L_{n+i})}$	$\frac{T(L_{n+i})}{L_{n+i}}$
2007	$10,000	$9,148	$1,204	$10,351	13.2%	1.04
2008	$7,161	$6,566	$1,009	$7,575	15.4%	1.06
2009	$5,105	$4,664	$815	$5,479	17.5%	1.07
2010	$3,568	$3,247	$632	$3,879	19.5%	1.09
2011	$2,467	$2,249	$441	$2,691	19.6%	1.09
2012	$1,717	$1,579	$309	$1,888	19.6%	1.10
2013	$1,206	$1,123	$191	$1,314	17.0%	1.09
2014	$752	$709	$114	$823	16.1%	1.09
2015	$442	$425	$54	$479	12.8%	1.08
2016	$160	$156	$14	$170	9.0%	1.06
2017	$0	$0	$0	$0	0.0%	n/a

1 n=2007, value of $0 \le i \le k = 10$ implied by valuation year
2 Table 2, Column (2)
3 Table 2, Column (4)
4 Table 6, Column (7)
5 Formula (2.1) generalized for $n+i$: (3)+(4)

Table 8 summarizes the calculation of the fair value of unpaid losses as of December 31, 2007 and subsequent December valuation dates. The fair value reserves are tabulated by valuation date in Column (5). These fair value estimates are based on capital calibration to the 99.5% confidence level combined with a market-clearing return on equity of 15% and market-clearing tax rate of 20% (corresponding to a pre-tax return roe_{PT} of 18.75%). The fair value of the unpaid losses as of December 31, 2007 is $10,351, which is 4% higher than the unpaid loss estimate as of that date. As the loss portfolio runs off, the ratio of the fair value reserve to unpaid losses can be expected to rise from 1.04 as of December 2007 to a peak of 1.10 as of December 2012 and then gradually decline to 1.06 in December 2016. These ratios are shown in Column (7). Column (6) shows the ratio of the risk margin component of the fair value reserve to the present value of the unpaid losses at each valuation date.

In our illustration the fair value reserve is much more sensitive to changes in interest rates than it is to changes in the pre-tax return requirement roe_{PT}. If the spot rate curve as of December 31, 2007 had been one hundred basis points lower at each point, the fair value reserve would have been $10,658 rather than $10,351, which corresponds to shift in the ratio of the fair value reserve to unpaid losses from 1.04 to 1.07. On the other hand, if roe_{PT} had been 17.75% instead of 18.75%, a decline of one hundred basis points, the fair value reserve would have declined from $10,351 to $10,273, which corresponds to a decline in the ratio of the fair value reserve to unpaid losses from 1.04 to 1.03. The change in fair value due to a one hundred basis point change in the risk free rate is about four times the change in fair value resulting from a one hundred basis point change in the required pre-tax return roe_{PT} [25]! Note also that a reduction in the risk-free rate increases the fair value reserve, while a reduction in roe_{PT} reduces it.

Table 9 shows the expected cash flows associated with the runoff of the fair value reserve of $10,351 as of December 31, 2007. Column (2) shows the underwriting assets corresponding to the fair value reserves as of December 31, 2007 and at successive December 31 valuation dates. Implicit in the fair value reserve calculations is the assumption that the fair value reserve amount will be invested in interest bearing assets consistent with the valuation formulas. Accordingly, the entries in Column (2) should be interpreted as invested asset amounts equal to the fair value reserves at each valuation date. Columns (3), (4) and (5) show the expected paid losses, net interest earned and cost of capital incurred during the one-year period following each valuation date. Column (6) shows the assets remaining at the end of each one-year period. Those ending amounts match the ending fair value reserve amounts shown in Column (7).

