The Analysis and Estimation of Loss & ALAE Variability: A Summary Report

CAS Working Party on Quantifying Variability in Reserve Estimates

Roger Hayne, co-chair
James Leise, co-chair
John T. Bonsignore
Yisheng Bu
Sandie Cagley
David R. Clark
James Christopher Guszcza
C.K. "Stan" Khury
Ravi Kumar
Atul Malhotra
Joseph O. Marker
Gary V. Nickerson
Bruce E. Ollodart
Dianne M. Phelps
Ralph Stephen Pulis
David L. Ruhm
David E. Sanders
Mark R. Shapland
Julie Sims
Greg Taylor
Gary G. Venter
Micah Grant Woolstenhulme

Review by Rodney Kreps

Abstract

Motivation. Casualty Actuaries have long been interested in the estimation of ultimate losses and ALAE. The potential variability of the ultimate outcome is critical to understanding the extent of the risks faced by the risk-bearing entity that either has adopted or is contemplating the adoption of loss and ALAE estimates. Over the years many people (actuaries and others) have made significant contributions to the literature and overall discussion of how to estimate the potential variability of ultimate losses, but there is no clear preferred method within the actuarial community. This research paper is an attempt to bring all of the historical research together in one cohesive document.

Method. The Working Party worked exclusively via e-mail and a private area of the CAS web site. After a joint effort to assemble an outline, the Working Party separated into subgroups, each assigned to prepare one of the sections of this paper.

Results. There are many approaches to estimating future payments for property and casualty liabilities, many of which have stochastic roots leading to not only an estimate of future payments but also of the distribution of those payments. However, we found no single method that is clearly superior. We have identified some areas of potential future research.

Conclusions. The actuarial profession does not yet have a single, all-inclusive method for estimating the distribution of future payments for property and casualty liabilities. Much work is yet to be done on the issue.

Availability. A copy of the Working Party's paper can be found on the CAS web site at http://www.casact.org/pubs/forum/05fforum/.

Keywords. Reserve Variability; Future Payment Variability; Generalized Linear Model; Delta Method; Over-Dispersed Poisson Model; Bootstrap; Bayesian Inference; Markov Chain Monte Carlo

1. INTRODUCTION

A risk bearing entity wishes to know its financial position on a particular date. In order to do this, among other items it must understand the future payments it will be liable to make for obligations existing at the date of the valuation. For an insurance situation, these future payments are not known with certainty at the time of the valuation.

The fundamental question that the risk bearing entity asks itself is:

Given any value (estimate of future payments) and our current state of knowledge, what is the probability that the final payments will be no larger than the given value?

The answer to this fundamental question can be provided by what is usually called the cumulative distribution function for the random variable of potential future payments. From this, one can easily determine the corresponding probability density function. We will call this probability density function "the distribution of future payments" at the valuation point. Although we might not always be successful, we try to maintain a distinction between future payments and "reserves". We try (though not always successfully) to use the term "reserves" for amounts booked in financial statements. We are focusing here on the total future payments and are not, at this time, considering issues of timing of those payments. Thus, our "distribution of future payments" should not be confused with issues relating to payout timing.

1.1 Research Context

It has long been recognized that traditional actuarial methods provide single point estimates of the amount of future payments. Those methods are generally deterministic and used alone do not provide any direct measure of how close one would expect that estimate to be to the final outcome, or even to the mean of possible final outcomes. Traditional actuarial reserve analyses recognize this shortcoming by applying a variety of different methods to derive multiple estimates of future payments. The range of such estimates is often used to give insight as to how "solid" the actuary's estimate selection is and may form the basis of the practitioner's own "range of reasonable estimates" for reserves. We note that such a range is often determined by considering the forecasts of a variety of deterministic "traditional" actuarial projection methods. Those methods usually only provide "estimates" of future payments without any additional statistical information. However, without such statistical quantification, we cannot determine how likely it is for the ultimate "realization" of future payments to be within that "range of reasonable estimates".

There has long been interest in translating the subjective "feel" for how "good" a liability estimate is to something more concrete, something that can be quantified by a probability distribution. This interest has led to more recent activity to cast the actuary's forecasting methods in a stochastic framework. A major benefit of such an approach is the existence of a specific statistical model and the possibility of estimating not only the expected value (statistical mean), but also the distribution of future payments or other summary statistics of the distribution.

Much work has been done, but in our view, the actuarial community does not yet have the answer to the fundamental question set out in Section 1 above. We believe that our community, and other users of actuarial forecasts, can well benefit from a work that summarizes the current state of knowledge of estimating the distribution of future payments.

1.2 Objective

It is the purpose of this paper to set forth the current state of knowledge regarding the estimation of this distribution. More specifically, the paper addresses the estimation of distributions to the extent that they can be quantified by models. There may be some loss liabilities that cannot be quantified by these models, including perhaps asbestosis liabilities and similar exposures, and these could considerably increase the uncertainty in the distribution beyond what would be calculated by the methods discussed.

From the outset, we draw sharp distinctions between the "distribution of future payments" as we have defined it here and other concepts such as "ranges of reasonable estimates" or the "appropriate" number to be used in a financial statement even if the full distribution of future payments is known with certainty. We believe that knowledge of the distribution is a prerequisite for any discussion of the variability of potential future payments, but is not sufficient to completely answer that question. In addition to the knowledge of that distribution, other factors come into play in the final "booking" of a liability number. Such factors include regulatory requirements, the view of the investment community, shareholder and policyholder considerations, to name just a few.

Though this paper is primarily aimed at the practicing actuary, a thorough understanding of the concepts we present will be necessary in order to appropriately interpret statements that attempt to quantify the uncertainty in estimates of incurred but yet unpaid losses. It is hoped that all audiences, including regulators, rating agencies, taxing authorities, shareholders, management, and actuaries, will benefit from a single vocabulary in describing and discussing uncertainty in estimates of future payments.

We note that the amount recorded on a financial statement as a provision for liabilities can be viewed against the landscape provided by the "distribution of future payments" as we have defined it. Given that distribution, and assuming that it is perfectly correct, it is an easy matter to see the likelihood that future payments will fall above or below the recorded amount and to calculate the expected (mean) financial consequence of any particular booked number. We can also find the probability that the actual or "realized" future payments will be in any given range of values with the distribution. In fact, percentiles of the distribution can be used to quantify ranges of reasonable estimates.

Armed with this tool, the practitioner can not only provide his or her "range of reasonable estimates", but he or she can quantify that range by saying, for example, that his or her range covers the area between the twenty-fifth and seventy-fifth percentiles of the distribution. By a p-percentile we mean the value such that there is a p percent probability of a lesser realization.

In addition it is easy to see how that amount compares to various statistics of the "distribution of future payments" such as the mean, mode, median, or other function of that distribution. It is not, however, the purpose of this paper to define the "appropriate" point along a distribution to be recorded in financial statements.

We stress the importance of this distinction between the distribution of future payments and the reserve number booked in a financial statement. The former provides a view of the range of possible outcomes and their likelihood (a landscape). Even if this distribution is completely known, it appears that current accounting guidance does not provide sufficient direction to arrive at a single "reserve" that should be booked.

Another distinction that we must make is between the distribution and a summary statistic of the distribution. Whereas a distribution describes a range of possible outcomes, a summary statistic is a particular value that conveys some information about the entire distribution. Examples of common summary statistics are the mean (average or expected value), mode (most likely value), and the median (the "middle" value or fiftieth percentile). In a situation where we are completely certain about a distribution then a well defined summary statistic such as the mean is completely known, even if the actual future outcome is at present unknown and unknowable. We note that accounting guidance for reserves seems to direct us to a point on the distribution (an estimate of the amount that will ultimately be paid) rather than to a particular summary statistic.

It thus appears necessary to rely on imprecise concepts such as "best estimate" or "range of reasonable estimates" when talking about an estimate of future payments in an accounting context. We note that the distribution of future payments has a specific statistical meaning and actually exists separately from the professional estimating that distribution, whereas the "range of reasonable estimates" is properly completely determined by the practitioner making the estimate and then only by the specific context (accounting definition) in which the reserves are being set. The "distribution of future payments" depends on neither the professional estimating it nor the methods used in that estimation. However, the methods used by the practitioner will affect his or her estimate of that distribution. In contrast, the "range of reasonable estimates" is completely determined by the practitioner and his or her methods and his or her interpretation of accounting guidance. The apparently vague accounting guidance as to the definition of "reserve" thus seems to make "reasonable" in this context subjective.

It appears that it is necessary to introduce the concept of "range of reasonable estimates" because the accounting guidance appears to require the booking of an estimate of future payments and because the actual amount of future payments is currently unknown. The "range of reasonable estimates" seems to be a surrogate for the more precise distribution of future payments often determined in reference to the projections or forecasts from a range of deterministic methods, and appears to be an attempt to communicate the dispersion of that distribution. The range itself remains subjective since "reasonable" itself is not defined and left up to the individual practitioner, though, as mentioned above, the practitioner can use percentiles in determining his or her range.

The discussion of how to incorporate the distribution of future payments into the final liability booked or into a "range of reasonable estimates" is probably not as advanced as the theory on calculating that distribution. Rather than risking the omission of a significant paper on the issue, and recognizing the ever-expanding scope of discussions of ranges and the amounts to be booked we do not provide specific references on this topic. We do note, however, that the entire CAS call for reserving papers in 1998 was on the subject of the actuary's best estimate of reserves. In addition, the 2003 call for reserving papers included the issue of range. The corresponding fall editions of the *Forum* contain the papers received as a result of these calls and can be the start of an interested reader's research on this topic.

This is not to say that the estimation of the distribution of future payments is a matter of science that can be done with precision. Much to the contrary, as we will discuss in this paper there is now no recognized way to estimate that distribution. All the known approaches have their strengths and weaknesses, but none completely assess all sources of uncertainty. It is quite possible that a complete solution to this problem is impossible, given the unknown and unknowable nature of insured liabilities. However, in discussing the uncertainty of future payments it is necessary that all parties know what various terms mean and how close to an ideal methodology a particular approach comes.

That is the primary purpose of this paper.

1.3 Sources of Uncertainty

Setting the objective as identifying the distribution of future payments allows us to specifically identify sources of uncertainty in those estimates. These sources of uncertainty should be kept in mind when evaluating any estimate of the distribution of future payments.

1.3.1 Process Uncertainty

In all but the most trivial estimation situations, the amount of future payments is not known with certainty. This uncertainty exists even if the practitioner is perfectly certain of the entire process generating future payments. An example of this process uncertainty is the uncertainty we face when trying to predict the outcomes of the roll of a fair die. We know that there are only six possible outcomes (one through six), each with the same likelihood. Even with this perfect knowledge of the underlying process, there is still unavoidable uncertainty as to what the next roll of the die will be. In insurance situations insurers try to aggregate a large number of independent risks so that the "law of large numbers" can be applied, reducing the uncertainty inherent in estimating the aggregate value of a large number of claims. However, even with such a large number of independent risks, process uncertainty still exists.

1.3.2 Parameter Uncertainty

Quite often a practitioner may elect to use a certain statistical distribution as a model for the distribution of future payments. Such distributions are often described in terms of a limited number of variables known as parameters. For example, the familiar normal distribution is completely determined by its mean and variance (two parameters). Even if the distribution is the correct one to use, the practitioner must still estimate the proper parameters. Parameter uncertainty refers to the uncertainty in the estimates of the parameters.

Returning to our die example, if we knew the future values we are to estimate were generated by the roll of a die, but we were uncertain as to whether or not the die were fair, this uncertainty would be an example of parameter uncertainty. We have the right "model" (roll of a die) but do not know the parameters (the chance of observing any given side). Often statistical estimation methods allow the practitioner to measure the amount of uncertainty inherent in particular parameter estimates.

1.3.3 Model or Specification Uncertainty

Probably the most difficult uncertainty to quantify in estimating the distribution of future payments lies in model or specification uncertainty. This is the uncertainty that the true process generating future payments actually conforms to a particular model selected. In nearly every stochastic model, the modeling begins by making the assumption that the underlying process follows the model. There is thus little possibility that the model itself can detect this source of uncertainty in the estimate of the distribution of future payments.

Taking our die analogy, an example of model uncertainty would be a situation where each roll is the roll of one of six "loaded" dice, with the choice of the particular die determined by the prior roll. Here no single loaded die model would accurately model the next roll.

There are numerous examples of model or specification uncertainty in traditional estimation techniques. Those techniques, as do most of the estimation methods currently in use, make the explicit assumption that past experience is a valid guide to future payments. A substantial portion of the paper by Berquist and Sherman[1] addresses ways to adjust traditional methods in situations where changes in the underlying environment invalidate that critical assumption. In effect, that paper provides ways to at least address the issue of model or specification error in traditional estimation analyses.

Trying a number of models and seeing which ones are most consistent with the data can also help reduce specification uncertainty.

Any estimate of the distribution of future payments should at least acknowledge this source of uncertainty, though its true measurement may be impossible.

---

[1] See Berquist and Sherman [5].

1.4 Outline

The remainder of this paper sets out the work of the Working Party on Quantifying Variability in Reserve Estimates. Section 2 discusses the scope of what we are attempting as well as provides a uniform glossary that we will use to communicate our results. Section 3 discusses criteria for reviewing models, while Section 4 gives a broad taxonomy of models currently in use. Section 5 discusses results of various models, while Section 6 points out some areas of future research. We finish with a list of caveats and limitations to this work in Section 7.

2. SCOPE, TERMINOLOGY, AND NOTATION

The purpose of this paper is to discuss, compare, and contrast – using a unified notation – existing ways of estimating the distribution of future payments and quantifying the variability of estimates of future loss and allocated loss adjustment expense payments for property and casualty insurance exposures. This paper does not give consideration to premiums or expenses contingent upon losses (such as those associated with reinsurance contracts or retrospectively rated policies), nor does this paper address issues associated with the timing of future payments like discounting.

It is not within the scope of this paper:

- to propose best practices for determining the distribution of future payments for loss and allocated loss adjustment expense; nor
- to recommend the level within the distribution of future payment estimates that should be recorded on a company's financial statements; nor
- to present original estimation methods and/or techniques. However, it is anticipated that this paper will be used as a platform to support future such research.

2.1 Terminology

Bootstrap Analysis:[2] The bootstrap is a resampling (see Resampling Methods below) technique in which N new samples are drawn from given observed data. Each sample is drawn with replacement and is the same size as the original sample. Bootstrapping is performed in order to study a statistic such as the mean of a variable. The statistic is

[2] See *S-Plus 6 for Windows Guide to Statistics*[55].

calculated for each of the N new samples, producing a bootstrap distribution for the statistic. The theory underlying bootstrapping describes how the bootstrap distribution can be used to make inferences about the statistic from the original distribution[3].

Decay model: A model in which the variable being analyzed declines over time. A common example from physical science is that of exponential decay, where the quantity $f(t)$ remaining at time t is the solution of the differential equation $df/dt = -\alpha f(t)$, where α is a constant.

Deterministic: This is a process whose outcome is known once the key parameters are specified. Examples are many of the laws of Newtonian Mechanics. Deterministic is an antonym of "stochastic."

Distribution of Future Payments: This term is used for the range of possible outcomes and their likelihood. In this paper the word "distribution" as applied to future payments means the distribution of the sum of all future payments rather than the time distribution of the individual payments.

Future Payment Estimation Model: See "Model."

Latent Liabilities: Present or potential liabilities due to emerge in the future which are not represented in historical data.

Liability: The actual amount that is owed and will ultimately be paid by a risk-bearing entity for claims incurred on or prior to a given accounting date.[4]

Mean Squared Error (*MSE*): The expected value of the squared difference between an estimator of a random variable and its true value is referred to as the *MSE*.

Mean Squared Error of Prediction (*MSEP*): The average of the squares of the differences between observations not used in model fitting and the corresponding values predicted by the model.

---

[3] See Efron, B. & Tibshirani, R.J.[15].

[4] While reserves and liabilities are sometimes used interchangeably, they are given separate definitions in this paper, and used differently throughout, to help clarify the concepts discussed.

