Flexible Factor Chain Ladder Model: A Stochastic Framework for Reasonable Link Ratio Selections

Emanuel Bardis, FCAS, MAAA; Ali Majidi; and Daniel Murphy, FCAS, MAAA

Abstract: The popular General/Property-Casualty Insurance chain ladder method was first expanded to include variance calculations by Mack [1]. As new research expands the chain ladder method's stochastic functionality, it is as important as ever to understand the assumptions underlying this fundamental approach and evaluate their appropriateness given the data. The purpose of this paper is to introduce more statistical rigor to this popular method and help bridge the gap between practice and statistical theory. We will expand the regression approach of Murphy[2] so that selected link ratios other than simple or volume weighted averages can be seen as optimizing a rigorous statistical model. We will derive formulas for the parameter risk and process risk of ultimate losses projected from such selected link ratios. We will discuss residual analysis and statistical measures for validating the selected factors. Using data previously analyzed in the literature, we will compare stochastic results from the popular application of the Mack formula to those based on our model. It is hoped that this paper will provide the actuarial practitioner with a statistically rigorous framework with which to measure objectively the appropriateness of the chain ladder deterministic and stochastic results, make more informed judgmental selections, and avoid injudicious conclusions based on potentially inappropriate assumptions.

Keywords: chain ladder; selection; Mack; Murphy; variance; reserve risk; residuals

Introduction

The Chain Ladder method is the most popular algorithm by which actuarial practitioners calculate a central estimate of the unpaid claim liability. Given the need of the actuarial profession to provide statistical descriptions, or models, of the loss development process, much research in the last two decades has been dedicated to framing this method within a statistical structure. This is the reason for the appeal of the stochastic formulas of Dr. Thomas Mack [1], who was the first to produce such a statistical model for the case of volume weighted average age-to-age factors (link ratios). Murphy [2] considers the chain ladder method as a special case of a more general linear regression approach. Zehnwirth [3] refers to this broader class of chain ladder models as the "extended link ratio family," but rejects that family on the grounds of insufficient predictive power and favors the "probability trend family" (PTF) instead. Using a Bayesian approach, Verall [4] incorporates judgment in a rigorous fashion to tackle the inflexibility of other methods (such as Zehnwirth's PTF). Unfortunately, the difficulty of verifying *a priori* link ratio distributions and the overall complexity of the MCMC (Markov Chain Monte Carlo) algorithm make this method difficult to implement in large enterprises.

The purpose of this paper is to bridge the gap between the stochastic underpinnings of the chain ladder method and its implementation in practice, i.e., when link ratios are selected based on judgment. We present a general chain ladder model that fulfills two key requirements:

1. Its central estimates are consistent[1] with chain ladder projections based on judgmentally selected factors[2], and
2. Its underlying assumptions and actuarial inputs are testable within a rigorously-defined statistical framework.[3]

The paper is organized as follows. In Section 1 we propose a flexible yet rigorous model of the chain ladder method built around the regression interpretation similar to Murphy [2] that satisfies the two requirements above. We call this model the Flexible Factor Model (FFM). In Section 2 we present formulas for the mean square error of chain ladder projections based on selected link ratios, as long as those selections are "reasonable" (defined below). In Section 3 we demonstrate how our model naturally embeds a process for visually and statistically testing the consistency of the actuary's selected link ratios with the development data within the triangle. Section 4 demonstrates these concepts with an example. In Section 5 we compare the FFM process, parameter and total risk estimates to the Mack [1] versions, showing that the Mack model is a special case of FFM when the actuary selects volume weighted link ratios. We also show how common use of "the Mack Method" can significantly understate potential variability. Section 6 is a summary that also includes thoughts for future research. We conclude with an Appendix of proofs of our major results.

1. A Chain Ladder Model for Flexibly-Selected Link Ratios

We start with the usual notation, where the observed cumulative paid losses[4] 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

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

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

This model is similar to the model proposed by Mack [1] and Murphy [2], with a more general and, as we shall see, more flexible error assumption (1). Assumption (2) is that the set $\{\varepsilon_{ik} \mid 1 \le i \le I, 1 \le k \le I + 1 - i\}$ of "noise given the Triangle $\boldsymbol{D}$" is comprised of independent

[1] By "consistent" we mean that the model's estimates will be the same as the estimates produced by the chain ladder's algorithmic steps.

[2] For example, when considering benchmarks in a reserve analysis.

[3] For example, to test, validate, and approve a company's internal model within the framework of Solvency II; see *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 - Solvency II* {*SEC(2007) 870*} {*SEC(2007) 871*, Article 43: Risk Management http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:52007PC0361:EN:NOT

[4] We refer to loss amounts as paid losses for consistency with prior literature. In fact, losses can be either paid or incurred amounts; can include or exclude adjustment expense; can even refer to claim counts. In short, the theory applies to any chain ladder estimable amount.

identical distributed (i.i.d.) normal[5] random variables; in particular we assume $E(\varepsilon_{ik}) = 0$ and $E(\varepsilon_{i,k}^2) = 1$. Making explicit the implicit assumption of the error term is crucial for providing a data set of residuals for model testing.

Under assumptions (1), (2) the best linear unbiased estimate of the link ratio, given the set of observations ***D,*** can be calculated as weighted averages of the observed link ratios:

$$\hat{f}_k(\alpha) = LR_k(\alpha) = \sum_{i=1}^{n-k} \frac{C_{i,k}^{1-\alpha}}{\sum_{j=1}^{n-k} C_{j,k}^{2-\alpha}} \cdot C_{i,k+1} = \sum_{i=1}^{n-k} w_{i,k}^{\alpha} \cdot F_{i,k} \tag{3}$$

where

$$F_{i,k} = \frac{C_{i,k+1}}{C_{i,k}}$$

is the accident year i link ratio from age k to age $k+1$, and the weights are functions of α:

$$w_{i,k}^{\alpha} := \frac{C_{i,k}^{2-\alpha}}{\sum_{j-1}^{n-k} C_{j,k}^{2-\alpha}}. \tag{4}$$

The optimal solution of model (1), (2) is specified by the parameters $(\hat{f}, \hat{\alpha})$ (the "model specification") where the solution for the values of the $\hat{\alpha}$'s will be discussed below. (Notational remark: The superscript α of w is not an exponent but emphasizes that the weights are a function of α.)

