Bootstrap Modeling: Beyond the Basics

Mark R. Shapland, FCAS, ASA, MAAA

Jessica (Weng Kah) Leong, FIAA, FCAS, MAAA

Abstract

Motivation. Bootstrapping is a very versatile model for estimating a distribution of possible outcomes for the unpaid claims, is relatively easy to use and explain to others, and can be readily "generalized" to be more flexible and combined with other related models that can be used to assess risk for a wide variety of enterprise risk management issues. While the CAS literature includes several papers that describe the bootstrap model, all of these papers are limited to the basic calculations of the model or focus on a particular aspect of the model. In contrast, this paper outlines the modifications to the basic algorithm that are required in order to put the bootstrap model into practical everyday use.

Method. This paper will start by pulling all of the issues from different papers into the complete basic bootstrap modeling framework using a standard notation. Then it will describe some of the enhancements required for practical usage and it will show how the output of the model can be easily "extended" to address other risk management issues. It will then expand the basic model and generalize the approach, as well as address many common modeling issues that arise during the diagnostic testing of the model parameters and assumptions. Finally, it will summarize testing of the model using simulated data and suggest possible areas for further research.

Results. The paper will illustrate the practical implementation of the bootstrap modeling framework as a powerful tool for estimating a distribution of unpaid claims.

Conclusions. The paper outlines the full versatility of the bootstrap model for the practicing actuary.

Availability. A set of companion Excel files are available at http://www.casact.org/pubs/forum/10fforum/, which contains the calculations illustrated in this paper as well as serving as a learning tool for the student or practicing actuary.

Keywords. Bootstrap, Over-Dispersed Poisson, Reserve Variability. Reserve Range, Distribution of Possible Outcomes.

1. INTRODUCTION

The term "bootstrap" has a colorful history that dates back to German folk tales of the 18th-century. It is aptly conveyed in the familiar cliché admonishing laggards to "pull oneself up by their own bootstraps." A physical paradox and virtual impossibility, the idea has nonetheless caught the imagination of scientists in a broad array of fields, including physics, biology and medical research, computer science, and statistics.

Bradley Efron, Chairman of the Department of Statistics at Stanford University, is most often associated as the source of expanding bootstrapping into the realm of statistics, with his notion of taking one available sample and using it to arrive at many others through resampling. His essential strategy involves duplicating the original sample and then treating the expanded sample that results from the process as a virtual population. Samples are then drawn with replacement from this population to verify the estimators.

In actuarial science, bootstrapping has become increasingly common in the process of loss reserving. The most commonly cited examples point to England and Verrall [9, 10], Pinheiro, et al. [25], and Kirschner, et al. [15], who suggest using a basic chain ladder technique to square a triangle of paid losses, repeating that randomly and stochastically over a large number of trials. The model generates a distribution of possible outcomes, rather than the chain ladder's typical point estimate, thus providing more information about the potential results. For example, without an estimated distribution it is impossible to directly estimate the amount of capital required[1] or how likely it is that the ultimate value of the claims will exceed a certain amount.

Another advantage of a bootstrap model is that it can be specifically tailored to the statistical features found in the data under analysis. This is particularly important as the results of any simulation model are only as good as the model used in the simulation process. If the model does not fit the data then the results of the simulation may not be a very good estimate of the distribution of possible outcomes. Like all models and methods, the quality of a bootstrap model depends on the quality of the assumptions. Thus, we will elaborate on the model diagnostics in Section 4.

A third advantage of a bootstrap model is that it can reflect the fact that insurance loss distributions are generally "skewed to the right." Rather than use the commonly recognized normal distribution (which is sometimes used as a simplifying assumption in other models), the bootstrap sampling process does not require a distributional assumption. Instead, the level of skewness in the underlying data is automatically reflected back into the resampled or pseudo data.

Another aspect of bootstrap models that could be considered a disadvantage is that they are more complex than other methods and thus more time consuming to create. However, once a flexible model has been developed they can be used as efficiently as most standard methods.

There are several disadvantages of bootstrap models that we will discuss in due course as we describe how this framework can be modified for a variety of practical uses.[2]

1.1 Objectives

The world of enterprise risk management is changing the horizon for actuaries. Understanding the central estimate for insurance claims is no longer adequate when managing risk. Actuaries must now measure and understand the distribution of the insurance claims in order to better understand and explain risk to management. On the pricing and dynamic risk modeling fronts, the actuarial

---

[1] Without an estimated distribution, required capital could be "estimated" using industry benchmark ratios or other rules of thumb, but these do not directly account for the specific risk profile under review.

[2] This section is based in large part on [22].

models have already embraced this new reality.

Unfortunately, in the reserving area the vast majority of actuaries are focused on deterministic point estimates for reserving. This is not surprising, as our primary standard of practice for reserving, ASOP 36, seems to be focused exclusively on deterministic point estimates and the regulators, via the actuarial opinion, are also focused on deterministic estimates. However, actuaries are free to estimate distributions instead of point estimates.[3] But nothing seems to be forcing the profession towards unpaid claim distributions.

This is changing due to a number of factors:

- the SEC is looking for more reserving risk information in the 10-K reports filed by publicly traded companies;
- all of the major rating agencies have built or are building dynamic risk models to help with their insurance rating process and welcome the input of company actuaries regarding unpaid claim distributions; and
- companies that use dynamic risk models to help their internal risk management processes need unpaid claim distributions.

One objective of this paper is to show how the bootstrap modeling framework can be used in practice, to help the wider adoption of unpaid claim distributions.

Another potential roadblock seems to be the notion that actuaries are still searching for the perfect model to describe "the" distribution of unpaid claims, as if imperfections in a model remove it from all consideration since it can't be "the one." This notion can also manifest itself when an actuary settles for a model that seems to work the best or is the easiest to use, or with the idea each model must be used in its entirety or not at all. Interestingly, this notion was dispelled long ago with respect to practice for deterministic point estimates as actuaries commonly use many different methods, which range from easy to complex, and judgmentally weight the results by accident year (i.e., use only parts of a method) to arrive at their best estimate. Thus, another objective of this paper is to show how stochastic reserving needs to be similar to deterministic reserving when it comes to analyzing and using the best parts of multiple models.

Finally, most of the papers describing stochastic models, including bootstrap models, tend to focus primarily on the theoretical aspects of the model while ignoring the data issues that commonly

---

3 Indeed, ASOP 43 opened the door a bit further by defining "actuarial central estimate" in such a way that it could include either deterministic point estimates or a first moment estimate from a distribution.

arise in practice. As a result, most models described in papers can be quite elegantly implemented yet can suffer from practical limitations such as only being useful for complete triangles or only for positive incremental values. This could also act as a deterrent by limiting the usability of a model to specific situations and by giving the impression that using the model is not worth the effort. Thus, while keeping as close to the theoretical foundation as possible, another objective of the paper is to illustrate how a variety of practical adjustments can be made to accommodate common data issues.

1.2 Outline

This paper will start by reviewing the notation from the CAS Working Party on Quantifying Variability in Reserve Estimates Summary Report [6] which we will use in this paper. Then we will illustrate and expand the foundation developed in other papers for the basic calculations of the bootstrap model, including showing how the GLM framework of the model can be "generalized" to include diagonal parameters. In order to be consistent with the theoretical foundation yet recognize practical needs, we will describe data issues that require enhancements to the basic algorithm. With a complete modeling framework established, we can then review the diagnostic tests to ensure that the model assumptions are consistent with the statistical features in the data. Should the assumptions appear inconsistent, we will suggest adjustments to the model that can be made.

Even though bootstrapping is a very versatile framework, it is still important to draw from the strengths of different models and weight distributions, similar to weighting point estimates, in order to get a best estimate of the distribution. Thus, we will briefly explore ways to combine the results of different models, including non-bootstrap models with bootstrap models. Since the analysis of enterprise risk involves all sources of risk, we will also explore correlation issues for the bootstrap model and then describe extensions to the model output and how they can be used for assessing risks in addition to reserve risk. In order to use the results with confidence, we will briefly discuss some findings related to testing of the model compared to another commonly used model (Mack). Finally, we will close with some possible areas for future research.

2. NOTATION

The papers that describe the basic bootstrap model use different notation, despite sharing common steps. Rather than pick the notation in one of the papers, we will use the notation from the CAS Working Party on Quantifying Variability in Reserve Estimates Summary Report [6] since it is intended to serve as a basis for further research and the bootstrap model is also described in that paper.

Many models visualize loss statistics as a two-dimensional array. The row dimension is the annual period by which the loss information is subtotaled, most commonly an accident year or policy year. For each accident period, w, the (w,d) element of the array is the total of the loss information as of development age d.[4] Here the development age is the accounting year[5] of the loss information expressed as the number of time periods after the accident or policy year. For example, the loss statistic for accident year 2 as of the end of calendar year 4 has development age 3 years.

For this discussion, we assume that the loss information available is an "upper triangular" subset of the two-dimensional array for rows $w = 1,2,\ldots,n$. For each row, w, the information is available for development ages 1 through $n-w+1$. If we think of year n as the most recent accounting year for which loss information is available, the triangle represents the loss information as of accounting dates 1 through n. The diagonal $k = w+d$ represents the loss information for each accident period w as of accounting year k.[6]

The paper uses the following notation for certain important loss statistics:

$c(w,d)$: cumulative loss from accident[7] year w as of age d. (Think w = "when" and d = "delay")

$c(w,n) = U(w)$: total loss from accident year w when claims are at ultimate values.

$R(w,d)$: future development after age d for accident year w, i.e., = $U(w) - c(w,d)$.

$q(w,d)$: incremental loss for accident year w from d - 1 to d.

$f(d)$: factor applied to $c(w,d)$ to estimate $q(w,d+1)$ or can be used more generally to indicate any factor relating to age d.

$F(d)$: factor applied to $c(w,d)$ to estimate $c(w,n)$ or can be used more generally to indicate any cumulative factor relating to age d.

$G(w)$: factor relating to accident year w – capitalized to designate ultimate loss level.

$h(w+d)$: factor relating to the diagonal k along which $w+d$ is constant.

---

4 Depending on the context, the (w,d) cell can represent the cumulative loss statistic as of development age d or the incremental amount occurring during the d th development period.

5 The development ages are assumed to be in yearly intervals for ease of discussion. However, they can be in different time units such as half-years, quarters, or months.

6 For a more complete explanation of this two-dimensional view of the loss information, see the *Foundations of Casualty Actuarial Science* [12], Chapter 5, particularly pages 210-226.

7 The use of accident year is also used for ease of discussion. All of the discussion could also apply to underwriting year, policy year, report year, etc.

$e(w,d)$: a mean zero random fluctuation which occurs at the w,d cell.

$E(x)$: the expectation of the random variable x.

$Var(x)$: the variance of the random variable x.

What are called factors here could also be summands, but if factors and summands are both used, some other notation for the additive terms would be needed. The notation does not distinguish paid vs. incurred, but if this is necessary, capitalized subscripts P and I could be used.

3. THE BOOTSTRAP MODEL

Even though many variations of the bootstrap model framework are possible, we will focus primarily on the most common example that essentially reproduces the basic chain ladder method. It will also be helpful to briefly review the assumptions that underpin the basic chain ladder method.

The foundation for any model is the data being modeled. Like many commonly used models, then, we will start with a triangle array of cumulative data:

		d					
		1	**2**	**3**	**…**	**n-1**	**n**
w	**1**	c(1,1)	c(1,2)	c(1,3)	…	c(1,n-1)	c(1,n)
	2	c(2,1)	c(2,2)	c(2,3)	…	c(2,n-1)	
	3	c(3,1)	c(3,2)	c(3,3)	…		
	…	…	…				
	n-1	c(n-1,1)	c(n-1,2)				
	n	c(n,1)					

A typical deterministic analysis of this data will start with an array of age-to-age ratios or development factors:

$$F(w,d)=\frac{c(w,d)}{c(w,d-1)}. \tag{3.1}$$

Then two key assumptions are made in order to make a projection of the known elements to their respective ultimate values. First, it is assumed that each accident year has the same development factor. Equivalently, for each $w=1,2,\ldots,n$:

$$F(w,d)=F(d).$$

Under this first assumption, one of the more popular estimators for the development factor is the weighted average:

$$\hat{F}(d)=\frac{\sum_{w=1}^{n-d+1}c(w,d)}{\sum_{w=1}^{n-d+1}c(w,d-1)}. \tag{3.2}$$

Certainly there are other popular estimators in use, but they are beyond our scope at this stage yet most are still consistent with our first assumption that each accident year has the same factor. Projections of the ultimate values, or $\hat{c}(w,n)$ for *w* = 1, 2, 3, … , *n*, are then computed using:

$$\hat{c}(w,n)=c(w,d)\prod_{i=d+1}^{n}\hat{F}(i). \tag{3.3}$$

This part of the claim projection algorithm relies explicitly on the second assumption, namely that each accident year has a parameter representing its relative level. These level parameters are the current cumulative values for each accident year, or $c(w,n-w+1)$. Of course variations on this second assumption are also common, but the point is that every model has explicit assumptions that are an integral part of understanding the quality of that model.

One variation on the second assumption is to assume that the accident years are completely homogeneous. In this case we would estimate the level parameter of the accident years using:

$$\frac{\sum_{w=1}^{n-d+1} c(w,d)}{n-d+1}. \tag{3.4}$$

Complete homogeneity implies that the observations $c(1,d)$, $c(2,d)$, …, $c(n-d+1,d)$ are generated by the same mechanism. Interestingly, the basic chain ladder algorithm explicitly assumes that the mechanisms generating the observations are NOT homogeneous and effectively that "pooling" of the data does not provide any increased efficiency.[8] In contrast, it could be argued that the Bornhuetter-Ferguson and Cape Cod methods are a "blend" of these two extremes as the homogeneity of the future expected result depends on the consistency of the a priori loss ratios and decay rate, respectively.

3.1 Origins of Bootstrapping

Possibly the earliest development of a stochastic model for the actuarial array of cumulative development data is attributed to Kremer [16]. The basic model described by Kremer can be defined by the multiplicative representation:

$$P(w,d) = G'(w) \times F'(d) \times e'(w,d). \tag{3.5}$$

Where: $G'(w)$ is a parameter representing the effect of accident year *w*,
$F'(d)$ is a parameter representing the effect of development period *d*, and
$e'(w,d)$ is a random error term.

Taking logarithms of both sides of equation (3.5), the model can be formulated as a two-way analysis of variance:

$$Y(w,d) = \log[P(w,d)] = \mu + G(w) + F(d) + e(w,d). \tag{3.6}$$

Where: μ is the overall mean effect on a log scale,
$G(w)$ is the residual effect due to accident year *w*,
$F(d)$ is the residual effect due to development period *d*,
$e(w,d)$ represent zero mean uncorrelated errors with $Var[e(w,d)] = \sigma^2$, and

$$\sum G(w) = \sum F(d) = 0. \tag{3.7}$$

This model is further described by England and Verrall [9] and Zehnwirth [39], so we will not elaborate further here. It should be noted, however, that the model in (3.6) can be extended by considering alternatives. This log-normal model, and generalizations thereof, has also been discussed in Zehnwirth [1, 40], Renshaw [30], Christofides [7], and Verrall [37, 38], among others.

---

8 For a more complete discussion of these assumptions of the basic chain ladder model see Zehnwirth [39].

3.2 The Over-Dispersed Poisson Model

The genesis of this model into a bootstrap framework originated with Renshaw and Verrall [31] when they proposed modeling the incremental claims $q(w,d)$ directly as the response, with the same linear predictor as Kremer [16], but using a generalized linear model (GLM) with a log-link function and an over-dispersed Poisson (ODP) error distribution. Then, England and Verrall [9] discuss how this model can be used to estimate parameters and use bootstrapping (sampling the residuals with replacement) to estimate the complete distribution. More formally, using the following:

$$E[q(w,d)] = m_{w,d} \text{ and } Var[q(w,d)] = \phi E[q(w,d)] = \phi m_{w,d}^{z} \tag{3.8}$$

$$\ln[m_{w,d}] = \eta_{w,d} \tag{3.9}$$

$$\eta_{w,d} = c + \alpha_w + \beta_d \text{, where: } w = 1, 2, \ldots, n; \ d = 1, 2, \ldots, n; \text{ and } \alpha_1 = \beta_1 = 0. \tag{3.10}$$

In this case the α parameters function as adjustments to the constant, *c*, level parameter and the β parameters adjust for the development trends after the first development period. The power, z, is used to specify the error distribution with $z = 0$ for normal, $z = 1$ for Poisson, $z = 2$ for Gamma and $z = 3$ for inverse Gaussian. Alternatively, we can remove the constant which will cause the α parameters to function as individual level parameters while the β parameters continue to adjust for the development trends after the first development period:

$$\eta_{w,d} = \alpha_w + \beta_d \text{, where: } w = 1, 2, \ldots, n; \text{ and } d = 2, \ldots, n. \tag{3.11}$$

Standard statistical software can be used to estimate parameters and goodness of fit measures. The parameter ϕ is a scale parameter that is estimated as part of the fitting procedure while setting the variance proportional to the mean (thus "over-dispersed" Poisson for $z = 1$). For educational purposes, we have included the calculations to solve these equations for a 10 x 10 triangle in the "Bootstrap Models.xls" file, but we will illustrate the calculations here for a 3 x 3 triangle for ease of exposition and in the "Simple GLM.xls" file. Consider the following incremental data triangle:

	1	**2**	**3**
1	$q(1,1)$	$q(1,2)$	$q(1,3)$
2	$q(2,1)$	$q(2,2)$	
3	$q(3,1)$		

In order to set up the GLM model to fit parameters to the data we need to do a log-link or transform which results in:

	1	**2**	**3**
1	$\ln[q(1,1)]$	$\ln[q(1,2)]$	$\ln[q(1,3)]$
2	$\ln[q(2,1)]$	$\ln[q(2,2)]$	
3	$\ln[q(3,1)]$		

The model is then specified using a system of equations with vectors of α_w and β_d parameters as follows:

$$\begin{aligned}
\ln[q(1,1)] &= 1\alpha_1 + 0\alpha_2 + 0\alpha_3 + 0\beta_2 + 0\beta_3 \\
\ln[q(2,1)] &= 0\alpha_1 + 1\alpha_2 + 0\alpha_3 + 0\beta_2 + 0\beta_3 \\
\ln[q(3,1)] &= 0\alpha_1 + 0\alpha_2 + 1\alpha_3 + 0\beta_2 + 0\beta_3 \\
\ln[q(1,2)] &= 1\alpha_1 + 0\alpha_2 + 0\alpha_3 + 1\beta_2 + 0\beta_3 \\
\ln[q(2,2)] &= 0\alpha_1 + 1\alpha_2 + 0\alpha_3 + 1\beta_2 + 0\beta_3 \\
\ln[q(1,3)] &= 1\alpha_1 + 0\alpha_2 + 0\alpha_3 + 1\beta_2 + 1\beta_3 .
\end{aligned} \tag{3.12}$$

Converting this to matrix notation we have:

$$\mathrm{Y} = \mathrm{X} \times \mathrm{A} \tag{3.13}$$

Where:

$$\mathrm{Y} = \begin{bmatrix} \ln[q(1,1)] & 0 & 0 & 0 & 0 & 0 \\ 0 & \ln[q(2,1)] & 0 & 0 & 0 & 0 \\ 0 & 0 & \ln[q(3,1)] & 0 & 0 & 0 \\ 0 & 0 & 0 & \ln[q(1,2)] & 0 & 0 \\ 0 & 0 & 0 & 0 & \ln[q(2,2)] & 0 \\ 0 & 0 & 0 & 0 & 0 & \ln[q(1,3)] \end{bmatrix}, \tag{3.14}$$

$$\mathrm{X} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 1 \end{bmatrix}, \text{ and} \tag{3.15}$$

$$\mathrm{A} = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \beta_2 \\ \beta_3 \end{bmatrix}. \tag{3.16}$$

In this form we can use the Newton-Raphson method[9] to solve for the parameters in the A vector that minimize the difference between the Y matrix and the W matrix:

[9] Other methods, such as orthogonal decomposition, can also be used to solve for the parameters.

$$W = \begin{bmatrix} \ln[m_{1,1}] & 0 & 0 & 0 & 0 & 0 \\ 0 & \ln[m_{2,1}] & 0 & 0 & 0 & 0 \\ 0 & 0 & \ln[m_{3,1}] & 0 & 0 & 0 \\ 0 & 0 & 0 & \ln[m_{1,2}] & 0 & 0 \\ 0 & 0 & 0 & 0 & \ln[m_{2,2}] & 0 \\ 0 & 0 & 0 & 0 & 0 & \ln[m_{1,3}] \end{bmatrix}. \tag{3.17}$$

Typically, X is known as the design matrix and W is known as the weight matrix. After solving the system of equations we will have:

$$\begin{aligned} \ln[m_{1,1}] &= \eta_{1,1} = \alpha_1 \\ \ln[m_{2,1}] &= \eta_{2,1} = \alpha_2 \\ \ln[m_{3,1}] &= \eta_{3,1} = \alpha_3 \\ \ln[m_{1,2}] &= \eta_{1,2} = \alpha_1 + \beta_2 \\ \ln[m_{2,2}] &= \eta_{2,2} = \alpha_2 + \beta_2 \\ \ln[m_{1,3}] &= \eta_{1,3} = \alpha_1 + \beta_2 + \beta_3 . \end{aligned} \tag{3.18}$$

This solution can then be shown as a triangle:

	1	**2**	**3**
1	$\ln[m_{1,1}]$	$\ln[m_{1,2}]$	$\ln[m_{1,3}]$
2	$\ln[m_{2,1}]$	$\ln[m_{2,2}]$	
3	$\ln[m_{3,1}]$		

These results can then be exponentiated to the fitted, or expected, incremental results of the GLM model:

	1	**2**	**3**
1	$m_{1,1}$	$m_{1,2}$	$m_{1,3}$
2	$m_{2,1}$	$m_{2,2}$	
3	$m_{3,1}$		

We will refer to this as the "GLM framework" and have illustrated this model for a simple 3 x 3 triangle in the "Simple GLM.xls" file. While the GLM framework is used to solve these equations for the fitted results, the usefulness of this framework is that the fitted results (with the Poisson error distribution assumption) will exactly equal the results that can be derived from volume-weighted average age-to-age ratios. That is, it can be reproduced by using the last cumulative diagonal, dividing backwards successively by each age-to-age factor and subtracting to get the fitted incremental results. We will refer to this method as the "simplified GLM". This has three very useful consequences.