Method: A systematic procedure for estimating future payments for loss and allocated loss adjustment expense. Methods are algorithms or series of steps followed to determine an estimate; they do not involve the use of any statistical assumptions that could be used to validate reasonableness or to calculate standard error. Well known examples include the chain-ladder (development factors) method or the Bornhuetter-Ferguson method. Within the context of this paper, "methods" refer to algorithms for calculating future payment estimates, not methods for estimating model parameters.

Model: A mathematical or empirical representation of how losses and allocated loss adjustment expenses emerge and develop. The model accounts for known and inferred properties and is used to project future emergence and development. An example of a mathematical model is a formulaic representation that provides the best fit for the available historical data. Mathematical models may be parametric (see below) or non-parametric. Mathematical models are known as "closed form" representations, meaning that they are represented by mathematical formulas. An example of an empirical representation of how losses and allocated loss adjustment expenses emerge and develop is the frequency distribution produced by the set of all reserve values generated by a particular application of the chain ladder method. Empirical distributions are, by construction, not in "closed form" as there is no underlying requirement that there be an underlying mathematical model.

Model (or Specification) Uncertainty: The risk, or variability, inherent in estimating the distribution of future payments for loss and allocated loss expense derived from the chance that the true process generating future payments does not conform to the particular model selected.[5]

Over-Dispersed Poisson Models (*ODP*):[6] Models for estimating future payments of claims in which the incremental claim payments $q(w,d)$ are "over-dispersed" Poisson random variables with:

---

[5] In common vernacular, actuaries and statisticians generally use the term "parameter uncertainty" to include both parameter uncertainty and model uncertainty as defined in this paper. The two risks are separated here in order to distinguish the portion that is readily measurable (assuming a given model) from the portion that is not. They are also separated to emphasize the fact that all models used by actuaries make assumptions about the claim process that are critical to the estimates they produce. See Shapland [58], p. 326.

[6] See England and Verrall [18], p.449.

$$\mathrm{E}[q(w,d)] = m_{wd} = x_w y_d \text{ and } \mathrm{Var}[q(w,d)] = \phi m_{wd}\text{, where } \phi > 1$$

and w is the accident period and d is the development period as defined in the Notation Section 2.2.

Example: Let Y be a Poisson random variable with mean and variance m/ϕ, where $\phi >$ 1. Then $X = \phi \cdot Y$ is an over-dispersed Poisson random variable with mean m and variance $\phi \cdot m$.

<u>Parameter Uncertainty</u>: The risk, or variability, in estimating the distribution of future payments for loss and allocated loss expense derived from the potential error in the estimated model parameters, assuming the process generating the claims is known (or assumed to be known). This type of uncertainty exists even if the process is known with certainty.

<u>Parametric Family of Distributions</u>: A collection of distribution functions where each member is specified by a fixed number of variables called parameters.[7] For example, the mean and variance specify each member of the family of univariate normal distributions.

<u>Parametric Model</u>: A statistical model where the random samples are assumed to be distributed according to a given parametric family of distributions. One goal of the modeling process is to determine the value of the parameters. Examples of parametric models include the Pareto and lognormal distributions.

<u>Prediction Error</u>: The square root of the *MSEP*. It is a measure of how well a model predicts observations not used in fitting the model.

<u>Process Uncertainty</u>: The risk, or variability, in estimating the distribution of future payments for loss and allocated loss adjustment expense resulting from the random nature of loss and allocated loss expense occurrence and settlement patterns. More generically, process uncertainty is the randomness of future outcomes given a known distribution of possible outcomes.[8]

---

[7] See Klugman, Panjer, and Willmot [34], page 45.

[8] For example, for a roll of a pair of "fair" dice, both the process and the possible outcomes are known in advance, yet the process uncertainty of the result from a specific roll of the dice still remains.

Pseudo-data: Generally refers to data that is "free data" in the sense that it can be obtained without additional experimental effort. The resampled data referred to in the Resampling Methods discussion below is an example.

Q-Q Plot: A quantile is the fraction (or percentage) of points below a given value. For example, the 0.1 (or 10%) quantile is the point at which 10% of the data fall below and 90% fall above that value. The Q-Q plot is a plot of the quantiles of one dataset against another (to test if they have the same distribution), or a dataset against a know distribution, such as the normal (to test if the data has the specified distribution).

Range of Reasonable Estimates: It is the range of estimates of the future payments, each estimate arising from a different, yet reasonable, model or method. Future payment estimates can also arise from knowledge other than that provided by the data. In contrast to the "distribution of future payments", the "range of reasonable estimates" is completely determined by the practitioner using all available input and applying professional judgment.

Resampling Methods:[9] In statistical analysis, the researcher is interested in obtaining not only a point estimate of a given statistic, but also an estimate of its variance and a confidence interval for the parameter's true value. Traditional statistics relies on the central limit theorem and normal approximations to make these estimates.

With the development of modern computers, researchers can use resampling methods to estimate standard errors, confidence intervals, and distributions for a statistic of interest. Resampling involves drawing a number of repeated samples, each sample itself drawn from the observed data. The statistic of interest is recalculated on the resampled data. The theory of resampling describes how the distribution of the statistic from the resampled data enables one to make inferences about the distribution of the statistic from the original data.

Reserve:[10] An amount selected for a specific purpose (for example, the amount to be carried in the liability section of a risk-bearing entity's balance sheet) which is a point estimate of the actual amount that is owed and will ultimately be paid by a risk-bearing entity

---

[9] This definition uses material from *S-Plus 6 for Windows Guide to Statistics, Volume 2,* Insightful Corporation, Seattle, Washington.

[10] While reserves and liabilities are sometimes used interchangeably, they are given separate definitions in this paper and used differently throughout, to help clarify the concepts discussed.

for claims incurred on or prior to a given accounting date. In the field of Finance, the term reserve refers to a segregation of retained earnings rather than an amount carried for a liability.

Risk (from the risk-bearing entity's point of view): The uncertainty[11] (deviation from expected) in both timing and amount of the future claim payment stream.[12,13] This definition is different from that in Finance, which defines risk[14] as the "measurable probability of losing or not gaining value."

Specification Uncertainty: See "Model Uncertainty."

Standard Deviation: The square root of the variance of a distribution or sample.

Standard Error: The estimated standard deviation of a probability distribution. When applied to the distribution of future payments, it includes both parameter uncertainty and process uncertainty.

Stochastic: Describing a process or variable that is random, that is, whose behavior follows the laws of probability theory. Stochastic is an antonym of "deterministic."

Variance of a Distribution: The expected value of the square of the difference between a random variable and the expected value of the random variable.

Variance of a Sample: The average of the sum of the squares of differences between sample values and the sample average. The sum of the squares can be divided by n or n-1, where n is the sample size.

2.2 Notation

---

[11] In section 3.6.1 of ASOP No. 36, sources of uncertainty are described and include the following: random chance; erratic historical development data; past and future changes in operations; changes in the external environment; changes in data, trends, development patterns and payment patterns; the emergence of unusual types or sizes of claims; shifts in types of reported claims or reporting patterns; and changes in claim frequency or severity.

[12] If the loss reserves are discounted, this would add an additional source of uncertainty to the expected value of the future payment stream. For purposes of the paper, "interest rate risk" will be ignored and reserves are assumed to be undiscounted.

[13] See Shapland [58], p. 325.

[14] Dictionary of Finance and Investment Terms, Sixth Edition (2003), Barron's Educational Series.

This paper describes many of the future payment estimation models in the actuarial literature. Many such 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.[15] Here the development age is the accounting year[16] 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 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 latest accounting year for which loss information is available, the triangle represents the loss information as of accounting dates 1 through n. The "diagonal" for which $w + d$ equals a constant, k, represents the loss information for each accident period w as of accounting year k.[17]

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

$c(w,d)$: cumulative loss from accident (or policy) year w as of age d. Think "when" and "delay."

$c(w,n) = U(w)$: total loss from accident year w when end of triangle reached.

$R(w,d)$: 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 more generally any factor relating to age d.

$F(d)$: factor applied to $c(w,d)$ to estimate $c(w,n)$ or more generally any

[15] 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.
[16] The development ages are assumed to be in yearly intervals for this discussion. However, they can be in different time units such as months.
[17] For a more complete explanation of this two-dimensional view of the loss information see the Foundations of Casualty Actuarial Science [21], Chapter 5, particularly pages 210-226.

	cumulative factor relating to age d.
$G(w)$:	factor relating to accident or policy year w – capitalized to designate ultimate loss level.
$h(w+d)$:	factor relating to the diagonal k along which $w+d$ is constant.
$e(w,d)$:	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.

Finally, we use many abbreviations throughout the remainder of this report. Most of these abbreviations are defined below.

AIC: Akaike Information Criteria
APD: Automobile Physical Damage
BIC: Bayesian Information Criteria
BF: Bornhuetter-Ferguson
BUGS: Bayesian Inference Using Gibbs Sampling
CL: Chain Ladder
CV: Coefficient of Variation
ELR: Expected Loss Ratio
EPV: Expected Process Variance
GB: Gunnar-Benktander
GLM: Generalized Linear Models
MCMC: Markov Chain Monte Carlo
MLE: Maximum Likelihood Estimate
MSE: Mean Squared Error
MSEP: Mean Squared Error of Prediction
ODP: Over-Dispersed Poisson
OLS: Ordinary Least Squares
SSE: Sum of Squared Errors
VHM: Variance of Hypothetical Mean

3. PRINCIPLES OF MODEL EVALUATION AND ESTIMATION OF FUTURE PAYMENT VARIABILITY

Historically, the problem of quantifying a probability distribution for a defined group of claim payments has been solved using "collective risk theory."[18] Actuaries have built many

[18] There are a number of good books and papers on the subject, including, but not limited to, Bühlmann [9], Gerber [22], and Seal [57].

sophisticated models based on this theory, but it is important to remember that each of these models makes assumptions about the processes that are driving claims and their settlement values. Some of the models make more simplifying assumptions than others, but none of them can ever completely capture all of the dynamics driving claims and their settlement values. In other words, none of them can ever completely eliminate model uncertainty.

While it is possible to estimate some portions of model uncertainty, developing criteria for evaluating different models will necessarily need to focus on parameter and process uncertainty.[19] Indeed, a fundamental question for evaluating a model is: "How well does it measure and reflect the uncertainty inherent in the data?" It is not simply a matter of calculating statistics to measure the uncertainty. The evaluation criteria must focus on *how well* the uncertainty is measured. Thus, another fundamental question is: "Does the model do a good job of capturing and replicating the statistical features found in the data?" Unfortunately, no single criterion will answer these questions.

As noted earlier, the goal of this paper is to set forth the current state of knowledge regarding the models used to estimate the distribution of future payments for a given block of claims (or equivalent). Many of the approaches to estimating a distribution of future payments involve fitting a statistical model to the available loss development data.[20] This will henceforth be called a *future payment estimation model* or *model*. A number of different modeling techniques can be used to fit statistical models to a dataset. Furthermore, any given technique can be used to specify a multitude of models. Therefore, the analyst needs to have available the tools and concepts needed to *evaluate* each candidate future payment estimation model. Based on these evaluations, the analyst can select the most appropriate models and modeling methodologies.

Section 3.1 will enumerate a number of principles and considerations (which we will collectively refer to as criteria) relevant to evaluating a future payment estimation model. Once a model has been specified there will typically be one or more techniques available for estimating the variability around the model's estimate of future payments. Section 3.2 will discuss three of these techniques.

[19] Shapland [58], p. 337.
[20] This is not limited to methods for evaluating loss development triangles.

3.1 Model Selection and Evaluation

Recall the three concepts of uncertainty discussed earlier: process, parameter, and model uncertainty. All three of these concepts are relevant for the purpose of estimating the variability of a model-based estimate of future payments. Of these three kinds of uncertainty, process uncertainty is often times (although not necessarily) the smallest when modeled statistically, yet the focus of the analyst should be to minimize the other two: parameter and model uncertainty. The goal of modeling insurance losses is ***not*** to minimize process uncertainty, as this is simply a reflection of the underlying process that is generating the claims. While some datasets exhibit a relatively small amount of process uncertainty, others can generate a large amount of process uncertainty. The goal of the analyst should be to select a statistical model(s), with the help of the criteria discussed below, which most accurately describes the process uncertainty in the data while also minimizing the parameter and model uncertainty.[21]

The general criteria for evaluating a model statistically can be quite numerous. Unfortunately, there is no single criterion that establishes a supreme model in every case. Instead, one must collectively review a variety of criteria in order to narrow the list to the best model(s) for each data set. Therefore, we present several of the most useful criteria for the practicing actuary. For ease of discussion, the criteria to be discussed have been segregated into three groups, listed roughly in order from the most "general" to the most "specific:"

- Criteria for selecting an appropriate modeling technique,
- Overall model reasonability checks, and
- Model goodness-of-fit and prediction error evaluation.

3.1.1. Criteria for Selecting an Appropriate Modeling Technique

The criteria for selecting a modeling technique are a blend of the pragmatic and the theoretical.

[21] The process of finding the "best" statistical model is a departure from the common practice of using multiple models to "define" a range by using the highs and lows from among the models used. It is also quite possible to end up with competing models that reflect different aspects of the historical information or different views on likely future outcomes.

Criterion 1: *Aims of the Analysis*. Will the procedure achieve the aims of the analysis? For example, if the analyst requires an estimate of the distribution of future payments, a stochastic future payment estimation model is likely to be preferred over a simpler, traditional estimation method such as the chain ladder.

Criterion 2: *Data Availability*. Does the analyst have access to the data elements required by the model and in sufficient quantity? Consideration should be given to whether the model under consideration requires unit record-level data or summarized "triangle" data, whether exogenous predictive information (such as historical inflation rates) is needed, and whether the data at hand has sufficient credibility for the model under consideration.

Criterion 3: *Non-Data Specific Model Evaluation*. The analyst should consider whether a particular model is appropriate based on general (non-data specific) background knowledge. Considerations include:

- Has this model been validated against historical data that is similar to the data at hand?
- Has this model been verified to perform well against a dataset that contains known results and that contains similar features to those expected to underlie the data to be analyzed?
- Are the assumptions of the model plausible given what is known about the process generating this data? Examples of such assumptions include the independence of accident years, similar development patterns across accident years, and constant claims (non-wage) inflation.

Criterion 4: *Cost/Benefit Considerations*. It is possible that two or more models of varying cost or complexity produce reasonable results. If this is the case, it is likely that the analyst would elect to use the simplest and cheapest of these models. If a more costly or complex model is expected to produce more complete or accurate results, then the analyst must decide whether the marginal accuracy justifies the marginal cost. Other considerations include

- Can the analysis be performed using widely available software, or would specialist software be required?
- How much analyst time and computer time does the procedure require?

- How difficult is it to describe the workings of the procedure to junior staff or the user of the model output?

3.1.2 Overall Model Reasonability Checks

By overall model reasonability checks, we mean "what measures can we use to judge the overall quality of the model?" For this, we suggest a number of criteria that can be used to test whether the summary statistics from the model are sound.[22] Two of the key statistics that can be produced for many models are the standard error of the distribution of future payments[23] and the coefficient of variation (*i.e.*, the standard error divided by the estimated mean).[24] While some of these criteria do not help distinguish between models, they do help determine if the overall model is sound and thus gets onto the "models to be analyzed" list.

Criterion 5: *Coefficient of Variation by Year.* For each (accident, policy or report) year, the coefficient of variation (estimated standard error as a percentage of estimated liabilities) should be the largest for the oldest (earliest) year and will, generally, get smaller for the more recent years.

Criterion 6: *Standard Error by Year.* For each (accident, policy or report) year, the standard error (on an absolute unit basis) should be the smallest for the oldest (earliest) year and will, generally, get larger for the more recent years.[25] To visualize this, remember that the liabilities for the oldest year represent the future payments in the tail only, while the liabilities for the most current year represent many more years of future payments including the tail. Even if payments from one year to the next are completely independent, the sum of many standard errors will be larger than the sum of fewer standard errors.