To illustrate, consider the "Distribution Free" chain ladder model introduced in Mack [8]. Mack's model is a special case of model (1), (2) with $\alpha_k=1$, $k=1,\ldots,I$. Mack proved that the weighted average link ratio estimators

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

are unbiased with the smallest variance[6]. Clearly these estimators are consistent with formula (3) with α=1 for all k.

[5] 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.

[6] Submission to the 1994 *Variability in Reserves Prize Program*: "Measuring the Variability of Chain Ladder Reserve Estimates," CAS Forum, 1994, Vol. 1, p. 141.

To reiterate, by selecting an α parameter we specify the *variance assumption* of model (1), (2). We shall see that for any "reasonably selected link ratio" (defined below) we can select an appropriate α parameter that will yield the selected link ratio as the best linear unbiased chain ladder estimator. That is what we mean when we say that by virtue of this simple error term extension, *model* (1), (2) embeds the traditional selected-factor-based chain ladder *method* in a statistical framework. We refer to model (1), (2) as the Flexible Factor Model (FFM).

We digress momentarily to distinguish between a model and a method. A model is a mathematical description of an observation, process or phenomenon, where "best fitted" parameters are based on the underlying data characteristics. A method on the other hand is an algorithm that produces estimates through a sequence of predetermined steps. Thus a method can always be used to calculate some estimates, whereas a model is based on assumptions that should be tested before its results are trusted. The traditional Chain Ladder *method* is consistent with many stochastic models that have been built around it, such as the Mack, Murphy, and over-dispersed Poisson models. In practice, however, actuaries *select* link ratios judgmentally because simple or volume weighted averages may be inappropriate in certain situations.[7] There is no doubt that such flexibility makes practical sense, but no matter how experienced an actuary is, the appropriateness of his/her judgment is always open to question. Under the *model* framework of this paper an actuary can respond to such challenges with objective, statistical justification.[8] We revisit this point in Section 3.

We present now our first major results.

Theorem 1.1: The "Reasonable" Link Ratio Function

Consider for a given triangle the corresponding link ratio function as in (3) and denote the set of all *reasonable* link ratios with $\mathrm{LR}_k(\Re) := \{\mathrm{LR}_k(\alpha) \mid \alpha \in \Re\}$ where $i_{\min,k}$ and $i_{\max,k}$ are the indices of $\min\{C_{j,k}, j < I-k\}$ and $\max\{C_{j,k}, j < I-k\}$ respectively. Then

1. If $c, d \in LR_k(\Re)$, then the whole interval $[c,d] \subseteq LR_k(\Re)$
2. $\mathrm{LR}_k(\alpha) \to F_{i_{\min},k}$ as $\alpha \to \infty$
3. $\mathrm{LR}_k(\alpha) \to F_{i_{\max},k}$ as $\alpha \to -\infty$
4. Every link ratio between the straight average, the weighted average and the link ratios corresponding to the minimum $\min\{C_{j,k}, j < I-k\}$ and maximum

[7] An experienced actuary recognizes, for example, trends in the triangles and adjusts the link ratios manually, perhaps influenced by a benchmark pattern.

[8] 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 identically distributed" assumption (2). The Bootstrapping results will be wrong if this assumption is violated.

$\max\{C_{j,k}, j < I-k\}$ of the loss amounts as of the previous maturity (i.e., the denominators of these link ratios) is reasonable.

Theorem 1.2: Existence of Optimal Alphas

Let $h_k \in \mathrm{LR}_k(\Re), k \le I-2$ be a set of reasonable link ratios (as defined in Theorem 1.1). Then for each k there is at least one α such that $\boldsymbol{h_k}$ is the ML-estimator of (1). We define the "optimal alpha" as

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

The condition k≤I-2 is stipulated because for the last development period (*k*=I-1) a regression approach will not work if there is only one observation. Proofs of Theorems 1.1 and 1.2 are in the appendix.

In other words, among all possible α, we take the one with smallest absolute value. If two possible α have exactly the same absolute value (i.e., more than one standard deviation FFM variance assumption is associated with the same link ratio), we choose the positive one. Thus, $\hat{\alpha}_k$ is well defined. Furthermore, values can be calculated using a solver[9].

Note that in the usual chain ladder model the standard deviation of the paid development process is assumed to be proportional to the square root of cumulative payments as of the beginning of the period. But why should this hold for all development years? Theorem 1.2 relaxes the volume-weighted requirement and shows that even with reasonable, judgmentally selected link ratios there exists an underlying statistical model with those selections being the optimal solutions.

2. Standard Error Formulas for the Flexible Factor Model

The Flexible Factor Model's link ratio parameters f_k can be estimated using weighted least squares regression. Let $\hat{f}_k$ denote those estimators, themselves random variables. Estimates of the conditional variance of those estimators, $\Delta^2(\hat{f}_k^2) := E(\hat{f}_k^2 \mid D) - \hat{f}_k^2$, and estimates $\hat{\sigma}_k$ of the scale parameters are standard outputs of regression software.[10]

Formulas for an Individual Accident Year

Consider an individual accident year (or "origin year") i as of its current age (or "lag") k. An estimate $\hat{C}_{i,k+1}$ of the mean value $\mu_{i,k+1}$ of the future loss $C_{i,k+1}$, given D, can be found by completing the square in the chain ladder sense. Assuming the estimates are unbiased, the mean square error of

[9] For example the Newton-Algorithm with starting point 0.
[10] The delta operator Δ denotes parameter risk. Excel's LINEST function refers to the statistics as "se1…the standard error of the coefficient" and "sey…the standard error for the y estimate," respectively.

the estimate, which by definition is the expected squared difference between the estimate and its target, is the sum of parameter risk and process risk:

$$mse(\hat{C}_{i,I}) = E((\hat{C}_{i,I} - C_{i,I})^2 \mid D) = (\hat{C}_{i,I} - \mathrm{E}(C_{i,I} \mid D))^2 + \mathrm{E}((C_{i,I} - \mathrm{E}(C_{i,I} \mid D))^2 \mid D)$$
$$= \Delta^2(C_{i,I}) + \Gamma^2(C_{i,I})$$

with Δ^2 and Γ^2 denoting the operators for parameter risk and process risk, respectively. Parameter risk and process risk can be calculated recursively according to the formulas below.