First, GLM portion of the algorithm can be replaced with a simpler link ratio algorithm while still being based on the underlying GLM framework. Second, the use of the age-to-age ratios serves as a "bridge" to the deterministic framework and allows the model to be more easily explainable to others. And, third, for the GLM algorithm the log-link process means that negative incremental values can often cause the algorithm to not have a solution, whereas using the link ratios will generally allow for a solution.[10]

With a model fitted to the data, the bootstrap process involves sampling with replacement from the residuals. England and Verrall [9] note that the deviance, Pearson, and Anscombe residuals could all be considered for this process, but the Pearson residuals are the most desirable since they are calculated consistently with the scale parameter. The unscaled Pearson residuals and scale parameter are calculated as follows:

$$r_{w,d} = \frac{q(w,d) - m_{w,d}}{\sqrt{m_{w,d}^{z}}} . \tag{3.19}$$

$$\phi = \frac{\sum r_{w,d}}{n - p} . \tag{3.20}$$

Where n = the number of data cells in the triangle and p = the number of parameters, which is typically equal to $2*n - 1$.[11] Sampling with replacement from the residuals can then be used to create new sample triangles of incremental values using formula 3.16. Sampling with replacement assumes that the residuals are independent and identically distributed, but it does not require the residuals to be normally distributed. Indeed, this is often cited as an advantage of the ODP bootstrap model since whatever distributional form the residuals have will flow through the simulation process. Some authors have referred to this a "semi-parametric" bootstrap model since we are not parameterizing the residuals.

$$q'(w,d) = r^{*} \times \sqrt{m_{w,d}^{z}} + m_{w,d} . \tag{3.21}$$

The sample triangle of incremental values can then be cumulated, new average age-to-age factors and loss development factors can be calculated for the sample and applied to calculate a point estimate for this data. This process could be described as getting a distribution of point estimates, which includes incorporating process variance and parameter variance in the simulation of the

---

[10] More specifically, individual negative cell values may not be a problem. If the total of all incremental cell values in a development column is negative, then the GLM algorithm will fail. This situation will not cause a problem fitting the model as a link ratio less than one will be perfectly useful. However, this may still cause other problems, which we will address in section 4.

[11] The number of parameters could be less than $2*n - 1$. For example, if the incremental values are zeros for the last three columns in a triangle then there will be three fewer β parameters since none are needed to fit to these zero values as the development process is completed already.

historical data. In England and Verrall [9] this is the end of the process, but at the end of the appendix they note that you should also multiply the resulting distribution by the degrees of freedom adjustment factor (3.22), to effectively allow for over-dispersion of the residuals in the sampling process.

$$f = \sqrt{\frac{n}{n-p}} \,. \tag{3.22}$$

Later, in England and Verrall [10], the authors note that the Pearson residuals (3.19) could be multiplied by the degrees of freedom adjustment factor (3.22) in order to correct for a bias in the residuals. They also expand the simulation process by adding process variance to the future incremental values from the point estimates. To add this process variance, they assume that each future incremental value $m_{w,d}$ is the mean and the mean times the scale parameter, $\phi m_{w,d}$, is the variance of a gamma distribution.[12] This revised model could now be described as estimating a distribution of possible outcomes, which incorporates process variance and parameter variance in the simulation of the historical and future data.

However, Pinheiro et al. [25, 26] noted that the bias correction for the residuals using the degrees of freedom adjustment factor (3.22) does not create standardized residuals, which is an important step for making sure that the residuals all have the same variance. In order to have standardized Pearson residuals, the GLM framework requires the use of a hat matrix adjustment factor.

$$H = X\left(X^T W X\right)^{-1} X^T W \,. \tag{3.23}$$

$$f_{w,d}^{H} = \sqrt{\frac{1}{1-H_{i,i}}} \,. \tag{3.24}$$

The hat matrix (3.23) is calculated using matrix multiplication of the design matrix (3.15) and the weight matrix (3.17). The hat matrix adjustment factor (3.24) uses the diagonal of the hat matrix. In Pinheiro, et al. [26] the authors note two important points about the bootstrap process as described by England and Verrall [9, 10]. First, the sampling of the residuals should not include any zero-value residuals, which are typically in the corners of the triangle.[13] The exclusion of the zero-value residuals is accounted for in the hat matrix adjustment factor (3.24), but another common explanation is that the zero-value cells will have some variance but we just don't know what it is yet so we should sample from the remaining residuals but not the zeros. Second, the hat matrix

[12] The Poisson distribution could be used, but it is considerably slower to simulate, so gamma is a close substitute that performs much faster in simulation.

[13] Technically, the two "corner" residuals are zero because they each have a parameter that is unique to that incremental value which causes the fitted incremental value to exactly equal the actual incremental value.

adjustment factor (3.24) is a replacement for the degrees of freedom factor (3.22), which improves the calculation of the residuals.[14]

Thus, the unscaled Pearson residuals (3.19) should be replaced by the standardized Pearson residuals:

$$r_{w,d}^{H} = \frac{q(w,d) - m_{w,d}}{\sqrt{m_{w,d}^{z}}} \times f_{w,d}^{H}. \tag{3.25}$$

However, the scale parameter (3.20) is still calculated as before, although the standardized Pearson residuals could be used to approximate the scale parameter as follows:

$$\phi^{H} = \frac{\sum r_{w,d}^{H}}{n}. \tag{3.26}$$

At this point we have a complete basic "ODP bootstrap" model, as it is often referred to, although various stages of this complete model have been in popular use and formally tested. It is also important to note that the two key assumptions mentioned earlier, each accident year has the same development factor and each accident year has a parameter representing its relative level, are equally applicable to this model.

In order for the reader to test out the different "combinations" of this modeling process the "Bootstrap Models.xls" file includes options to allow these historical algorithms to be simulated. Our purpose in describing this evolution of the bootstrap model framework is threefold: first, to allow the interested reader to better understand the details of the algorithm and how these papers have contributed to the model framework; second, to illustrate the value of collaborative research via different published papers and the contributions of different authors; and, third, to provide a solid basis for us to continue the evolutionary process.

3.3 Variations on the ODP Model

When estimating insurance risk it is generally considered desirable to focus on the claim payment stream in order to measure the variability of the actual cash flows that directly affect the bottom line. Clearly, changes in case reserves and IBNR reserves will also impact the bottom line, but to a considerable extent the changes in IBNR are intended to counter the impact of the changes in case reserves. To some degree, then, the total reserve movements can act to mask the underlying changes due to cash flows. On the other hand, the case reserves represent potential future payments so we

---

[14] This second point was not addressed clearly in Pinheiro et al. [25], but as the authors updated and clarified the paper in [26] this issue was more clearly addressed.

should not just ignore them and focus exclusively on paid data.

3.3.1 Bootstrapping the incurred loss triangle

The ODP model, as described, can be used to model both paid and incurred data. However, the resulting distribution from using incurred data will be possible outcomes of the IBNR so they will not be directly comparable to the distribution of possible outcomes of the total unpaid (i.e., from using paid data). A convenient way of converting the results of an incurred data model to a payment stream is to use payment patterns applied to the ultimate value of the incurred claims. This is consistent with how a deterministic incurred ultimate can be converted using a paid development pattern. If a paid data model is run in parallel with the incurred data model the possible outcomes from the paid data model can be used to convert incurred ultimate values to a payment pattern for each iteration (and for each accident year individually).

The "Bootstrap Models.xls" file illustrates this concept. It is worth noting, however, that this process allows the "added value" of using the case reserves to help predict the ultimate results to work its way into the calculations, thus perhaps improving the estimates, while still focusing on the payment stream for measuring risk. In effect, it allows a distribution of IBNR to become a distribution of IBNR and case reserves. This process could be made more sophisticated by correlating the residual sampling and/or process variance portions of the parallel models. Correlations must be considered if, for example, you wanted the iterations showing long payment streams to be compared with the iterations with high incurred results. It is also possible to use other modeling algorithms such as the Munich chain ladder (see [27]), although that is beyond the scope of this paper.

3.3.2 Bootstrapping the Bornhuetter-Ferguson and Cape Cod models

Another common issue with using the ODP bootstrap process is that iterations for the latest few accident years can produce results with more variance than you would expect given what you simulated for the earlier accident years. This is usually due to the fact that age-to-age factors are used to extrapolate the sampled values prior to adding process variance, which is completely analogous to one of the weaknesses of the deterministic paid chain ladder method.

As for the deterministic chain ladder method, the ODP bootstrap process can be modified by changing the extrapolation of future incremental values by using the Bornhuetter-Ferguson or generalized Cape Cod algorithms, among others. These deterministic methods can be converted into stochastic models while still using the underlying ODP assumptions and process, and that the deterministic assumptions of these methods can also be converted to stochastic assumptions. For example, instead of simply using a vector of deterministic a priori loss ratios for the Bornhuetter-Ferguson model, we could add a vector of standard deviations to go with these means, assume a distribution and simulate a different a priori loss ratio for every iteration of the model. Finally, it is worth noting that these "new" models can be set up separately for paid and incurred data and that the paid and incurred assumptions should be internally consistent with each other and with other models, as they should be for deterministic methods.

The "Bootstrap Models.xls" file also illustrates the Bornhuetter-Ferguson and Cape Cod models.

3.4 Generalizing the ODP Model

Using deterministic algorithms to enhance the flexibility of the basic ODP bootstrap process is a straightforward way to create additional models and to overcome many of the limitations of using bootstrapping. However, some limitations are more difficult to overcome just by using these algorithms. For example, calendar-year effects can be adjusted using a Berquist-Sherman algorithm but it is hard to make the assumptions more stochastic.

Rather than add essentially deterministic algorithms to a stochastic model, another approach is to go back to the original GLM framework and generalize the basic model. Returning to formulas (3.8) to (3.11), the GLM framework does not require a certain number of parameters so we are actually free to specify only as many parameters as we need to get a robust model. Indeed, it is ONLY when we specify a parameter for EVERY accident year and EVERY development year and specify a Poisson error distribution that we end up exactly replicating the volume weighted average age-to-age factors that allow us to substitute the deterministic algorithm instead of solving the GLM fit.

Thus, using the original GLM framework we can specify a model with only a few parameters, but there are two drawbacks to doing so. First, the GLM must be solved for each iteration of the bootstrap model (which may slow down the simulation process) and, second, the model is no longer directly explainable to others using age-to-age factors.[15] While the impact of these drawbacks should be considered, the potential benefits of using the GLM framework can be much greater.

First, having fewer parameters will help avoid the potential of over-parameterizing the model.[16] For example, if we use only one accident year parameter then the model specified using a system of equations is as follows (which is analogous to formula 3.12):

---

[15] However, age-to-age factors could be calculated for the fitted data to compare to the actual age-to-age factors and used as an aid in explaining the model to others.

[16] Over-parameterization is a common criticism of the ODP bootstrap model. This will be addressed more completely in Section 5.

$$\begin{aligned}
\ln[q(1,1)] &= 1\alpha_1 + 0\beta_2 + 0\beta_3 \\
\ln[q(2,1)] &= 1\alpha_1 + 0\beta_2 + 0\beta_3 \\
\ln[q(3,1)] &= 1\alpha_1 + 0\beta_2 + 0\beta_3 \\
\ln[q(1,2)] &= 1\alpha_1 + 1\beta_2 + 0\beta_3 \\
\ln[q(2,2)] &= 1\alpha_1 + 1\beta_2 + 0\beta_3 \\
\ln[q(1,3)] &= 1\alpha_1 + 1\beta_2 + 1\beta_3
\end{aligned} \tag{3.27}$$

In this case we will only have one level parameter and *n-1* development trend parameters, but it will only be coincidence that we would end up with the equivalent of average age-to-age factors. Interestingly, this model parameterization moves us away from one of the common basic assumptions (i.e., each accident year has its own level) and substitutes the assumption that all accident years are homogeneous.

Another example of using fewer parameters would be to only use one development year parameter (while continuing to use an accident-year parameter for each year), which would equate to the following system of equations:

$$\begin{aligned}
\ln[q(1,1)] &= 1\alpha_1 + 0\alpha_2 + 0\alpha_3 + 0\beta_2 \\
\ln[q(2,1)] &= 0\alpha_1 + 1\alpha_2 + 0\alpha_3 + 0\beta_2 \\
\ln[q(3,1)] &= 0\alpha_1 + 0\alpha_2 + 1\alpha_3 + 0\beta_2 \\
\ln[q(1,2)] &= 1\alpha_1 + 0\alpha_2 + 0\alpha_3 + 1\beta_2 \\
\ln[q(2,2)] &= 0\alpha_1 + 1\alpha_2 + 0\alpha_3 + 1\beta_2 \\
\ln[q(1,3)] &= 1\alpha_1 + 0\alpha_2 + 0\alpha_3 + 2\beta_2
\end{aligned} \tag{3.28}$$

In this example the model parameterization would continue to follow the two common assumptions (i.e., each accident year has its own level and uses the same development factor), although again it would be pure coincidence to end up with the equivalent of average age-to-age factors.[17] It is also interesting to note that for both of these two examples there will be one additional non-zero residual that can be used in the simulations because in each case one of the incremental values no longer has a unique parameter – i.e., for (3.27) $q(3,1)$ is no longer uniquely defined by α_3, and for (3.28) $q(1,3)$ is no longer uniquely defined by β_3.

This flexibility allows the modeler to use enough parameters to capture the statistically relevant level and trend changes in the data without forcing a specific number of parameters.[18]

The second benefit, and depending on the data perhaps the most significant, is that this

[17] If we were to generalize the development factor assumption to focus on the number of parameters instead, then we would have only one parameter instead of a different parameter for each development period.
[18] How to determine which parameters are statistically relevant will be discussed in Section 5.

framework allows us the ability to add parameters for calendar-year trends. Adding diagonal parameters to (3.11) we now have:

$$\eta_{w,d} = \alpha_w + \beta_d + \gamma_k \text{, where: } w = 1, 2, \ldots, n;\ d = 2, \ldots, n; \text{ and } k = 2, \ldots, n. \tag{3.29}$$

A complete system of equations for the (3.29) framework would look like the following:

$$\begin{aligned}
\ln[q(1,1)] &= 1\alpha_1 + 0\alpha_2 + 0\alpha_3 + 0\beta_2 + 0\beta_3 + 0\gamma_2 + 0\gamma_3 \\
\ln[q(2,1)] &= 0\alpha_1 + 1\alpha_2 + 0\alpha_3 + 0\beta_2 + 0\beta_3 + 1\gamma_2 + 0\gamma_3 \\
\ln[q(3,1)] &= 0\alpha_1 + 0\alpha_2 + 1\alpha_3 + 0\beta_2 + 0\beta_3 + 1\gamma_2 + 1\gamma_3 \\
\ln[q(1,2)] &= 1\alpha_1 + 0\alpha_2 + 0\alpha_3 + 1\beta_2 + 0\beta_3 + 1\gamma_2 + 0\gamma_3 \\
\ln[q(2,2)] &= 0\alpha_1 + 1\alpha_2 + 0\alpha_3 + 1\beta_2 + 0\beta_3 + 1\gamma_2 + 1\gamma_3 \\
\ln[q(1,3)] &= 1\alpha_1 + 0\alpha_2 + 0\alpha_3 + 1\beta_2 + 1\beta_3 + 1\gamma_2 + 1\gamma_3
\end{aligned} \tag{3.30}$$

However, there is no unique solution for a system with seven parameters and six equations, so some of these parameters will need to be removed. A logical starting point would be to start with a model with one accident year (level) parameter, one development trend parameter and one calendar trend parameter and then add or remove parameters as needed. The system of equations for this basic model is as follows:

$$\begin{aligned}
\ln[q(1,1)] &= 1\alpha_1 + 0\beta_2 + 0\gamma_2 \\
\ln[q(2,1)] &= 1\alpha_1 + 0\beta_2 + 1\gamma_2 \\
\ln[q(3,1)] &= 1\alpha_1 + 0\beta_2 + 2\gamma_2 \\
\ln[q(1,2)] &= 1\alpha_1 + 1\beta_2 + 1\gamma_2 \\
\ln[q(2,2)] &= 1\alpha_1 + 1\beta_2 + 2\gamma_2 \\
\ln[q(1,3)] &= 1\alpha_1 + 2\beta_2 + 2\gamma_2
\end{aligned} \tag{3.31}$$

A fourth benefit of the GLM framework is that it can be used to model data shapes other than triangles. For example, missing incremental data for the first few diagonals would mean that the cumulative values could not be calculated and the remaining values in those first few rows would not be useful for the simplified GLM. However, since the GLM framework uses the incremental values the entire trapezoid can be used to fit the model parameters.[19]

It should also be noted that the GLM framework allows the future expected values to be directly estimated from the parameters of model for each sample triangle in the bootstrap simulation process. However, we must solve the GLM within each iteration for the same parameters as we originally set up for the model rather than using age-to-age factors to project future expected values.

The additional modeling power that the flexible GLM framework adds to the actuary's toolkit

[19] We will examine this issue in more detail in Section 4.

cannot be overemphasized. Not only does it allow one to move away from the two basic assumptions of a deterministic chain ladder method, it allows for the ability to match the model parameters to the statistical features you find in the data and to extrapolate those features. For example, modeling with fewer development trend parameters means that the last parameter can be assumed to continue past the end of the triangle which will give the modeler a "tail" of the incremental values beyond the end of the triangle without the need for a specific tail factor.

While we have continued to illustrate the GLM framework in the body of the paper with a 3 x 3 triangle, also included in the companion Excel files are a set of "Simple GLM 6___.xls" files that illustrate the calculations for these different models using a 6 x 6 triangle. Also, the "Bootstrap Models.xls" file contains a "flexible" model for a 10 x 10 triangle that can be used to specify any combination of accident year, development year, and calendar year parameters, including setting parameters to zero. The flexible GLM model is akin to the incremental log model described in Barnett and Zehnwirth [1], so we will leave it to the reader to explore this flexibility by using the Excel file.

4. PRACTICAL ISSUES

Now that we have expanded the basic ODP bootstrap model in a variety of ways, we also want to address some of the key assumptions of the ODP model and some common data issues.

4.1 Negative Incremental Values

As noted in Section 3.2, because of the log-link used in the GLM framework the incremental values must be greater than zero in order to parameterize a model. However, a slight modification to the log-link function will help this common problem become a little less restrictive. If we use (4.1) as the log-link function, then individual negative values are only an issue if the total of all incremental values in a development column is negative, as the GLM algorithm will not be able to find a solution in that case.

$$\begin{gathered} \ln[q(w,d)] \text{ for } q(w,d) > 0, \\ 0 \text{ for } q(w,d) = 0, \\ -\ln[abs\{q(w,d)\}] \text{ for } q(w,d) < 0. \end{gathered} \tag{4.1}$$

Using (4.1) in the GLM framework will help in many situations, but it is quite common for entire development columns of incremental values to be negative, especially for incurred data. To give the GLM framework the ability to solve for a solution in this case we need to make another modification to the basic model to include a constant.

$$\ln[m_{w,d}] + \psi = \eta_{w,d} \tag{4.2}$$

Whenever a column or columns of incremental values sum to a negative value, we can find the largest negative[20] in the triangle, add the absolute value of the largest negative to every incremental value in the triangle, set ψ equal to the largest negative, and solve the GLM using formulas (3.10), (3.11), or (3.29). Then when we use (4.2) to calculate the fitted incremental values, the constant ψ is used to reduce each fitted incremental value by the largest negative.