Criterion 7: *Overall Coefficient of Variation.* The coefficient of variation (standard error as a percentage of estimated liabilities) should be smaller for all (accident, policy or report) years combined than for any individual year.

---

[22] Shapland [58], pp. 334-337.

[23] The standard error for an unknown distribution is analogous to the standard deviation for a known distribution.

[24] These standard error concepts assume that the underlying exposures are relatively stable from year to year – *i.e.,* no radical changes. In practice, random changes do occur from one year to the next which could cause the actual standard errors to deviate from these concepts somewhat. In other words, these concepts will generally hold true, but should not be considered hard and fast rules in every case.

[25] For example, the total reserves for 1990 might be 100 with a standard error of 100 (coefficient of variation is 100%), while the total reserves for 2000 might be 1,000 with a standard error of 300 (coefficient of variation is 30%).

Criterion 8: *Overall Standard Error.* The standard error (on an absolute unit basis) should be larger for all (accident, policy or report) years combined than for any individual year.[26]

Criterion 9: *Correlated Standard Error & Coefficient of Variation.* The standard error should be smaller for all lines of business combined than the sum of the individual lines of business – on both an absolute unit basis and as a percentage of total liabilities (*i.e.*, coefficient of variation).

Criterion 10: *Reasonability of Model Parameters and Development Patterns.* For all modeling techniques the estimated parameters should be checked for consistency with actuarially informed common sense. In particular the signs and relative magnitudes of the parameters should be checked against common sense. Similarly, the loss development patterns implicit in the model's parameters should be checked for reasonability and consistency with one's expectations.

Criterion 11: *Consistency of Simulated Data with Actual Data.* Whenever simulated data is created based on a particular model, it should exhibit the same statistical properties as the real data. In other words, the simulated data should be statistically indistinguishable from real data.

Criterion 12: *Model Completeness and Consistency.* It is possible that other data elements or background knowledge could be integrated with the model results, thereby resulting in a more accurate prediction. For example, one might wish to incorporate one's knowledge of a changing inflation rate or claims settlement practice into the model. Similarly, one's prior expectations of an accident year's ultimate loss ratio could be integrated into the analysis through Bornhuetter-Ferguson or Bayesian methodology.

A significant portion of any liability estimate is the portion of the assumptions that lay beyond the actual data triangle. The assumptions for future development, trends, normality, etc. should be consistent with the modeled historical assumptions. This is not to say that assumptions cannot change going forward; they can. This is simply to say that they should do so in an explainable manner that is consistent with the modeled historical assumptions.

3.1.3 Model Goodness-of-Fit and Prediction Error Evaluation

[26] Strictly speaking, this criterion assumes that the individual years are not negatively correlated.

By model goodness-of-fit and prediction error evaluation, we mean "what measures can we use to judge whether a model is capturing the statistical features in the data?" In other words, does the model provide a good fit to the data compared to other models? For this, we suggest a number of criteria that can be used to test statistical goodness of fit and the general model assumptions.

<u>Criterion 13</u>: *Validity of Link Ratios.* Venter[27] shows that link ratios are a form of regression and how they can be tested statistically. All models based on link ratios need to be tested in order to validate the entire approach. Standard statistical methods for testing regression models can be used for this and for regression models of future payments in general.

<u>Criterion 14</u>: *Standardization of Residuals.* It is most useful to analyze a model's "standardized" or "normalized" residuals. A standardized residual is the difference between a data point's actual value and modeled value, divided by an estimate of the value's standard deviation. Ideally, such residuals will be normally distributed, with a mean of zero and standard deviation of one.

Many (if nearly all) models of the loss process make assumptions about the underlying distribution of the losses. In general, they either make a simplifying assumption that the losses themselves or their logarithms are normally distributed or that the remaining "noise" after the underlying distribution has been modeled and parameterized is normally distributed.[28] A model's standardized residuals should be checked for normality. Outliers and heteroscedasticity[29] should be analyzed with particular care. Normality can be checked, for example, by producing a Q-Q plot. Alternately, a histogram of the standardized residuals can be produced, along with a superimposed standard normal distribution. If desired, the kernel density estimation technique can be applied to the histogram of standardized residuals in order to produce a smoothed estimate of the residuals' distribution. This distribution estimate can then be visually compared with the superimposed standard normal distribution.

---

[27] See Venter [71].

[28] Not all models assume normality in the residuals. For example, *GLM* models can model the data structure without assuming a form for the distribution.

[29] A model's standard residuals are "homoscedastic" when they are equal, or have a similar spread, for all variables. A model's standard residuals are "heteroscedastic" when then have a different spread for some variables. A plot of the residuals will usually allow the user to determine their scedasticity. Most standard formulas assume homoscedasticity, so when heteroscedasticity is present, the standard error estimates will usually be biased to the low side.

Criterion 15: *Analysis of Residual Patterns*. In addition to the normality and outlier checks, residuals can be checked against various dimensions of interest. In particular it is good practice to plot standardized residuals against the following x-dimensions:

- Development period;
- Accident period;
- Calendar period; and
- Fitted value.

Ideally, the residuals at each value of the x dimension of interest will be randomly scattered around zero. Non-random patterns might indicate the need for additional parameters or an alternate model.

Criterion 16: *Prediction Error and Out-of-Sample Data*. Perhaps the best way to evaluate any predictive model is to test the accuracy of its predictions on data that was not used to fit the model. In an extreme case, one can fit a model containing x parameters to a loss development array containing x data points. The fit will be perfect, and therefore the residuals will all be zero. In this extreme case, all of the residual analysis tests (Criteria 14 and 15) will be trivially satisfied. However, it is unlikely that such a model would make good predictions going forward. In cases such as this, the model is said to "over-fit" the data.

One way to guard against over-fit is to set aside part of one's data in the model fitting process, and use this data to evaluate the model's predictive accuracy. Such a dataset is called a "holdout sample" or "out-of-sample data". For example, one might set aside the most recent one or two calendar periods ("diagonals") of data from one's loss triangle. The model can be used to provide predicted values for each holdout data point, and these predicted values can be compared with the actual values.

Criterion 17: *Goodness-of-Fit Measures*. In addition to using holdout data, one can evaluate competing models by using various goodness-of-fit measures. The purpose of model selection is to find the model that best fits the available data, with model complexity being appropriately penalized. Such measures therefore analytically approximate validation on out-of-sample data. They do so by combining some measure of the model's overall "error" (using a loss function such as squared error loss or log-likelihood) and an offsetting penalty for the number of model parameters relative to the number of data points available.

Goodness-of-fit measures include:

- Adjusted sum of squared errors (*SSE*): SSE is defined as the sum of the squares of the differences between the modeled loss and the actual loss. Adjusted SSE equals SSE divided by $(n-K)^2$, where n is the number of data points and K is the number of parameters in the model.[30]
- Akaike Information Criterion (*AIC*): The AIC states that one competing model is better than another if it has a lower value of $-2\log(l)+2K$. $\log(l)$ denotes the log of the maximum likelihood.
- Bayesian Information Criterion (*BIC*): The BIC states that one competing model is better than another if it has a lower value of $-2\log(l)+\log(n)K$.

Each of these concepts provides a quantitative measure that ideally enables one to find an optimal tradeoff between minimizing model bias and predictive variance.

Criterion 18: *Ockham's Razor and the Principle of Parsimony*. This is a philosophical principle. When choosing between competing models, the principle of parsimony states that all else being equal, the simpler model is preferable. While it is important to find the best model and add enough parameters to capture the salient features in the data, it is equally important not to over-parameterize.

Criterion 19: *Predictive Variability*. What one ultimately wants is an estimate of future payments involving as little uncertainty as possible. Furthermore, one would like to quantify the uncertainty in one's future payment estimate. Ideally this would take the form of providing the probability distribution of the future payment estimate. An alternate approach would be to estimate the standard error of the future payment estimate. Section 3.2 outlines three general approaches to estimating this variability.

Criterion 20: *Model Validation*. Another way to validate a model is to systematically remove the last several diagonals from the triangle and make the same forecast of ultimate values without the excluded data. This post-sample predictive testing, or validation, is important for determining if the model is stable or not.

[30] This measure was suggested by Venter [72].

3.2 Methods for Evaluating Variability

3.2.1 Possible Approaches

The methods used to calculate distributions of future payments are grouped into three general categories: analytical evaluation of incremental data, bootstrap simulations and Bayesian models.

3.2.2 Analytical Evaluation

This subsection outlines the procedures for measuring variability in respect of future payments. Such variability estimation can be implemented for future payment estimates that are to emerge in each of the future periods, for each of the accident years, and for all the accident years combined. Note that the analytical approaches described here are only for a single line of business; in other words, no correlations among multiple lines of business will be taken into account here in evaluating future payment variability. The procedure outline presented below is largely based upon Clark[31] and England and Verrall[32] .

1. Data Requirement. The variability of future payment estimates can be estimated from a data triangle of incremental payments. Let $q(w,d)$ denote the incremental payment for accident year w and development year d, and m_{wd} the expected value of $q(w,d)$. A distributional form is chosen for $q(w,d)$, which could be an over-dispersed Poisson, negative binomial, gamma, or many others.
2. A structural form is chosen for m_{wd}, which could be either non-linear in the parameters or modeled in a generalized linear model.
   a) With a generalized linear model, a link function needs to be specified for the relationship between m_{wd} and the parameters.
   b) While modeling m_{wd} in a non-linear model, the emergence of incremental payments needs to be modeled by selecting an appropriate reserve estimation method. Section 4 surveys various methods used to obtain an estimate of future payments.
3. The parameter estimation for the linear or non-linear model requires setup of a

[31] See Clark [10].
[32] See England and Verrall [18].

maximum likelihood function and maximization of the function with respect to relevant parameters. For the generalized linear model, most statistical software packages have built-in procedures to do the estimation, and the user only needs to choose the link function and the distributional form. For the estimation of the non-linear model, a functional form should be specified for the percentage loss emergence.

4. The variability of future payment estimates can be measured by the variance of the distribution of future payments, which is denoted by $\mathrm{Var_f}[\hat{q}(w,d)]$. As stated earlier, the variance of the distribution of future payments for accident year w and total future payments, denoted respectively by $\mathrm{Var_f}[\hat{q}(w,*)]$ and $\mathrm{Var_f}[\hat{q}(*,*)]$, can be evaluated within the framework of the above stated parametric models. Several points should be noted here.

   a) The variance of the distribution of future payments is decomposed into process variance and the variance of parameter estimates, or mathematically, $\mathrm{Var_f}[\hat{q}(*,*)] \approx \mathrm{Var}[q(*,*)] + \mathrm{Var}[\hat{q}(*,*)]$.

   b) The calculation of the variance of the distribution of accident year future payment estimates should take into account any correlations between the predicted values for different development periods of the same accident year, in addition to the variance of each of the individual predicted values.

   c) The variance of the distribution of the total future payments is the sum of the variances for each accident year future payment estimate and the covariances between accident year future payment estimates.

The variance of the distribution of future payments can be numerically derived through some approximation method. Appendix A gives the analytical forms for these variances for which approximation through the delta method is used in the derivation.

3.2.3 Bootstrap Evaluation

The residuals saved from estimating the generalized linear models or nonlinear models can be used for the bootstrap simulation to obtain the distribution of future payments. For instance, one way of bootstrapping is sampling with replacement from the scaled Pearson residuals and constructing a large number of (equal to the number of simulations, N)

pseudo past triangles.[33] For each of the N pseudo loss triangles, their corresponding lower triangles of future incremental losses are estimated by following the procedures outlined in Section 3.2.2. For each accident year, the mean of future payments, the parameter variance and the process variance can then be calculated from the N future triangles. Note that the parameter variance thus derived from the simulation needs to be adjusted by a factor equal to $n/(n-p)$. England and Verrall describe the calculation of the standard error[34] of the bootstrap future payment distribution. The mean and standard error obtained from the bootstrapping should then be compared to the corresponding values calculated through the analytical approach to check for errors.

A simplified bootstrap simulation procedure that yields identical results has also been discussed in England and Verrall[35]. The authors propose using the standard chain-ladder method in the simulation to obtain the future incremental loss triangles (the lower triangles) as well as the past triangles (the upper triangles) instead of going through the complicated procedures of solving the maximum likelihood functions of the over-dispersed Poisson models. The detailed bootstrap procedure is outlined in Appendix 3 of England and Verrall [36].As compared with the analytical approach, one obvious advantage of the bootstrap simulation is that it not only gives the future payment means and standard errors but also provides the distribution of future payments. The percentile distribution of future payments and the histogram of overall future payments and future payments for each accident year can easily be obtained from the simulated pseudo data sets.

3.2.4 Bayesian Evaluation

A promising, though less frequently discussed, approach to estimating future payments and future payment variability is the use of Bayesian modeling. At a high level, Bayesian modeling can be viewed as an extension of classical or "frequentist" modeling in which the analyst is willing to consider distributions on the parameters of one's statistical model.

Let us sketch the outlines of the frequentist modeling paradigm. Suppose one has a candidate model $p(q|\theta)$ for the terms $q(w,d)$ in a loss development array. $q(w,d)$ denotes the incremental losses for accident year w from development period $d-1$ to d and θ

[33] See England and Verrall [16].
[34] See England and Verrall [16,18]; England and Verrall call this the prediction error.
[35] See England and Verrall [16,18].
[36] See England and Verrall [18].

denotes the vector of parameters to be estimated from the available data $\{q(1,1), q(1,2),\ldots, q(1,n), q(2,1),\ldots, q(2,n-1),\ldots, q(n,1)\}$. Suppose maximum likelihood is used to derive the estimate θ_{MLE} of θ. The missing terms of the array (*i.e.*, the elements of the future payments), $R = \{R_{2,n}, R_{3,n-1}, .R_{3,n}, ..., R_{n,2}, ..., R_{n,n}\}$, can be then estimated by calculating $\{p(R_{w,d} | \theta_{MLE})\}$.

The frequentist paradigm therefore takes the parameterized model $p(q|\theta)$ as fundamental. The data q are used to estimate θ, and this estimate is then used, via the model formula, to make forecasts or inferences.

The Bayesian paradigm expands this conceptual framework by treating the parameter vector θ as a further set of random variables. Therefore just as the (observed) random variables q admit of the probability distribution $p(q|\theta)$, the (unobserved) random variables θ admit of a further probability distribution $p(\theta)$. $p(\theta)$ is known as a *prior probability distribution*.

The key insight of the Bayesian paradigm is that the data *q* can be used to refine or update the prior distribution $p(\theta)$ to a *posterior* distribution $p'(\theta)$. This updating is performed via *Bayes Theorem*.

$$p'(\theta) \equiv p(\theta \mid q) = \frac{p(q \mid \theta) p(\theta)}{\int p(q \mid \theta) p(\theta) d\theta} \propto p(q \mid \theta) p(\theta) \tag{3.1}$$

Notice that the first factor on the right side of the equation is the statistical model from the frequentist paradigm. This statistical model is also known as the *likelihood function*. Rather than filtering the data q through the model $p(R|\theta)$ to produce a point estimate of θ, the data is used to refine the *distribution* of θ via Bayes Theorem.

The posterior distribution can in turn be used to generate the distribution of future claims, R:

$$p(R \mid q) = \int p(R \mid \theta) p(\theta \mid q) d\theta \tag{3.2}$$

A concrete example of Bayesian loss estimation is provided by Verrall[37]. The "frequentist" model Verrall begins with is the over-dispersed Poisson (ODP) model described in England and Verrall [38]:

[37] See England and Verrall [17].
[38] See England and Verrall [17].

$$q(w,d) \propto_{iid} \text{ODP}_{\varphi}(x_d y_w) \tag{3.3}$$

where $\sum_d y_d = 1$. The parameter vectors, $x = \{x_1, x_2, ..., x_n\}$ and, $y = \{y_1, y_2, ..., y_n\}$, represent the rows (accident years) and columns (development periods) respectively of the loss array. Note that the mean and variance of $q(w,d)$ equal $x_w y_d$ and $\varphi x_w y_d$ respectively. φ is known as the *dispersion parameter*.