Parameter Risk: Variance of the estimate of the mean future value of loss

For the first period after the current diagonal,

$$\Delta^2(\hat{C}_{i,k+1}) = C_{i,k}^2 \Delta^2(\hat{f}_k)$$

since $C_{i,k}^2$ is a constant.

For s=2, 3, …,

$$\Delta^2(\hat{C}_{i,k+s}) = \mu_{i,k+s-1}^2 \Delta^2(\hat{f}_{k+s-1}) + f_{k+s-1}^2 \Delta^2(\hat{C}_{i,k+s-1}) + \Delta^2(\hat{f}_{k+s-1}) \Delta^2(\hat{C}_{i,k+s-1}) .$$

Note: The formulas above agree with "the Mack Formula" for α = 0, 1, 2 [7] with the exception of the third term (the product of the variances) in the parameter risk formula.

Process Risk: Variance of the deviation of future value of loss from its mean

For the first period after the current diagonal, $\Gamma^2(C_{i,k+1}) = C_{i,k}^{\alpha_k} \sigma_k^2$. For subsequent periods

$$\Gamma^2(C_{i,k+s}) = \left(E(C_{i,k+s-1}) \middle| D\right)^{\alpha_{k+s-1}} \cdot \Psi\left(\alpha, cv(C_{i,k+s-1})\right) \cdot \sigma_{k+s-1}^2 + f_{k+s-1}^2 \cdot \Gamma^2(C_{i,k+s-1})$$

where Ψ is a function of α and of the coefficient of variation κ of the future losses. Estimates of the expected values $E(C_{i,k+s-1})$ come from the chain ladder's "squaring-the-triangle" process. The f_{k+s-1} are the selected link ratios, and $\Gamma^2\left(C_{i,k+s-1}\right)$ is the process risk as of the previous age.

Under the assumption that the "noise" $\varepsilon_{i,k}$ is normally distributed, it is straightforward to show that the function Ψ is a polynomial in κ for positive integer (n) values of α:[11]

---

[11] Expand the Taylor series of f(x)=x^n around μ, and use the fact that odd central moments of a normal random variable are zero and even central moments are related to σ according to the formula $E(X^n) = \frac{n!}{2^{n/2}(n/2)!}(\sigma)^n$. See http://en.wikipedia.org/wiki/Normal_distribution#Moments.
An exact value for Ψ(α,κ) for non-integral positive values of α is difficult to present in closed form because E(X^α) is undefined when the probability that X<0 is non-zero (for example, E(√X) is undefined for X<0). For triangles of property/casualty losses with small coefficient of variation, reasonable approximations for such real "moments" are available using simulation. Details available upon request of the authors.

$$\Psi(n,\kappa) = \sum_{\substack{j=0 \\ j\text{ even}}}^{n} \frac{1\cdot n\cdot(n-1)\cdots(n-(j-1))}{2^{j/2}\,(j/2)!}\kappa^{j}\ .$$

For $\alpha > 0$ but not an integer, we recommend linearly interpolating $\Psi(\alpha,\kappa)$ between $\Psi(\lfloor\alpha\rfloor,\kappa)$ and $\Psi(\lceil\alpha\rceil,\kappa)$ where $\lfloor\alpha\rfloor$ denotes the floor function of α, which is the largest integer $\leq \alpha$, while $\lceil\alpha\rceil$ denotes the ceiling function of α, which is the smallest integer $\geq \alpha$.[12] For negative values of α, note that such a selection would imply an actuarial assumption that the variability of loss at the end of a development period is *inversely* proportional to the value of loss at the beginning of the period, an unusual assumption for General/Property-Casualty insurance. Nevertheless, if the data and the selection indicate a negative α, we recommend using simulation to calculate $\Psi(\alpha,\kappa)$. Such simulations could be performed with Excel or another programming language. An example using R is provided as Appendix B. For illustration, Figure 5 in Appendix A graphs simulated values of Ψ as a function of α for different coefficients of variation. Notice that Ψ is a convex function, so linear interpolations for positive α will be conservative approximations.

To calculate the parameter and process risk quantities, we need to estimate Δ^2 and Γ^2. We follow the traditional statistics approach here, replacing all unknown quantities by their corresponding estimates.[13]

Formulas for All Accident Years Combined

Recursive variance formulas for all accident years combined become slightly more complicated because at each new age an additional accident year is included.

For ages j = 2, 3, …, let $X_j = \sum_{i=I-j+2}^{I} C_{i,j}$ be the sum of the future values of losses for accident years that have not yet matured to age j (the most recent accident year is denoted by I). Let $M_j = \sum_{i=I-j+2}^{I} \mu_i$ denote the expected value of X_j and let $\hat{X}_j = \sum_{i=I-j+2}^{I} \hat{C}_{i,j}$ be its chain ladder estimate.

Parameter Risk: Variance of the estimate $\hat{X}_j$

For j=2, only the most recent accident year is included in the total, so the parameter risk of $\hat{X}_2$ is equal to $\Delta^2(\hat{X}_2) = \Delta^2(\hat{f}_1)\cdot C_{I,1}^2$.

For j=3, 4, …,

$$\Delta^2(\hat{X}_j) = (M_{j-1} + C_{I-j+2,j-1})^2\,\Delta^2(\hat{f}_{j-1}) + f_{j-1}^2\,\Delta^2(\hat{X}_{j-1}) + \Delta^2(\hat{f}_{j-1})\,\Delta^2(\hat{X}_{j-1}).$$

---

12 Our tests have shown that for small κ the FLOOR and CEILING functions yields not significantly different results.

13 See [6] for a discussion of resampling.

Process Risk: Variance of X_j

Under Assumption (2), all accident years are independent; therefore, the process variance of the sum of the future values as of a given age is the sum of the process variances:

$$\Gamma^2(X_j) = \sum_{i=I-j+2}^{I} \Gamma^2(C_{i,j}) .$$

As before, the formulas above agree with the "Mack Formula" for α = 0, 1, 2 with the exception of the third term (the product of the variances) in the parameter risk formula.