The combination of formulas (4.1) and (4.2) allow the GLM framework to handle all negative incremental values, which overcomes a common criticism of the ODP bootstrap. Incidentally, these formulas can also be used to allow the incremental log model described by Barnett and Zehnwirth [1] to handle negative incremental values.

When using the age-to-age factors to simplify the ODP bootstrap simulation process, the solution to negative incremental values needs to focus on the residuals and sampled incremental values since an age-to-age factor less than 1.00 will create negative incremental values in the fitted values. More specifically, we need to modify formulas (3.19) and (3.21) as follows:

$$r_{w,d} = \frac{q(w,d) - m_{w,d}}{\sqrt{abs\{m_{w,d}\}}}. \tag{4.3}$$

$$q'(w,d) = r^* \times \sqrt{abs\{m_{w,d}\}} + m_{w,d}. \tag{4.4}$$

While the fitted incremental values and residuals using the age-to-age simplification will generally not match the GLM framework solution using (4.1) and (4.2) they should be reasonably close. While the "purists" may object to these practical solutions, we must keep in mind that every model is an approximation of reality so our goal is to find reasonably close models rather than only restrict ourselves to "pure" models. After all, the assumptions of the "pure" models are themselves approximations.

4.1.1 Negative values during simulation

Even though we have solved problems with negative values when parameterizing a model, negative values can still affect the process variance in the simulation process. When each future incremental value (using $m_{w,d}$ as the mean and the mean times the scale parameter, $\phi m_{w,d}$, as the variance) is sampled from a gamma distribution to add process variance, the parameters of a gamma distribution must be positive. In this case we have two options for using the gamma distribution to

---

[20] The largest negative value can either be the largest negative among the sums of development columns (in which case there may still be individual negative values in the adjusted triangle) or the largest negative incremental value in the triangle.

simulate from a negative incremental value, $m_{w,d}$.

$$-Gamma\left[abs\{m_{w,d}\},\phi abs\{m_{w,d}\}\right] \quad (4.5)$$

$$Gamma\left[abs\{m_{w,d}\},\phi abs\{m_{w,d}\}\right]+2m_{w,d} \quad (4.6)$$

Using formula (4.5) is more intuitive as we are using absolute values to simulate from a gamma distribution and then changing the sign of the result. However, since the gamma distribution is skewed to the right, the resulting distribution using (4.5) will be skewed to the left. Using formula (4.6) is a little less intuitive, but seems more logical since subtracting twice the mean, $m_{w,d}$, will result in a distribution with a mean of $m_{w,d}$ while keeping it skewed to the right (since $m_{w,d}$ is negative).

Negative incremental values can also cause extreme outcomes. This is most prevalent when resampled triangles are created with negative incremental losses in the first few development periods, causing one column of cumulative values to sum close to zero and then next column sum to a much larger number and, consequentially, age-to-age factors that are extremely large. This can result in one or two extreme iterations in a simulation (for example, outcomes that are multiples of 1,000s of the central estimate). These extreme outcomes cannot be ignored, even if the high percentiles are not of interest, because they are likely to significantly affect the mean of the distribution.

In these instances, you have several options. You can 1) remove these iterations from your simulation and replace them with new iterations, 2) recalibrate your model, 3) limit incremental values to zero, or 4) use more than one model.

The first option is to identify the extreme iterations and remove them from your results. Care must be taken that only truly unreasonable extreme iterations are removed, so that the resulting distribution does not understate the probability of extreme outcomes.

The second option is to recalibrate the model to fix this issue. First you must identify the source of the negative incremental losses. For example, it may be from the first row in your triangle, which was the first year the product was written, and therefore exhibit sparse data with negative incremental amounts. One option is to remove this row from the triangle if it is causing extreme results and does not improve the parameterization of the model.

The third option is to limit incremental losses to zero, where any negative incremental is replaced with a zero incremental. This can be done in many ways. Negative incremental values can be replaced with zeros in the original data triangles. Negative incremental values can be kept in the original data triangles, but replaced with zeros if they appear in the sampled triangles. Negative incremental losses can be kept in the historical sampled triangle but replaced with zeros in the projected future incremental losses. Finally, negative incremental values can be replaced with zeros based on which development column they are in (this option is used in the "Bootstrap Models.xls" file). Judgment is required when deciding amongst these options.

The most theoretically sound method to deal with negative incremental values is to consider the source of these losses. If they are caused by reinsurance or salvage and subrogation, then you can model the losses gross of salvage and subrogation, model the salvage and subrogation separately, and combine the iterations assuming 100% correlation.

4.2 Non-Zero Sum of Residuals

The residuals that are calculated in the bootstrap model are essentially error terms, and should be identically distributed with a mean of zero. Generally, however the average of all the residuals is non-zero. The residuals are random observations of the true residual distribution, so this observation is not necessarily incompatible with the true residual distribution having a mean of zero. The real issue is whether these residuals should be adjusted so that their average is zero. For example, if the average of the residuals is positive, then re-sampling from the residual pool will not only add variability to the resampled incremental losses, but may increase the resampled incremental losses such that the average of the resampled loss will be greater than the fitted loss.

The reason why residuals may not sum to zero is due to differing magnitudes of losses in each accident year. If the magnitude of losses is higher for a particular accident year that shows higher development than the weighted average, then the average of all the residuals will be negative. If the magnitude of losses is lower for a particular accident year that shows higher development than the weighted average, then the average of all the residuals will be positive.

It can be argued that the non-zero average of residuals is a characteristic of the data set, and therefore should not be removed. However, if a zero residual average is desired, then one option is the addition of a single constant to all residuals, such that the sum of the shifted residuals is zero.

4.3 Using an *N*-Year Weighted Average

The basic ODP bootstrap model can be simplified by using volume-weighted average age-to-age factors for all years in the triangle. It is quite common, however, for actuaries to use weighted averages that are less than for all years. Thus, it is also important to be able to adjust the ODP bootstrap model to use N-year average age-to-age factors.

For the GLM framework, we can use N years of data by excluding the first few diagonals in the triangle so that we only use N+1 diagonals (since an N-year average uses N+1 diagonals) to parameterize the model. The shape of the data to be modeled essentially becomes a trapezoid instead of a triangle, the excluded diagonals are given zero weight in the model and we have fewer calendar year trend parameters if we are using formula (3.29). When running the bootstrap simulations we will only need to sample residuals for the trapezoid that we used to parameterize the model as that is all that will be needed to estimate parameters for each iteration.

Using the simplified GLM we can also calculate N-year average factors instead of all-year factors and exclude the first few diagonals when calculating residuals. However, when running the bootstrap simulations we would still need to sample residuals for the entire triangle so that we can calculate cumulative values. To be consistent with the assumptions of the simplified GLM in this case, we would still want to use N-year average factors for projecting the future expected values.

The calculations for the GLM framework are illustrated in the companion "Simple GLM 6 with 3yr avg.xls" file. Note that because the GLM framework estimates parameters for the incremental data, the fitted values will no longer match the fitted values from the simplified GLM using volume-weighted average age-to-age factors. However, the fitted values are generally close so the simplified GLM will still be a reasonable approximation to the GLM framework.

4.4 Missing Values

Sometimes the loss triangle will have missing values. For example, values may be missing from the middle of the triangle. Another example is a triangle that is missing the oldest diagonals, if loss data was somehow lost or not kept in the early years of writing the book of business.

If values are missing, then the following calculations will be affected:

- Loss development factors
- Fitted triangle – if the missing value lies on the last diagonal
- Residuals
- Degrees of freedom

There are several solutions. The missing value may be estimated using the surrounding values. Or, the loss development factors can be modified to exclude the missing value, and there will not be a corresponding residual for this missing value. Subsequently, when triangles are resampled, the simulated incremental corresponding to the missing value should not be resampled to reproduce the uncertainty in the original dataset.

If the missing value lies on the last diagonal, the fitted triangle cannot be calculated in the usual way. A solution is to estimate the value, or use the value in the second to last diagonal to construct the fitted triangle. These are not strictly mathematically correct solutions, and judgment will be needed as to their affect on the resulting distribution.

4.5 Outliers

There may be extreme or incorrect values in the original triangle dataset that would be considered outliers. These may not be representative of the variability of the dataset in the future and, if so, the modeler may want to remove their impact from the model.

There are several solutions. If these values formed the first row of the data triangle, which is common, then this whole first row could be deleted, and the model run on a smaller triangle. Alternatively, these values could be removed, and dealt with in the same manner as missing values. Another alternative is to identify outliers and exclude them from the average age-to-age factors (either the numerator, denominator, or both) and residual calculations, as when dealing with missing values, but re-sample the corresponding incremental when simulating triangles.

The calculations for the GLM framework are illustrated in the companion "Simple GLM 6 with Outlier.xls" file. Again the GLM framework fitted values will no longer exactly match the fitted values from the simplified GLM using volume weighted average age-to-age factors.

4.6 Heteroscedasticity

As noted earlier, the ODP model is based on the assumption that the Pearson residuals are independent and identically distributed. It is this assumption that allows the model to take a residual from one development period/accident period and apply it to the fitted loss in any other development period/accident period, to produce the sampled values. In statistical terms this is referred to as homoscedasticity and it is important that this assumption is validated.

A problem is commonly observed when some development periods have residuals that appear to be more variable than others – i.e., they appear to have different distributions or variances. If this observation is correct, then we have multiple distributions within the residuals (statistically referred to as heteroscedasticity) and it is no longer possible to take a residual from one development/accident period and deem it suitable to be applied to any other development/accident period. In making this assessment, you must account for the credibility of the observed difference, and also to note that there are fewer residuals as the development years become older, so comparing development years is difficult, particularly near the tail-end of the triangle.[21]

To adjust for heteroscedasticity in your data there are at least two options, 1) stratified sampling, or 2) calculating variance parameters. Stratified sampling is accomplished by organizing the development periods by group with homogeneous variances within each group and then sampling with replacement only from the residuals in each group. While this process is straightforward and easy to accomplish, quite often some groups may only have a few residuals in them, which limits the amount of variability in the possible outcomes.

The second option is to sort the development periods into groups with homogeneous variances and calculate the standard deviation of the residuals in each of the "hetero" groups. Then calculate h_i, which is the hetero-adjustment factor, for each group, i:

$$h_i = \frac{Max[stdev(r^i_{w,d})]}{stdev(r^i_{w,d})}. \tag{4.7}$$

All residuals in group i are multiplied by h_i.

$$r^{iH}_{w,d} = \frac{q(w,d) - m_{w,d}}{\sqrt{m_{w,d}}} \times f^H_{w,d} \times h^i. \tag{4.8}$$

Now all groups have the same standard deviation and we can sample with replacement from among all $r^{iH}_{w,d}$. The original distribution of residuals has been altered, but this can be remedied. When the residuals are resampled, the residual is divided by the hetero-adjustment factor that applies to the development year of the incremental value, as shown in (4.9).

$$q^i\,'(w,d) = \frac{r^*}{h^i} \times \sqrt{m_{w,d}} + m_{w,d}. \tag{4.9}$$

By doing this, the heteroscedastic variances we observed in the data are replicated when the sample triangles are created, but we are able to freely resample with replacement from the entire pool of residuals. Also note that we have added more parameters so this will affect the degrees of freedom, which impacts the scale parameter (3.20) and the degrees of freedom adjustment factor (3.22). Finally, the hetero group parameters should also be used to adjust the variance when simulating the future process variance.

It is possible to modify the GLM framework to also include "hetero group" parameters, but that is beyond the scope of this paper.

---

[21] We will illustrate how to use residual graphs and other statistical tests to evaluate heteroscedasticity in Section 5.

4.7 Heteroecthesious Data

The basic ODP bootstrap model requires both a symmetrical shape (e.g., annual by annual, quarterly by quarterly, etc. triangles) and homoecthesious data (i.e., similar exposures).[22] As discussed above, using an *N*-year weighted average in the simplified GLM model or adjusting to a trapezoid shape allow us to "relax" the requirement of a symmetrical shape. Other non-symmetrical shapes (e.g., annual x quarterly data) can also be modeled with either the simplified GLM or GLM framework, but they will not be discussed in detail in this paper.

Most often, the actuary will encounter heteroecthesious data (i.e., incomplete or uneven exposures) at interim evaluation dates, with the two most common data triangles being either a partial first development period or a partial last calendar period. For example, with annual data evaluated as of June 30, partial first development period data would have development periods ending at 6, 18, 30, etc. months, while partial last calendar period data would have development periods as of 12, 24, 36, etc. months for all of the data in the triangle except the last diagonal, which would have development periods as of 6, 18, 30, etc. months. In either case, not all of the data in the triangle has full annual exposures – i.e., it is heteroecthesious data.

4.7.1 Partial first development period data

For partial first development period data, the first development column has a different exposure period than the rest of the columns (e.g., in the earlier example the first column has six months of development exposure while the rest have 12). In a deterministic analysis this is not a problem as the age-to-age factors will reflect the change in exposure. For parameterizing an ODP bootstrap model, it also turns to be a moot issue. In addition, since the Pearson residuals use the square root of the fitted value to make them all "exposure independent" that part of an ODP bootstrap model is likewise unaffected.

The only adjustment for this type of heteroecthesious data is the projection of future incremental values. In a deterministic analysis, the most recent accident year needs to be adjusted to remove exposures beyond the evaluation date. For example, continuing the previous example the development periods at 18 months and later are all for an entire year of exposure whereas the six month column is only for six months of exposure. Thus, the 6-18 month age-to-age factor will effectively extrapolate the first six months of exposure in the latest accident year to a full accident year's exposure. Accordingly, it is common practice to reduce the projected future payments by half

---

22 To our knowledge, the terms *homoecthesious* and *heteroecthesious* are new. They are a combination of the Greek *homos* (or ὁμός) meaning the same or *hetero* (or ἑτερο) meaning different and the Greek *ekthesē* (or ἔκθεση) meaning exposure.

to remove the exposure from June 30 to December 31.

The simulation process for the ODP bootstrap model can be adjusted similarly to the way a deterministic analysis would be adjusted. After the age-to-age factors from each sample triangle are used to project the future incremental values the last accident year's values can be reduced (in the previous example by 50%) to remove the future exposure and then process variance can be simulated as before. Alternatively, the future incremental values can be reduced after the process variance step.

4.7.2 Partial last calendar period data

For partial last calendar period data, most of the data in the triangle has annual exposures and annual development periods, except for the last diagonal which, continuing our example, only has a six-month development period (and a six-month exposure period for the bottom cell). For a deterministic analysis, it is quite common in this situation to exclude the last diagonal when calculating average age-to-age factors, interpolate those factors for the exposures in the last diagonal and use the interpolated factors to project the future values. In addition, the last accident year will also need to have the future incremental values reduced to remove exposures beyond the evaluation date.

Similarly to the adjustments for partial first development period data, we could adjust the calculations and steps in the simplified GLM model, but adjustments to the GLM framework are more problematic. Instead of ignoring the last diagonal during the parameterization of the model, an alternative is to adjust or annualize the exposures in the last diagonal to make them consistent with the rest of the triangle.

During the bootstrap simulation process, age-to-age factors can be calculated from the fully annualized sample triangles and interpolated. Then, the last diagonal from the sample triangle can be adjusted to de-annualize the incremental values in the last diagonal – i.e., reversing the annualization of the original last diagonal. The new cumulative values can be multiplied by the interpolated age-to-age factors to project future values. Again, the future incremental values for the last accident year must be reduced (in the previous example by 50%) to remove the future exposure.[23]

4.8 Exposure Adjustment

Another common issue in real data is exposures that have changed dramatically over the years.

---

[23] These heteroecthesious data issues are not illustrated in the "Bootstrap Models.xls" file.

For example, in a line of business that has experienced rapid growth or is being run off. If the earned exposures exist for this data, then a useful option for the ODP bootstrap model is to divide all of the claim data by the exposures for each accident year – i.e., effectively using pure premium development instead of total loss development. Quite often this will improve the fit of the model to the data.

During the bootstrap simulation process, all of the calculations would be done using the exposure-adjusted data and only after the process variance step has been completed would you multiply the results by the exposures by year to restate them in terms of total values again.

4.9 Parametric Bootstrapping

Because the number of data points used to parameterize the ODP bootstrap model are limited (in the case of a 10x10 triangle to 53 residuals), it is hard to determine whether the most extreme observation is a one-in-100 or a one-in-1,000 event (or simply, in this example, a one-in-53 event). Of course, the nature of the extreme observations in the data will also affect the level of extreme simulations in the results. Judgment is involved here, but the modeler will either need to be satisfied with the level of extreme simulations in the results or modify the bootstrap algorithm.

One way to overcome a lack of extreme residuals for the ODP bootstrap model would be to parameterize a distribution for the residuals and resample using the distribution (e.g., use a normal distribution if the residuals are normally distributed). This option for "sampling residuals" is beyond the scope of the companion Excel files, but this is commonly referred to as parametric bootstrapping.

5. DIAGNOSTICS

The quality of a bootstrap model depends on the quality of the underlying assumptions. When any model fails to "fit" the data, it cannot produce a good estimate of the distribution of possible outcomes.[24]

One of the advantages of the ODP bootstrap model is how readily it can be tailored to some of the statistical features of the data using the GLM framework and considerations described in the previous two sections. The CAS Working Party, in the third section of their report on quantifying variability in reserve estimates [6], identified 20 criteria or diagnostic tools for gauging the quality of

---

[24] While the examples are different, significant portions of sections 5 and 6 are based on [22] and [14].

a stochastic model. The Working Party also noted that, in trying to determine the optimal "fit" of a model, or indeed an optimal model, no single diagnostic tool or group of tools can be considered definitive. Depending on the statistical features found in the data, a variety of diagnostic tools are necessary to best judge the quality of the model assumptions and to change or adjust the parameters of the model. In this sense, the diagnostic tools are used to help find the models that ultimately provide the best fit to the data. We will discuss some of these tools in detail in this paper.

The key diagnostic tests are designed for three purposes: to test various assumptions in the model, to gauge the quality of the model fit, or to help guide the adjustment of model parameters. Some tests may be considered relative in nature, enabling results from one set of model parameters to be compared to those of another, for a specific model. In turn, by analyzing these results a modeler may then be able to improve the fit of the model. For the most part, however, the tests generally can't be used to compare different models. The objective, consistent with the goals of a deterministic analysis, is not to find the one best model, but rather a set of reasonable models.

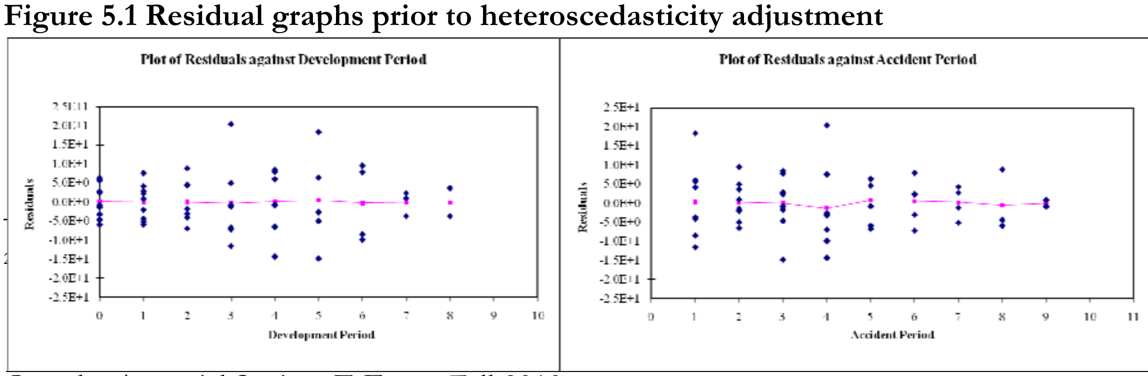
Some diagnostic measures include statistical tests, providing a pass/fail determination for some aspects of the model assumptions. This can be useful even though a "fail" does not necessarily invalidate an entire model; it only points to areas where improvements can be made to the model or its parameterization. The goal is to find the sets of models and parameters that will yield the most realistic, most consistent simulations, based on statistical features found in the data.