Within the frequentist paradigm, maximum likelihood theory (in particular the theory of generalized linear models) can be used to estimate the parameter vector $\theta = (x, y, \varphi)$. These parameters in turn are used to estimate the unknown elements of the loss array: $R_{w,d} = x_w y_d$. England and Verrall also demonstrate how to analytically derive confidence intervals around the sum of the future payment estimates. Note that this is a complex derivation that only results in variance information about the distribution of future payments.

Verrall [39] extends this frequentist model to a Bayesian model by introducing prior distributions on the row and column parameters x and y. (Note that a prior distribution could also be placed on φ but Verrall chooses to use a plug-in estimate for simplicity.)

The data $q = \{q(1,1), \ldots, q(n,1)\}$ are used to obtain a posterior distribution of (x, y):

$$p(x, y \mid q, \varphi) \propto \prod_{w=1}^{n} \prod_{d=1}^{n-i+1} \text{ODP}_{\varphi}[q(w,d) \mid x, y] \prod_{i=1}^{n} p(x_i) p(y_i)\,. \tag{3.4}$$

The posterior distribution in turn determines the distributions of the unknown elements of the loss array:

$$p(R_{w,d} \mid q) = \int p(R_{w,d} \mid x, y, \varphi) p(x, y \mid q, \varphi) dx dy\,. \tag{3.5}$$

To summarize, Verrall develops both a frequentist and Bayesian *ODP* model of a loss array. The frequentist approach uses the data and the *ODP* model to generate point estimates of future payments and (with some labor) confidence intervals around these point estimates. In the Bayesian approach, he introduces prior distributions of the *ODP* model parameters (x, y). Bayes Theorem is applied to the known elements q of the future payment array to generate the posterior distribution of (x, y). This posterior distribution in turn determines the distributions of future payments $\{R_{w,d}\}$. For example, the adoption of a gamma distribution prior on each *x* parameter, in conjunction with the *ODP* conditional likelihood, is shown to yield a gamma posterior distribution, and a posterior mean of future

[39] See England and Verrall [17].

$\{R_{w,d}\}$ that may be interpreted as a Bornhuetter-Ferguson estimate.

Ultimately, one would like to calculate the mean and various percentiles of the distribution of the total future payments $R = \sum R_{w,d}$. [Let $p(R|q)$ denote the distribution of the total future payments.]

Unfortunately, it is typically impossible to calculate such quantities analytically. Even calculation of the posterior of a single $R_{w,d}$ will not usually be possible. Because of the numerical difficulty involved, Bayesian methodology remained at a relative impasse for several decades. However, recent developments in Monte Carlo integration have made it practical to approximate the mean and percentiles of the distribution of future payments with a high degree of accuracy.

The basic idea of Monte Carlo integration is to generate a large sample of draws from the posterior distribution $p(\theta|q)$. This sample of draws allows one to easily approximate any quantity that depends on the posterior density. To illustrate, suppose we have generated 10,000 draws from Verrall's posterior density $p(x, y|q)$. (Reference to the dispersion parameter φ will henceforth be suppressed.) That is, we have a sample of 10,000 values of $\theta = (x, y)$. Let $\theta^{(1)}, \ldots, \theta^{(10,000)}$ denote these 10,000 estimates of θ. For each one of these values $\theta^{(k)}$, we can readily compute each unknown value $\{R_{w,d}\}$ of the loss array (recall that $R_{w,d} = x_w y_d$) and add them together: $R^{(k)} = \sum R_{w,d}{}^{(k)}$.

In this way, we have generated 10,000 draws from the distribution of the total future payments. The average value of these 10,000 draws constitutes an estimate of the future payments:

$$Future\,Payments = \frac{1}{10{,}000}\sum_{k=1}^{10{,}000} R^{(k)} \approx \int p(R\,|\,\theta)p(\theta\,|\,q)d\theta \tag{3.6}$$

Similarly, the empirical 5th and 95th percentiles of this simulated distribution $\{R^{(k)}\}$ constitute one of many possible variability estimates.

The surprising ease of these calculations is due the fact that we were able to generate the draws $\theta^{(1)}, \ldots, \theta^{(10,000)}$ from the posterior distribution $p(\theta|q)$. This sampling of the posterior distribution is accomplished by *Markov Chain Monte Carlo* ($MCMC$) simulation. $MCMC$ techniques are recipes for constructing a Markov chain of random variables $\theta^{(0)}$, $\theta^{(1)}$, $\theta^{(2)}, \ldots$ that in the limit "forget" their arbitrary starting value $\theta^{(0)}$ and converge to the stationary distribution $p(\theta|q)$. Two commonly used $MCMC$ techniques are the Hastings-

Metropolis Sampler and the Gibbs Sampler. Details of these *MCMC* techniques will be omitted for brevity of exposition, but can be found in most modern introductions to Bayesian modeling.

To summarize: one way of using Bayesian methodology to estimate the distribution of future payments is to begin with a frequentist model (such as the England-Verrall *ODP* model) of one's loss array. One then supplements this model by assigning prior distributions to many or all of the parameters of the model. Next, a *MCMC* technique such as Gibbs Sampler can be used to generate an empirical posterior distribution of the model parameters. Finally, this distribution of parameters can be plugged into the model to generate the corresponding distribution of future payments. In short, *MCMC* integration makes it possible to estimate not only the expected value and variability of future payments, but the actual distribution of future payments.[40]

3.3 Feasibility and Merits of Each Approach

3.3.1 Analytical Approach

There are several techniques available for model evaluation. Some of the testing procedures have been suggested in Venter [41]. The first one is to test the significance of parameter estimates. Secondly, residuals can be used to test the validity of model assumptions in various ways. The residuals can be plotted against the development period, the accident year, the calendar year of emergence, or any other variable of interest. The validity of model assumptions requires that the residuals appear to be randomly distributed around the zero line. Any anomalous residual plot is an indication that some of the model assumptions are incorrect or the model is misspecified. Thirdly, the goodness fit of the model can be tested by using the *AIC* and *BIC* criteria.

For the generalized linear model, the table also reports the scaled deviance and scaled Pearson chi-square, which are directly obtained from the computer-generated output. These two scaled statistics, under certain regularity conditions, have a limiting chi-square

---

[40] Another example of Bayesian revision was given by Taylor, McGuire and Greenfield (2003) in an ASTIN Colloquium keynote address (see www.economics.unimelb.edu.au/actwww/wps2004/No113.pdf). This paper dealt with loss estimation regression models.

[41] See Venter [72].

distribution, with degrees of freedom equal to the number of observations minus the number of parameters estimated. A scaled deviance close to one may be an indication of a good model fit. However, the examination of the deviance for model fitness should always be accompanied by the examination of residuals. As an illustrative example, Appendix B uses a sample of incurred loss data to estimate the variability of expected future payments and discusses the goodness fit of the model.

Besides parameter uncertainties and process disturbances, model misspecification may exist, which should be reflected in the anomaly of the residual plots. For instance, leaving out the calendar effect in the estimation could be a misspecification, considering the data triangle used normally spans over a considerably long period of time. As a remedy, the data elements in the loss triangle can be adjusted by some appropriate measures so that the specification error coming from the calendar effect can be effectively removed. If the calendar effect is caused by inflation, all the incremental loss data can be deflated to a common basis before the model is estimated. On the other hand, some model specification tests (for example, the WALD statistics) can also be used in examining whether the calendar effect can be treated as a nuisance parameter.

3.3.2 Bootstrap Approach

The bootstrap was described in Section 3.2.3 in connection with an over-dispersed Poisson model. It is seen there to be a numerical procedure, algebraically simple, in concept at least.

The procedure may be generalized to any (non-Bayesian) model structure.[42] It produces an estimate of the whole distribution of future payments, rather than just a small number of summary statistics.

Though the procedure is conceptually simple, it can involve some practical complexities. For example, it assumes that all residuals are unbiased. This may be difficult to achieve precisely with a suitably parsimonious model. Small regions of bias in the triangle of residuals can be highly disturbing to bootstrap results.

Difficulties can also arise when the raw observations, and therefore the residuals, are drawn from a long-tailed distribution. There are no difficulties from a theoretical

[42] For detail, see Taylor [65], Chapter 11.

standpoint, provided that all residuals are equi-distributed. In practice, however, this will often not be so. Instead one may face residuals which are all long-tailed, but from somewhat different distributions.

3.3.3 Marks & Chain Monte Carlo (*MCMC*) Approach

As discussed in section 3.2.4, the Bayesian approach to loss estimation produces the distribution of future payments, not merely information about the mean and variance. While it is in practice impossible to analytically derive the distribution of future payments, it is readily possible to approximate this distribution though *MCMC* simulation.

A high-level statistical programming language for Bayesian modeling with *MCMC* is BUGS: Bayesian Inference Using Gibbs Sampling. The BUGS language is implemented in the freely available WinBUGS software package developed by the Biostatistics Unit at Cambridge University. Thus, both the methodology and necessary computing environment are now readily available to the analyst who wishes to apply Bayesian methodology to loss estimation problems.

Another merit of the Bayes/*MCMC* approach is that it provides an open-ended modeling environment in which the analyst can integrate (possible vague or qualitative) prior knowledge or beliefs with his or her stochastic model of the loss development process. Verrall's *ODP* model exemplifies this. The England-Verrall frequentist *ODP* model is similar (though not identical) to the classic chain-ladder model. Verrall's Bayesian extension of this model provides a rigorous way to incorporate one's prior beliefs about one or more accident years' ultimate losses into the *ODP* (chain-ladder) modeling framework. As Verrall points out, his Bayesian *ODP* model is therefore analogous to the classic Bornhuetter-Ferguson technique.

It should be emphasized that Verrall's Bayesian *ODP* model is not the only Bayesian model of loss development currently available. Other relevant contributions to date include De Alba[43] , Ntzoufras and Dellaportas[44] and Scollnik [45]. Nor does Verrall's presentation illustrate the only way of integrating one's prior beliefs with a model of the loss development process.

[43] See DeAlba [12].
[44] See Ntzoufras and Dellaportas [49].
[45] See Scollnik [56].

The Bayes/MCMC loss estimation framework, based on simulation of the distribution of future payments, is a low cost application, providing rigorous incorporation of prior beliefs. Though relatively new to the actuarial community, it appears to have considerable promise.

3.4 Categorization of Models

There are a number of properties of future payment estimation models that have a bearing on the choice of procedure for evaluation of variability. These are discussed in this section. The categorization of models so arrived at here differs from those appearing in Sections 4.5 and 4.6, which is more concerned with their properties relating to estimation of the mean of the distribution of future payments.

3.4.1 Bayesian and Non-Bayesian

The future payment estimation model may be Bayesian or non- Bayesian. Examples appear in Sections 3.2.4 and 3.2.3 respectively.

In the case of a non-Bayesian model, variability will be estimated by reference to the residuals derived from the data points and the corresponding fitted values according to the model. These residuals may be manipulated by analytical means, or by bootstrapping.

The variability within a Bayesian model contains additional mathematical structure as it relates to the Bayesian distribution of future payments, reflecting the prior distribution as well as the data points. In principle, the variance of the distribution may be derived from its analytical form but, as pointed out in Section 3.2.4, this will not be practical in many cases. The *MCMC* approach described in Section 3.2.4 will then be the natural one.

3.4.2 Simple and Complex

At a fundamental level, a loss estimation procedure is a mapping from a set of data points to the mean of the future payments. Mathematically, this mapping will be quite complex, even for the simpler estimation procedures.

The corresponding procedure for estimating variability is a mapping from the data points to the variance of the future payments, and is more complex again. Precise evaluation of this variance will not be practical except in the simplest of models.

Equation (A.16) (found in Appendix A) provides an example of an approximation in a specific, rather simple, case. As illustrated there, this result requires two ingredients:

- Evaluation of the partial derivatives of the model's forecasts with respect to its parameters; and
- The covariance matrix associated with the estimates of those parameters.

The difficulty in evaluation of these quantities will increase rapidly with increasing complexity of model structure.

The weight of algebra in even only moderately complex models may be such as to defeat feasibility of this analytical approach. This is exemplified by the method of Mack[46], which generates a complex expression for the estimated variance of future payments calculated according to the simple chain ladder method. In most cases, it may be necessary to resort to bootstrapping (Section 3.3.2) for estimation of variances.

Moreover, even when the variance of future payments may be estimated analytically, it does not provide information on the thickness of the tails of the distribution of future payments. Again, the bootstrap may prove useful in providing an estimate of the entire distribution of future payments.

3.4.3 Models with Multiple Sub-Models

Some models are composed of two or more distinct sub-models. Examples given by Taylor[47] include the Payments per Closed Claim[48] and Projected Case Estimates models. The first of these, for example, comprises:

- A model of claim closure counts; and
- A model of sizes of closures.

In such cases, estimation of the variability of future payments will require consideration of variability within each of the sub-models. It is evident from the comment in Section 3.4.2 that there is likely to be substantial difficulty in attempting to pursue this analytically. The bootstrap is likely to provide the most practical approach.

The bootstrap would need to be applied separately to each sub-model, and the sub-models then combined. In the Payments per Closed Claim example above this would produce say m realizations of forecast claim closure count arrays $\{f_j(w,d),\ j=1,\dots,m\}$,

[46] See Mack [37].
[47] See Taylor [65], Chapter 4.
[48] Also referred to as Payments per Claim Finalized.

and m realizations of forecast size arrays $\{s_j(w,d),\ j=1,\ldots,m\}$. These are then combined to produce forecast paid loss arrays $\{q_j(w,d),\ j=1,\ldots,m\}$ where $q_j(w,d)=f_j(w,d)\cdot s_j(w,d)$.

3.4.4 Independence of Data Observations

Care needs to be taken to ensure that estimates of variability account correctly for any dependencies between data items incorporated in the model specification. The most pervasive form of dependency in future payment estimation models arises in relation to cumulative data. For example, since

$$c(w,d+k)=c(w,d)+q(w,d+1)+q(w,d+2)+\ldots+q(w,d+k), \tag{3.7}$$

it follows that

$$\mathrm{Cov}[c(w,d),c(w,d+k)]=\mathrm{Var}[c(w,d)], \tag{3.8}$$

when all incremental paid losses are stochastically independent.

The bootstrap procedure described in Section 3.2.3 relies on the stochastic independence of the residuals that it permutes in the production of pseudo-data sets. The residual corresponding to the j-th observation Y_j is of the form

$$R_j=(Y_j-\hat{Y}_j)\big/\sigma_j \tag{3.9}$$

where $\hat{Y}_j$ is the value fitted to Y_j by the model and $\sigma_j^2=\mathrm{Var}[Y_j]$.

Generally, the set of residuals $\{R_j\}$ will not be mutually stochastically independent even if $\{Y_j\}$ is, since $\hat{Y}_j$ is a function of all the Y_k. However, if there are many observations, each $\hat{Y}_j$ will depend only slightly on any one Y_k. Then $\{R_j\}$ will be "nearly independent" and the bootstrap may be applied at least without gross violation of its assumptions.

This will not be so, however, if the Y_j represent cumulative data, *e.g.* $Y_j=c(w,d)$. Then, with an alternative but obvious labeling of observations, (3.8) implies that $\mathrm{Cov}[R_{w,d},R_{w,d+k}]$ is likely to be strongly non-zero. Direct application of the bootstrap to models of cumulative data will therefore usually be inappropriate.

It will often be reasonable, however, to retain the model based on cumulative data but to bootstrap by permuting the corresponding incremental residuals

$$R_{w,d} = \{[c(w,d) - c(w,d-1)] - [\hat{c}(w,d) - \hat{c}(w,d-1)]\} / \tau_j \ , \qquad (3.10)$$

where $\tau_j^2 = \mathrm{Var}[c(w,d) - c(w,d-1)] = \mathrm{Var}[q(w,d)]$.

4. METHODS AND MODELS

In this section we distinguish estimation models from estimation methods, and describe many of the estimation models in the actuarial literature.