3. Residuals and Model Selection

In the traditional world, an actuary's methods and selections are defended by his/her expertise and experience. In a modeling world, mathematical and graphical tools can provide more objective ways to defend one's selections and to communicate one's conclusions. One of the most important diagnostic and validation tools are residuals, which are in general the difference between a data set and its "formulaic representation." For FFM, the formulaic representation of the data is given by the model specification $(\hat{f},\hat{\alpha})$ and the corresponding residuals are defined as

$$r_{i,k} := r_{i,k}(\hat{f},\hat{\alpha}) := (C_{ik+1} - \hat{f}_k C_{ik}) / (\hat{\sigma}_k C_{i,k}^{\hat{\alpha}_k/2}) \tag{5}$$

Now, given a set of selected link ratios, how does the actuary confidently defend the resulting estimate of the unpaid claim liability? The first step is to demonstrate that the corresponding α parameters lead to residuals $\{r_{i,k}\}$ that "look like noise." This "noise hypothesis" – i.e., the residuals are independent and identically distributed normal random variables – can be tested visually (e.g., QQ-plots) as well as with "hard" statistics (e.g., Shapiro-Francia-test for normality [5]). If the test fails and one chooses to adjust the selections, how does one know if the new set of link ratios are "better" than the initial selections?

The raising and answering of these questions within a reserve analysis is encapsulated by an analytical flow which we call the "actuarial validation cycle," illustrated in Figure 1 below:

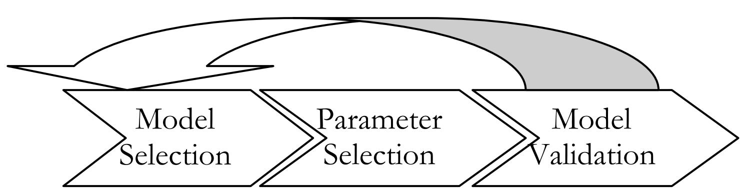

Figure 1: Actuarial validation cycle

The actuarial validation cycle underscores the idea that models offer proposals to understand the data structure.

Figure 1 illustrates that a failed validation step leads to re-selecting the initial model. Selecting a model other than FFM is certainly an option, but that decision and its implications are beyond the scope of this paper.

Assumptions *per se* can be argued but not tested. Assumptions formed as *hypotheses*, however, can be mathematically tested by their implied residuals. The FFM approach to the traditional chain ladder practice of selecting link ratios is one way to test and validate those selections objectively. There may be others. To cite George Box, "Essentially, all models are wrong, but some are useful."

4. An Example

We consider the triangle in Table 1 of RAA data quite well analyzed in Mack [1], Zehnwirth [3] and elsewhere in the literature. We consider it here within the FFM framework to illustrate possible iterations through the actuarial validation cycle of **Figure 1**.

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

First we declare our *goal*, which is to find a model that describes our data within a certain level of confidence.

- Model Selection: We start with the FFM chain ladder model, which means that we believe cumulative losses behave according to the equation

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

- Parameter Selection: This means that we choose a set of link ratios, then calculate the corresponding α, which in turn determines the variance assumption. We start here by selecting the simple averages shown above.

- Model Validation: Now we test the corresponding residuals shown in Table 2.

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							

In the graph below, the residuals appear fairly random. A few of the residuals (the red ones) are outliers.

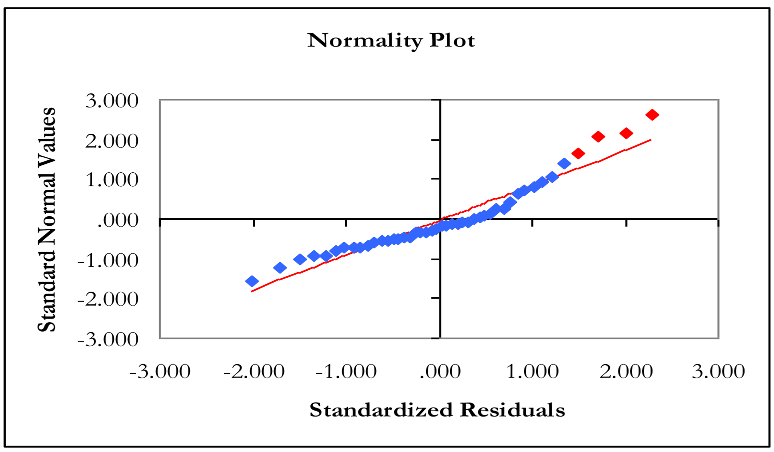

Figure 2: Residuals based on simple average link ratio selections versus quantiles of the normal distribution (red line)

Besides the visual diagnostic above, we want to check the "noise hypothesis" with an objective statistical test. Here the Shapiro-Francia P-Value is 0.26% which suggests that the assumption of normality of the residuals is rejected at the 5.0% confidence level. (When the P-Value is less than one's predetermined confidence level, the null hypothesis – i.e., that the residuals are i.i.d. normal – should be rejected.) This means we need to go back to step one.

- Model (Re-)Selection: Within the scope of this paper we stay within the FFM framework.

- Parameter (Re-)Selection: The first few link ratios produce outliers in our first iteration, so we might change the first three selected link ratios to volume weighted averages. In our second iteration 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: By comparing Figure 2 and Figure 3, we see that the re-selected link ratios lead to residuals that have a much better appearance of being a random sample from a normal distribution. The Shapiro-Francia test delivers a P-Value of 12.0%, so given our 5% confidence level we would accept this model, the selected parameters, the corresponding liability estimates, and the standard errors.