To illustrate some of the diagnostic tests for the ODP bootstrap model we will consider data from England and Verrall [9].[25]

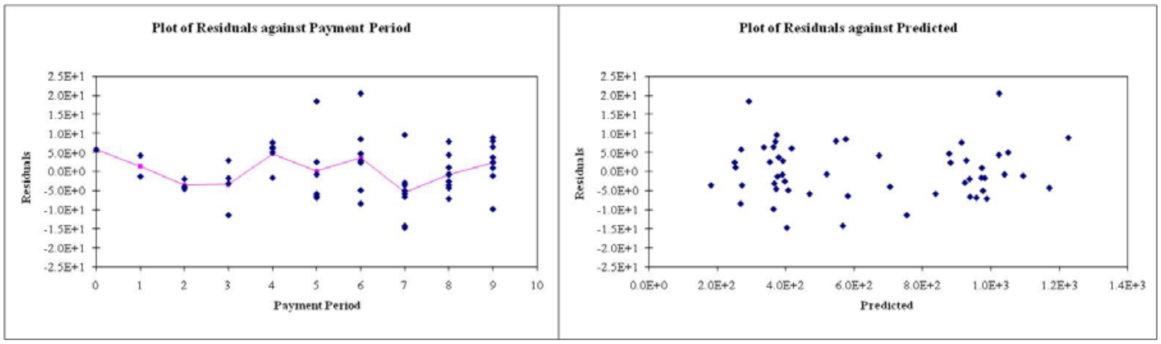
5.1 Residual graphs

The ODP bootstrap model does not require a specific type of distribution for the residuals, but they are assumed to be independent and identically distributed. Because residuals will be sampled with replacement during the simulations, this requirement becomes important and thus it is necessary to test this assumption. A look at graphs of residuals is a good way to do this.

Figure 5.1 Residual graphs prior to heteroscedasticity adjustment



2 … is data is



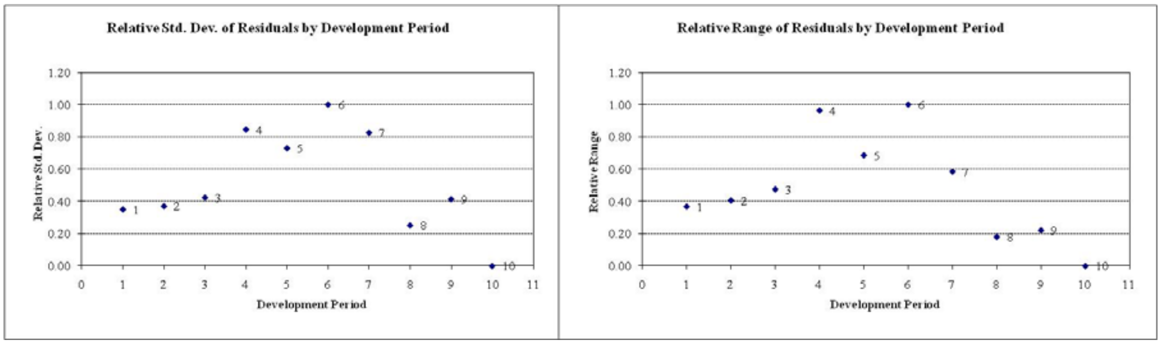
Going clock-wise, and starting from the top-left-hand corner, the graphs in Figure 5.1 show the residuals (blue dots) by development period, accident period, and calendar period and against the fitted incremental loss (in the lower-right-hand corner). In addition, the graphs include a trend line (in pink) that highlights the averages for each period.

At first glance, the residuals in the graphs appear reasonably random, indicating the model is likely a good fit of the data. But a closer look may also reveal potential features in the data that, with the benefit of further analysis, may indicate ways to improve the model fit.

The graphs in Figure 5.1 do not appear to indicate issues with trends, even if the trends for the development and accident periods are both essentially straight. That's because the simplified GLM specifies a parameter for every row and column of the triangle. The development-period graph does, however, reveal a potential heteroscedasticity issue associated with the data. Heteroscedasticity is when random variables have different variances. Note how the upper left graph appears to show a variance of the residuals in the first three periods that differs from those of the middle four or last two periods.

Adjustments for heteroscedasticity can be made with the “Bootstrap Models.xls” file, which enables us to recognize groups of development periods and then adjust the residuals to a common standard deviation value. As an aid to visualizing how to group the development periods into “hetero” groups, graphs of the standard deviation and range relativities can then be developed. Figure 5.2 represents pre-adjusted relativities for the residuals shown in Figure 5.1.

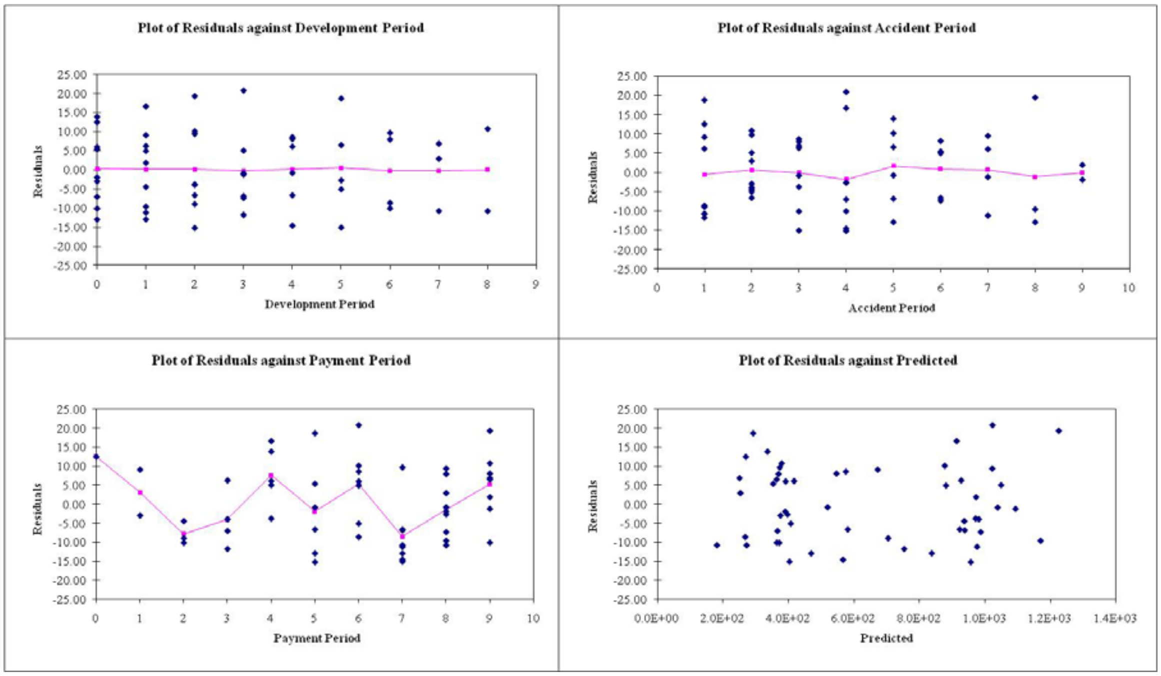
Figure 5.2 Residual relativities prior to heteroscedasticity adjustment



The relativities illustrated in Figure 5.2 help to clarify the veracity of this test, indicating that the residuals in the first three periods are different from those in the middle four or the last two. However, further testing will be required to assess the optimal groups, which can be performed using the other diagnostic tests noted below.

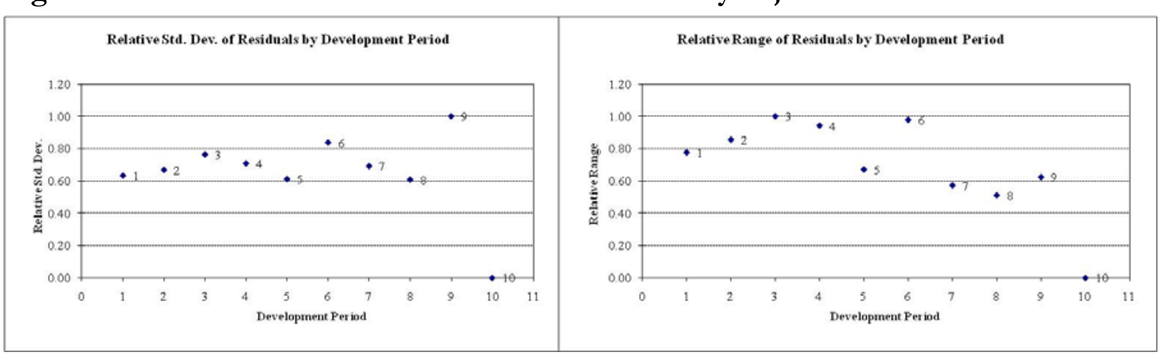
The residual plots in Figure 5.3 originate from the same data model after setting up "hetero" groups for the same array: the first three, middle four, and last two development periods, respectively. Determining whether this "hetero" grouping has improved the model fit will require review of other diagnostic tests.

Figure 5.3 Residual graphs after heteroscedasticity adjustment



Comparing the residual plots in Figures 5.1 and 5.3 does show that the general "shape" of the residuals has not changed and the "randomness" is still consistent. But the residuals now appear to exhibit the same standard deviation, or homoscedasticity. More consistent relativities may also be seen in a comparison of the residual relativities in Figures 5.2 and 5.4.

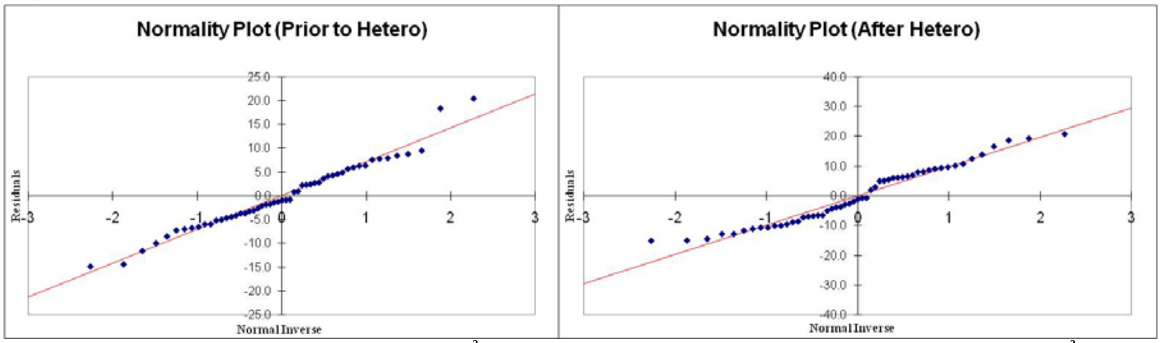
Figure 5.4 Residual relativities after heteroscedasticity adjustment



5.2 Normality test

The ODP bootstrap model does not depend on the residuals being normally distributed, but even so, comparing residuals against a normal distribution remains a useful test, enabling comparison of parameter sets and gauging skewness of the residuals. This test uses both graphs and calculated test values. Figure 5.5 is based on the same heteroscedasticity groups used earlier.

Figure 5.5 Normality plots prior to and after heteroscedasticity adjustment



N = 53 P-Value = 20.5% R^2 = 97.2%
Normal: MU = 0.04, Sigma = 7.09 AIC = 206.1, BIC = 93.1

N = 53 P-Value = 9.9% R^2 = 96.4%
Normal: MU = 0.07, Sigma = 9.85 AIC = 257.8, BIC = 148.8

Even before the heteroscedasticity adjustment, the residual plots appear close to normally distributed, with the data points tightly distributed around the diagonal line. The p-value, a statistical pass-fail test for normality, came in at 20.5%, which far exceeds the value generally considered a "passing" score of the normality test, which is greater than 5.0%.[26] The graphs in Figure 5.5 also show N (the number of data points) and the R^2 test. After the hetero adjustment, the p-value and R^2 don't appear to improve, which indicates that the tested "hetero" groups have not made the residual distribution more normally distributed.

While the p-value and R^2 tests are straightforward and easy to apply, neither adjusts for additional parameters used in the model, a critical limitation. Two other tests, the Akaike Information Criteria (AIC) and the Bayesian Information Criteria (BIC), address this limitation, using the difference between each residual and its normal counterpart from the normality plot to calculate the Residual Sum Squared (RSS) and include a penalty for additional parameters, as shown in (5.1) and (5.2), respectively.[27]

$$AIC = 2\times p + n\times\left[\ln\left(\frac{2\times\pi\times RSS}{n}\right)+1\right] \tag{5.1}$$

$$BIC = n\times\ln\left(\frac{RSS}{n}\right) + p\times\ln(n) \tag{5.2}$$

A smaller value for the AIC and BIC tests indicate residuals that fit a normal distribution more

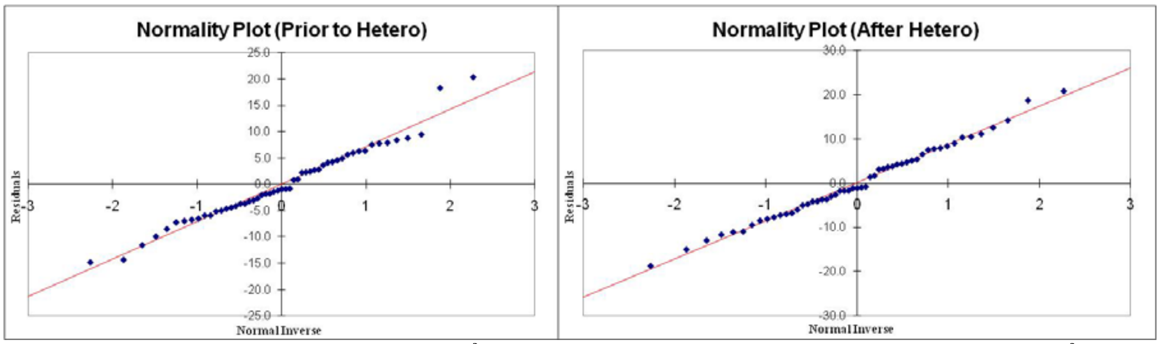
---

[26] Remember that this doesn't indicate whether the bootstrap model itself passes or fails – the bootstrap model doesn't require the residuals to be normally distributed. While not included in the "Bootstrap Models.xls" file, as discussed in section 4.9, it could be used to determine whether to switch to a parametric bootstrap process using a normal distribution.

[27] There are different versions of the AIC and BIC formula from various authors and sources, but the general idea of each version is consistent.

closely, and this improvement in fit overcomes the penalty of adding a parameter. With some trial and error, a better "hetero" grouping was found with the normality results shown in Figure 5.6.[28] For the new "hetero" groups, all of the statistical tests improved dramatically.

Figure 5.6 Normality plots prior to and after heteroscedasticity adjustment



N = 53 P-Value = 20.5% R^2 = 97.2%
Normal: MU = 0.04, Sigma = 7.09 AIC = 206.1, BIC = 93.1

N = 53 P-Value = 95.4% R^2 = 99.2%
Normal: MU = 0.05, Sigma = 8.51 AIC = 160.8, BIC = 51.7

5.3 Outliers

Identifying outliers in the data provides another useful test in determining model fit. Outliers can be represented graphically in a box-whisker plot, which shows the inter-quartile range (the 25th to 75th percentiles) and the median (50th percentile) of the residuals—the so-called box. The whiskers then extend to the largest values within three times this inter-quartile range. Values beyond the whiskers may generally be considered outliers and are identified individually with a point.

[28] In the "Bootstrap Models.xls" file the England and Verrall data was entered as both paid and incurred. The first set of "hetero" groups are illustrated for the "incurred" data and the second set of "hetero" groups are illustrated for the "paid" data.

Figure 5.7 Box-Whisker Plots Prior to and After Heteroscedasticity Adjustment

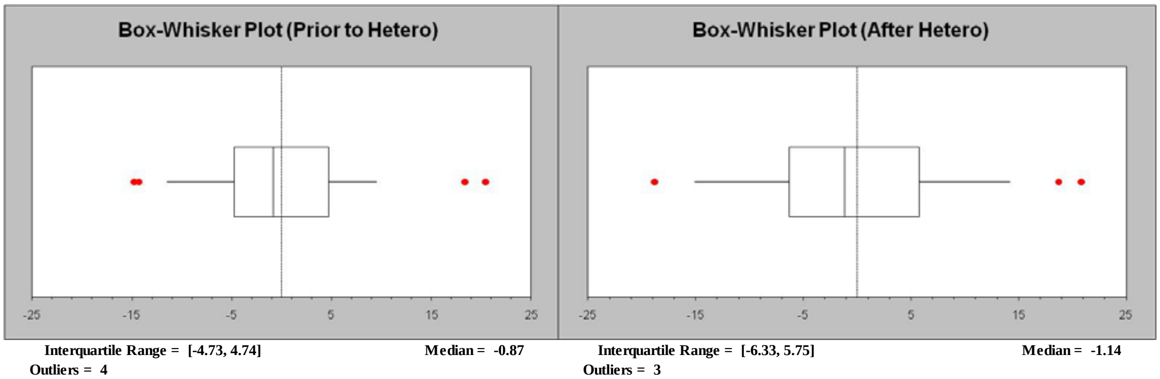


Figure 5.7 shows an example of the residuals for the second set of "hetero" groups (Figure 5.6). A pre-hetero adjustment plot returns four outliers (red dots) in the data model, corresponding to the two highest and two lowest values in the previous graphs in Figures 5.1, 5.3, 5.5, and 5.6.

Even after the hetero adjustment, the residuals still appear to contain three outliers. Now comes a very delicate and often tricky matter of actuarial judgment. If the data in those cells genuinely represent events that cannot be expected to happen again, the outliers may be removed from the model (by giving them zero weight). But extreme caution should be taken even when the removal of outliers seems warranted. The possibility always remains that apparent outliers may actually represent realistic extreme values, which, of course, are critically important to include as part of any sound analysis.

Additionally, when residuals are not normally distributed a significant number of "outliers" tend to result, which may be only an artifact of the function of the distributional shape of the residuals. Again, it is preferable to let these stand in order to enable the simulation process to replicate this shape.

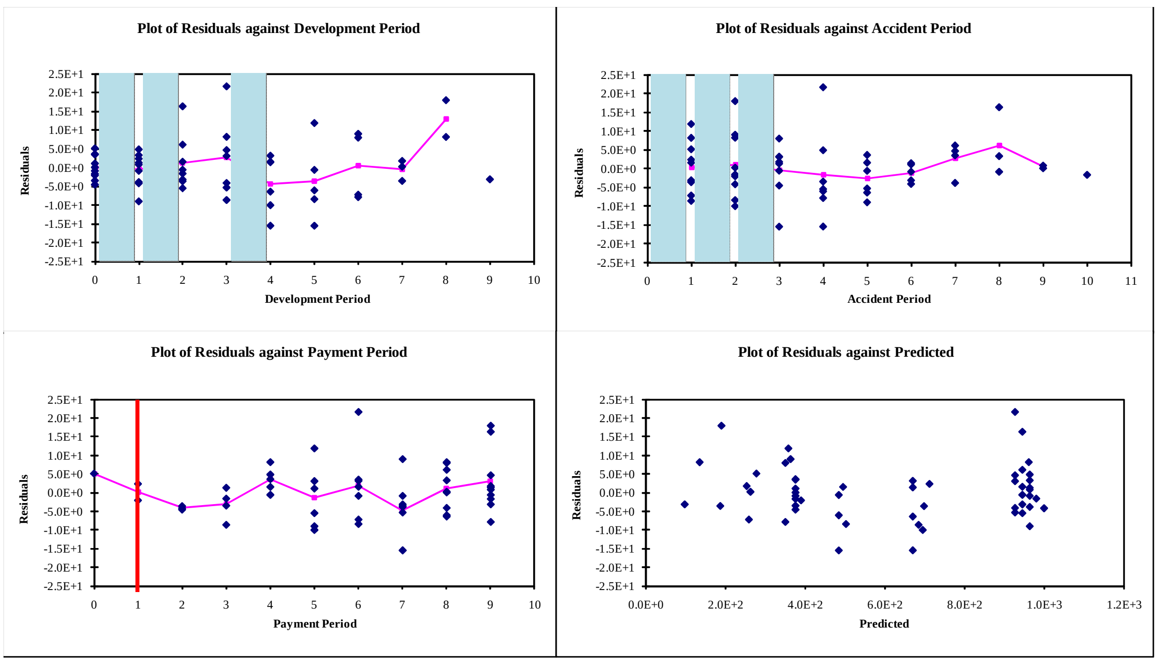
While the three diagnostic tests shown above demonstrate techniques commonly used with most types of models, they are not the only tests available. Next, we'll take a look at the flexibility of the GLM framework and some of the diagnostic elements of the simulation results. For a more extensive list of other tests available, see the report, CAS Working Party on Quantifying Variability in Reserve Estimates [6].

5.4 Parameter adjustment

As noted in section 5.1 the relatively straight average lines in the development and accident period graphs are a reflection of having a parameter for every accident and development period. In some instances, this is also an indication that the model may be over parameterized. Using the "flexible" model in the "Bootstrap Models.xls" file we can illustrate the power of removing some of the parameters.