4.1 Notation

This section uses the following notation, which is more completely described in Section 2.2:

$c(w,d)$:	cumulative loss from accident (or policy) year w as of age d.
$c(w,n) = U(w)$:	total loss from accident year w when end of triangle reached .
$R(w,d)$:	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 other incremental information for period $d+1$.
$F(d)$:	factor applied to $c(w,d)$ to estimate $c(w,n)$ or other cumulative information relating to age d.
$G(w)$:	factor relating to accident or policy year w – capitalized to designate ultimate loss level.
$h(w+d)$:	factor relating to the diagonal k along which $w+d$ is constant.
$e(w,d)$:	a mean zero random fluctuation which occurs at the w,d cell.

4.2 Methods

A method is an algorithm or recipe – a series of steps that are followed to give an estimate of future payments. The well-known chain ladder (CL) and Bornhuetter-Ferguson (BF) methods are examples. A more intricate method, suggested by Gunnar Benktander (GB) in the April 1976 issue of The Actuarial Review, uses a weighted average of the CL and BF estimates within the BF procedure. For a paid loss application, let $F(d)$ be the average proportion of ultimate claims paid through age d, and U_0 be a prior estimate of $U(w)$. Then the estimates of $U(w)$ after observing $c(w,d)$ are:

$$U_{BF}(w) = c(w,d) + [1 - F(d)] \cdot U_0 \tag{4.1}$$

$$U_{CL}(w) = c(w,d) / F(d) \tag{4.2}$$

$$\begin{aligned} U_{GB}(w) &= c(w,d) + [1 - F(d)] \cdot [F(d) \cdot U_{CL}(w) + \{1 - F(d)\} \cdot U_0(w)] \\ U_{GB}(w) &= c(w,d) + [1 - F(d)] \cdot U_{BF}(w) \end{aligned} \tag{4.3}$$

Thus the original estimate U_0 in BF is replaced by a weighted average of the CL estimated ultimate and the BF prior ultimate losses, where the weight on CL is $F(d)$. This is the same as replacing U_0 with $U_{BF}(w)$ so is also called iterated BF. It is not hard to see that the expected future development from this method is a weighted average of the future development from the CL and BF methods, again with weight $F(d)$ on CL.

CL, BF, and GB are thus three methods of future payment estimation that have been specified here up to the calculation of $F(d)$ and U_0. These calculations would have to be defined to make the methods into complete algorithms. Since they are methods, they show how to do the calculations but do not detail any statistical assumptions that might be tested or used to calculate standard errors.

4.3 A Method for Estimating Ranges

One way to calculate a range around estimated ultimate losses would be to proceed as follows:

1. For each age d, calculate age d to age $d+1$ loss development factors $f(d)$ as the average such factor over all accident years available and multiply these to get the age to ultimate factors $F(d)$.
2. For each d, sum the squared deviations of the age d individual accident year factors from $f(d)$. With n factors in the column, divide by $n-1$ to estimate the average squared deviation, then multiply by $n/(n-1)$ to adjust for uncertainty about $f(d)$.

Call the result $s^2(d)$. Set $s^2(n) = s^2(n-1)$.

3. Calculate $S^2(d)$, the estimated variance of the age-to-ultimate factor $F(d)$, working backwards from $S^2(n) = s^2(n)$ using the formula for the variance of the product of two independent variates, so $S^2(d) = f(d)^2 S^2(d+1) + F(d+1)^2 s^2(d) + s^2(d) S^2(d+1)$.

4. Estimate the expected ultimate loss for each accident year w by multiplying $c(w,d)$ from the latest diagonal by $F(d)$ and the variance for the accident year as $c(w,d)^2 S^2(d)$.

5. Sum the estimated accident year losses and variances over all accident years, and assume the sum is lognormally distributed with mean and variance equal to the summed means and variances.

6. Use that lognormal distribution to estimate percentiles of outcomes of the ultimate losses.

As with methods in general, this one tells you how to do the calculation, but does not provide any statistical assumptions that could be used to validate its reasonableness.

Simulation could also be used as a method for calculating future payment ranges. For instance, Patel and Raws[49] discuss an approach to this. The paper describes a procedure for generating future payment ranges using a combination of actuarial judgment and statistical simulation. In its application, the paper assumes a company writing multiple lines of business over multiple accident years. It is assumed that ultimate loss estimates have been generated by a variety of standard actuarial methodologies for each line of business/accident year. The paper then describes how an actuary might use this range of estimates, applying judgment to choose a loss distribution (and the associated specifying parameters) by line of business/accident year. Simulation techniques are then applied using the selected distributions to generate a range of future payments across all accident years/lines of business (i.e., a range of aggregate future payments). The paper examines three specific applications of this process.

---

[49] See Patel and Raws [50].

4.4 Models

A model specifies statistical assumptions about the loss process, usually leaving some parameters to be estimated. Then estimating the parameters gives an estimate of the ultimate losses and some statistical properties of that estimate. There are various methods that could be used for estimating the parameters, such as maximum likelihood and various robust estimators, but unless otherwise noted, "methods" here will refer to algorithms for calculating loss future payments, not methods for estimating model parameters.

Mack presents[50] a loss development model to address issues of weighted averages of CL and BF estimators. He assumes that the payout pattern is already known with $F(d)$ the proportion of ultimate losses paid by age d, and looks at how to evaluate the accuracy of the CL and BF estimators that use these factors and how they can best be weighted together. In the current notation, he defines:

$U_0(w)$ = prior expected value for $U(w)$ with $E[U_0(w)] = E[U(w)]$.

$U_0(w)$ is assumed to be independent of $U(w)$, $c(w,d)$ and $R(w)$.

$$E[c(w,d)/U(w)\,|\,U(w)] = F(d) \tag{4.4}$$

$$Var[c(w,d)/U(w)\,|\,U(w)] = B(U(w))[F(d)(1-F(d))], \tag{4.5}$$

where B is assumed constant over d's.

$$A(U(w)) = U(w)^2 B(U(w)) \tag{4.6}$$

Mack suggests that $B(U(w))$ could, for example, be assumed to be a constant or a factor times $U(w)$. Either way, the accident year's difference in its proportion of losses paid by age d from the long-term average $F(d)$ is highest near the middle of the payout pattern, where $F(d)(1-F(d))$ is highest. The CL estimate gets better for mature ages as the annual variation of the payout portion goes down and losses are grossed up by a lower factor $1/F(d)$. In fact, dividing the definition of B by $(F(d))^2$:

$$Var[U_{CL}(w)/U(w)\,|\,U(w)] = B(U(w))[1-F(d)]/F(d) \tag{4.7}$$

which decreases in $F(d)$.

The accuracy of the BF estimate also improves over time since the factor $1-F(d)$ on U_0 gets smaller. The expected squared error $E(U_0-U)^2$ does not change with age, however. Note that $E(U_0-U)^2 = E(U_0-EU_0+EU-U)^2 = Var(U_0)+Var(U)$ by the

[50] See Mack [38, 42].

first two assumptions.

Mack considers credibility weighted estimators of CL and BF for $R(w)$ of the form:

$$R_Z(w,d) = [1-F(d)][Z(w,d)c(w,d)/F(d) + [1-Z(w,d)]U_0(w)], \tag{4.8}$$

which has the lowest error for any given w when d is on the last diagonal.

He finds that the mean squared error (MSE) is minimized by taking $Z(w,d) = F(d)/[F(d)+K(w)]$, where:

$$K(w) = \frac{E[A(U(w))]}{Var(U_0(w)) + Var(U(w)) - E[A(U(w))]} \qquad \text{If } K<0\text{, set it to } 0\text{, so } Z=1 \tag{4.9}$$

Setting $Y(w,d) = [1-F(d)]E[A(U(w))]$, Mack then finds the mean squared errors of some possible estimators as:

$$MSE(R_{BF}(w,d)) = Y(w,d)[1+[1-F(d)]/K(w)] \tag{4.10}$$

$$MSE(R_{CL}(w,d)) = Y(w,d)/F(d)] \tag{4.11}$$

$$MSE(R_Z(w,d)) = Y(w,d)\{Z(w,d)^2[1-F(d)]/F(d)+1+[1-Z(w,d)]^2[1-F(d)]/K(w)\} \tag{4.12}$$

The latter formula gives the CL and BF formulas when $Z=1$ or 0. Mack shows that the MSE of the BF method is less than that for CL exactly when $F(d)<K(w)$ for w and d on the latest diagonal, which gives a criterion for the best age to switch from BF to CL. However, the credibility estimator is better still. Mack suggests that the GB method, which does not use the optimal Z's but is easy to calculate, is better than either CL or BF in most practical cases.

To apply this model, the parameters have to be estimated. U_0 and $Var(U_0)$ are from outside data, perhaps from ratemaking. $Var(U)$ could come from a historical loss ratio distribution. The F's and A's are assumed known, but as often occurs in credibility theory they are usually not, and could be estimated from historical loss development data. The MSE's above would then be the conditional MSE's given the payout pattern, so the unconditional MSE's would be their expectation over the distribution of the expected payout pattern – not the distribution of the annual payout pattern around the average, but the uncertainty in the average. This would probably have little effect on the relative MSE's for the different methods, but the difference between $1/E[F(d)]$ and $E[1/F(d)]$ could have an effect for some distributions of $F(d)$.

This example of a method and a related model illustrates what a model provides: testable assumptions about the claim development process, parameters to be estimated, and a mean estimate and measures of deviation from the mean. This model is different from many, however, as the model assumptions allow future payment estimators of various forms, and the assumptions and the accuracy of each estimator at various ages can be calculated. Other models specify a claim development process and look for the best estimator meeting certain criteria, regardless of form. For instance, if the distributions of the observations are specified, the maximum likelihood estimator of the future payments might be sought.

4.5 Types of Models

Here is a structure for classifying estimation models:

The first split is models based on individual claims histories *vs.* models based on triangles.

For models based on triangles, it is possible to model a triangle as a function of itself plus other triangles, as examples, paid a function of paid and incurred, or auto property damage a function of auto liability and auto physical damage triangles. So the first split of triangle models is by models of single triangles *vs.* models simultaneously incorporating multiple triangles.

For models of single triangles an important distinction is conditional *vs.* unconditional models. For both of them, the parameters are estimated using the data in the triangle, but for conditional models the data in the triangle is part of the set of independent variables used in the expression of the model of future loss emergence, like:

$$q(w,d+1) = c(w,d)f(d) + e(w,d+1) \tag{4.13}$$

This model has one parameter for each age, as the factors are applied directly to losses. For unconditional models the data in the triangle is not an independent variable in the model equation for future development such as:

$$q(w,d+1) = G(w)f(d) + e(w,d+1) \tag{4.14}$$

This has a parameter for each age and one for each accident year as well.

Link ratios can be expressed as a conditional model. The 1972 Bornhuetter-Ferguson method can be expressed as an unconditional model, where $G(w)$ is the expected losses for accident year w from pricing. Other models estimate $G(w)$ from the data. It is not unusual to find conditional and unconditional models that will give the same estimate of the mean total incurred.

Another distinction is whether or not there are diagonal terms in the model, like:

$$q(w,d+1) = c(w,d)f(d)h(w+d+1) + e(w,d+1) \tag{4.15}$$

or:

$$q(w,d+1) = G(w)f(d)h(w+d+1) + e(w,d+1) \tag{4.16}$$

Another distinction is whether or not the model is parametric. To get future payment ranges in the end, we need some parametric assumption, but this is not necessarily true for getting estimates of standard deviations. The Mack and Murphy chain ladder models discussed below are expressed as non-parametric, for instance. Of course, we could always argue that using squared error implicitly assumes normal distributions, or at least gives us the same answers as assuming normal distributions, but we can still call these approaches non-parametric.

Models can also be distinguished by whether they have fixed parameters or varying parameters. Varying parameter models let you have different parameters for each accident year or lag, but the degree to which the parameters can change from year to year is constrained by some kind of parameter variance limitation. However, this is possible for any model, so it is not used as a categorizing variable for models but rather as a model building tool that can be used in various types of models.

4.6 Some Estimation Models

Once a model postulates a process that generates loss development, estimation of the parameters of that process will provide estimates of means and distributions of future payments. This is shown in some detail for a few models, but is implicit in all of them. Models are presented below according to the classification scheme outlined.

4.6.1 Single Triangle Models

4.6.1.1 Single Triangle, Conditional, Non-parametric, No Diagonal Terms

Conditional models estimate future development conditional on the losses emerged so far. Basically if the expression for future development explicitly refers to emerged losses, it is a conditional model. The history of development factors is not entirely clear, but they go back at least to Thomas F. Tarbell[51] . Thomas Mack and Daniel Murphy put development factors into a conditional non-parametric framework in the early 1990's.

4.6.1.1.1 Mack's Model

Mack[52], develops formulas for estimating the standard errors of the chain ladder future payment distributions. In developing the formulas, Mack makes three key assumptions:

1. $E[c(w,d+1)\,|\,c(w,1),...,c(w,d)] = f(d)c(w,d)$, *i.e.*, the chain ladder model applies,
2. $\{c(v,1),c(v,2),...,c(v,n)\}$ is independent of $\{c(w,1),c(w,2),...,c(w,n)\}$ for $v \neq w$, and
3. $Var[c(w,d+1)\,|\,c(w,1),...,c(w,d)] = c(w,d)Var[f(d)]$.

It's clear from these assumptions that this formulation of the chain ladder is estimating future payments conditional on the triangle of observations. To calculate the estimated standard error of the future payment distributions, perform the following steps:

1. Calculate the weighted average development factors:

$$f(d) = \sum_w c(w,d+1) \Big/ \sum_w c(w,d) \tag{4.17}$$

2. Calculate the weighted variances of the development factors:

$$Var[f(d)] = [1/n+d-1)]\sum_w c(w,d)[c(w,d+1)/c(w,d) - f(d)]^2 \tag{4.18}$$

3. Estimate the variance of an accident year future payment distribution:

$$Var_f[R(w,n-w+1)] = \sum_{d=n-w+1}^{n-1} \frac{Var[f(d)]}{f(d)^2}\left(\frac{1}{c(w,d)} + \frac{1}{\sum_{j=1}^{n-d} c(j,d)}\right) \tag{4.19}$$

4. Estimate the variance of the all accident years future payment distribution:

---

[51] See Tarbell [59].
[52] See Mack [37].

$$Var_f\left[\sum_{w=2}^{n} R(w,n-w+1)\right] = \sum_{w=2}^{n}\left\{Var_f[R(w,n-w+1)] + c(w,n)\sum_{j=i+1}^{n} c(j,n)\sum_{d=n+1-w}^{n-1}\frac{2Var[f(d)]/f(d)^2}{\sum_{l=1}^{n-d} c(l,d)}\right\} \tag{4.20}$$

4.6.1.1.2 Murphy's Models

Murphy[53] describes five conditional, non-parametric chain ladder models. After making certain assumptions about the error term, he applies least squares regression theory to estimate optimal link ratios. One of the models allows for an intercept, but the remaining four are strictly multiplicative with the differences between them arising from how the error term relates to emerged losses.

Three of the multiplicative models are of the form $c(w,d+1) = f(d)c(w,d) + c(w,d)^{i/2}e(w,d+1)$ where i = 0, 1 or 2. Of particular interest is the model defined by $i = 1$, namely, $c(w,d+1) = f(d)c(w,d) + c(w,d)^{1/2}e(w,d+1)$. Dividing each side by $c(w,d)^{1/2}$ transforms the equation into a simple linear regression of $c(w,d+1)/c(w,d)^{1/2}$ onto $c(w,d)^{1/2}$. The least squares estimate of $f(d)$ simplifies to $\sum_w c(w,d+1)\Big/\sum_w c(w/d)$, the weighted average development factor. Murphy labels this model WAD.

In addition to the calculation of model parameters, Murphy shows variance estimates for an individual accident year and for all years combined. In both cases, the estimated variance of the estimated future payment is calculated as the sum of parameter variance (the variability of the estimated future payment about its true mean) and process variance (the variability of the actual future payment about its true mean). Both variance pieces are developed recursively.