[25] (10,658-10,351)/(10,273-10,351)=-3.94

TABLE 9 Fair Value Reserve Expected Run-off Cash Flows As of December 31, 2007						
(1)	(2)	(3)	(4)	(5)	(6)	(7)
Year Ending 12/07+ i	Expected Beginning U/W Assets	Expected Paid Losses in Next Year	Expected Net Interest Earned in Next Year	Expected Cost of Capital in Next Year	Expected Ending U/W Assets	Expected Ending Fair Value Reserve
$n+i$ [1]	$T(L_{n+i})$ [2]	P_{n+i+1} [3]	[4]	[5]	[6]	$T(L_{n+i+1})$ [7]
2007	$10,351	($2,839)	$298	($235)	$7,575	$7,575
2008	$7,575	($2,055)	$181	($222)	$5,479	$5,479
2009	$5,479	($1,538)	$146	($209)	$3,879	$3,879
2010	$3,879	($1,100)	$126	($215)	$2,691	$2,691
2011	$2,691	($750)	$99	($152)	$1,888	$1,888
2012	$1,888	($511)	$68	($130)	$1,314	$1,314
2013	$1,314	($454)	$49	($86)	$823	$823
2014	$823	($310)	$31	($65)	$479	$479
2015	$479	($283)	$16	($43)	$170	$170
2016	$170	($160)	$5	($15)	$0	$0
2017	$0	$0	$0	$0	$0	$0

1 n=2007, value of $0 \le i \le k = 10$ implied by valuation year
2 Equal to fair value reserve: Table 8, Column (5)
3 Table 2, Column (3) one row down, expressed as negative number
4 $((2)+0.5\times(3))\times r_{i:1}$, $r_{i:1}$ from Table 1, Column (3)
5 Table 6, Column (6), expressed as negative number
6 (2)+(3)+(4)+(5)
7 Table 8, Column (5) one row down

For example, the December 31, 2007 underwriting assets of $10,351 corresponding to the fair value reserve of the same amount are expected to be reduced over the following year by paid losses of $2,839 (Column (3)) and cost of capital $235 (Column (4)) and increased by $298 of net interest earned (Column (5)), resulting in a balance of $7,575 after one year

(Column (6)). That balance matches the fair value reserve amount as of December 31, 2008 of $7,575 shown in Column (7).

Table 10 illustrates the adequacy of the required capital to ensure fair value funding of unpaid losses at the 99.5% confidence level over each successive one-year time horizon. Columns (2) through (5) are analogous to the same columns of Table 9, and, in fact, for the year beginning December 31, 2007 the entries in Columns (2) and (5) are identical. However, the paid loss amount shown in Column (3) is the paid loss portion of the 99.5% confidence level hindsight estimate one year out (rather than the expected value amount shown in Table 9) and the net interest earned shown in Column (4) reflects that higher paid loss amount. Column (6) shows the accumulated value of the capital assets after one year. Column (7) shows the year-end value of the combined underwriting and capital assets. Column (8) shows the fair value of the unpaid losses embedded in the 99.5% confidence level hindsight estimate.

For example, in the year beginning December 31, 2007 the paid loss portion of the 99.5% confidence level one-year hindsight estimate is $3,174 (vs. the $2,839 in the expected case). Interest earned is slightly lower due to the higher loss payment ($293 vs. $298). The cost of capital is the same $235 as in the expected case. The value of the capital assets at the end of the year is $1,809 ($1,524 × 1.1875). The ending value of the combined underwriting and capital assets after one year is $9,045, which matches the fair value of the unpaid loss portion of the 99.5% confidence level hindsight estimate as of December 31, 2008, which is shown in Column (8).

Table 10 shows that at each successive valuation date through December 31, 2016, the combined underwriting and capital assets are adequate to meet the fair value funding requirement at the 99.5% confidence level. In practical terms that means that sufficient assets are available to fund both the 99.5% confidence level loss obligations as they become payable and the cost of the capital required to support the unpaid losses at that level throughout the run-off period. Because the fair value reserve includes a risk margin sufficient to pay the market cost of capital, the insurer should be able to raise additional capital, if necessary, or, alternatively, a regulator should be able to arrange for a transfer of the unpaid losses to a third party reinsurer with spare capital.