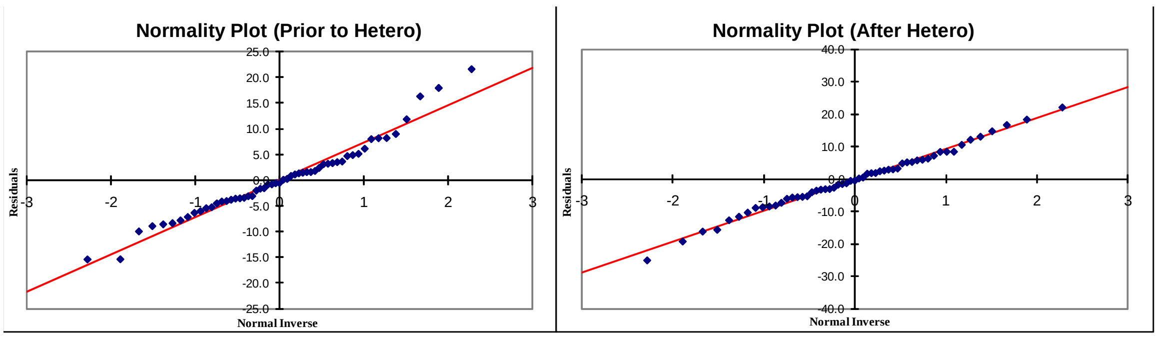
Starting with the "basic" model which includes only one parameter for accident, development and calendar periods (i.e., only one α, β and γ parameter), with a little trial and error we can find a reasonably good fit to the data using only three accident, three development and no calendar parameters. Adding blue bars to signify a parameter and red bars to signify no parameter (i.e., parameter of zero), the residual graphs for the "flexible" model are shown in figure 5.8.

Figure 5.8 Residual graphs for "flexible" model



Using the second set of "hetero" groups we can also check the normality graphs and statistics in figure 5.9 and outliers in figure 5.10. Comparing the statistics to the simplified GLM values shown in figures 5.6 and 5.7, some values improved while others did not. However, the values are not significantly different, yet the "flexible" model is far more parsimonious.

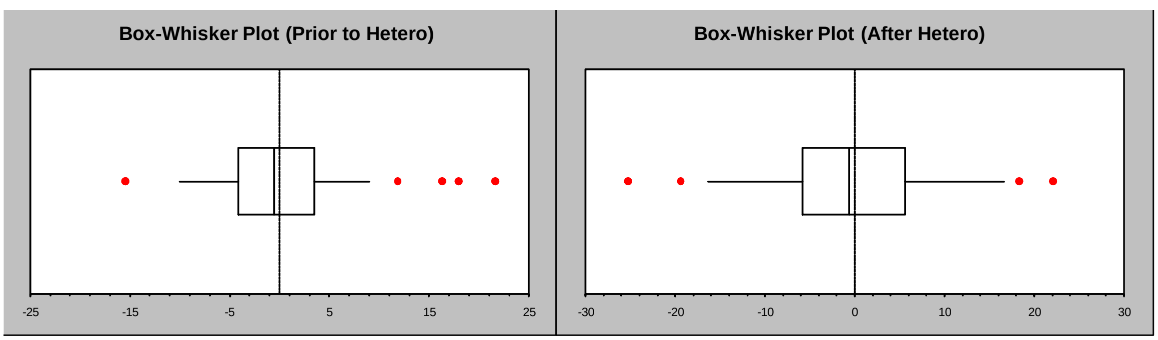
Figure 5.9 Normality plots for "flexible" model



N = 55 P-Value = 7.6% R^2 = 96.2%
Normal: MU = 0.07, Sigma = 7.26 AIC = 204.9, BIC = 60.9

N = 55 P-Value = 94.8% R^2 = 99.2%
Normal: MU = -0.31, Sigma = 9.40 AIC = 150.3, BIC = 10.2

Figure 5.10 Box-Whisker plots for "flexible" model



Interquartile Range = [-4.16, 3.46] Median = -0.58
Outliers = 6

Interquartile Range = [-5.78, 5.61] Median = -0.59
Outliers = 4

5.5 Model results

Once diagnostics have been reviewed, simulations should be run for each model. These simulation results may often provide an additional diagnostic tool to aid in evaluation of the model. As one example, we will review the results for the England and Verrall data using the simplified GLM model. The estimated-unpaid results shown in Figure 5.11 were simulated using 1,000 iterations with the hetero adjustments from Figure 5.6.

Figure 5.11 Estimated-Unpaid Model Results

England & Verrall Data
Accident Year Unpaid
Paid Chain Ladder Model

Accident Year	To Date	Mean Unpaid	Standard Error	Coefficient of Variation	Minimum	Maximum	50.0% Percentile	75.0% Percentile	95.0% Percentile	99.0% Percentile
1999	3,901	-	-		-	-	-	-	-	-
2000	5,339	93	125	134.0%	(377)	900	62	156	306	502
2001	4,909	479	246	51.3%	(115)	1,694	447	615	940	1,186
2002	4,588	723	276	38.2%	(51)	1,892	691	899	1,220	1,515
2003	3,873	984	293	29.7%	267	2,160	976	1,176	1,453	1,802
2004	3,692	1,430	366	25.6%	434	2,888	1,400	1,670	2,072	2,405
2005	3,483	2,183	484	22.2%	896	3,812	2,140	2,497	3,038	3,483
2006	2,864	3,909	749	19.2%	1,793	6,482	3,875	4,402	5,175	5,935
2007	1,363	4,261	830	19.5%	1,757	7,865	4,221	4,789	5,700	6,321
2008	344	4,672	1,839	39.4%	617	11,009	4,523	5,853	7,878	9,509
Totals	34,358	18,737	2,769	14.8%	11,019	29,190	18,647	20,533	23,611	25,486

5.5.1 Estimated-Unpaid Results

It's recommended to start diagnostic review of the estimated-unpaid table with the standard error (standard deviation) and coefficient of variation (standard error divided by the mean), shown in Figure 5.11. Keep in mind that the standard error should increase when moving from the oldest years to the most recent years, as the standard errors (value scale) should follow the magnitude of the mean of unpaid estimates. In Figure 5.11, the standard errors conform to this pattern. At the same time, the standard error for the total of all years should be larger than any individual year.

Also, the coefficients of variation should generally decrease when moving from the oldest years to the more recent years and the coefficient of variation for all years combined should be less than for any individual year. With the exception of the 2008 accident year, the coefficients of variation in Figure 5.11 seem to also conform, although some random fluctuations may be seen.

The main reason for the decrease in the coefficient of variation has to do with the independence in the incremental claim-payment stream. Because the oldest accident year typically has only a few incremental payments remaining, or even just one, the variability is nearly all reflected in the coefficient. For more current accident years, random variations in the future incremental payment stream may tend to offset one another, thereby reducing the variability of the total unpaid loss.

While the coefficients of variation should go down, they could also start to rise again in the most recent years, which can been seen in Figure 5.11 for 2008. Such a reversal could result from a couple of issues:

- With an increasing number of parameters used in the model, the parameter uncertainty tends to increase when moving from the oldest years to the more recent years. In the most recent years, parameter uncertainty can grow to "overpower" process uncertainty, which may cause the coefficient of variation to start rising again. At a minimum, increasing parameter uncertainty will slow the rate of decrease in the coefficient of variation.
- The model may be overestimating the uncertainty in recent accident years if the increase is significant. In that case, the Bornhuetter-Ferguson or Cape Cod model may need to be used instead of a chain-ladder model.

Keep in mind also that the standard error or coefficient of variation for the total of all accident years will be less than the sum of the standard error or coefficient of variation for the individual years. This is because the model assumes that accident years are independent.

Minimum and maximum results are the next diagnostic element in our analysis of the estimated-unpaid claims in Figure 5.11, representing the smallest and largest values from all iterations of the simulation. These values will need to be reviewed in order to determine their veracity. If any of them seem implausible, the model assumptions would need to be reviewed. Their effects could materially alter the mean indication.

5.5.2 Mean and Standard Deviation of Incremental Values

The mean and standard deviation of every incremental value from the simulation process also provide useful diagnostic results, enabling us to dig deeper into potential coefficient of variation issues that may be found in the estimated-unpaid results. Consider, for example, the mean and standard deviation results shown in Figures 5.12 and 5.13, respectively.

Figure 5.12 Mean of incremental values

England & Verrall Data
Accident Year Incremental Values by Development Period
Paid Chain Ladder Model

Accident	Mean Values									
Year	**12**	**24**	**36**	**48**	**60**	**72**	**84**	**96**	**108**	**120**
1999	266	675	694	767	421	294	267	180	274	67
2000	375	945	973	1,030	588	400	376	251	383	93
2001	372	926	987	1,040	572	406	373	249	385	94
2002	369	916	967	1,037	576	395	366	253	380	90
2003	333	837	893	936	508	362	334	222	342	86
2004	351	876	943	983	546	384	354	237	362	93
2005	395	973	1,028	1,093	606	425	389	266	400	97
2006	463	1,165	1,218	1,297	721	511	472	315	476	116
2007	393	964	1,020	1,075	601	422	388	262	396	97
2008	340	861	913	974	543	359	345	233	361	84

The mean values in Figure 5.12 appear consistent throughout and support the increases in estimated unpaid by accident year that are shown in Figure 5.11. In fact, the future mean values, which lay beyond the stepped diagonal line in Figure 5.12, sum to the results in Figure 5.11. The standard deviation values in Figure 5.13, however, only appear consistent up to 2007; 2008 has larger standard deviations, which again are consistent with the standard deviations seen in Figure 5.11. But contrariwise the standard deviations can't be added because the standard deviations in Figure 5.11 represent those for aggregated incremental values by accident year, which are less than perfectly correlated.

Figure 5.13 Standard deviation of incremental values

England & Verrall Data
Accident Year Incremental Values by Development Period
Paid Chain Ladder Model

Accident	Standard Error Values									
Year	**12**	**24**	**36**	**48**	**60**	**72**	**84**	**96**	**108**	**120**
1999	106	120	233	232	138	143	103	88	137	69
2000	126	142	264	272	153	173	128	106	162	125
2001	131	134	260	281	156	174	127	102	205	129
2002	126	139	256	274	150	171	124	115	195	128
2003	122	131	246	249	148	158	125	107	178	118
2004	128	134	252	259	151	167	132	108	187	127
2005	133	144	281	283	180	190	142	123	207	134
2006	141	161	302	356	202	207	168	140	232	153
2007	124	142	297	326	185	180	138	115	208	124
2008	120	340	410	458	256	210	190	137	230	121

6. USING MULTIPLE MODELS

So far we have focused only on one model. In practice, multiple stochastic models should be used in the same way that multiple methods should be used in a deterministic analysis. First the results for each model must be reviewed and finalized, after an iterative process of diagnostic testing and reviewing model output. Then these results can be combined by assigning a weight to the results of each model.

Two primary methods exist for combining the results for multiple models:

- **Run models with the same random variables.** For this algorithm, every model uses the exact same random variables. In the "Bootstrap Models.xls" file, the random values are simulated before they are used to simulate results, which means that this algorithm may be accomplished by reusing the same set of random variables for each model. At the end, the incremental values for each model, for each iteration by accident year (that have a partial weight), can be weighted together.
- **Run models with independent random variables.** For this algorithm, every model is run with its own random variables. In the "Bootstrap Models.xls" file, the random values are simulated before they are used to simulate results, which means that this algorithm may be accomplished by simulating a new set of random variables for each model. At the end, the weights are used to randomly select a model for each iteration by accident year so that the result is a weighted "mixture" of models.

Both algorithms are similar to the process of weighting the results of different deterministic methods to arrive at an actuarial best estimate. The process of weighting the results of different stochastic models produces an actuarial best estimate of a distribution.

The second method of combining multiple models can be illustrated using combined Schedule P data for five top 50 companies.[29] Data for all Schedule P lines with 10 years of history may be found in the "Industry Data.xls" file, but we will confine our examination to Parts A, B, and C. For each line of business we ran simplified GLM models for paid and incurred data (labeled Chain Ladder), as well as paid and incurred data for the Bornhuetter-Ferguson and Cape Cod models described in section 3.3. For this section, we will only focus on the results for Part A (Homeowners/Farm owners).

By comparing the results for all six models (or fewer, depending on how many are used)[30] a qualitative assessment of the relative merits of each model may be determined. Bayesian methods

---

[29] The five companies represent large, medium and smaller companies that have been combined to maintain anonymity. For each Part, a unique set of five companies were used.
[30] Other models in addition to a bootstrap model could also be included in the weighting process.

can be used to determine weighting based on the quality of each model's forecasts. The weights can be determined separately for each year. The table in Figure 6.1 shows an example of weights for the Part A data.[31] The weighted results are displayed in the "Best Estimate" column of Figure 6.2. As a parallel to a deterministic analysis, the means from the six models could be considered a reasonable range (i.e., from $4,059 to $5,242).

Figure 6.1 Model weights by accident year

Accident	Model Weights by Accident Year						
	Chain Ladder		Bornhuetter-Ferguson		Cape Cod		
Year	Paid	Incurred	Paid	Incurred	Paid	Incurred	TOTAL
1999	50.0%	50.0%					100.0%
2000	50.0%	50.0%					100.0%
2001	50.0%	50.0%					100.0%
2002	50.0%	50.0%					100.0%
2003	50.0%	50.0%					100.0%
2004	50.0%	50.0%					100.0%
2005	50.0%	50.0%					100.0%
2006	12.5%	12.5%	18.8%	18.8%	18.8%	18.8%	100.0%
2007	12.5%	12.5%	18.8%	18.8%	18.8%	18.8%	100.0%
2008	12.5%	12.5%	18.8%	18.8%	18.8%	18.8%	100.0%

Figure 6.2 Summary of results by model

Five Top 50 Companies
Schedule P, Part A -- Homeowners / Farmowners (in 000,000's)
Summary of Results by Model

Accident	Mean Estimated Unpaid						
	Chain Ladder		Bornhuetter Ferguson		Cape Cod		Best Est.
Year	Paid	Incurred	Paid	Incurred	Paid	Incurred	(Weighted)
1999	-	-	-	-	-	-	-
2000	2	1	1	2	2	2	1
2001	38	36	25	25	25	32	37
2002	42	40	36	36	36	42	41
2003	57	60	56	57	57	66	59
2004	98	98	94	92	92	106	99
2005	212	219	164	166	166	189	218
2006	290	292	327	318	318	371	339
2007	677	665	715	701	701	823	739
2008	3,826	3,826	2,642	2,840	2,840	3,324	3,192
Totals	5,242	5,239	4,059	4,236	4,236	4,953	4,726

With our focus on the entire distribution, the weights by year were used to randomly sample the specified percentage of iterations from each model. A more complete set of the results for the "weighted" iterations can be created similar to the tables shown in section 5. The companion "Best Estimate.xls" file can be used to weight six different models together in order to calculate a weighted

[31] For simplicity, the weights are judgmental and not derived using Bayesian methods.

best estimate. An example for Part A is shown in the table in Figure 6.3.

Figure 6.3 Estimated-unpaid model results (best estimate)

Five Top 50 Companies
Schedule P, Part A -- Homeowners / Farmowners (in 000,000's)
Accident Year Unpaid
Best Estimate (Weighted)

Accident Year	Paid To Date	Mean Unpaid	Standard Error	Coefficient of Variation	Minimum	Maximum	50.0% Percentile	75.0% Percentile	95.0% Percentile	99.0% Percentile
1999	5,234	-	-		-	-	-	-	-	-
2000	6,470	1	11	745.3%	(52)	84	0	2	19	51
2001	7,848	37	39	104.5%	(68)	263	28	56	112	164
2002	7,020	41	36	86.7%	(48)	230	33	58	108	155
2003	7,291	59	41	69.1%	(41)	276	49	78	136	191
2004	8,134	99	49	49.1%	(14)	377	90	121	188	259
2005	10,800	218	78	36.0%	28	666	209	259	359	457
2006	7,522	339	129	38.1%	37	1,227	321	402	570	739
2007	7,968	739	259	35.1%	112	1,981	722	875	1,196	1,557
2008	9,309	3,192	920	28.8%	1,090	11,122	3,128	3,629	4,792	5,722
Totals	77,596	4,726	999	21.1%	2,528	13,422	4,632	5,209	6,554	7,442
Normal Dist.		4,726	999	21.1%			4,726	5,400	6,369	7,050
logNormal Dist.		4,725	968	20.5%			4,628	5,307	6,461	7,419
Gamma Dist.		4,726	999	21.1%			4,656	5,356	6,480	7,354
TVaR							5,454	6,003	7,311	8,915
Normal TVaR							5,523	5,996	6,786	7,388
logNormal TVaR							5,484	6,021	7,054	7,964
Gamma TVaR							5,518	6,049	7,018	7,824

6.1 Additional Useful Output

Three rows of percentile numbers for the normal, lognormal, and gamma distributions, which have been fitted to the total unpaid-claim distribution, may be seen at the bottom of the table in Figure 6.3. These fitted mean, standard deviation, and selected percentiles are in their respective columns; the smoothed results can be used to assess the quality of fit, parameterize a DFA model, or used to estimate extreme values,[32] among other applications.

Four rows of numbers indicating the Tail Value at Risk (TVaR), defined as the average of all of the simulated values equal to or greater than the percentile value, may also be seen at the bottom of Figure 6.3. For example, in this table, the 99th percentile value for the total unpaid claims for all accident years combined is 7,442, while the average of all simulated values that are greater than or equal to 7,442 is 8,915. The Normal TVaR, Lognormal TVaR, and Gamma TVaR rows are calculated similarly, except that they use the respective fitted distributions in the calculations rather than actual simulated values from the model.

An analysis of the TVaR values is likely to help clarify a critical issue: if the actual outcome exceeds the X percentile value, how much will it exceed that value on average? This type of assessment can have important implications related to risk-based capital calculations and other technical aspects of enterprise risk management. But it is worth noting that the purpose of the

[32] Of course the use of the extreme values assumes that the models are reliable.

normal, lognormal, and gamma TVaR numbers is to provide "smoothed" values—that is, that some of the random statistical noise is essentially prevented from distorting the calculations.

6.2 Estimated Cash Flow Results

An ODP bootstrap model's output may also be reviewed by calendar year (or by future diagonal), as shown in the table in Figure 6.4. A comparison of the values in Figures 6.3 and 6.4 indicates that the total rows are identical, because summing the future payments horizontally or diagonally will produce the same total. Similar diagnostic issues (as discussed in Section 5) may be reviewed in the table in Figure 6.4, with the exception of the relative values of the standard errors and coefficients of variation moving in opposite directions for calendar years compared to accident years. This phenomenon makes sense on an intuitive level when one considers that "final" payments, projected to the furthest point in the future, should actually be the smallest, yet relatively most uncertain.

Figure 6.4 Estimated Cash Flow (best estimate)

Five Top 50 Companies
Schedule P, Part A -- Homeowners / Farmowners (in 000,000's)
Calendar Year Unpaid
Best Estimate (Weighted)

Calendar Year	Mean Unpaid	Standard Error	Coefficient of Variation	Minimum	Maximum	50.0% Percentile	75.0% Percentile	95.0% Percentile	99.0% Percentile
2009	3,093	726	23.5%	1,445	9,809	3,024	3,428	4,355	5,101
2010	799	186	23.3%	312	2,057	786	900	1,125	1,329
2011	362	97	26.7%	124	856	356	422	528	601
2012	191	63	32.9%	52	507	183	224	312	386
2013	118	52	44.0%	(14)	430	110	144	212	285
2014	64	34	52.8%	(61)	205	60	80	127	175
2015	50	36	71.1%	(14)	332	42	67	116	191
2016	41	39	95.9%	(93)	296	31	56	112	177
2017	7	17	257.3%	(60)	175	0	9	40	64
2018	-	-		-	-	-	-	-	-
Totals	4,726	999	21.1%	2,528	13,422	4,632	5,209	6,554	7,442

6.3 Estimated Ultimate Loss Ratio Results

Another output table, Figure 6.5, shows the estimated ultimate-loss ratios by accident year. Unlike the estimated-unpaid and estimated-cash-flow tables, the values in this table are calculated using all simulated values, not just the values beyond the end of the historical triangle. Because the simulated sample triangles represent additional possibilities of what could have happened in the past, even as the "squaring of the triangle" and process variance represent what could happen as those same past values are played out into the future, we are in possession of sufficient information to enable us to estimate the complete variability in the loss ratio from day one until all claims are completely paid and settled for each accident year.[33]

Figure 6.5 Estimated-loss-ratio (best estimate)

Five Top 50 Companies
Schedule P, Part A -- Homeowners / Farmowners (in 000,000's)
Accident Year Ultimate Loss Ratios
Best Estimate (Weighted)

Accident Year	Mean Loss Ratio	Standard Error	Coefficient of Variation	Minimum	Maximum	50.0% Percentile	75.0% Percentile	95.0% Percentile	99.0% Percentile
1999	66.3%	23.9%	36.0%	-1.4%	155.5%	65.5%	71.1%	118.7%	146.5%
2000	78.4%	24.1%	30.8%	-0.7%	189.8%	77.6%	83.9%	123.8%	157.3%
2001	87.9%	25.5%	29.0%	12.9%	260.1%	88.5%	94.1%	136.1%	175.3%
2002	72.2%	21.9%	30.3%	-31.1%	170.8%	71.6%	76.3%	117.4%	143.5%
2003	64.7%	19.2%	29.7%	15.1%	227.3%	63.4%	68.3%	104.6%	125.7%
2004	64.1%	17.3%	27.1%	-5.8%	130.7%	62.9%	67.1%	102.1%	118.6%
2005	80.3%	18.8%	23.4%	16.4%	165.7%	79.1%	84.5%	119.6%	139.1%
2006	55.1%	16.3%	29.5%	7.9%	205.9%	53.8%	57.6%	89.7%	106.1%
2007	56.7%	16.2%	28.6%	10.1%	123.8%	56.8%	60.7%	89.0%	106.4%
2008	83.6%	20.6%	24.6%	33.1%	307.1%	81.8%	87.9%	123.5%	150.9%
Totals	70.1%	6.7%	9.5%	50.0%	114.7%	69.9%	74.1%	81.0%	87.7%

The use of all simulated values indicates that the standard errors in Figure 6.5 should be proportionate to the means, while the coefficients of variation should be relatively constant by accident year. In terms of diagnostics, any increases in standard error and coefficient of variation for the most recent years would be consistent with the reasons previously cited in Section 5.4 for the estimated-unpaid tables. Risk management-wise, the loss ratio distributions have important implications for projecting pricing risk.