Continuing to assume the WAD model, Murphy's method for developing a range about the estimated future payments for accident year w is as follows:

For $d = n - w + 1$ to $n - 1$, estimate: the age-to-age factor $f(d)$; the expected cumulative losses at each future period starting with $c(w,d+1) = f(d)c(w,d)$; and the variance of the link ratio $Var[f(d)] = MSE(d)/\sum_w c(w,d)$, where $MSE(d)$ is the mean

[53] See Murphy [46].

squared error from the regression on column d.

Beginning with $d = n - w + 1$ and stopping at $d = n - 1$, estimate the parameter variance recursively:
$Var[c(w,d+1)] = Var[f(d)c(w,d)] = c(w,d)^2 Var[f(d)] + Var[c(w,d)][f(d)^2 + Var[f(d)]]$. For $d = n - w + 1$, the formula simplifies to $c(w,n-w+1)^2 Var[f(n-w+1)]$ because $Var[c(w,n-w+1)] = 0$. The parameter variance for the future payments is the value when $d = n - 1$.

Beginning with $d = n - w + 1$ and stopping at $d = n - 1$, estimate the process variance recursively:
$Var[c(w,d+1) \mid c(w,n-w+1)] = c(w,d)^2 MSE(d) + f(d)^2 Var[c(w,d) \mid c(w,n-w+1)]$.
This simplifies to $c(w,n-w+1)^2 MSE(n-w+1)$ for $d = n - w + 1$. The process variance for the future payments is the value when $d = n - 1$.

Add the parameter and process variance to find the total variance.

Given the total variance, make an assumption about the distribution of the error term and derive a confidence interval. A common assumption is that the errors are distributed normally so that a one-sided confidence interval at the α level equals $R(w,n-w+1) + t_\alpha$ (process variance + parameter variance)$^{1/2}$ where t is from a t-distribution with the appropriate degrees of freedom.

4.6.1.1.3 Other Conditional, Non-parametric Models

1. Murphy also suggests adding an intercept to the chain ladder, so

$$q(w,d+1) = f(d)c(w,d) + j(d) + c(w,d)^{i/2} e(w,d+1) \tag{4.21}$$

2. The Mack model of credibility weighting CL and BF estimates models the ultimate losses conditional on the observations on the latest diagonal. It can be expressed as:

$$c(w,n) = Z(w,d)c(w,d)/F(d) + [1 - Z(w,d)][c(w,d) + [1 - F(d)]G(w)] + e(w,n), \tag{4.22}$$

where $G(w)$ is the prior estimate of ultimate losses for year w.

The idea of the credibility model discussed above is to minimize the variance of $e(w,n)$.

3. Robbin[54] and Venter[55] provide another credibility model for weighting together

[54] See Robbin [54].
[55] See Venter [70].

different estimates of claim count development and total loss development, respectively. Besides the CL and BF estimates of future loss emergence, they also give weight to the pegged estimate $U_0 - c(w,d)$, which does not update the total incurred as losses develop. A credibility weighting similar to Mack's is produced, but some weight goes to the pegged estimate.

4. A credibility model is proposed by Neuhaus[56]. He uses the Bühlmann-Straub credibility model, so assumes there is a parameter W for the accident year that determines the distribution. Then the assumptions are:

1. For all d, the $q(w,d)|W$ are independent;
2. $E[q(w,d)/f(d)|W] = m(W)$, so dividing by $f(d)$ grosses up incremental losses to ultimate; and
3. $Var[q(w,d)/f(d)|W] = s^2(W)$.

These lead to $Var[c(w,d)|W] = F(d)s^2(W)$, which is different than Mack's credibility model. Nonetheless, the credibility formulas turn out to be the same, although estimation of some parameters could be different than suggested by Mack.

4.6.1.2 Single Triangle, Conditional, Parametric, No Diagonal Terms

Any of the non-parametric models could have parametric assumptions introduced. Thus you could have the model:

$$q(w,d+1) = f(d)c(w,d) + c(w,d)^{i/2}e(w,d+1) \tag{4.23}$$

and assume that e is normal or t-distributed with mean 0, or follows a positive distribution shifted by its mean, like lognormal minus its mean, loglogistic minus its mean, *etc.*

Gogol[57] introduces a Bayesian estimation using lognormal distributions. He assumes:

1. $c(w,n) = U$ ~ lognormal(μ, σ^2); and
2. $c(w,n)|U$ ~ lognormal(ν, τ^2)

---

[56] See Neuhaus [47].
[57] See Gogol [26].

He then shows that $U \mid c(w,n) \sim \text{lognormal}(\mu_1, \sigma_1^{\,2})$, where μ_1 and σ_1 can be estimated from the data. This involves a credibility weighting of chain ladder and prior estimates.

4.6.1.3 Single Triangle, Unconditional, Non-parametric, No Diagonal Terms

Unconditional models estimate future development as a function of parameters of the model, with no reference to losses emerged to date. Typically the loss history in the claims triangle will be used to estimate the parameters, however.

The prototype of unconditional methods is the Bornhuetter and Ferguson.[58] As a model, this method can be expressed as:

$$q(w,d) = G(w)f(d) + e(w,d)\,, \tag{4.24}$$

where $f(d)$ is the percentage of losses paid from age $d-1$ to age d and $G(w)$ is the prior estimate of ultimate losses for the year.

Since $G(w)$ is given for each accident year, estimation of $f(d)$ to minimize $\sum e(w,d)^2$ is just a no-constant regression, so then $f(d)$ is estimated as $\sum_w q(w,d) / \sum_w G(w)$.

A popular variant is to estimate $G(w)$ from the triangle as well as $f(d)$. Strangely enough, such models have been given names like stochastic chain ladder, even though they do not estimate future development conditional on $c(w,d)$. It would be more historically accurate to call them stochastic BF models. One variant is:

$$q(w,d) = G(w)f(d)\exp[e(w,d)]\,. \tag{4.25}$$

Since all factors are presumably positive, taking the log gives:

$$\ln q(w,d) = \ln G(w) + \ln f(d) + e(w,d)\,. \tag{4.26}$$

This system is actually over-determined in that adding a number x to every $\ln G(w)$ and subtracting it from every $\ln f(d)$ would give the same estimates. To fully determine the system, one variable has to be set to a constant. One way to estimate the parameters is to minimize the sum of the squares of the $e(w,d)$'s separately for each row and each column of the triangle, which produces a system of $2n$ linear equations that can be solved for the $\ln f$'s and $\ln G$'s.

It is not too much more difficult to make the error additive, so

$$q(w,d) = G(w)f(d) + \exp[e(w,d)]\,. \tag{4.27}$$

[58] See Bornhuetter and Ferguson [7].

Then losses do not have to be positive, but solving the system (still minimizing the row and column squared error sums) ends up with $2n$ non-linear equations. These can be solved iteratively, as in the Bailey minimum bias procedure. Venter[59] gives an example of this procedure.

Another variant is to assume the expected values of all the accident years are at the same level. This might hold for a triangle of on-level loss ratios, for example. Then the model would be:

$$q(w,d) = Gf(d) + e(w,d)\,. \tag{4.28}$$

Thus, there is no dependence on w except in the error term – all the accident years are at the same level G. Since $Gf(d)$ is constant by column, the projected future incremental loss emergence is constant for each column. Thus, this model is sometimes called the additive chain ladder, although it is actually an unconditional model. Also, it is sometimes called Cape Cod, as a method by that name can be used to estimate G and the $f(d)$'s. The minimal least squares solution for each column would just set $Gf(d)$ to the average of the column, with G arbitrary.

These models can all be modified by making the error term a function of the mean. One variation is to make the variance of the error term proportional to the mean. This can be done to any of the above models, *e.g.*,

$$\ln c(w,d) = \ln G(w) + \ln f(d) + e(w,d)[G(w)f(d)]^{1/2} \tag{4.29}$$

$$q(w,d) = G(w)f(d) + e(w,d)[G(w)f(d)]^{1/2} \tag{4.30}$$

$$q(w,d) = Gf(d) + e(w,d)[G(w)f(d)]^{1/2} \tag{4.31}$$

Other powers of the mean could be used as a factor on the error term as well.

Even though some unconditional models reproduce chain ladder estimates, they can be distinguished from the chain ladder in their residuals. The chain ladder estimates each incremental cell as a factor times the previous cumulative. The unconditional models estimate the same cell as a factor times a single level for the accident year. Depending on the data, one or the other could give a better explanation for the triangle being fitted.

4.6.1.4 Single Triangle, Unconditional, Parametric, No Diagonal Terms

The model $q(w,d) = G(w)f(d) + e(w,d)$ can also be parametric, with the distribution

[59] See Venter [71].

of e specified. For instance Hachemeister and Stanard[60] take the case where $G(w)f(d)$ is the mean of a Poisson distribution, so $q(w,d)$ is Poisson distributed with that mean. They show that the *MLE* estimates reproduce the chain ladder future payments[61]. Here $e(w,d)$ can be considered shifted Poisson, i.e., a Poisson distribution less its mean. Another option would be to have $e(w,d)$ a constant times a shifted Poisson, to change the error variance:

$$q(w,d) = G(w)f(d) + e(w,d) \quad \text{where} \quad e(w,d) \sim k \; [\text{Poisson}[G(w)f(d)] - G(w)f(d)]$$

Renshaw and Verrall[62] discuss the over-dispersed Poisson, and show that it also gives the same estimate as the chain ladder. Negative binomial can be used here instead of Poisson. Actually, any of the non-parametric unconditional models can be made parametric just by making a distributional assumption. A typical example is:

$$\ln c(w,d) = \ln G(w) + \ln f(d) + e(w,d) \text{ where } e \text{ is normal}(0,\sigma^2).$$

This was proposed for instance in Kremer[63].

Often the unconditional parametric models assume that the incremental observations are independent even within an accident year. This helps with the estimation, but may be unrealistic.

4.6.1.5 Adding in Diagonal Terms

Parametric *vs.* non-parametric is a less significant distinction than it may appear, in that non-parametric models are usually fit by least squares, and so are equivalent to assuming normality. Conditional *vs.* unconditional may seem not too important in that they both estimate the parameters from the data, and we can set up unconditional models to reproduce chain ladder estimates. However, the statistical properties of the models are quite different. Conditional models gross up emerged losses to estimate future incurred while unconditional models postulate emergence as a percentage of hypothetical mean ultimates. The process of loss emergence is different for the two – they would be simulated differently for example – and the tests of goodness-of-fit of the models and the estimated variances could be quite

[60] See Hachemeister and Stanard [27].

[61] This was submitted to ASTIN Bulletin but never published, purportedly because the reviewers felt that the results were already well known at that time. It was published in a German textbook by Kremer in 1985 and by Mack in an appendix in ASTIN 1991 and by others.

[62] See Renshaw and Verrall [53].

[63] See Kremer [35].

different as well. Venter[64] gives an example where the fit of an unconditional model is quite a bit better than a fit of a conditional model to the same data. For that data the losses emerging are better explained as a percent of a constant ultimate than as a percent of the losses already emerged. Other data could give the opposite conclusion.

A distinction that is significant both apparently and in practice is whether or not calendar year effects are included in the model. These could result if inflation after the loss date affects eventual loss payments, for example. They could also come from claim department activity that makes some calendar years high and others low. Calendar year terms can be added to both conditional and unconditional models. We use $h(w+d)$ as a calendar year effect, since $w+d$ is constant on a diagonal. For example:

$$q(w,d) = G(w)f(d)h(w+d)\exp[e(w,d)]\,; \tag{4.32}$$

$$q(w,d) = G(w)f(d)h(w+d) + e(w,d)\,; \text{ and} \tag{4.33}$$

$$q(w,d+1) = c(w,d)f(d)h(w+d+1) + c(w,d)^{1/2}e(w,d+1)\,. \tag{4.34}$$

The calendar year factors can be estimated by linear or non-linear regression. Venter[65] provides some examples. However a significant issue arises: calendar year inflation induces an inflation effect in both accident year and age of claim directions, so could be difficult to separate from them. In fact one of the early papers on calendar year effects was Taylor[66]. Although Taylor refers to some earlier works, his paper was the first look at estimating calendar year effects for many actuaries, and in fact models with such effects came to be known as separation models. Taylor's model is basically Cape Cod plus inflation:

$$q(w,d) = Gf(d)h(w+d) + e(w,d) \tag{4.35}$$

That is, the data is assumed to be normalized so that there are no systematic accident year effects except as induced by calendar year inflation. This could be extended to adding accident year parameters, perhaps selectively for years with significant deviation from the original model.

An alternative would be to start with a general conditional or unconditional row-column model and add in calendar year effects for diagonals that deviate significantly from the fitted. This could pick up high or low diagonals that have been affected by specific issues for one or two years – a new computer system in the claims department for example. This would

64 See Venter [71].
65 See Venter [69].
66 See Taylor [17].

remove the distortion such issues could produce on the parameters of the original model, for example.

If the original data is adjusted for exposure changes but not price changes, the model of Cape Cod plus inflation could show the overall effect of calendar year price changes on the losses. If this in itself explains the differences across accident years, it would be evidence that inflation, in fact, is working in the calendar year direction. Years with high inflation, for example, would show up high along the whole diagonal. On the other hand, if inflation affects a line mainly across accident years, high inflation years would affect the accident year only. This distinction should show up in the residuals of the model.

Calendar year effects present issues for squaring the triangle also. The age effects are quantified to the end of the square but the diagonal effects are only measured up to the last diagonal. If only a few specific diagonals have been picked up as unusually high or low, then perhaps no further projection in the calendar year direction would be called for. This would be a finding that future inflation does not affect open claims – all inflation effects have been accounted for by the accident year factors. Even in this case, however, future deviations up or down could affect the variance of results.

On the other hand, if the model fitting has found significant ongoing calendar year trend, this should be projected in filling out the future results, both for the mean and deviation from the mean. The estimated trend for the latest diagonal could be the mean trend projected, but this could be modified by econometric analysis of expected inflation.

4.6.1.6 Restricting Parameter Variation

Much of the recent literature has addressed methods for reducing the number of parameters in a model by restricting parameter variation. The typical conditional model starts out with $n-1$ parameters for $n(n-1)/2$ data points, while the unconditional model would have $2n-1$ parameters. Cape Cod reduces this down to n by forcing all accident years to be the same. Adding calendar year inflation could double this. Besides reducing the degrees of freedom, many parameters end up being statistically insignificant. For instance, it is not unusual for all ages after about three to have $f(d)$ less than its standard error.

The simplest way to constrict the parameters is to force some of them to be the same. Perhaps none of the factors $f(11)$ to $f(19)$ are significant, but they all are close enough to each other that if you allow just one parameter for all those ages it will be significant. The same thing can be done with calendar years or with accident years, which would be in the direction of Cape Cod, but not all the way there. (A related alternative would be to force a trend line, so all the parameters in a given range fall on the line, using up only two degrees of freedom, or a curve can be used instead of a line, such as a power curve. That might also be convenient for projecting age factors beyond the triangle.

An example is provided by Barnett and Zehnwirth[67] who discuss the CL model with an accident year trend:

$$q(w, d+1) = c(w, d) f(d) + a(d) + b(d) w + c(w, d)^{1/2} e(w, d+1) . \qquad (4.36)$$

Here $a(d) + b(d)w$ represents a constant part of development plus the level of accident year w.

There are also smoothing techniques that can be used to restrict the degree that a parameter can differ from the one next to it. One simple version of this was presented by Gerber and Jones.[68] If the true parameter is changing each period by a random amount with variance a, and the direct estimation procedure has a variance v around the current true parameter, then the smoothed estimate is an update from the previous smoothed estimate based on a credibility weighting of the latest direct estimate and the previous smoothed estimate. If the credibility of the i^{th} direct estimate is Z_i, then the $(i+1)^{th}$ direct estimate's credibility satisfies $1/Z_{i+1} = 1 + 1/(Z_i + a/v)$, while the smoothed estimate from period i gets weight $1 - Z_{i+1}$ to produce the smoothed estimate for period $i+1$. Here Z starts at 1, since there is no previous smoothing for the first point, and goes down to a limit of $(J^2/4 + J)^{1/2} - J/2$,where $J = a/v$.