TABLE 10

Adequacy of Capital to Ensure Fair Value Reserve Funding
99.5% Confidence Level
Expected as of December 31, 2007

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Year Ending 12/07+ *i*	Expected Beginning U/W Assets	Expected Paid Losses in 99.5% Level Hindsight	Expected Net Interest Earned in Next Year	Expected Pre-Tax Cost of Capital in Next Year	Expected Ending Capital Assets	Expected Ending U/W + Capital Assets	99.5% Level Ending Fair Value Reserve
$n+i$ [1]	$T(L_{n+i})$ [2]	P_{n+i+1} [3]	[4]	[5]	[6]	[7]	[8]
2007	$10,351	($3,174)	$293	($235)	$1,809	$9,045	$9,045
2008	$7,575	($2,307)	$177	($222)	$1,646	$6,870	$6,870
2009	$5,479	($1,866)	$141	($209)	$1,585	$5,130	$5,130
2010	$3,879	($1,470)	$119	($215)	$1,704	$4,018	$4,018
2011	$2,691	($1,011)	$93	($152)	$1,241	$2,863	$2,863
2012	$1,888	($778)	$63	($130)	$1,061	$2,103	$2,103
2013	$1,314	($679)	$44	($86)	$713	$1,306	$1,306
2014	$823	($575)	$25	($65)	$546	$753	$753
2015	$479	($549)	$10	($43)	$366	$263	$263
2016	$170	($284)	$1	($15)	$128	$0	$0

1 n=2007, value of $0 \le i \le k = 10$ implied by valuation year
2 Equal to fair value reserve: Table 8, Column (5)
3 Paid loss component of 99.5% confidence level one-year hindsight estimate:
[Table 4, Column (3) – Table 5, Column (3)] / (1 + 0.5 × Table 1, Column (3)), expressed as negative number
4 ((2)+0.5×(3))× $r_{i:1}$, $r_{i:1}$ from Table 1, Column (3)
5 Table 6, Column (6), expressed as negative number
6 Table 7, Column (3) × 1.1875: $(C^R_{n+i} \times (1 + roe_{PT}))$
7 (2)+(3)+(4)+(5)+(6)
8 Table 5, Column (3) × (1 + Table 8, Column(6) one row down)

4. SUMMARY AND CONCLUSIONS

In this paper we have derived and illustrated a comprehensive framework for the determination of the fair value reserve for unpaid losses that is consistent with a capital requirement established with the objective of ensuring adequate loss and cost-of-capital funding at the α confidence level for each successive year of the run-off period. That framework supports the consistent quantification of the required capital, the implied cost-of-capital-based risk margin and the fair value reserve from the expected volatility, payment and other characteristics of an unpaid loss portfolio.

Because the fair value reserve at time n is a function of capital, which in turn is a function of the sequence of expected fair value reserves in the run-off period, which are functions of future required capital, and so on, it is necessary to determine the required capital and the fair value reserve using an integrated recursive procedure. The key ingredients required for execution of that procedure are 1) the market-clearing cost of capital, and 2) α-quantile estimates from the distribution of the one-year hindsight loss estimate at each run-off period annual valuation date, as well as knowledge of the time n risk-free yield curve and the expected unpaid loss run-off pattern.

In our illustration we used a market-clearing pre-tax cost of capital roe_{PT} of 18.75%, reflecting an after-tax return on equity assumption of 15% and a tax rate of 20%. Further research is needed on the question of the true market-clearing cost of capital in this context. Conceptually, it is appealing to seek to infer the required after-tax return on equity from observed market returns. However, there are at least two issues which complicate such an analysis.

First, in response to demands by reinsurance buyers for high quality security, active reinsurers have historically held capital far beyond the regulatory minimum level. We suspect that the reinsurance market does not compensate reinsurers for holding that additional capital at the same rate as for the base capital tranche corresponding to the regulatory requirement. If that is true, then unadjusted cost of capital estimates inferred from market returns on held capital will understate the actual cost of capital on the basic Solvency II capital tranche, unless a way can be found to determine and correct for differential market returns by capital tranche.