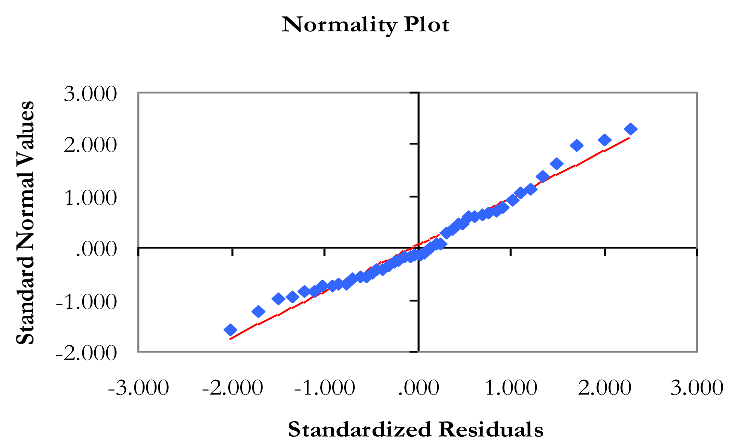

Figure 3: Residuals based on the re-selected link ratios versus quantiles of the normal distribution

5. Comparing Uncertainty Estimates

In practice today it is not uncommon to find coefficients of variation (CV's) based on Mack's volume-weighted-average standard error formulas applied to chain ladder projections based on selected link ratios that are not volume weighted averages. The resulting uncertainty estimates can be suspect due to this fundamental inconsistency. The FFM model eliminates this inconsistency by making sure that all reasonably selected link ratios are best linear unbiased estimators for an underlying model of the data, and the uncertainty estimates resulting from the FFM formulas are consistent with those selections.

To illustrate this point, Table 3 below compares the standard errors of the Mack and FFM models for the RAA triangle (Section 4) when the selected factors are the volume weighted averages.

Table 3 - Standard Errors based on volume weighted average factors

AY	FFM results			Mack results		
	Process risk	Parameter risk	Total risk	Process risk	Parameter risk	Total risk
i=2	150	142	206	150	142	206
i=3	470	410	623	470	410	623
i=4	549	507	747	549	507	747
i=5	1,227	809	1,470	1,227	809	1,469
i=6	1,824	826	2,002	1,824	825	2,002
i=7	2,042	844	2,209	2,042	844	2,209
i=8	4,947	2,058	5,359	4,947	2,057	5,358
i=9	6,035	1,925	6,334	6,035	1,921	6,333
i=10	23,464	7,325	24,581	23,464	7,276	24,566
Total:	24,920	10,193	26,924	24,920	10,153	26,909

The total risk CV of the liability for all accident years combined is 51.6%. The similarity of results is not surprising since, as we saw in Section 1, Mack's model is a special case of the FFM.[14]

Now, suppose one selected the following link ratios for the RAA data:

Table 4 – Alternative link ratio selections based on judgment

1 to 2	2 to 3	3 to 4	4 to 5	5 to 6	6 to 7	7 to 8	8 to 9	9 to 10
3.500	1.750	1.275	1.175	1.112	1.040	1.035	1.018	1.009

The FFM formulas result in a total risk overall CV of the liability for all accident years combined of 63.8%. To impute the Mack-formula based CV of 51.6% to the estimated liability from these link ratios would understate the total risk by about 19% (1-.516/.638).

6. Conclusion and Further Research

Given reasonably selected link ratios, we have shown how the Flexible Factor Model

- reproduces the point estimates of the traditional chain ladder methodology,
- determines estimates of risk consistent with those point estimates,
- offers a framework for statistically objective diagnostic and validation tools, and
- enhances the analytical reserving work flow.

[14] As mentioned above, the uncertainty estimators of the two models will agree with the exclusion of the third term (the product of the variances) from the parameter risk formulas of the FFM.

Development of similar results for other deterministic methods – such as Bornhuetter-Ferguson, Cape-Cod, and Munich Chain Ladder – seems feasible. Various bootstrapping techniques could be conducted on the FFM residuals, emphasizing the role residuals play in assuring meaningful results. A Bayesian approach could prove fruitful, where one defines a "prior" for the α_k and derives the *aposteriori* distribution for the variance assumption. However, if FFM is too simplistic to model the data appropriately, a natural next step would be to introduce an intercept term to the regression model as suggested in Murphy [2].

Appendix A

Proof of Theorem 1.1 (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 $C_{i_{\min},k} < C_{j,k}, (j \le I - k)$. 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
$$\boldsymbol{w}_{\boldsymbol{i}_{\min},\boldsymbol{k}}^{\alpha} = \boldsymbol{C}_{\boldsymbol{i}_{\min},\boldsymbol{k}}^{2-\alpha} / \sum_{j=1}^{n-k} \boldsymbol{C}_{j,k}^{2-\alpha} = \boldsymbol{C}_{\boldsymbol{i}_{\min},\boldsymbol{k}}^{2} / \sum_{j=1}^{n-k} \boldsymbol{C}_{j,k}^{2} \cdot (\boldsymbol{C}_{\boldsymbol{i}_{\min},\boldsymbol{k}} / \boldsymbol{C}_{j,k})^{\alpha} .$$
Obviously all $(C_{i_{\min},k} / C_{j,k}) < 1, j \ne i_{\min}$, thus all terms converge to 0 except for $j = i_{\min}$, so that we see $\sum_{j=1}^{n-k} \boldsymbol{C}_{j,k}^{2} \cdot (\boldsymbol{C}_{\boldsymbol{i}_{\min},\boldsymbol{k}} / \boldsymbol{C}_{j,k})^{\alpha} \to \boldsymbol{C}_{\boldsymbol{i}_{\min},\boldsymbol{k}}^{2}$ 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 $\boldsymbol{F}_{\boldsymbol{i}_{\min},\boldsymbol{k}} = \boldsymbol{F}_{\boldsymbol{i}_{\max},\boldsymbol{k}} = 2.5$. 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 shown in the accompanying graph.