This is a simple example of a credibility smoothing procedure called the Kalman filter, and is also related to exponential smoothing. This filter and a generalization of it aimed at generalized linear models are discussed in Taylor, McGuire and Greenfield[69].

The variances a and v do not have to be constant. For example, if the regression diagnostics suggest that the parameter has changed a lot from one period to the next, a high value of the change variance a can be postulated at that point, which would give high credibility to the direct estimate of the parameter and low weight to the previous smoothed estimate. Also smoothing can work in two directions – past and future. The smoothing

---

67 See Barnett and Zehnwirth [2].
68 See Gerber and Jones [23].
69 See Taylor, McGuire and Greenfield [67].

could be reversed at the last point and continue backward as if it was still going forward, and even reverse again at the beginning, etc. Whenever the process stops, the smoothed estimates are the final estimates of the parameters.

Zehnwirth and Barnett use parameter restrictions like these in a family of models that basically adds calendar year trend to Kremer's lognormal model. The main idea is to have parameters to model effects in each of the three directions (accident year, age of development, and calendar year). The model uses logarithms of incremental data. This postulates that trends are linear on a logarithmic scale and are easier to discern in incremental data. Also the incremental data at each period is the new information that needs to be modeled, where the cumulative losses are a mixture of new and old information.

The following modeling schema was presented at the 2002 CLRS.[70] In our notation:

$$\ln q(w,d) = G(w) + \sum_{i=1}^{d} f(i) + \sum_{j=2}^{w+d} h(j) + e(w,d). \tag{4.37}$$

Here $G(w)$ is the level of accident year w, $f(i)$ is the single period development, and $h(j)$ is the increase in calendar year cost levels in one year. This is called a modeling schemata and not a model because it would be inappropriate to just estimate the G, f, and h parameters by *MLE*. There is multi-collinearity between the calendar year and accident year effects, so direct estimation would not be meaningful. It may be reasonable, however, to have constant or gradually changing trends within some time frames, perhaps with jumps to new levels when regime change takes place. A process of model identification, estimation, and validation is needed to find meaningful parameters that work together to model the data triangle within this schema. The model is not fully specified until the parameter restrictions have been established.

The age d factor could be represented as $F(d)$ instead of the sum of the individual $f(i)$'s up to age d, but doing it this way makes a difference in the application of parameter restrictions. For instance, $f(i)$ could be constant for several consecutive ages, which would (assuming a negative trend) represent an exponential decline in payments before application of calendar year trend. If the triangle ends with a constant value of $f(i)$ for the last few ages, this could be projected beyond the triangle to continue the pattern.

Similarly $H(w+d)$ could be used to represent the sum of the $h(j)$ up to $w+d$. This

[70] It is available at www.casact.org/coneduc/clrs/2002/handouts/barnett1.pdf.

would be a cumulative factor, but again it is probably more intuitive to apply parameter restrictions to the individual annual trends $h(j)$.

Another way of framing this scheme would be to make the parameters average trend factors:

$$\ln q(w,d) = p(w) + dq(d) + (w+d)r(w+d) + e(w,d), \tag{4.38}$$

where $q(d)$ is the average development age trend through age d; the development age trend represents the expected change from one development age to another.

$r(w+d=t)$ is the average diagonal (payment year) trend; the payment year trend represents the mean of the (random) trend between payment year $w-1$ and w. If the data is inflation adjusted for price or wage inflation, then the trends along the payment year usually represent social inflation, and e(w,d), the error term, is distributed normal($0,\sigma^2$).

In any of these setups, before a general framework for the model is decided upon, a preliminary analysis can be performed on the loss array to determine the existence of trends. It is difficult to determine the existence of trends separately in each of the three directions. A preliminary determination of existence of trends in any particular direction can be established by fitting a single development factor model such as:

$$\ln q(w,d) = p + dq(d) + e(w,d), \tag{4.39}$$

where $p(w)$ is constant across accident years and $r(w+d)=0$.

To estimate the observed log incremental losses and then charting and examining residuals sorted by development year, accident year, payment year, and fitted values. If there is a trend in a particular direction, it will be shown by the distribution of residuals by that direction. Then based on the results of this preliminary analysis the original scheme described by Zehnwirth and Barnett can be further specified depending on the observed trends in any particular direction.

The accident year trend and the development age trend are essentially considered independent of each other, as their trend vectors are orthogonal. However, the payment year trend vector is not orthogonal to either the accident year direction or the development year direction. That is, a trend in the payment year direction is also projected onto the development age and accident year directions. Similarly, accident year trends are projected onto payment year trends. The relationship between the three directions is as follows:

diagonal (payment year) trend = accident year trend + development age trend.

The model scheme with cumulative trends can be re-expressed as:

$$\ln q(w,d) = [p(w) + wr(w+d)] + d[q(d) + r(w+d)] + e(w,d). \tag{4.40}$$

The components in braces can be considered accident year and age components, but due to the calendar year trend, the accident year level at a cell depends on the time since the accident, and the average development at any age is affected by accident year. This shows that the calendar year trend $r(w+d)$ projects in both other directions.

The proposed model can also be modified if the loss triangle array being modeled exhibits different trends during different periods of time. In such a scenario we could divide the data into blocks of time periods so that each block of data exhibits homogeneous trends and then model each of these blocks of data separately or have additional parameters to model differing trends for different time periods.

Modeling log incremental puts certain limitations on the model, such as incremental amounts being estimated cannot be 0 or negative. As a result, this model may be suitable for modeling paid and case reserve amounts instead of paid and incurred amounts, which adds the advantage of possibly being able to test case reserve adequacy if it is thought to be changing.

An additional issue associated with multi-parameter models such as the one described above is the multi-collinearity between independent variables described above. One way to get around this issue is to use the Kalman filter or exponential smoothing to restrict the parameters.

Other methods of smoothing could be used such as cubic splines. England and Verrall[71] give a framework for some parametric unconditional models using spine smoothing:

$$E[q(w,d)] = m(w,d) \text{ and} \tag{4.41}$$

$$Var[q(w,d)] = km(w,d)^i, \tag{4.42}$$

where the distribution of $q(w,d)$ is either normal, Poisson, gamma, or inverse gaussian for i = 0, 1, 2, 3 respectively.

$$\ln m(w,d) = (w+d)b + c + s_w(w) + s_d(d) + s_d(\ln d). \tag{4.43}$$

Here the s's denote smoothing functions. With no smoothing there is a parameter for

[71] See England and Verrall [17].

each accident year and age. With infinite smoothing there is one parameter for all accident years combined and age factors are forced to fall on a curve that is a linear combination of d and $\ln d$ (*i.e.*, a so-called Hoerl curve). The degree of smoothing can be controlled separately for w and d.

This is a framework for models that fall into the category of generalized linear models. It does not include truly non-linear models like:

$$q(w,d) = G(w)f(d)h(w+d) + e(w,d). \tag{4.44}$$

However, such models are not difficult to fit with modern search techniques, even with e following a t-distribution, log-t, or other distributions not in the $i = 0$ to 3 list. The model can also be fit non-parametrically to minimize the squared-error sums of each row, column, and diagonal of the triangle by the iterative procedure of Bailey minimum bias. Given starting values for the parameters, the iterative equations for the next values are:

$$h(w+d=t) = \sum_{w+d=t} q(w,d)G(w)f(d) / \sum_{w+d=t} G(w)^2 f(d)^2 ; \tag{4.45}$$

$$G(w) = \sum_d q(w,d)h(w+d)f(d) / \sum_d h(w+d)^2 f(d)^2 \text{ ; and} \tag{4.46}$$

$$f(d) = \sum_w q(w,d)h(w+d)G(w) / \sum_w h(w+d)^2 G(w)^2 . \tag{4.47}$$

Convergence is usually fairly fast. Thus, with a large enough triangle this model has problems neither with the non-linearity nor with the degrees of freedom. One problem it does have is that inflation is being measured with both accident year and calendar year parameters, whose effects overlap. Another is that many of the parameters will not be significant. Thus some parameter restriction methods will usually have to be employed. A starting point might be to use a single value of G for all years, then look at the residuals to see if more G's are needed.

With smoothing procedures like filters there is an issue of how many parameters are in the model, so the degrees of freedom can be computed. One suggestion, for instance proposed by Ye[72], is to define the generalized degrees of freedom used up (*i.e.*, number of parameters in the model) as the sum over all the observations of the derivative of the fitted value at that observation with respect to the observation. This can be approximated, for instance, by making a small change at an observation and seeing how much the fitted point changes, and repeating for all observations. As an example, suppose you fit a cubic polynomial to four points. The polynomial will go through all four. Changing any of the

[72] See Ye [80].

points by a small amount will change the fitted values by the same amount, so the sum of the derivatives will be four. Thus, all degrees of freedom are used up. One way to think of this is that each data point gets a degree of freedom initially, which gives it power to pull the model towards itself. If it can completely control the model, so any change in the point changes the fitted value by the same amount, the model has used up its entire degree of freedom. If it can only pull the fitted value by half of the change, the model has only used ½ of its degree of freedom, etc.

4.6.2 Multiple Triangle Models

The initial multiple triangle models were just splits of one set of loss data into frequency and severity components. For instance, Fisher and Lange[73] use a report year approach to develop claim payout patterns and size of claim by settlement date.

Some more recent innovations were presented at the 2003 ASTIN Colloquium. For instance, Quarg[74] showed that the paid to incurred ratio at any point in development contains information relevant to future development in both the paid and incurred triangles. Conversely, if paid is high compared to incurred, then higher incurred and/or lower paid development is likely. If incurred is high compared to paid, then higher paid and/or lower incurred development is likely. Thomas Mack in his discussion of this paper suggested multiple regression models for both paid and incurred. Using subscripts P and I to denote paid and incurred, such a model is:

$$q_I(w,d+1) = f_{II}(d)c_I(w,d) + f_{PI}(d)c_P(w,d) + e_I(w,d+1)\text{; and} \tag{4.48}$$

$$q_P(w,d+1) = f_{IP}(d)c_I(w,d) + f_{PP}(d)c_P(w,d) + e_P(w,d+1)\,. \tag{4.49}$$

All the f's would be expected to be positive, since higher paid leads to higher future incurred and higher incurred leads to higher future paid.

Correlation in development between lines of business can be handled similarly. If lines tend to develop in a correlated fashion, then information from one line can improve the estimates from another. The above multiple regression model could be used, where instead of paid and incurred, the triangles would represent different lines of business. Another paper at the 2003 ASTIN Colloquium which discussed correlation issues is Gillet and Serra[75] .

[73] See Fisher and Lange [19].
[74] See Quarg [51].
[75] See Gillet and Serra [25].

4.6.3 Models of Individual Claims Histories

Modeling individual claim development can give an alternative view to future payment estimation by triangles of sums of claims, but it can also help with layer pricing, net reserves after reinsurance, and distributions of ultimate claims. Models proposed include the transition matrix model and conditional distributions.

4.6.3.1 Transition Matrix

The transition matrix method was introduced by Hachemeister[76]. A follow up paper by Hesselager[77] also discusses this method. The method is pretty simple but a good deal of individual claims data is needed. Claims are put into categories that include size ranges, status as to open, closed, unreported, percentage paid, etc. Then probabilities are calculated of claims in one category moving to another category at the next evaluation. These probabilities can be arranged in a matrix based on the combination of starting and ending categories. Then the vector of claims by category can be multiplied by this matrix to get the vector of claims by category at the next evaluation. Applying this many times can get to ultimate – although different matrices at different stages of development may be needed. Several companies that have tried this approach seem to feel it works well.

4.6.3.2 Conditional Development

Conditional development, broadly speaking, tries to find the conditional distribution of sizes a claim may have given what we know about it today.

NCCI uses a variation of this method for excess pricing studies. They have a large number of claims that are not reported after age five, so they need a 5th to ultimate development procedure that includes the spread in claim sizes that takes place during this development. In a study of a sample of claims with a longer history available, they use maximum likelihood to fit a distribution to the individual claim development factors for future development for claims open at 5th. For claims that close by the later evaluation, the development factor is known so the contribution to the likelihood function from that claim is the probability density at its development factor. For claims that are still open, all we know for sure is that the development factor is greater than the ratio of the latest paid to date

[76] See Hachemeister [28].
[77] See Hesselager [32].

amount to the incurred amount at 5th for the claim. So the contribution to the likelihood function for such a claim is $1-F(x)$, where x is the paid-now-to-old-incurred ratio.

NCCI concluded in one study that for 5th to ultimate an inverse gamma distribution with CV of 90% best modeled the development factor distribution. Rumor has it that recent studies have suggested a lower CV, perhaps as low as 40%, but it is not publicly known if the development dates are comparable between studies. To apply this approach, in general we would need to model other reports than 5th. Presumably the losses spread more when they are less developed. Perhaps an inverse gamma could be used with a CV that reduces for more mature losses.

Another problem with this method is that it does not use the latest incurred information on open claims. One way to do this would be to replace the open claim by a number of claims that range from the paid amount on upward, and scale the log likelihood function so that all of these together represent a single claim. The California workers' compensation rating bureau uses the current incurred development factor for open claims as the censorship point, which is not strictly what we know, but it does reflect the latest estimate and could be a reasonable approximation.

Some published papers that discuss conditional development include Taylor, McGuire and Greenfield [78] mentioned above and Norberg's[79] papers. Also a major study of French motor vehicle claims found that claims development varies by claim size, with a Weibull distribution giving a good approximation to the conditional development given the claim size.

Taylor et. al. model claim severity by a fairly heavy-tailed claim distribution where the mean claim size is conditioned on the time of claim occurrence, the time of settlement, and operational time, which is the proportion of total claims that closed before it did. Inflation is modeled both as a function of accident date and settlement date, and the mean claim size effects in both directions are also functions of the order of settlement. This allows development to vary by claims characteristics, but does not provide for a single claim developing into a range, which is what is desired for estimation of ultimate severity distributions.

---

[78] See Taylor, McGuire and Greenfield [67].
[79] See Norbeg [48].

5. COMPARE, CONTRAST AND DISCUSS RESULTS

This section is an illustration of how we might evaluate the techniques discussed in Section 4. It is important to appreciate that the results of any evaluation will depend on the data used. A technique is only "good" with respect to particular data. A technique is "useful" if there is a wide range of data for which it is "good". To decide whether a technique is useful, it must be evaluated on a large number of typical datasets, which is outside of the scope of this paper. Certainly there are no published studies of this type, and there is only anecdotal evidence on what features of models are important.

We have chosen two techniques to illustrate the application of the evaluation criteria of Section 3.1: the estimated range (*ER*) method of Section 4.3 and the over-dispersed Poisson model (*ODP*) of Section 4.6.1.4. The (*ODP*) model is a special case of the generalized linear model with log link function described in Section 3.2.2. An example of its application is given in Appendix B. Other techniques are mentioned in passing but are not systematically evaluated.

5.1 Criteria for Selecting an Appropriate Modeling Technique

5.1.1 Criterion 1: Aims of the Analysis

Most of the techniques discussed in Section 4 will provide at least the mean and standard error of the distribution of ultimate losses or future payments. The *ER* method and the *ODP* model both do this. In addition, the *ER* method allows us to estimate the percentiles of these distributions. The *ODP* model has no explicit distribution associated with it, so a distributional assumption would be required to estimate percentiles. Distribution-free models such as Mack's model in Section 4.6.1.1.1 produce only means and standard errors, but if percentiles are required, distributional assumptions can be added, as in Murphy's models of Section 4.6.1.1.2.

As discussed in Section 3.2.2, the standard error must include both process variance and parameter uncertainty. For a model, it is usually clear how this should be done (although it may be necessary to make some simplifications for computational convenience). For a method, however, the only way to verify that the calculation of the standard error is correct is to check an equivalent model, or to validate it on a large number of triangles.