A second complication relates to the market-clearing tax rate. U.S. reinsurers face a 35% statutory rate, while off-shore reinsurers face much lower statutory rates. However, U.S. reinsurers often pay less than the statutory rate and off-shore reinsurers often pay more. For example, Bermuda reinsurers, subject to a statutory rate of zero at home, typically pay income and excise taxes on some of their U.S. business. The key issue is the *effective* tax rate. The Economist magazine has reported U.S. and OECD-average effective corporate tax rates of 24% and 20%, respectively [2]. However, the Economist-cited study did not examine the effective rates specifically applicable to reinsurers, and those rates might differ from the corporate average. Clearly, further research on the market-clearing tax rate is warranted.

Discussion of how to model the behavior of the successive one-year hindsight loss estimates of unpaid losses throughout the run-off period is beyond the scope of this paper. Clearly, results from such modeling are critical to the application of the framework we have presented and further research in that area would be welcome.

For our illustration, in order to estimate plausible one-year hindsight loss estimate distributions, we analyzed the historical volatility and correlation of one-year loss development by Annual Statement Schedule P line of business and by age reported by the insurer selected for this example. After selecting volatility and correlation parameters, we modeled the one-year development behavior of the illustrative insurer's reserves for all lines. There are other and perhaps better ways of estimating one-year loss reserve development distributions[26].

Fair value reserves are an essential component of insurance company economic capital modeling. As we have shown in this paper, economic capital is also an essential component of fair value reserving. The two are inextricably linked.

An insurer's available economic capital is the difference between its actual fair value assets and its fair value liabilities. Its required economic capital is the amount consistent with a target such as that embedded in the Solvency II directive, where the total capital requirement addresses the risks arising not only from unpaid losses but all other balance sheet and underwriting risks as well. While the focus of this paper has been on the amount of capital required to support fair value loss reserves in isolation, the concepts presented here clearly have application to the economic capital requirements arising from those other risks and indeed the entire insurance enterprise.

APPENDIX A

Deriving Forward Rates from the Spot Rate Curve

We can identify the set of required forward rates by decomposing the yield curve into forward rate components. For example, the two-year spot rate r_2 as of time n is an average

[26] For example, see Appendix C of [6].

rate for the two-year period to maturity, comprising a rate of r_1 for the first year and a rate of $r_{1:1} = \dfrac{(1+r_2)^2}{1+r_1} - 1$ for the second year. $r_{1:1}$ is the one-year forward rate implied by the spot rate curve as of time *n* for the one-year maturity. Likewise, the three-year spot rate r_3 as of time *n* can be decomposed into three discrete one-year rates r_1, $r_{1:1}$, and $r_{2:1} = \dfrac{(1+r_3)^3}{(1+r_2)^2} - 1$ corresponding to the first, second and third years, respectively, of the three year term to maturity. In general, $r_{f:1} = \dfrac{(1+r_{f+1})^{f+1}}{(1+r_f)^f} - 1$ is the *f*-year forward rate implied by the time *n* yield curve for the one-year maturity.

The discount factor $v_m^m = (1+r_m)^{-1}$ implied by the *m*-year maturity (*m* an integer) rate r_m can also be expressed in terms of forward discount factors for the one-year maturity:

$$v_m^m = v_1 \cdot v_{1:1} \cdot v_{2:1} \cdots v_{m-1:1} \tag{A.1}$$

Generally, we can determine any *f*-year forward rate implied by the time *n* yield curve for any *m*-year maturity (including non-integer values of *f* and *m*) as follows:

$$r_{f:m} = \left(\frac{(1+r_{f+m})^{f+m}}{(1+r_f)^f} \right)^{1/m} - 1 \tag{A.2}$$

For example, using Formula (3.2) we can decompose the one-year rate r_1 into the six-month rate $r_{0.5}$ and the six-month forward six-month rate $r_{0.5:0.5} = \left(\dfrac{1+r_1}{(1+r_{0.5})^{0.5}} \right)^2 - 1$. In similar fashion we can also determine forward rates $r_{1.5:0.5}$, $r_{2.5:0.5}$, $r_{3.5:0.5}, \ldots, r_{k-0.5:0.5}$. Note that, in practice, we don't always have rates for maturities at odd intervals such as $r_{1.5}$, $r_{2.5}$, $r_{3.5}, \ldots$, $r_{k-0.5}$ and, in such cases, interpolation is necessary to obtain estimates of such rates.