Table 5: Link Ratio Example

152	380	**2.500**
185	449	2.425
217	537	2.478
250	550	2.201
262	655	**2.500**
235	466	1.985
207	411	1.989
185	372	2.011
Simple Average:	α=2	**2.261**
VW Average:	α=1	**2.258**
	α=0	**2.243**

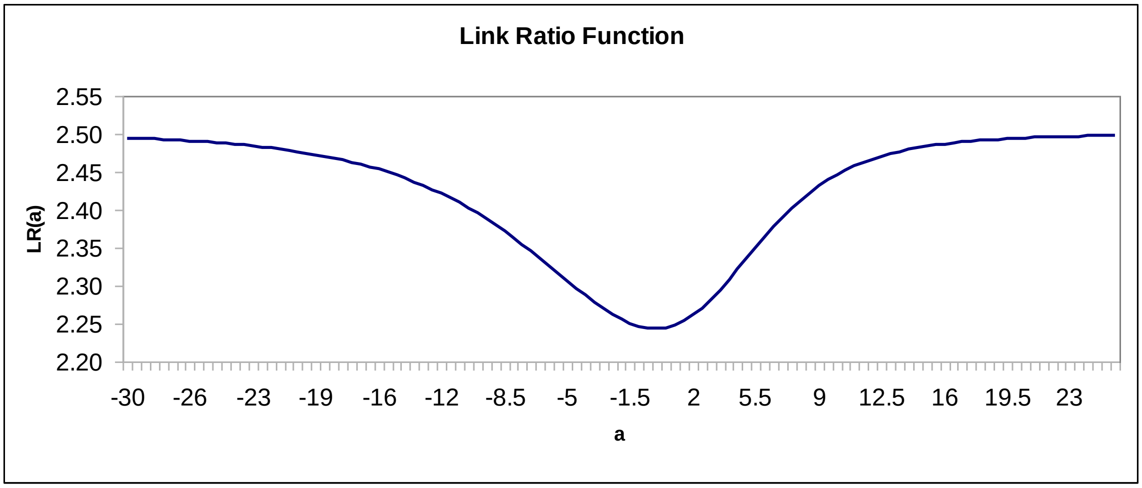

Figure 4: Reasonable link ratios derived from the Link Ratio Function

Proof of Theorem 1.2

Using Theorem 1.1 we observe that the set $\{\alpha \in \Re \mid h_k = \mathrm{LR}(\alpha)\}$ is not empty. Furthermore we note that $h_k = \mathrm{LR}(\alpha) \Leftrightarrow (h_k \cdot \sum_{j=1}^{n-k} C_{j,k}^{2-\alpha} - \sum_{i=1}^{n-k} C_{i+1,k} C_{i,k}^{1-\alpha}) = 0$, which can be solved with an appropriate numerical solver algorithm. In particular the Newton-Raphson algorithm can be easily employed. If we consider the h_k equation described above as a function of α, noted as f(α), the Newton-Raphson algorithm calculates an appropriate α that serves as a root of the equation, i.e. f(α)=0. The approximation of the root is achieved by calculating successive tangents of f(α) by generating the sequence $\{p_n\}$ defined by:

$$p_n = p_{n-1} - \frac{f(p_{n-1})}{f'(p_{n-1})} \text{ (for } n \geq 1\text{)}.$$

More than one solution can be produced by the application of the Newton-Rahpson algorithm. Consider again the example in Table 5 above where we get two solutions for the link ratio 2.400: -10.5 and 7.5, thus we set the variance estimation to max(-10.5, 7.5)=7.5.

Proof of the Parameter Risk Formulas – single accident year

For the first period after the current diagonal, $\hat{C}_{i,k+1} = \hat{f}_k C_{i,k}$, so $\Delta^2(\hat{C}_{i,k+1}) = C_{i,k}^2 \Delta^2(\hat{f}_k^2)$ since $C_{i,k}^2$ is a constant.

For s>1 periods after the current diagonal, $\hat{C}_{i,k+s} = \hat{f}_{k+s-1}\hat{C}_{i,k+s-1}$, so based on the "law of total variance":

$$\Delta^2(\hat{C}_{i,k+s}) = E(Var(\hat{C}_{i,k+s} \,|\, \hat{C}_{i,k+s-1})) + Var(E(\hat{C}_{i,k+s} \,|\, \hat{C}_{i,k+s-1}))$$

$$= E(\hat{C}_{i,k+s-1}^2 Var(\hat{f}_{k+s-1})) + Var(\hat{C}_{i,k+s-1} E(\hat{f}_{k+s-1}))$$

$$= Var(\hat{f}_{k+s-1}) E(\hat{C}_{i,k+s-1}^2) + \mathrm{Var}(\hat{\mathrm{C}}_{\mathrm{i,k+s-1}} f_{k+s-1})$$

$$= Var(\hat{f}_{k+s-1})\left(Var(\hat{C}_{i,k+s-1}) + E^2(\hat{C}_{i,k+s-1})\right) + f_{k+s-1}^2 Var(\hat{C}_{1,k+s-1})$$

$$= \mu_{i,k+s-1}^2 \Delta^2(\hat{f}_{k+s-1}) + f_{k+s-1}^2 \Delta^2(\hat{C}_{i,k+s-1}) + \Delta^2(\hat{f}_{k+s-1})\Delta^2(\hat{C}_{i,k+s-1}).$$

Proof of the Process Risk Formulas – single accident year

For the first period after the current diagonal, $\Gamma(C_{i,k+1}) = C_{i,k}^{\alpha_k}\sigma_k^2$. For *s*>1 periods after the current diagonal, process risk can be calculated recursively according to the formula:

$$\Gamma^2(C_{i,k+s}) = f_{k+s-1}^2 \cdot \Gamma^2(C_{i,k+s-1}) + E(C_{i,k+s-1}^{\alpha_{k+s-1}} \,|\, D)\sigma_{k+s-1}^2$$

Proof:

For the first period after its current age (s=1) the process risk for $C_{i,k+1}$ comes directly from assumption (1):

$$\Gamma^2(C_{i,k+1}) = C_{i,k}^{\alpha_k}\sigma_k^2 \qquad (5)$$

because $C_{i,k}^{\alpha_k}$ is a known constant.