6.4 Distribution Graphs

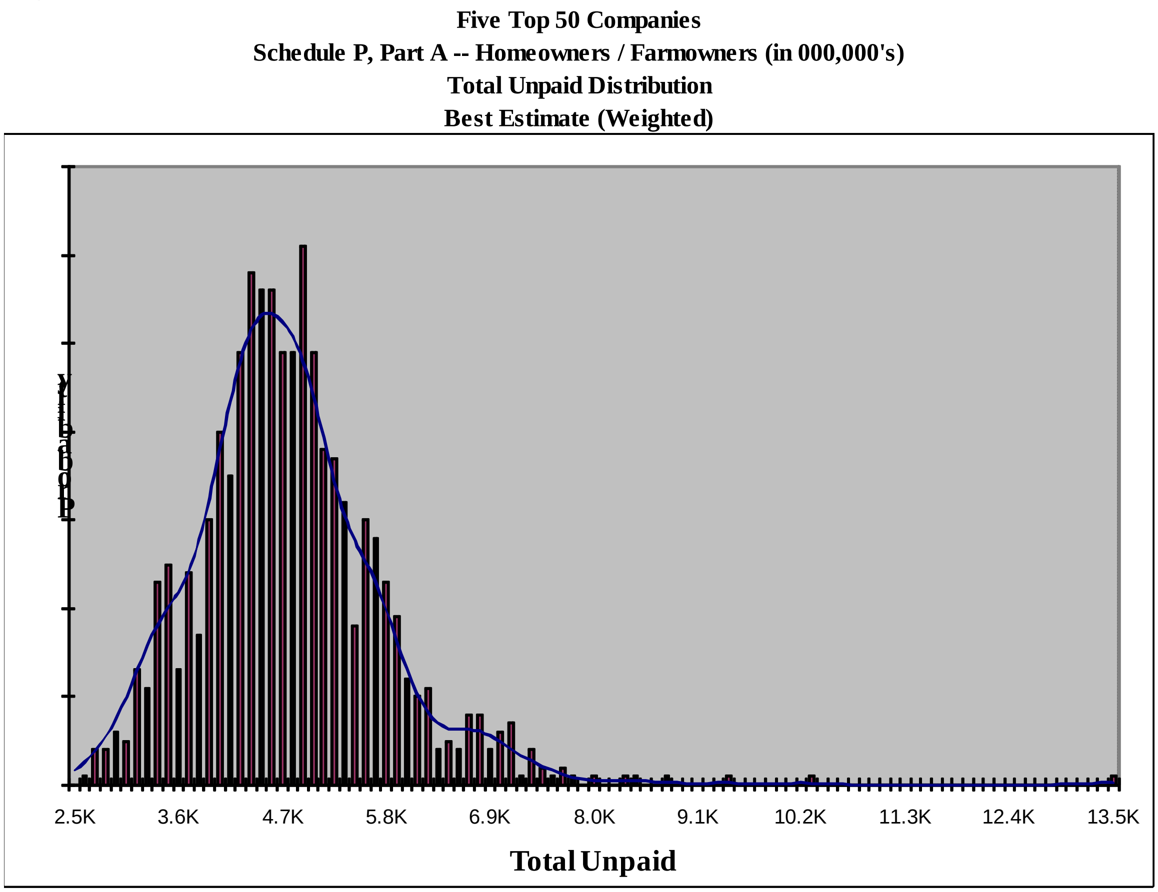
The final model output to consider is a histogram of the estimated-unpaid amounts for the total of all accident years combined, as shown in the graph in Figure 6.6. This total-unpaid-distribution histogram was created by dividing the range of all values generated from the simulation into 100

[33] If we are only interested in the "remaining" volatility in the loss ratio, then the values in the estimated-unpaid table (Figure 6.3) can be added to the cumulative paid values by year and divided by the premiums.

buckets of equal size and then counting the number of simulations that fall within each bucket. Dividing the number of simulations in each bucket by the total number of simulations (1,000 in this case) enables us to arrive at the frequency or probability for each bucket or bar in the graph.

Because the simulation results typically appear jagged, as they do in Figure 6.6, a Kernel density function (the blue line) is also used to calculate a smoothed distribution fit to the histogram values.[34] A Kernel density function may be conceptualized as a weighted average of values close to each point in the jagged distribution, with systematically less weight being given to values furthest from the points evaluated.[35]

Figure 6.6 Total Unpaid Claims Distribution



Another useful strategy for graphing the total unpaid distribution may be accomplished by creating a summary of the six model distributions used to determine the weighted "best estimate"

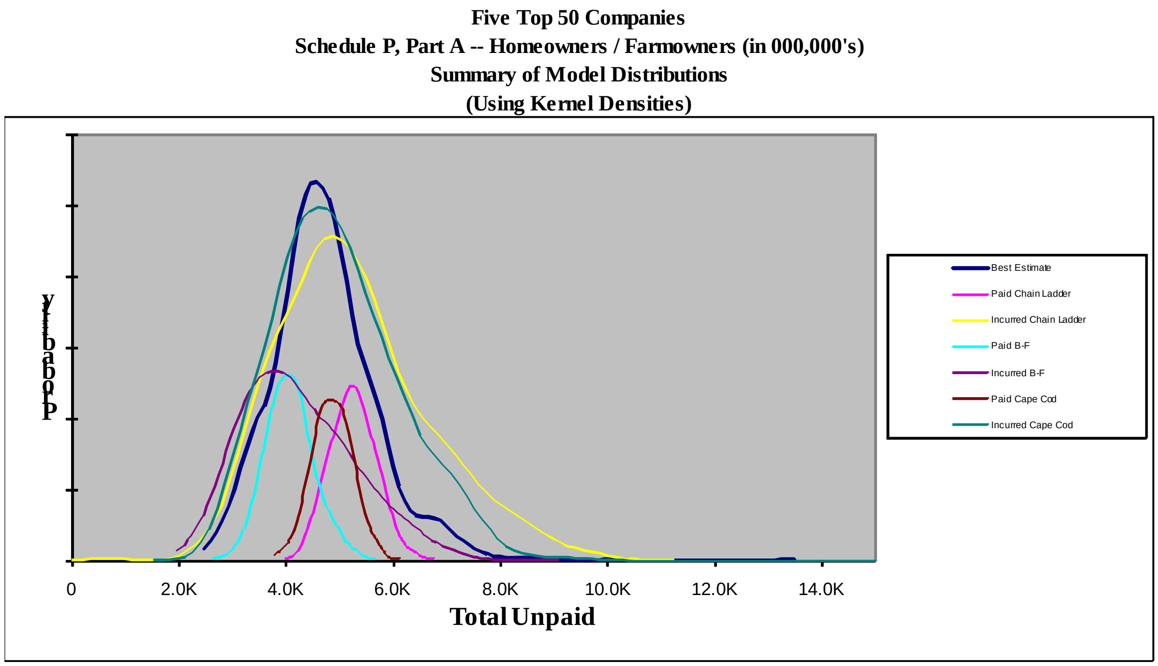
---

[34] Essentially, a Kernel density function will estimate each point in the distribution by weighting all of the values near that point, with less weight given the further the other points are from each respective point.

[35] For a more detailed discussion of Kernel density functions, see Wand & Jones, *Kernel Smoothing*, Chapman & Hall, 1995.

and distribution. An example of this graph using the kernel density functions is shown in Figure 6.7.

Figure 6.7 Summary of model distributions



The corresponding tables and graphs for the Part B and Part C results are shown in Appendices A and B, respectively.

6.5 Correlation

Results for an entire business unit can be estimated, after each business segment has been analyzed and weighted into best estimates, using aggregation. This represents another area where caution is warranted. The procedure is not a simple matter of "adding up" the distributions for each segment. In order to estimate the distribution of possible outcomes for the company as a whole a process that incorporates the correlation of results among segments must be used.[36]

Simulating correlated variables is commonly accomplished with a multivariate distribution whose parameters and correlations have been previously specified. This type of simulation is most easily applied when distributions are uniformly identical and known in advance (for example, all derived from a multivariate normal distribution). Unfortunately, these conditions do not exist for the ODP bootstrap model, a process that does not allow us to know the characteristics of distributions in

[36] This section assumed the reader is familiar with correlation.

advance. If their shapes turn out, indeed, to be different, then another approach will be needed.

Two useful correlation processes for the bootstrap model are location mapping and re-sorting.[37]

With location mapping, each iteration will include sampling residuals for the first segment and then going back to note the location in the original residual triangle of each sampled residual.[38] Each of the other segments is sampled using the residuals at the same locations for their respective residual triangles. Thus, the correlation of the original residuals is preserved in the sampling process.

The location-mapping process is easily implemented in Excel and does not require the need to estimate a correlation matrix. There are, however, two drawbacks to this process. First, it requires all of the business segments to come with data triangles that are precisely the same size with no missing values or outliers when comparing each location of the residuals.[39] Second, the correlation of the original residuals is used in the model, and no other correlation assumptions can be used for stress testing the aggregate results.

The second correlation process, re-sorting, can be accomplished with algorithms such as Iman-Conover or Copulas, among others. The primary advantages of re-sorting include:

- The triangles for each segment may have different shapes and sizes
- Different correlation assumptions may be employed
- Different correlation algorithms may also have other beneficial impacts on the aggregate distribution

For example, using a *t*-distribution Copula with low degrees of freedom rather than a normal-distribution Copula, will effectively "strengthen" the focus of the correlation in the tail of the distribution. This type of consideration is important for risk-based capital and other risk modeling issues.

To induce correlation among different segments in the bootstrap model, a calculation of the correlation matrix using Spearman's Rank Order and use of re-sorting based on the ranks of the total unpaid claims for all accident years combined may be done. The calculated correlations for Parts A, B, and C based on the paid residuals after hetero adjustments may be seen in the table in

---

[37] For a useful reference see Kirschner, et al. [15].

[38] For example, in the "Bootstrap Models.xls" file the locations of the sampled residuals are shown in Step 15, which could be replicated iteration by iteration for each business segment.

[39] It is possible to fill in "missing" residuals in another segment using a randomly selected residual from elsewhere in the triangle, but in order to maintain the same amount of correlation the selection of the other residual would need to account for the correlation between the residuals, which complicates the process.

Figure 6.8.

Figure 6.8 Estimated Correlation and P-values

Rank Correlation of Residuals after Hetero Adjustment - Paid

LOB	1	2	3
1	1.00	0.52	0.23
2	0.52	1.00	0.25
3	0.23	0.25	1.00

P-Values of Rank Correlation of Residuals after Hetero Adjustment - Paid

LOB	1	2	3
1	0.00	0.00	0.09
2	0.00	0.00	0.08
3	0.09	0.08	0.00

Using these correlation coefficients, the "Aggregate Estimate.xls" file, and the simulation data for Parts A, B, and C, we can then calculate the aggregate results for the three lines of business that are summarized in the table in Figure 6.9. A more complete set of tables for the aggregate results is shown in Appendix C.

Figure 6.9 Aggregate estimated unpaid

Five Top 50 Companies
Aggregate All Lines of Business
Accident Year Unpaid

Accident Year	Paid To Date	Mean Unpaid	Standard Error	Coefficient of Variation	Minimum	Maximum	50.0% Percentile	75.0% Percentile	95.0% Percentile	99.0% Percentile
1999	18,613	-	-		-	-	-	-	-	-
2000	20,618	31	12	37.5%	(21)	117	30	34	50	77
2001	22,866	115	40	34.8%	12	354	108	137	189	234
2002	22,842	211	43	20.3%	107	419	205	234	293	333
2003	22,351	387	52	13.4%	221	660	381	415	478	547
2004	22,422	741	86	11.5%	439	1,097	735	791	894	981
2005	24,350	1,514	150	9.9%	874	2,062	1,507	1,600	1,788	1,911
2006	19,973	2,958	264	8.9%	1,944	4,153	2,945	3,087	3,427	3,753
2007	18,919	5,533	475	8.6%	3,623	7,612	5,506	5,778	6,356	6,845
2008	15,961	12,565	1,195	9.5%	8,649	20,314	12,526	13,230	14,542	15,970
Totals	208,915	24,056	1,324	5.5%	19,572	32,759	24,008	24,852	26,193	27,644

Note that using residuals to correlate the lines of business, as in the location mapping method, and measuring the correlation between residuals, as in the re-sorting method, are both liable to create correlations that are close to zero. For reserve risk, the correlation that is desired is between the total unpaid amounts for two segments. The correlation that is being measured is the correlation between each incremental future loss amount, given the underlying model describing the overall trends in the data. This may or may not be a reasonable approximation.

Correlation is often thought of as being much stronger than "close to zero." For pricing risk, the correlation that is desired is between the loss ratio movements by accident year between two segments. This correlation is not as likely to be close to zero, so correlation of loss ratios (e.g., for the data in Figure 6.5) is often done with a different correlation assumption compared to reserving risk.

7. MODEL TESTING

Work on testing stochastic unpaid claim estimation models is still in its infancy. Most papers on stochastic models display results, and some even compare a few different models, but they tend to be void of any statistical evidence regarding how well the model in question predicts the underlying distribution. This is quite understandable since we don't know what the underlying distribution is, so with real data the best we can hope for is to retrospectively test a very old data set to see how well a model predicted the actual outcome.[40]

Testing a few old data sets is better than not, but ideally we would need many similar data sets to perform meaningful tests. One recent paper authored by the General Insurance Reserving Oversight Committee (GI ROC) in their papers for the General Insurance Research Organizing (GIRO) conference in 2007 titled "Best Estimates and Reserving Uncertainty" [28] and their updated in 2008 titled "Reserving Uncertainty" [29] took a first step in performing more meaningful statistical testing of a variety of models.

A large number of models were reviewed and tested in these studies, but one of the most interesting portions of the studies were done by comparing the unpaid liability distributions created by the Mack and ODP bootstrap model against the "true" artificially generated unpaid loss percentiles. To accomplish these tests, artificial datasets were constructed so that all of the Mack and ODP bootstrap assumptions, respectively, are satisfied. While the artificial datasets were recognized as not necessarily realistic, the "true" results are known so the Working Parties were able to test to see how well each model performed against datasets that could be considered "perfect".

7.1 Mack model results

To test the Mack model, incremental losses were simulated for a 10 x 10 square of data based on the assumptions of the Mack model. For the 30,000 datasets simulated, the upper triangles were used and the Mack model was applied to estimate the expected results and various percentiles. The actual results (lower triangle) for each iteration were then compared to the Mack estimates to see

[40] For example, data for accident years 1990 to 2000 could be completely settled and all results known as of 2010. Thus, we could use the triangle as it existed at year end 2000 to test how well a model predicts the final results.

how often they exceeded each tested percentile. If the model is working well, then the actual results should exceed the estimated percentiles one minus the percentile percent of the time – e.g., for the 90th percentile, the actual results should exceed the estimated 10% of the time.

In the test, the proportion of simulated scenarios in which the "true" outcome exceeded the 99th percentile of the Mack method's results was around 8-13%. If the Mack method's distribution was accurate, this should be 1%. However, it appears that the distribution created by the Mack method underestimates tail events.

7.2 Bootstrap model results

To test the ODP bootstrap model, incremental losses were simulated for a 10 x 10 square of data based on the assumptions of the ODP bootstrap model. For the 30,000 datasets simulated, the upper triangles were used and the OPD bootstrap model from England and Verrall [9 and 10] were used to estimate the expected results and various percentiles. Similarly, the proportion of simulated scenarios in which the "true" outcome exceeded the 99th percentile of the Bootstrap method's results was around 2.6-3.1%.

Thus, the bootstrap model performed better than the Mack model for "perfect" data, even though the results for both models were somewhat deficient in the sense that they both seem to underpredict the extremes of the "true" distribution. In fairness, it should be noted however, that the ODP bootstrap model that was tested did not include many of the "advancements" described in section 3.2.

7.3 Future testing

The testing done for GIRO was a significant improvement over simply looking at results for different models, without knowing anything about the "true" underlying distribution. The next step in the testing process will be to test models against "true" results for realistic data instead of "perfect" data. The CAS Loss Simulation Model Working Party is testing a model that will create datasets from the claim transaction level up. The goal is to create thousands of datasets based on characteristics of real data that can be used for testing various models.

8. FUTURE RESEARCH

With testing of stochastic models in its infancy, much work in the area of future research is needed. We only offer a few such areas.

- Expand testing of the ODP bootstrap model with realistic data using the CAS loss simulation model.
- Expand the ODP bootstrap model in other ways, for example use of the Munich chain ladder with an incurred/paid set of triangles, or the use of claim counts and average severities.
- Research other risk analysis measures and how the ODP bootstrap model can be used for enterprise risk management.
- Research how the ODP bootstrap model can be used for Solvency II requirements in Europe and the International Accounting Standards.
- Research into the most difficult parameter to estimate: the correlation matrix.

9. CONCLUSIONS

With this paper we endeavored to show how the ODP bootstrap model can be used in a variety of practical ways, and to illustrate the diagnostic tools the actuary needs to assess whether the model is working well. By doing so, we believe that this toolset can become an integral part of the actuaries regular estimation of unpaid claim liabilities, rather than just a "black box" to be used only if necessary.

Acknowledgment

The authors gratefully acknowledge the many authors listed in the References (and others not listed) that contributed to the foundation of the ODP bootstrap model, without which our research would not have been possible.

Supplementary Material

There are several companion files designed to give the reader a deeper understanding of the concepts discussed in the paper. The files are all in the "Beyond the Basics.zip" file. The files are:

Model Instructions.doc – this file contains a written description of how to use the primary bootstrap modeling files.

<u>Primary bootstrap modeling files</u>:

Industry Data.xls – this file contains Schedule P data by line of business for the entire U.S. industry and five of the top 50 companies, for each LOB that has 10 years of data.

Bootstrap Model.xls – this file contains the detailed model steps described in this paper as well as various modeling options and diagnostic tests. Data can be entered and simulations run and saved for use in calculating a weighted best estimate.

Best Estimate.xls – this file can be used to weight the results from six different models to get a "best estimate" of the distribution of possible outcomes.

Aggregate Estimate.xls – this file can be used correlate the best estimate results from 3 LOBs/segments.

Correlation Ranks.xls – this file contains the ranks used to correlate results by LOB/segment.

<u>Simple example calculation files</u>:

Simple GLM.xls – this file illustrates the calculation of the GLM framework for a simple 3 x 3 triangle.

Simple GLM 6.xls – this file illustrates the calculation of the GLM framework for a simple 6 x 6 triangle.

Simple GLM 6 with Outlier.xls – this file illustrates how the calculation of the GLM framework for a simple 6 x 6 triangle is adjusted for an outlier.

Simple GLM 6 with 3yr avg.xls – this file illustrates how the calculation of the GLM framework for a simple 6 x 6 triangle is adjusted to only use the equivalent of a three-year average (i.e., the last four diagonals).

Simple GLM 6 with 1 Acc Yr Parameter.xls – this file illustrates the calculation of the GLM framework using only one accident year (level) parameter, a development year trend parameter for every year and no calendar year trend parameter for a simple 6 x 6 triangle.

Simple GLM 6 with 1 Dev Yr Parameter.xls – this file illustrates the calculation of the GLM framework using only one development year trend parameter, an accident year (level) parameter for every year and no calendar year trend parameter for a simple 6 x 6 triangle.