We can also determine forward rates for multi-year maturities from the forward rates for one-year maturities. For example, given the one-year and two-year forward rates $r_{1:1}$ and $r_{2:1}$ for the one-year maturity, we can determine the one-year forward rate for the two-year maturity as $r_{1:2} = \left((1+r_{1:1}) \cdot (1+r_{2:1})\right)^{1/2}$. Similarly, the one-year forward discount factor can be expressed in terms of the one-year and two-year forward discount factors for the one-year maturity: $v_{1:2}^2 = v_{1:1} \cdot v_{2:1}$. In general, the formula for the *f*-year forward discount factor for the *m*-year maturity (*m* an integer) can be expressed as:

$$v_{f:m}^m = v_{f:1} \cdot v_{f+1:1} \cdot v_{f+2:1} \cdots v_{f+m-1:1}$$

$$= \prod_{i=0}^{m-1} v_{f+i:1} \tag{A.3}$$

APPENDIX B

Proof of Formula (2.6): $\boldsymbol{R'_n = v_1 \cdot ((\frac{roe}{1-tax} - r_1) \cdot C_n^R + R'_{n+1})}$

Formula (2.5) expresses $T(L_n)$ as follows:

$$\begin{aligned} R'_n &= (\frac{roe}{1-tax} - r_1) \cdot C_n^R \cdot v_1 + (\frac{roe}{1-tax} - r_{1:1}) \cdot C_{n+1}^R \cdot v_2^2 + (\frac{roe}{1-tax} - r_{2:1}) \cdot C_{n+2}^R \cdot v_3^3 \\ &+ \ldots + (\frac{roe}{1-tax} - r_{k-1:1}) \cdot C_{n+k-1}^R \cdot v_k^k \end{aligned} \tag{2.5}$$

If we replace the multi-year risk-free discount factors $v_2^2, v_3^3,\ v_4^4, \ldots, v_k^k$ with equivalent factors based on forward rates for one-year money, we can rewrite Formula (2.5) as:

$$\begin{aligned} R'_n = &\left(\frac{roe}{1-tax} - r_1\right) \cdot C_n^R \cdot v_1 + \left(\frac{roe}{1-tax} - r_{1:1}\right) \cdot C_{n+1}^R \cdot v_1 \cdot v_{1:1} \\ &+ \left(\frac{roe}{1-tax} - r_{2:1}\right) \cdot C_{n+2}^R \cdot v_1 \cdot v_{1:1} \cdot v_{2:1} + \ldots + \left(\frac{roe}{1-tax} - r_{k-1:1}\right) \cdot C_{n+k-1}^R \cdot v_1 \cdot v_{1:1} \cdot v_{2:1} \cdots v_{k-1:1} \end{aligned}$$

Factoring out the one-year discount factor v_1 from all of the terms, we obtain:

$$\begin{aligned} R'_n = v_1 \cdot \Bigg(&\left(\frac{roe}{1-tax} - r\right)_1 \cdot C_n^R + \left(\frac{roe}{1-tax} - r_{1:1}\right) \cdot C_{n+1}^R \cdot v_{1:1} \\ &+ \left(\frac{roe}{1-tax} - r_{2:1}\right) \cdot C_{n+2}^R \cdot v_{1:1} \cdot v_{2:1} + \ldots + \left(\frac{roe}{1-tax} - r_{k-1:1}\right) \cdot C_{n+k-1}^R \cdot v_{1:1} \cdot v_{2:1} \cdots v_{k-1:1} \Bigg) \end{aligned}$$

and, finally:

$$R'_n = v_1 \cdot \left(\left(\frac{roe}{1-tax} - r_1 \right) \cdot C_n^R + R'_{n+1} \right), \tag{2.6}$$

where

$$\begin{aligned} R'_{n+1} = &\left(\frac{roe}{1-tax} - r_{1:1}\right) \cdot C_{n+1}^R \cdot v_{1:1} + \left(\frac{roe}{1-tax} - r_{2:1}\right) \cdot C_{n+2}^R \cdot v_{1:1} \cdot v_{2:1} \\ &+ \left(\frac{roe}{1-tax} - r_{3:1}\right) \cdot C_{n+3}^R \cdot v_{1:1} \cdot v_{2:1} \cdot v_{3:1} + \ldots + \left(\frac{roe}{1-tax} - r_{k-1:1}\right) \cdot C_{n+k-1}^R \cdot v_{1:1} \cdot v_{2:1} \cdot v_{3:1} \cdots v_{k-1:1} \end{aligned}$$