For s>1 we again rely on the "law of total variance":

$$\Gamma^2(C_{i,k+s}) = E(Var(C_{i,k+s} \,|\, D)) + Var(E(C_{i,k+s} \,|\, D))$$

$$= E(C_{i,k+s-1}^{a_{k+s-1}}\sigma_{k+s-1}^2 \,|\, D) + Var(E(f_{k+s-1}C_{i,k+s-1}) \,|\, D)$$

$$= E(C_{i,k+s-1}^{a_{k+s-1}} \,|\, D)\sigma_{k+s-1}^2 + f_{k+s-1}^2\Gamma^2(C_{i,k+s-1})$$

As explained in the text we favor approximating $E(C_{i,k+s-1}^{\alpha_{k+s-1}} \,|\, D)$ in practice with $\left(E(C_{i,k+s-1} \,|\, D)\right)^{\alpha_{k+s-1}} \cdot \Psi$, where factor Ψ is a function of α and the coefficient of variation κ.

For estimates of Γ^2, we replace all unknown quantities by their best estimates: f_k by $\hat{f}_k$, σ_k by $\hat{\sigma}_k$, etc. Again we note here that $\hat{\sigma}_k^2$ and $\hat{f}_k^2$ both depend on $\hat{\alpha}_k$. However we drop the functional notation $\hat{\sigma}_k^2(\hat{\alpha}_k)$ and $\hat{f}_k^2(\hat{\alpha}_k)$ for convenience of presentation.

Proof of the Parameter Risk Formulas – all accident years combined

For j=3, 4,…, $\hat{X}_j = \hat{f}_{j-1} \cdot \left(\hat{X}_{j-1} + C_{I-j+2,j-1}\right)$, where I-j+2 is the only accident year that has matured as of age j-1. By employing the "law of total variance" mentioned above, we have:

$$\Delta^2(\hat{X}_j) = E(Var(\hat{X}_j \mid \hat{X}_{j-1})) + Var(E(\hat{X}_j \mid \hat{X}_{j-1}))$$
$$= E(Var(\hat{f}_{j-1}(\hat{X}_{j-1} + C_{I-j+2,j-1}) \mid \hat{X}_{j-1})) + Var(E(\hat{f}_{j-1}(\hat{X}_{j-1} + C_{I-j+2,j-1}) \mid \hat{X}_{j-1}))$$
$$= E((\hat{X}_{j-1} + C_{I-j+2,j-1})^2 Var(\hat{f}_{j-1} \mid \hat{X}_{j-1})) + Var((\hat{X}_{j-1} + C_{I-j+2,j-1}) E(\hat{f}_{j-1} \mid \hat{X}_{j-1}))$$
$$= \Delta^2(\hat{f}_{j-1}) E((\hat{X}_{j-1} + C_{I-j+2,j-1})^2) + Var(f_{j-1}(\hat{X}_{j-1} + C_{I-j+2,j-1}))$$
$$= \Delta^2(\hat{f}_{j-1}) \left\{ Var(\hat{X}_{j-1}) + E^2(\hat{X}_{j-1} + C_{i-j+2,j-1}) \right\} + f_{j-1}^2 Var(\hat{X}_{j-1})$$
$$= (M_{j-1} + C_{i-j+2,j-1})^2 \Delta^2(\hat{f}_{j-1}) + f_{j-1}^2 \Delta^2(\hat{X}_{j-1}) + \Delta^2(\hat{f}_{j-1}) \Delta^2(\hat{X}_{j-1}),$$

since $C_{I-j+2,j-1}$ is a constant.

Proof of the Process Risk Formulas – all accident years combined

The formula for process risk is straightforward since all accident years are assumed to be independent and the process variance of the sum of the losses for all accident years is the sum of the process variance of each accident year.

Process Risk Function Ψ based on simulations

Results of simulations calculating the function Ψ a s a ratio of $E(X^\alpha)$ over $(E(X))^\alpha$ for a truncated-normal random variable X are shown in Figure 5 below.

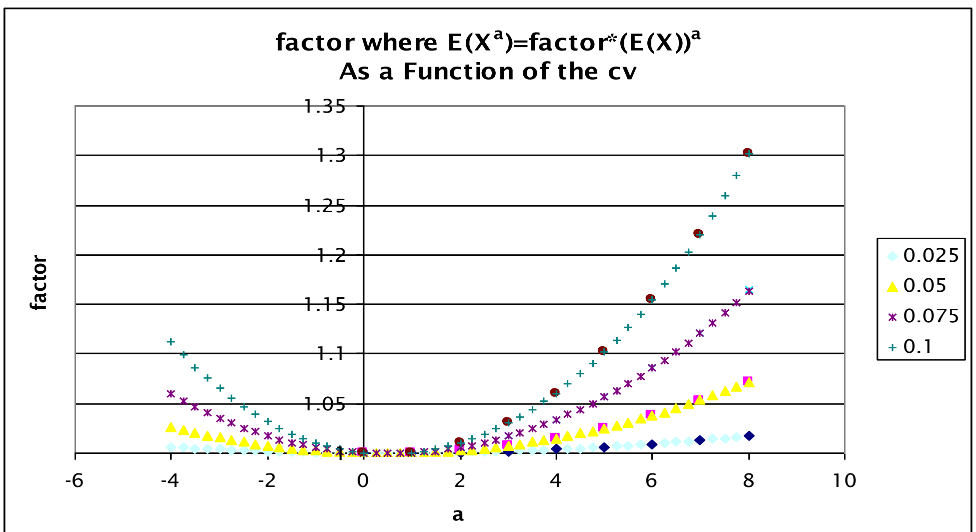

Figure 5: Function Ψ based on simulated values

Appendix B

Estimating the Ψ Function in R[15]

To approximate $\Psi(\alpha, cv) = E(X^{\alpha}) / E(X)^{\alpha}$ we simulate many normal variates **X** with coefficient of variation *cv*, and calculate $mean(X^{\alpha}) / (mean(X))^{\alpha}$ for our α of interest. Technically, X^{α} is only defined on the positive support of *X* when α is not a whole number, so we employ rejection sampling[16] to ensure that **X** consists of positive values only. The R function in Figure 6 creates a sample of size 10000 of positive-only pseudo-normal random variates.[17]

```
rnorm.positive <- function(mu,sigma) {
    sampl <- rnorm(10000,mu,sigma) # simulated normals
    negative <- sampl<=0  # flag the negative values
    while (any(negative)) { # resample those cells
        sampl[negative] <- rnorm(sum(negative),mu,sigma)
        negative <- sampl<=0
        }
    sampl               # the final sample
    }
```