Simple GLM 6 with 1 Acc Yr & 1 Dev Yr Parameter.xls – this file illustrates the calculation of the GLM framework using only one accident year (level) parameter, one development year trend parameter and no calendar year trend parameter for a simple 6 x 6 triangle.

Simple GLM 6 with 1 Acc Yr 1 Dev Yr & 1 Cal Yr Parameter.xls – this file illustrates the calculation of the GLM framework using only one accident year (level) parameter, one development year trend parameter and one calendar year trend parameter for a simple 6 x 6 triangle.

Appendix A – Schedule P, Part B Results

In this appendix the results for Schedule P, Part B (Private Passenger Auto Liability) are shown.

Figure A.1 Estimated-unpaid model results (best estimate)

Five Top 50 Companies
Schedule P, Part B -- Private Passenger Auto Liability (in 000,000's)
Accident Year Unpaid
Best Estimate (Weighted)

Accident Year	Paid To Date	Mean Unpaid	Standard Error	Coefficient of Variation	Minimum	Maximum	50.0% Percentile	75.0% Percentile	95.0% Percentile	99.0% Percentile
1999	11,816	-	-		-	-	-	-	-	-
2000	12,679	27	4	14.8%	14	41	27	29	33	37
2001	13,631	66	8	12.4%	35	96	66	70	81	88
2002	14,472	142	21	14.9%	73	225	141	153	178	201
2003	13,717	270	32	11.7%	146	390	269	286	324	361
2004	13,090	525	68	12.9%	277	767	526	559	641	709
2005	12,490	1,048	127	12.2%	553	1,503	1,048	1,100	1,278	1,387
2006	11,598	2,148	222	10.4%	1,124	3,066	2,150	2,249	2,511	2,865
2007	10,306	3,960	383	9.7%	2,115	5,421	3,962	4,103	4,611	5,158
2008	6,357	8,195	778	9.5%	4,554	11,486	8,174	8,549	9,434	10,682
Totals	120,157	16,380	898	5.5%	12,811	19,377	16,341	16,836	17,863	18,955
Normal Dist.		16,380	898	5.5%			16,380	16,986	17,857	18,469
logNormal Dist.		16,380	904	5.5%			16,355	16,975	17,909	18,595
Gamma Dist.		16,380	898	5.5%			16,364	16,976	17,884	18,541

Figure A.2 Estimated cash flow (best estimate)

Five Top 50 Companies
Schedule P, Part B -- Private Passenger Auto Liability (in 000,000's)
Calendar Year Unpaid
Best Estimate (Weighted)

Calendar Year	Mean Unpaid	Standard Error	Coefficient of Variation	Minimum	Maximum	50.0% Percentile	75.0% Percentile	95.0% Percentile	99.0% Percentile
2009	8,090	459	5.7%	6,153	9,603	8,081	8,335	8,843	9,338
2010	3,944	225	5.7%	3,127	4,787	3,935	4,079	4,311	4,560
2011	2,162	132	6.1%	1,586	2,672	2,165	2,239	2,376	2,514
2012	1,125	77	6.9%	864	1,450	1,124	1,171	1,252	1,326
2013	546	42	7.7%	404	697	545	571	617	672
2014	275	20	7.4%	205	371	274	288	310	327
2015	137	15	11.2%	97	192	137	146	162	179
2016	71	6	8.2%	50	93	71	74	80	86
2017	30	3	11.0%	15	41	31	32	35	38
2018	-	-		-	-	-	-	-	-
Totals	16,380	898	5.5%	12,811	19,377	16,341	16,836	17,863	18,955

Figure A.3 Estimated-loss-ratio (best estimate)

Five Top 50 Companies
Schedule P, Part B -- Private Passenger Auto Liability (in 000,000's)
Accident Year Ultimate Loss Ratios
Best Estimate (Weighted)

Accident Year	Mean Loss Ratio	Standard Error	Coefficient of Variation	Minimum	Maximum	50.0% Percentile	75.0% Percentile	95.0% Percentile	99.0% Percentile
1999	75.6%	9.2%	12.1%	38.2%	103.8%	75.6%	77.5%	92.8%	100.2%
2000	81.8%	9.7%	11.9%	43.8%	112.6%	81.9%	83.9%	99.7%	107.4%
2001	83.5%	9.6%	11.5%	47.6%	116.1%	83.4%	85.6%	101.0%	110.3%
2002	79.4%	9.0%	11.4%	45.3%	108.7%	79.4%	81.3%	97.0%	103.1%
2003	68.9%	7.2%	10.5%	39.3%	94.7%	68.7%	70.7%	82.7%	89.0%
2004	65.7%	7.4%	11.3%	36.3%	89.1%	65.5%	67.3%	80.1%	85.6%
2005	66.5%	7.6%	11.4%	36.2%	91.6%	66.3%	68.1%	80.9%	87.3%
2006	66.4%	5.8%	8.7%	33.6%	92.3%	66.4%	67.3%	76.6%	86.4%
2007	70.1%	6.2%	8.8%	40.4%	95.8%	69.9%	71.0%	81.1%	90.6%
2008	71.1%	6.4%	8.9%	41.5%	98.6%	71.2%	73.0%	81.3%	91.2%
Totals	72.3%	2.4%	3.3%	63.7%	80.6%	72.2%	73.8%	76.3%	78.0%

Figure A.4 Mean of incremental values

Five Top 50 Companies
Schedule P, Part B -- Private Passenger Auto Liability (in 000,000's)
Accident Year Incremental Values by Development Period
Best Estimate (Weighted)

Accident	Mean Values									
Year	12	24	36	48	60	72	84	96	108	120
1999	5,257	3,374	1,464	850	460	225	113	59	32	25
2000	5,625	3,613	1,566	908	490	240	121	61	34	27
2001	6,086	3,906	1,690	981	531	261	131	67	37	29
2002	6,489	4,168	1,805	1,044	567	279	140	71	39	31
2003	6,233	3,996	1,730	1,003	544	268	134	68	38	30
2004	6,073	3,894	1,689	978	528	261	131	67	37	29
2005	6,035	3,869	1,679	973	527	259	130	66	37	29
2006	6,050	3,882	1,685	1,032	571	271	138	66	40	30
2007	6,301	4,042	1,798	1,037	577	272	139	66	41	30
2008	6,361	4,202	1,811	1,048	581	276	140	66	41	30

Figure A.5 Standard deviation of incremental values

Five Top 50 Companies
Schedule P, Part B -- Private Passenger Auto Liability (in 000,000's)
Accident Year Incremental Values by Development Period
Best Estimate (Weighted)

Accident	Standard Error Values									
Year	12	24	36	48	60	72	84	96	108	120
1999	643	417	188	109	62	35	14	14	4	3
2000	677	437	195	115	66	36	15	14	5	4
2001	708	456	201	119	70	38	16	14	5	4
2002	745	481	215	127	72	40	16	16	5	4
2003	663	423	188	115	65	38	15	15	5	4
2004	691	441	201	122	67	39	15	17	5	4
2005	692	448	200	119	69	39	16	15	5	4
2006	530	348	156	112	70	36	14	14	5	4
2007	556	363	178	109	67	37	14	13	5	3
2008	571	406	179	109	66	37	14	14	5	3

Figure A.6 Total unpaid claims distribution

Five Top 50 Companies
Schedule P, Part B -- Private Passenger Auto Liability (in 000,000's)
Total Unpaid Distribution
Best Estimate (Weighted)

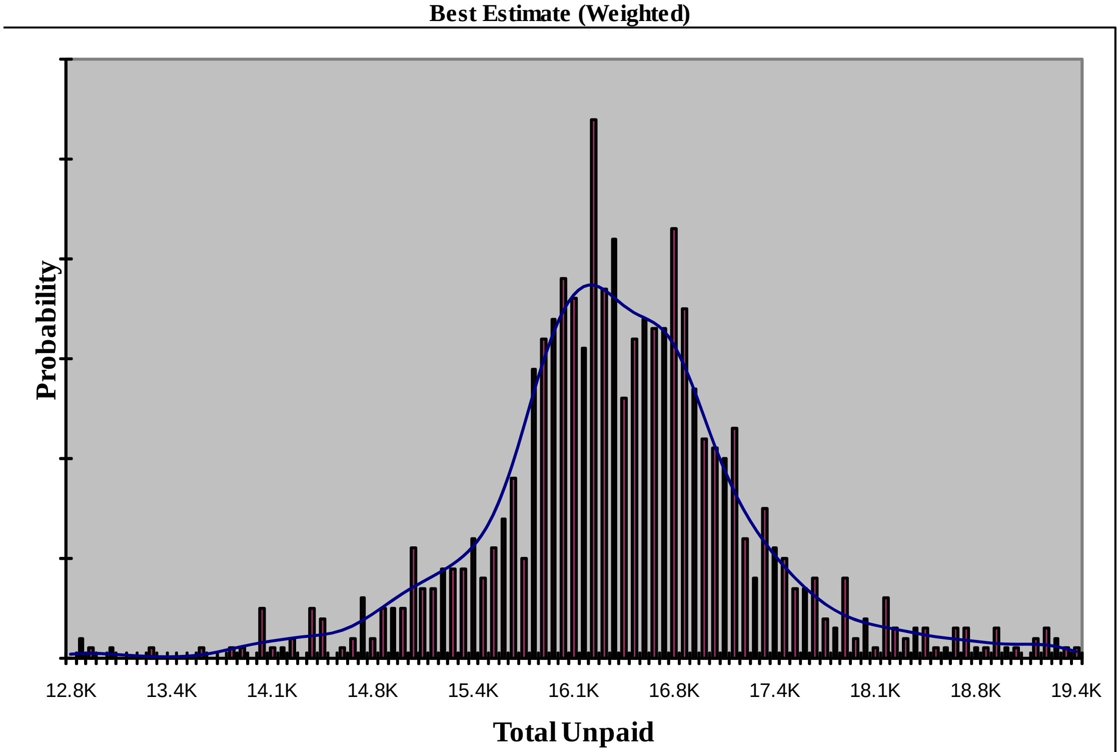
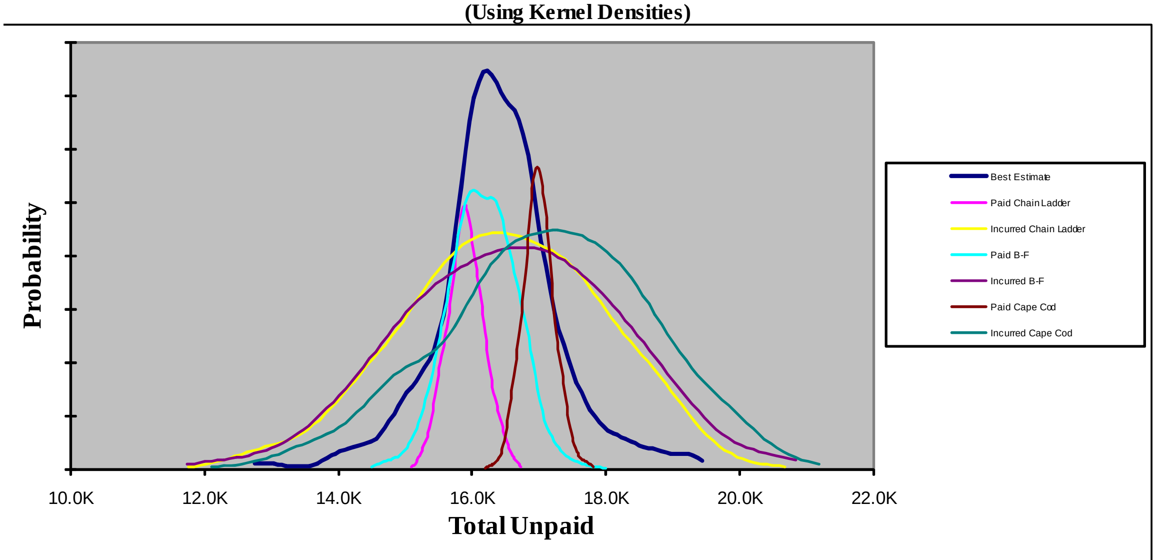


Figure A.7 Summary of model distributions

Five Top 50 Companies
Schedule P, Part B -- Private Passenger Auto Liability (in 000,000's)
Summary of Model Distributions
(Using Kernel Densities)



Appendix B – Schedule P, Part C Results

In this appendix the results for Schedule P, Part C (Commercial Auto Liability) are shown.

Figure B.1 Estimated-unpaid model results (best estimate)

Five Top 50 Companies
Schedule P, Part C -- Commercial Auto Liability (in 000,000's)
Accident Year Unpaid
Best Estimate (Weighted)

Accident Year	Paid To Date	Mean Unpaid	Standard Error	Coefficient of Variation	Minimum	Maximum	50.0% Percentile	75.0% Percentile	95.0% Percentile	99.0% Percentile
1999	1,563	-	-		-	-	-	-	-	-
2000	1,469	3	2	70.8%	(1)	12	2	4	6	8
2001	1,387	12	4	31.4%	3	31	12	14	19	22
2002	1,350	28	5	19.5%	12	53	28	31	37	41
2003	1,342	58	8	13.9%	33	84	59	64	71	79
2004	1,198	116	17	14.9%	61	191	115	127	146	158
2005	1,061	249	34	13.8%	151	334	250	272	304	322
2006	853	472	56	11.9%	323	628	479	516	553	577
2007	645	834	73	8.8%	605	1,015	844	891	937	965
2008	294	1,178	106	9.0%	904	1,484	1,181	1,262	1,337	1,366
Totals	11,162	2,950	149	5.0%	2,434	3,363	2,949	3,055	3,186	3,276
Normal Dist.		2,950	149	5.0%			2,950	3,050	3,194	3,295
logNormal Dist.		2,950	150	5.1%			2,946	3,048	3,202	3,314
Gamma Dist.		2,950	149	5.0%			2,947	3,048	3,198	3,306

Figure B.2 Estimated cash flow (best estimate)

Five Top 50 Companies
Schedule P, Part C -- Commercial Auto Liability (in 000,000's)
Calendar Year Unpaid
Best Estimate (Weighted)

Calendar Year	Mean Unpaid	Standard Error	Coefficient of Variation	Minimum	Maximum	50.0% Percentile	75.0% Percentile	95.0% Percentile	99.0% Percentile
2009	1,171	65	5.5%	974	1,374	1,172	1,214	1,280	1,321
2010	806	46	5.7%	657	960	806	838	882	911
2011	488	35	7.1%	364	595	490	512	544	571
2012	256	27	10.7%	174	343	255	274	303	324
2013	125	15	12.2%	73	177	124	136	150	160
2014	58	8	13.7%	35	90	57	63	71	76
2015	30	5	15.8%	17	47	30	33	39	42
2016	14	3	24.6%	4	25	13	16	19	22
2017	3	2	55.4%	(0)	12	3	4	6	7
2018	-	-		-	-	-	-	-	-
Totals	2,950	149	5.0%	2,434	3,363	2,949	3,055	3,186	3,276

Figure B.3 Estimated-loss-ratio (best estimate)

Five Top 50 Companies
Schedule P, Part C -- Commercial Auto Liability (in 000,000's)
Accident Year Ultimate Loss Ratios
Best Estimate (Weighted)

Accident Year	Mean Loss Ratio	Standard Error	Coefficient of Variation	Minimum	Maximum	50.0% Percentile	75.0% Percentile	95.0% Percentile	99.0% Percentile
1999	89.5%	3.2%	3.5%	80.2%	99.5%	89.5%	91.8%	94.5%	96.3%
2000	81.3%	2.8%	3.5%	72.9%	90.1%	81.2%	83.3%	86.0%	87.4%
2001	73.1%	2.6%	3.5%	63.8%	81.8%	73.2%	74.8%	77.4%	79.2%
2002	60.6%	2.1%	3.5%	53.7%	66.2%	60.6%	62.1%	64.1%	65.5%
2003	55.5%	1.9%	3.5%	48.9%	61.4%	55.5%	56.9%	58.7%	59.8%
2004	53.8%	2.1%	3.9%	47.5%	60.2%	53.8%	55.3%	57.3%	58.5%
2005	51.5%	2.2%	4.3%	43.1%	57.9%	51.6%	53.0%	55.1%	56.5%
2006	53.7%	2.9%	5.3%	43.5%	62.1%	53.9%	55.7%	58.1%	59.8%
2007	59.6%	3.6%	6.1%	46.9%	68.6%	59.9%	62.4%	65.0%	66.6%
2008	61.8%	4.6%	7.5%	49.4%	75.3%	62.1%	65.3%	68.7%	70.2%
Totals	62.5%	0.9%	1.5%	59.6%	65.3%	62.5%	63.1%	64.0%	64.6%

Figure B.4 Mean of incremental values

Five Top 50 Companies
Schedule P, Part C -- Commercial Auto Liability (in 000,000's)
Accident Year Incremental Values by Development Period
Best Estimate (Weighted)

Accident	Mean Values									
Year	12	24	36	48	60	72	84	96	108	120
1999	332	384	345	244	135	64	29	17	12	3
2000	312	360	326	229	127	61	27	15	11	3
2001	297	343	311	218	121	57	26	15	9	3
2002	292	339	305	214	118	56	25	16	10	3
2003	296	343	309	218	119	58	28	17	11	3
2004	276	322	288	203	111	62	26	15	10	3
2005	270	312	280	199	128	64	27	16	10	3
2006	265	308	278	224	128	65	27	16	10	3
2007	299	348	331	239	136	68	29	17	11	3
2008	294	370	320	231	132	67	27	16	11	3

Figure B.5 Standard deviation of incremental values

Five Top 50 Companies
Schedule P, Part C -- Commercial Auto Liability (in 000,000's)
Accident Year Incremental Values by Development Period
Best Estimate (Weighted)

Accident	Standard Error Values									
Year	12	24	36	48	60	72	84	96	108	120
1999	18	35	18	15	21	14	5	4	3	2
2000	17	34	18	15	19	13	5	4	3	2
2001	17	34	17	15	20	13	5	4	3	2
2002	16	32	17	14	19	14	5	4	3	1
2003	17	32	17	14	20	13	6	4	3	2
2004	16	32	16	14	19	14	6	3	3	1
2005	16	32	16	14	22	15	6	4	3	2
2006	16	31	16	25	22	15	6	4	3	1
2007	17	35	29	22	22	14	7	4	3	2
2008	17	45	27	21	21	13	6	4	3	2

Figure B.6 Total unpaid claims distribution

Five Top 50 Companies
Schedule P, Part C -- Commercial Auto Liability (in 000,000's)
Total Unpaid Distribution
Best Estimate (Weighted)

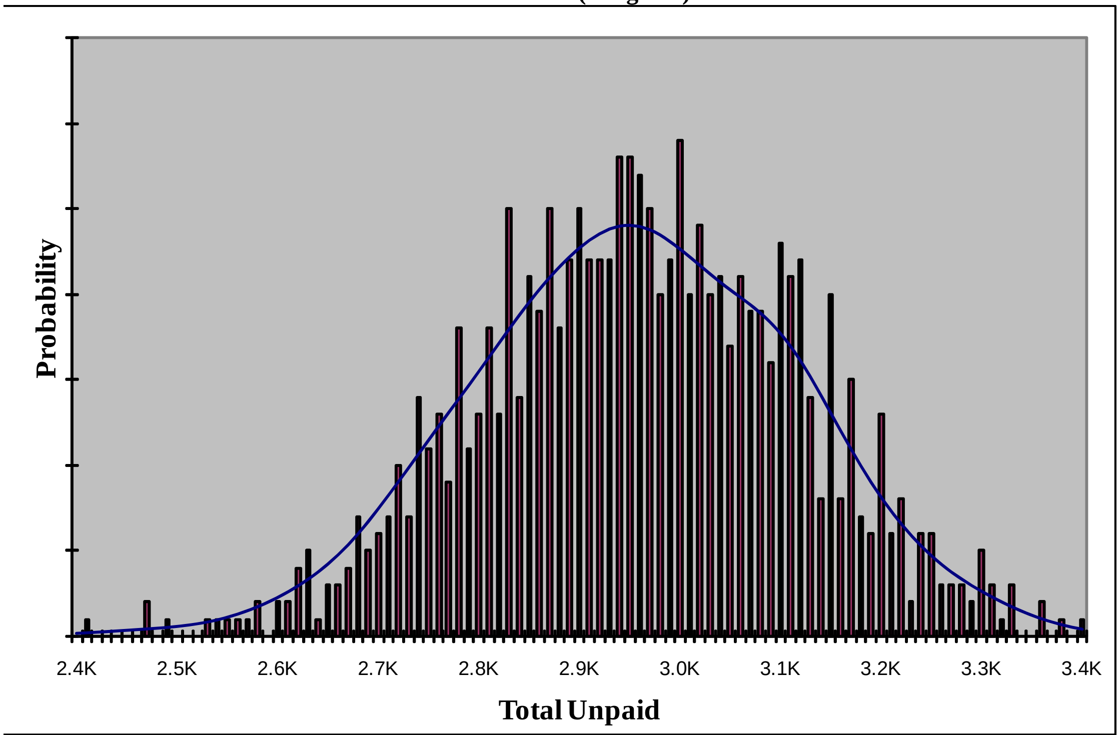
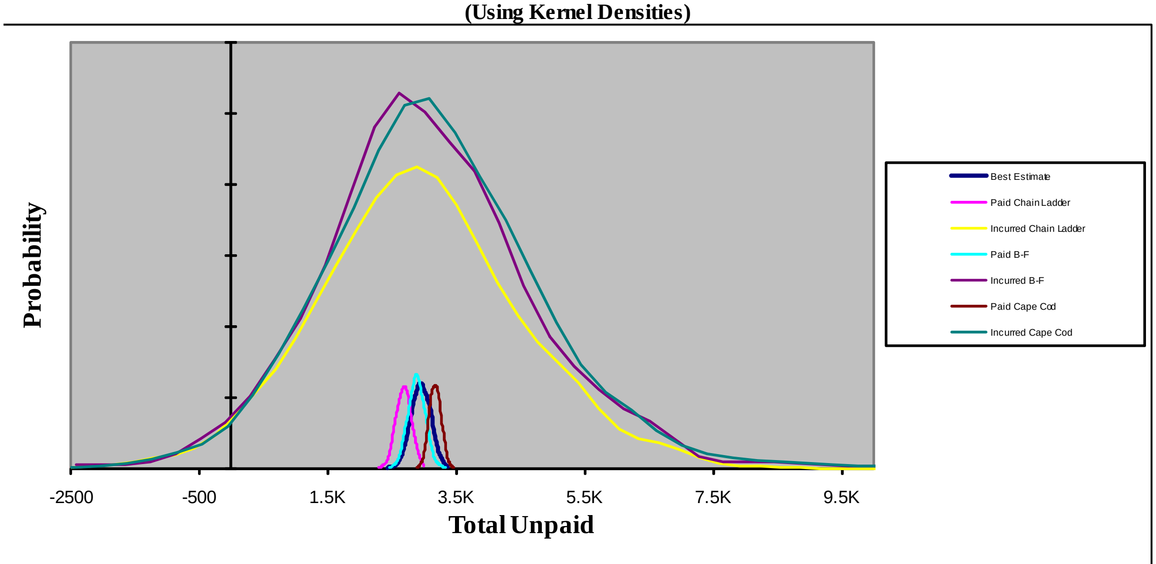


Figure B.7 Summary of model distributions

Five Top 50 Companies
Schedule P, Part C -- Commercial Auto Liability (in 000,000's)
Summary of Model Distributions
(Using Kernel Densities)



Appendix C – Aggregate Results

In this appendix the results for the correlated aggregate of the three Schedule P lines of business (Parts A, B, and C) are shown, using the correlation calculated from the paid data after adjustment for heteroscedasticity.