R'_{n+1} can be characterized as the time *n* estimate of the present value risk charge required at time *n*+1. In general, R'_{n+i}, the time *n* estimate of the present value risk charge required at time $n+i$, can be expressed for $1 \le i \le k-1$ as:

$$\begin{aligned} R'_{n+i} = &\left(\frac{roe}{1-tax} - r_{i:1}\right) \cdot C^{R}_{n+i} \cdot v_{i:1} + \left(\frac{roe}{1-tax} - r_{i+1:1}\right) \cdot C^{R}_{n+i+1} \cdot v_{i:1} \cdot v_{i+1:1} \\ &+ \left(\frac{roe}{1-tax} - r_{i+2:1}\right) \cdot C^{R}_{n+i+2} \cdot v_{i:1} \cdot v_{i+1:1} \cdot v_{i+2:1} + \ldots \\ &+ \left(\frac{roe}{1-tax} - r_{k-1:1}\right) \cdot C^{R}_{n+k-1} \cdot v_{i:1} \cdot v_{i+1:1} \cdot v_{i+2:1} \cdots v_{k-1:1} \end{aligned}$$

or, more succinctly, as

$$R'_{n+i} = v_{i:1} \cdot \left(\left(\frac{roe}{1-tax} - r_{i:1}\right) \cdot C^{R}_{n+i} + \cdot R_{n+i+1}\right) \tag{2.7}$$

Abbreviations and Notations

α	= confidence level (probability) that insolvency can be avoided
c_n	= ratio of required capital to unpaid losses at time n : C_n^R / L_n
C_n^R	= $c_n \cdot L_n$ = required capital at time n
f	= subscript denoting the time (years) to a forward contract delivery date
f_{n+1}	= fraction by which the time n+1 unpaid losses embedded in the one-year hindsight estimate at α confidence level exceeds the expected time n+1 one-year hindsight estimate $h_{n+1} = l_{n+1} + p_{n+1}$ = random variable, at time n, for one-year hindsight losses as of time n+1, given L_n
i	= integer subscript denoting a number of years beyond the initial valuation date at time n, $0 \le i \le k-1$
k	= integer number of years of loss payments beyond time n
L_n	= unpaid losses at time n
l_{n+1}	= random variable, at time n, for unpaid losses as of time n+1, given L_n
$L_{n+1} + P_{n+1}$	= one year hindsight estimate of L_n at time n+1
m	= integer subscript denoting the time (years) to maturity of a bond
n	= integer subscript denoting the first of a sequence of annual loss reserve valuation dates (time n+i is i years later)
P_{n+1}	= paid losses between time n and n+1
p_{n+1}	= random variable, at time n, for paid losses between time n and n+1, given L_n
$Prob(\cdot)$	= probability operator
$PV(\cdot)$	= risk-free present value operator
$PV(h_{n+1})$	$= PV(L_{n+1}) + p_{n+1} \cdot (1 + \frac{1}{2} r)$ = random variable, at time n, for the present value of h_{n+1} as of time n+1, given L_n
$PV(L_{n+1} \mid VaR_\alpha(PV(h_{n+1}))$	= present value of the unpaid loss component of the one-year hindsight loss estimate a the α confidence level
R'_n	= risk-free present value of future risk charges associated with unpaid losses L_n at time n
r	= risk-free annual interest rate assuming a flat yield curve

$r_{f:m}$ = risk-free annual f-year forward interest rate on the m-year maturity bond for the period from time $n+f$ to $n+f+m$

r_m = risk-free annual interest rate for the m-year maturity bond for the period from time n to $n+m$

r'_n = $c_n \cdot (\frac{roe}{1-tax} - r)$ = annual risk charge expressed as a rate of return on L_n

roe = annualized required after-tax return on equity (capital)

roe_{PT} = annualized required pre-tax return on equity (capital)