Figure 6: Generate a sample of size 10000 of positive pseudo-normal random variates

Unfortunately we cannot simply set μ=1, σ=*cv* and use the resulting random sample because the sample cv may significantly differ from the input *cv,* especially for large values of the target *cv*. To illustrate, when the target *cv* is small, say 0.2,

```
> set.seed(2009) # so results can be duplicated
> X<-rnorm.positive(1,0.2)
> cv<-sd(X)/mean(X) # the sample cv
[1] 0.2004325
```

the sample `cv` is close to the input σ=*cv*, but for a larger target such as 0.8

```
> set.seed(2009)
> X<-rnorm.positive(1,0.8)
> cv<-sd(X)/mean(X)
[1] 0. 5818142
```

the sample `cv` is far from the input σ=*cv*. Therefore, for any given target *cv* we first find a (`mu,sigma`) pair such that the `rnorm.positive` function builds a sample whose sample `cv` is as close as possible to the target *cv*. To find that pair, we use R's `optim` function. For example, in the "R session" below we see that for a target *cv* of 0.8, a sample of 10000 positive pseudo-normal variates **X** generated from μ = -1.196564 and σ = 4.676847 will have a sample coefficient of variation close to that target:

---

[15] R is a statistical computing and graphics software environment widely used for academic and commercial research, and supported by a worldwide community. R is available for free at http://www.r-project.org.
[16] For an example of a similar application of rejection sampling, see http://www.biostat.wustl.edu/archives/html/s-news/2001-04/msg00033.html
[17] In R, text following the #-sign are comments.

```
> set.seed(2009)
> S<-optim(c(1,.8), # vector of mu, sigma starting values
     sample.cv.distance, # function to minimize
     gr=NULL, # no gradient function provided to optim
     0.8, # desired target cv needed by sample.cv.distance
     method="BFGS") # quasi-Newton method works well here
> S$par
[1] -1.196564  4.676847
> X.8<-rnorm.positive(-1.196564,4.676847)
> sd(X.8)/mean(X.8)
[1] 0.787364 # sample cv is close to 0.8 target
```

For those new to R, the above warrants some explanation. `optim` tries to minimize the function `sample.cv.distance` (Figure 7) by an intelligent search through all possible (μ,σ) pairs

```
sample.cv.distance <- function(musigma, targetcv) {
    mu <- musigma[1]     # mu=1st element of musigma vector
    sigma <- musigma[2]  # sigma=2nd element
    if (sigma<=0) return(100) # to avoid sigma<=0 solutions
    y <- rnorm.positive(musigma[1],musigma[2]) # the sample
    abs(sd(y)/mean(y)-targetcv) # the distance
    }
```

Figure 7: function we want `optim` to minimize
(Note: It is possible for `optim` to try a `musigma` pair containing a negative value for `sigma`; `sample.cv.distance` penalizes such out-of-bounds tries by returning a large "distance".)

starting with (1,.8). Each time `sample.cv.distance` is called, it generates an `rnorm.positive` sample – depending on the `musigma` vector that `optim` sends it – and returns the distance between the `cv` of that sample and the target *cv* (0.8 in the session above). When `optim` decides it has found the minimum possible distance, it returns the (μ,σ) solution vector in its `$par` component, referenced as `S$par` in the session above.

We now use our sample **X.8** generated in the session above to estimate Ψ(α,*cv*=0.8) for negative values of α. For example,

```
> mean(X.8^(-.5))/(mean(X.8))^(-.5)
[1] 1.309064
```

shows that Ψ(-0.5,0.8) is approximately 1.31 and

```
> mean(X.8^(-2))/(mean(X.8))^(-2)
[1] 1682.209
```

shows that Ψ(-2,0.8) is a staggering 1682. We could use **X.8** to estimate Ψ(α,0.8) for positive α too, rather than linearly interpolate between integer values per Section 2. For example, if α=1.5 – for a selected ATA between the weighted (α=1) and simple (α=2) averages – the estimate of Ψ(1.5,0.8) is

```
> mean(X.8^1.5)/(mean(X.8))^1.5
[1] 1.128022
```

which is less than the linearly interpolated value, 1.32.[18]

[18] From the formula for Ψ(*n*,*κ*) in Section 2, Ψ(1,0.8)=1 and Ψ(2,0.8)=1+(0.8)^2=1.64.

References

[1] Buchwalder, M., H. Bühlmann, M. Merz, and M.V. Wüthrich, (2006). "The Mean Square Error of Prediction in the Chain Ladder Reserving Method (Mack and Murphy Revisited)," *ASTIN Bulletin* (Preprint)*,* **2006**, Vol. 36, No. 2, pp. 521–542.

[2] Mack, Thomas, "Distribution-Free Calculation of the Standard Error of Chain Ladder Reserve Estimates," *ASTIN Bulletin*, **1993**, Vol. 23, No. 2, pp. 213-225.

[3] Mack, Thomas, "Measuring the Variability of Chain Ladder Reserve Estimates," CAS *Forum*, Spring 1994, Vol. 1, pp. 101-182.

[4] Mack, Thomas, "The Standard Error of Chain Ladder Reserve Estimates: Recursive Calculation and Inclusion of a Tail Factor," *ASTIN Bulletin*, **1999**, Vol. 29, No. 2, pp. 361-366.

[5] Murphy, Daniel, "Unbiased Loss Development Factors," *PCAS*, **1994**, Vol. LXXXI, No. 154, pp. 154-222.

[6] Royston, P., "A Pocket-Calculator Algorithm for the Shapiro-Francia Test for Normality: An Application to Medicine," *Statistics in Medicine*, **1993**, Vol. 12, No. 2, pp.181-184.

[7] Verall, Richard J., "A Bayesian Generalized Linear Model for The Bornhuetter-Ferguson Method of Claims Reserving," *North American Actuarial Journal*, **2004**, Vol. 8, No. 3, pp. 67-89.

[8] Zenwirth, Ben and Glenn Barnett, "Best Estimate Reserving," *PCAS*, **2000**, Vol. LXXXVII, No. 167, pp. 245-321.