Figure A.1 Estimated-unpaid model results (best estimate)

Five Top 50 Companies
Aggregate All Lines of Business
Accident Year Unpaid

Accident Year	Paid To Date	Mean Unpaid	Standard Error	Coefficient of Variation	Minimum	Maximum	50.0% Percentile	75.0% Percentile	95.0% Percentile	99.0% Percentile
1999	18,613	-	-		-	-	-	-	-	-
2000	20,618	31	12	38.2%	(26)	111	30	35	49	78
2001	22,866	115	40	34.7%	5	348	106	136	190	240
2002	22,842	211	43	20.5%	109	397	205	234	295	348
2003	22,351	387	51	13.2%	230	624	381	416	482	532
2004	22,422	741	86	11.6%	432	1,080	732	788	883	1,003
2005	24,350	1,514	156	10.3%	876	2,079	1,506	1,601	1,779	1,908
2006	19,973	2,958	267	9.0%	1,771	3,970	2,942	3,092	3,428	3,704
2007	18,919	5,533	487	8.8%	3,472	7,657	5,525	5,770	6,402	6,981
2008	15,961	12,565	1,410	11.2%	7,894	21,492	12,527	13,260	14,919	16,794
Totals	208,915	24,056	1,644	6.8%	18,197	34,272	23,963	25,008	26,726	28,724
Normal Dist.		24,056	1,644	6.8%			24,056	25,164	26,760	27,880
logNormal Dist.		24,055	1,635	6.8%			24,000	25,124	26,835	28,105
Gamma Dist.		24,056	1,644	6.8%			24,018	25,143	26,822	28,044

Figure A.2 Estimated cash flow (best estimate)

Five Top 50 Companies
Aggregate All Lines of Business
Calendar Year Unpaid

Calendar Year	Mean Unpaid	Standard Error	Coefficient of Variation	Minimum	Maximum	50.0% Percentile	75.0% Percentile	95.0% Percentile	99.0% Percentile
2009	12,354	1,018	8.2%	9,070	19,805	12,305	12,931	13,951	15,231
2010	5,549	348	6.3%	4,293	7,187	5,535	5,768	6,123	6,499
2011	3,012	188	6.2%	2,349	3,740	3,012	3,129	3,329	3,491
2012	1,572	114	7.3%	1,262	2,009	1,563	1,641	1,769	1,865
2013	789	73	9.3%	583	1,117	785	830	913	1,019
2014	397	42	10.5%	260	552	395	420	470	512
2015	217	39	18.1%	133	505	211	234	289	351
2016	125	40	32.3%	(13)	396	116	142	200	266
2017	40	18	44.6%	(25)	208	36	42	71	98
2018	-	-		-	-	-	-	-	-
Totals	24,056	1,644	6.8%	18,197	34,272	23,963	25,008	26,726	28,724

Figure A.3 Estimated loss ratio (best estimate)

Five Top 50 Companies
Aggregate All Lines of Business
Accident Year Ultimate Loss Ratios

Accident Year	Mean Loss Ratio	Standard Error	Coefficient of Variation	Minimum	Maximum	50.0% Percentile	75.0% Percentile	95.0% Percentile	99.0% Percentile
1999	73.7%	9.5%	12.9%	39.8%	110.8%	73.4%	77.6%	91.2%	99.7%
2000	80.7%	9.6%	11.9%	50.0%	122.2%	80.5%	84.5%	97.4%	107.9%
2001	84.2%	10.1%	12.0%	53.4%	139.7%	84.2%	88.4%	101.0%	112.9%
2002	75.7%	9.0%	11.9%	42.2%	119.1%	75.6%	79.2%	92.6%	100.5%
2003	66.5%	7.8%	11.7%	40.7%	116.9%	66.2%	70.0%	80.3%	89.2%
2004	64.3%	7.4%	11.6%	37.4%	97.3%	64.0%	67.8%	78.4%	86.0%
2005	70.7%	8.4%	11.9%	41.1%	104.5%	70.2%	74.0%	86.5%	94.6%
2006	61.2%	6.9%	11.3%	38.4%	119.4%	60.8%	63.1%	74.3%	82.4%
2007	64.0%	7.5%	11.7%	40.4%	93.5%	63.9%	66.4%	78.7%	86.4%
2008	75.5%	9.8%	13.0%	46.7%	168.1%	74.8%	78.4%	92.6%	107.2%
Totals	70.8%	2.9%	4.0%	62.0%	88.1%	70.7%	72.5%	75.7%	78.0%

Figure A.4 Mean of incremental values

Five Top 50 Companies
Aggregate All Lines of Business
Accident Year Incremental Values by Development Period

Accident	Mean Values									
Year	12	24	36	48	60	72	84	96	108	120
1999	9,292	4,873	2,023	1,183	635	310	153	81	67	30
2000	10,524	5,354	2,154	1,254	666	326	162	84	75	31
2001	11,868	5,902	2,324	1,333	713	348	174	90	82	34
2002	11,809	6,018	2,398	1,382	740	363	181	94	82	35
2003	11,790	5,934	2,343	1,355	722	354	178	93	82	34
2004	12,243	5,985	2,313	1,327	703	356	175	92	85	33
2005	14,191	6,560	2,419	1,369	741	367	181	94	98	32
2006	11,951	5,878	2,290	1,418	775	366	186	92	81	39
2007	12,654	6,215	2,525	1,440	793	371	188	93	83	39
2008	16,129	6,928	2,585	1,463	801	378	191	94	86	40

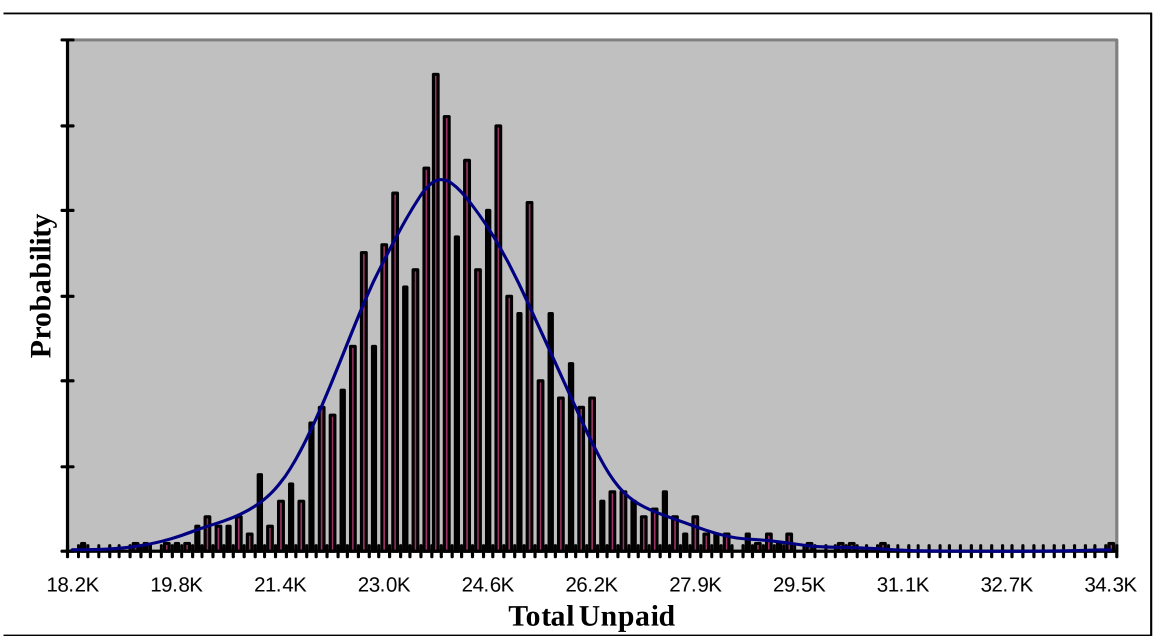
Figure A.5 Standard deviation of incremental values

Five Top 50 Companies
Aggregate All Lines of Business
Accident Year Incremental Values by Development Period

Accident	Standard Deviation Values									
Year	12	24	36	48	60	72	84	96	108	120
1999	1,503	607	216	123	72	39	16	15	27	7
2000	1,549	615	224	132	76	40	17	14	28	12
2001	1,755	681	242	137	81	43	18	15	37	14
2002	1,704	702	252	145	81	44	18	17	34	12
2003	1,714	657	230	135	76	42	17	16	35	15
2004	1,743	662	231	138	79	43	18	17	38	13
2005	1,998	762	253	148	83	44	19	16	47	16
2006	1,745	624	201	141	86	40	19	15	29	16
2007	1,850	678	246	138	85	42	18	14	32	17
2008	2,534	918	270	147	88	45	20	15	37	18

Figure A.6 Total unpaid claims distribution

Five Top 50 Companies
Aggregate All Lines of Business
Total Unpaid Distribution



REFERENCES

[1] Barnett, Glen and Ben Zehnwirth. 2000. Best Estimates for Reserves. *PCAS* LXXXVII, 2: 245-321.

[2] Berquist, James R., and Richard E. Sherman. 1977. Loss Reserve Adequacy Testing: A Comprehensive, Systematic Approach. *PCAS* LXIV: 123-184.

[3] Björkwall, Susanna. 2009. Bootstrapping for Claims Reserve Uncertainty in General Insurance. Mathematical Statistics, Stockholm University. Research Report 2009:3, Licenciate thesis. http://www2.math.su.se/matstat/reports/seriea/2009/rep3/report.pdf.

[4] Björkwall, Susanna, Ola Hössjer and Esbjörn Ohlsson. 2009. Non-parametric and Parametric Bootstrap Techniques for Age-to-Age Development Factor Methods in Stochastic Claims Reserving. *Scandinavian Actuarial Journal*, 4: 306-331.

[5] Bornhuetter, Ronald and Ronald Ferguson. 1972. The Actuary and IBNR. *PCAS* LIX: 181-195.

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

[7] Christofides, S. 1990. Regression models based on log-incremental payments. *Claims Reserving Manual*, vol. 2. Institute of Actuaries, London.

[8] Efron, Bradley. 1979. Bootstrap Methods: Another Look at the Jackknife. *The Anals of Statistics*, 7-1: 1-26.

[9] England, Peter D. and Richard J. Verrall. 1999. Analytic and Bootstrap Estimates of Prediction Errors in Claims Reserving. *Insurance: Mathematics and Economics*, 25: 281-293.

[10] England, Peter D. and Richard J. Verrall. 2002. Stochastic Claims Reserving in General Insurance. *British Actuarial Journal*, 8-3: 443-544.

[11] Freedman, D.A. 1981. Bootstrapping Regression Models. *The Anals of Statistics*, 9-6: 1218-1228.

[12] *Foundations of Casualty Actuarial Science*, 4th ed. 2001. Arlington, Va.: Casualty Actuarial Society.

[13] Hayne, Roger M. 2008. A Stochastic Framework for Incremental Average Reserve Models. *CAS E-Forum* (Fall): 174-195.

[14] International Actuarial Association. 2010. Stochastic Modeling – Theory and Reality from an Actuarial Perspective. Available from www.actuaries.org/stochastic.

[15] Kirschner, Gerald S., Colin Kerley, and Belinda Isaacs. 2008. Two Approaches to Calculating Correlated Reserve Indications Across Multiple Lines of Business. *Variance* (Spring), 1-2: 15-38.

[16] Kremer, E. 1982. IBNR claims and the two way model of ANOVA, *Scandinavian Actuarial Journal*: 47-55.

[17] Mack, Thomas. 1993. Distribution Free Calculation of the Standard Error of Chain Ladder Reserve Estimates. *ASTIN Bulletin*, 23-2: 213-225.

[18] Mack, Thomas. 1999. The Standard Error of Chain Ladder Reserve Estimates: Recursive Calculation and Inclusion of a Tail Factor. *ASTIN Bulletin*, 29-2: 361-366.

[19] Mack, Thomas and Gary Venter. 2000. A Comparison of Stochastic Models that Reproduce Chain Ladder Reserve Estimates. *Insurance: Mathematics and Economics*, 26: 101-107.

[20] McCullagh, P. and J. Nelder. 1989. *Generalized Linear Models*, 2nd ed. Chapman and Hall.

[21] Merz, Michael and Mario V. Wüthrich. 2008. Modeling the Claims Development Result For Solvency Purposes. *Casualty Actuarial Society E-Forum, Fall*: 542-568.

[22] Milliman. 2009. Using the Milliman Reserve Variability Model. Version 1.4.

[23] Moulton, Lawrence H. and Scott L. Zeger. 1991. Bootstrapping Generalized Linear Models. *Computational Statistics and Data Analysis* 11, 53–63.

[24] Murphy, Daniel M. 1994. Unbiased Loss Development Factors. *PCAS* LXXXI: 154-222.

[25] Pinheiro, Paulo J. R., João Manuel Andrade e Silva and Maria de Lourdes Centeno. 2001. Bootstrap Methodology in Claim Reserving. *ASTIN Colloquium*: 1-13.

[26] Pinheiro, Paulo J. R., João Manuel Andrade e Silva and Maria de Lourdes Centeno. 2003. Bootstrap Methodology in Claim Reserving. *The Journal of Risk and Insurance*, 70: 701-714.

[27] Quarg, Gerhard and Thomas Mack. 2008. Munich Chain Ladder: A Reserving Method that Reduces the Gap between IBNR Projections Based on Paid Losses and IBNR Projections Based on Incurred Losses. *Variance* (Fall), 2-2: 266-299.

[28] ROC/GIRO Working Party. 2007. Best Estimates and Reserving Uncertainty. Institute of Actuaries.

[29] ROC/GIRO Working Party. 2008. Reserving Uncertainty. Institute of Actuaries.

[30] Renshaw, A.E., 1989. Chain ladder and interactive modelling (claims reserving and GLIM). *Journal of the Institute of*

Actuaries 116 (III), 559–587.
[31] Renshaw, A.E. and R.J. Verrall. 1994. A stochastic model underlying the chain ladder technique. Proceedings XXV *ASTIN Colloquium*, Cannes.
[32] Ruhm, David L. and Donald F. Mango. 2003. A Method of Implementing Myers-Read Capital Allocation in Simulation. *CAS Forum* (Fall), 451-458.
[33] Shapland, Mark R. 2007. Loss Reserve Estimates: A Statistical Approach for Determining "Reasonableness". *Variance* (Spring), 1-1: 120-148.
[34] Struzzieri, Paul J. and Paul R. Hussian. 1998. Using Best Practices to Determine a Best Reserve Estimate. *CAS Forum* (Fall): 353-413.
[35] Venter, Gary G. 1998. Testing the Assumptions of Age-to-Age Factors. *PCAS* LXXXV: 807-47.
[36] Venter, Gary G. 2003. A Survey of Capital Allocation Methods with Commentary Topic 3: Risk Control. *ASTIN Colloquium*,
[37] Verrall, Richard J. 1991. On the estimation of reserves from loglinear models. *Insurance: Mathematics and Economics* 10, 75–80.
[38] Verrall, Richard J. 2004. A Bayesian Generalized Linear Model for the Bornhuetter-Ferguson Method of Claims Reserving. *North American Actuarial Journal*, 8-3: 67-89.
[39] Zehnwirth, Ben, 1989. The Chain Ladder Technique — A Stochastic Model. *Claims Reserving Manual*, vol. 2. Institute of Actuaries, London.
[40] Zehnwirth, Ben. 1994. Probabilistic Development Factor Models with Applications to Loss Reserve Variability, Prediction Intervals and Risk Based Capital. *CAS Forum* (Spring), 2: 447-606.

Abbreviations and notations

Collect here in alphabetical order all abbreviations and notations used in the paper

AIC: Akaike Information Criteria
APD: Automobile Physical Damage
BIC: Bayesian Information Criteria
BF: Bornhuetter-Ferguson
CC: Cape Cod
CL: Chain Ladder
CoV: Coefficient of Variation
DFA, Dynamic Financial Analysis
ELR: Expected Loss Ratio
GLM: Generalized Linear Models
ERM, Enterprise Risk Management
MLE: Maximum Likelihood Estimate
ODP: Over-Dispersed Poisson
OLS: Ordinary Least Squares
RSS: Residual Sum Squared
SSE: Sum of Squared Errors

Biographies of the Authors

Mark R. Shapland is Consulting Actuary in Milliman's Atlanta office where he is responsible for various stochastic reserving projects, including modeling of asbestos liabilities, and is a key member of the Property & Casualty Insurance Software (PCIS) development team. He has a B.S. degree in Actuarial Science from the University of Nebraska-Lincoln. He is a Fellow of the Casualty Actuarial Society, an Associate of the Society of Actuaries and a Member of the American Academy of Actuaries. He was the leader of Section 3 of the Reserve Variability Working Party and is currently the Chair of the CAS Committee on Reserves, co-chair of the Tail Factor Working Party, and co-chair of the Loss Simulation Model Working Party. He is also a co-presenter of the CAS Reserve Variability Limited Attendance Seminar and has spoken frequently on this subject both within the CAS and internationally. He can be contacted at mark.shapland@milliman.com.

Jessica (Weng Kah) Leong is a Consulting Actuary in the New York office of Milliman. In this role, she has helped clients develop reserve distributions for the purposes of market value financial reporting, capital adequacy, capital allocation, portfolio transfer and enterprise risk management modeling. Jessica holds a Bachelor of Commerce degree from the University of Melbourne, Australia. She is a Fellow of the Institute of Actuaries of Australia, a Fellow of the Casualty Actuarial Society, a Member of the American Academy of Actuaries and an Associate of the Institute of Actuaries (UK). Jessica is serving on the Casualty Actuarial Society's Committee on Reserves, International Committee and the Casualty Loss Reserving Seminar Planning Committee, as well as the Enterprise Risk Management Capability Sub-committee of the Institute of Actuaries of Australia. She can be contacted at jessica.leong@milliman.com.