S_{n+1} = $(T(L_n) + C_n^R) \cdot (1+r)$ = accumulated value at time $n+1$ of initial assets equal to time n capital and loss reserve fair value plus interest

tax = income tax rate

$T(\cdot)$ = fair value at time n of unpaid losses L_n

$T(L_n)$ = fair value at time n of unpaid losses L_n

$T(l_{n+1})$ = random variable, at time n, for fair value at time $n+1$ of unpaid losses, given L_n

$T(L_{n+1} + P_{n+1})$ = $T(L_{n+1}) + P_{n+1} \cdot (1 + \frac{1}{2}r)$ = fair value at time $n+1$ of one-year hindsight estimate of L_n

T_{n+1}^{-1} = inverse distribution function of t_{n+1}

$T(P_{n+1})$ = $P_{n+1} \cdot (1 + \frac{1}{2}r)$ = fair value at time $n+1$ of paid losses P_{n+1}

$T(p_{n+1})$ = random variable, at time n, for fair value at time $n+1$ of paid losses between time n and $n+1$, given L_n

t_{n+1} = $T(l_{n+1} + p_{n+1})$ = $T(l_{n+1}) + p_{n+1} \cdot (1 + \frac{1}{2}r)$ = random variable, at time n, for fair value at time $n+1$ of one-year hindsight estimate of L_n

v = $(1+r)^{-1}$ = one-year risk-free discount factor assuming a flat yield curve

$v_{f:m}$ = $(1+r_{f:m})^{-1}$ = one-year risk-free discount factor corresponding to $r_{f:m}$

v_m = $(1+r_m)^{-1}$ = one-year risk-free discount factor r_m

$VaR_\alpha(t_{n+1})$ = Value-at-Risk with respect to t_{n+1} at the α confidence level

$VaR_\alpha(PV(h_{n+1}))$ = Value-at-Risk with respect to $PV(h_{n+1})$ at the α confidence level

REFERENCES

[1] Conger, Robert F., James D. Hurley and Stephen P. Lowe, "How Might the Presentation of Liabilities at Fair Value Have Affected the Reported Results of U.S. Property and Casualty Insurers", *Fair Value of P&C Liabilities: Practical Implications*, Arlington, VA, Casualty Actuarial Society, **2004**, 1-114. http://www.casact.org/pubs/fairvalue/

[2] *The Economist*, "Tax reform: Overhauling the old jalopy", Economist Newspaper Limited, New York, August 2, **2007**. http://www.economist.com/finance/displaystory.cfm?story_id=9596317

[3] Feldblum, Sholom, "Fair Value Accounting for Property-Casualty Insurance Liabilities", *Casualty Actuarial Society Discussion Paper Program*, Volume: Spring, **2006**, 1-36, http://www.casact.org/pubs/dpp/dpp06/06dpp01.pdf

[4] "Gearing up for Solvency II", PricewaterhouseCoopers, **2007**. http://www.pwc.com/extweb/pwcpublications.nsf/docid/3CC2D8034A0F95D78525734700504E62/$File/PwC_gearing_up_for_Solvency_II.pdf

[5] Gutterman, Sam, Mark W. Littmann, Marcus Tarrant and Dan Thomas, "An Investigation of Practical Matters Related to Implementing Fair Value Accounting for Property/Casualty Loss Reserves", *Fair Value of P&C Liabilities: Practical Implications*, Arlington, VA, Casualty Actuarial Society, **2004**, 115-175. http://www.casact.org/pubs/fairvalue/115.pdf

[6] Wacek, Michael G., "Consistent Measurement of Property-Casualty Risk-Based Capital Adequacy", *Casualty Actuarial Society Forum*, Volume: Spring, **2007**, 107-163. http://www.casact.org/pubs/forum/07spforum/07Sp113.pdf

Note: An Excel spreadsheet supporting the calculation of the values of Tables 1 through 10 is available at http://www.casact.org/library/index.cfm?fa=caveat.