In the case of the *ER* method, there is good reason to believe that it may understate the parameter uncertainty of the total – there is an allowance in the standard errors of the individual development factors for parameter uncertainty, but there is no allowance for correlation between the accident years resulting from parameter uncertainty. This is likely to be important when the expected future payments are large for several accident years and the triangle is relatively small (so the parameter uncertainty may be large).

It is particularly important to test distributional assumptions (see Criterion 14) if estimates of percentiles are required. This is not possible for methods (unless there is an equivalent model), which means there is considerable risk in relying on a method like the *ER* to estimate percentiles.

5.1.2 Criterion 2: Data Availability

Both the *ER* method and the *ODP* model require only a triangle of data.

Zero or negative data may create issues for the *ODP* model (there are similar issues with all models that contain logs of means or of data). The *ODP* model is unable to estimate its full set of parameters when any of the row or column sums is zero or negative. Some software packages may not allow any data to be zero or negative, so those values would have to be omitted altogether from the calculation (this is done in the example in Appendix B).

In that case, there are several possible assumptions we could make for these missing parameters (usually they are development year parameters). For example, we could assume that the mean and variance of the corresponding incrementals are zero (which corresponds to the chain ladder estimates). Alternatively, we could take the more conservative approach of setting the missing parameters to the lowest or last development year parameter. The example in Appendix B, which is based on incurred loss data, uses the last development year parameter in place of missing development year parameters. Zero or negative values are less likely to occur in paid loss data and so will be less of an issue. Figure 5.5 illustrates the sensitivity of parameter estimates to the treatment of negatives and zeroes.

The quality of the available data should always be considered. If only one triangle is available, this is not an issue, but when more than one triangle is available, a careful assessment should be made of whether one or the other appears to give more reliable forecasts.

When there is enough paid loss data available to give an adequate estimate of the losses in the later development periods, it may be the case that the predictive quality of the paid loss data is better than that of the incurred loss data, as they are not influenced by the subjectivity (changing company policy, individual preferences) that can affect case reserve estimates. In addition, some actuaries have suggested that the use of distributions of future payments for risk analysis (*e.g.*, risk based capital) should focus on paid loss data in order to keep the subjectivity of case reserve estimates from biasing the risk measure.[80] If a method requires projections of the number of claims closed, for example, it should be checked that their stability has not been affected by changes within the company.

5.1.3 Criterion 3: Non-Data Specific Modeling Technique Evaluation

To validate a technique against historical data, we would need many sets of data where the "rectangle" had already been completed. Each value of actual ultimate losses for the rectangle would correspond to a percentile of the forecast distribution for the method applied to the upper triangle. The distribution of these percentiles over many datasets should be a sample from a uniform distribution on [0%,100%]. As far as we are aware, no such validation has been published, for any of the methods or models of Section 4. An example of how to validate a single dataset is provided in the discussion of Criterion 16 below (Section 5.3.4).

An alternative is to use simulated data. This of course only tells us how the technique behaves on data resembling the simulated data, but it may still be useful in identifying deficiencies of a model and their practical impact.

With a method such as the *ER*, it is not possible to test the assumptions against what is known about the process generating the data, as there are no explicit assumptions. However, of the models in Section 4, the closest to this method would be Murphy's model with $i = 2$, as this model's estimates of ratios are based on averaging the individual ratios. Then assessing the reasonability of this method/model would include evaluating the assumptions of Murphy's model.

The *ODP* model assumes (as do many of the Section 4 models) that the pattern of development is the same in all accident years, that there is no or constant inflation, and there is no dependence between accident years. Past experience may have shown whether these

[80] For example, see Shapland [58], p. 336 and the definition of Risk in Section 2.

assumptions are, or are not, usually satisfied.

Another test of the assumptions of a model for loss data is its behavior under scaling and inflation. Clearly if loss data is multiplied by 1000, the forecast probability distribution should be scaled by the same multiplier (this would not be true of count data, however). For example, the Poisson generalized linear model (*i.e.*, without the over-dispersion factor) does not scale, so it should not be used for loss data, although it may be appropriate for count data. Similarly, if the loss data is inflated by 10% per annum, the forecast probability distribution should be inflated by the same rate. The *ODP* model does not satisfy this property.

5.1.4 Criterion 4: Cost/Benefit Considerations

Like most methods, the *ER* method has a low cost, because it can be implemented relatively easily in a spreadsheet. It is possible that appropriate diagnostics could be designed that would indicate when the method could safely be used, but we are not aware of any available at present. It appears that the standard errors of individual accident year totals may be reasonable if the underlying estimation method for ratios is sound (which of course needs to be verified). However, the standard error of the total of all accident years may be significantly underestimated due to the potential for parameter uncertainty or inflation correlating over accident years.

The *ODP* model (and related generalized linear models) is sufficiently complicated for its implementation in a spreadsheet to require careful validation against a statistical package. Because it may be numerically unstable in some cases, it would be unwise to rely on a "do-it-yourself" implementation, so specialist statistical software is probably required. Some learning time and customization of the software (for example, to calculate standard errors of distributions of future payments) would be necessary unless purpose-built software was purchased. The usual form of the *ODP* model does not allow the modeling of superimposed inflation and is over-parameterized, although it could be extended to remedy these defects (see Criterion 18).

5.2 Overall Model Reasonability Checks

5.2.1 Criterion 5: Coefficient of Variation by Year

For the *ER* method, the requirement that the coefficient of variation (CV) of the liabilities is generally smaller for later accident periods means $S^2(d)/[F(d)-1]$ should be an increasing function of d. Some algebra shows that this will be the case when $f(d)>1$, $F(d+1)>1$ and $s^2(d)$ is "small enough."

In the example of Appendix B (incurred loss data with 40 periods, denoted below as IL40), many of the later development factors are less than one. Note that in all examples of the application of *ER* we will use arithmetic averages to calculate ratios, as this is the most logical match to this method. For the last seven accident periods, $f(d)>1$ and $F(d+1)>1$. For these periods, CV for accident liability totals decreases with increasing accident period, as this criterion says should generally be the case. For the corresponding paid loss data (denoted below as PL40), there are two periods where $s^2(d)$ is large enough to make the CV not monotonically decreasing (see Figure 5.1). It is likely that the analyst would consider these large discontinuities implausible, casting doubt on the reasonability of this model for this data.

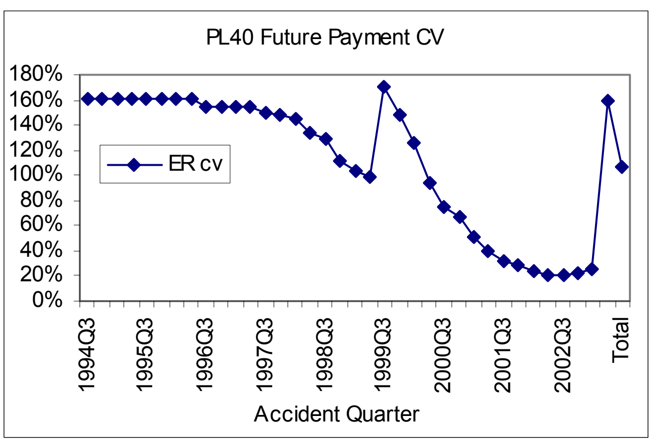


Figure 5.1 Coefficient of variation of liabilities versus accident quarter for the estimated range method applied to the paid loss data PL40

The results of applying both the *ER* method and *ODP* model to the data in Taylor and

Ashe[81] (denoted below as TA83) are shown in Figure 5.2. This data has relatively stable exposures from year to year, so it would be expected to satisfy this check. However, the CV increases with increasing accident year for both techniques in some periods. It appears that this is due to over-parameterization in the case of the *ODP* model (see Criterion 18).

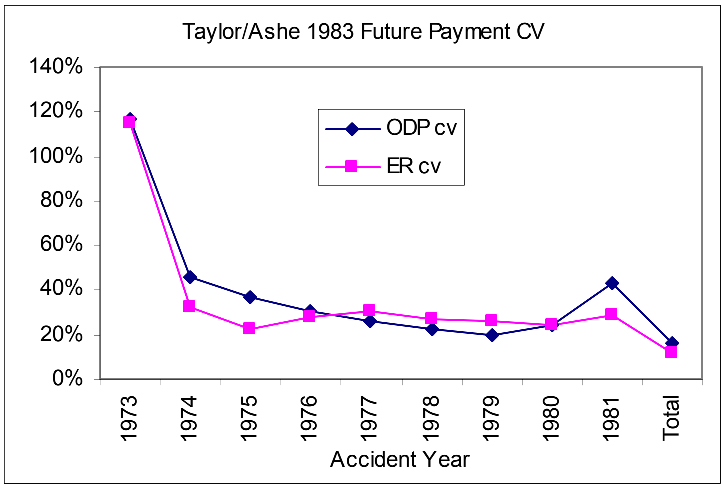


Figure 5.2 Coefficient of variation of liabilities versus accident year for the estimated range method and the over-dispersed Poisson model applied to the Taylor/Ashe [82]data

5.2.2 Criterion 6: Standard Error by Year

For the *ER* method, the requirement that the standard error is generally largest for later accident periods means that $c(w,d)S(d)$ should be greater than $c(w-1,d+1)S(d+1)$. This will certainly be the case if $c(w-1,d+1) \leq f(d)c(w,d)$, and this is likely to hold if the underlying exposures are relatively stable from year to year.

This is the case for the *ER* method with the IL40 and PL40 data – the standard error always increases if the ultimate increases, and sometimes it increases even when the ultimate decreases, particularly in the later accident periods when the variability in the corresponding development period is larger. It is also the case for the TA83 data, which has a more uniform exposure – it has monotonically increasing standard errors for both techniques (see Figure 5.3).

[81] See Taylor and Ashe [63].
[82] See Taylor and Ashe [63].

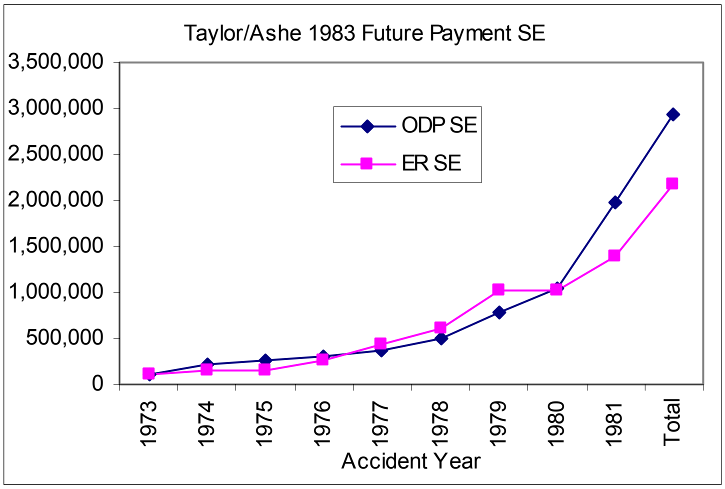


Figure 5.3 Standard error versus accident year for the estimated range method and over-dispersed Poisson model applied to the Taylor/Ashe data

5.2.3 Criterion 7: Overall Coefficient of Variation

The requirement that the CV of total liabilities be smaller than for any accident period does not hold when one period has a very large CV of liabilities compared to the other periods. This can happen with the ER method when one development period has a higher variability in its development factors than the others. For example, this requirement does not hold when the ER method is applied to the PL40 data, where the CV of the last accident period is very high (see Figure 5.1).

On the other hand, if one period has a very small CV compared to the other periods, this requirement may fail. For example, the earliest accident periods of the IL40 data have a CV of zero under the ER method, so the CV of the total liabilities is larger than this.

If the periods with zero or negative CV are excluded, this criterion holds when the *ER* method is applied to the IL40 data. This criterion also holds for both the *ER* method and the *ODP* model when they are applied to the TA83 data (see Figure 5.2).

5.2.4 Criterion 8: Overall Standard Error

The requirement that the standard error of the total be larger than for any accident period will always hold for the *ER* method, as the variance of the total is the sum of the individual variances. Figure 5.3 shows that it holds when the *ODP* model is applied to the TA83 data.

5.2.5 Criterion 9: Correlated Standard Error & Coefficient of Variation

There is no description for the *ER* method and the *ODP* model of how the results of different triangles should be combined, so it is not possible to test this criterion for these techniques.

5.2.6 Criterion 10: Reasonability of Model Parameters and Development Patterns

As the *ER* method only relates to the calculation of standard errors, there is little that "common sense" can say about its results, other than that they should behave according to the criteria above and that the CV would be expected to vary smoothly between accident years. A pattern such as that shown in Figure 5.1 appears to violate reasonability. It is a result of the fact that the estimate of the standard error in this method is very sensitive to outliers. For example, a single high value in accident quarter 1Q1998, development quarter 17, produces a high standard error in the accident quarter totals from 3Q1999 onwards. Several low values in the first development quarter are the main reason for the very high standard error in the last accident quarter, 2Q2003.

The parameters for the *ODP* model are easiest to assess for reasonability as fitted values, either on the original $ scale, or on the log scale. The accident period parameters for the ODP model fitted to the PL40 data are shown in Figure 5.4 on the $ scale. The factor of two between the last two parameters is surprising, as the actual data value for the last accident quarter is only 40% higher than the average of the previous four values in the same development period. However, the fitted value in the last accident period is always equal to the actual value when using the ODP model, so the last fitted value is obliged to be 333.

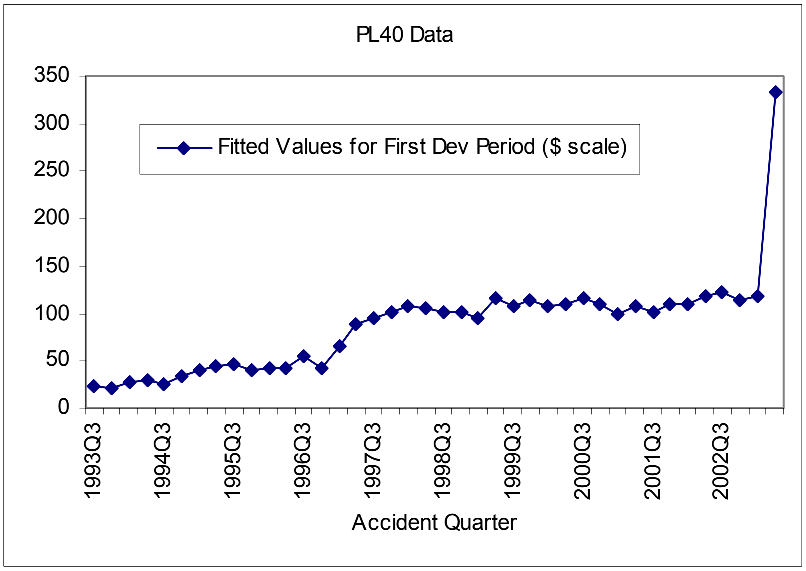


Figure 5.4 Fitted values ($ scale) for the first development quarter versus accident quarter for the over-dispersed Poisson model applied to the PL40 data

The corresponding development period parameters are shown in Figure 5.5, this time on a log scale, so that the smaller values can be seen clearly. Between development periods 5 and 15 there is a reasonably linear trend, corresponding to an exponential decay. After that, it would appear that the parameters are just "noise", and this is verified by comparing the parameter estimates with their standard errors. This is not surprising, as there is very little data greater than zero after development period 17. In fact, there is no data greater than zero at all in periods 26-28 and 30-39. The assumption was made that the parameter in those development periods should be set to the last parameter that could be estimated. From Figure 5.5, it appears that this will over-estimate the forecast in those periods, so an alternative assumption might give a better result.

It is possible to include negative and zero values in the estimation provided that the sum for the development period is positive. The dashed line in Figure 5.5 shows that this has a significant effect on some of the estimates. It would appear that the linear trend from development period 5 extends as far as development period 25, after which the data is essentially zero. It seems likely that it would be possible to find an adequately fitting model with fewer parameters (see Criterion 18).

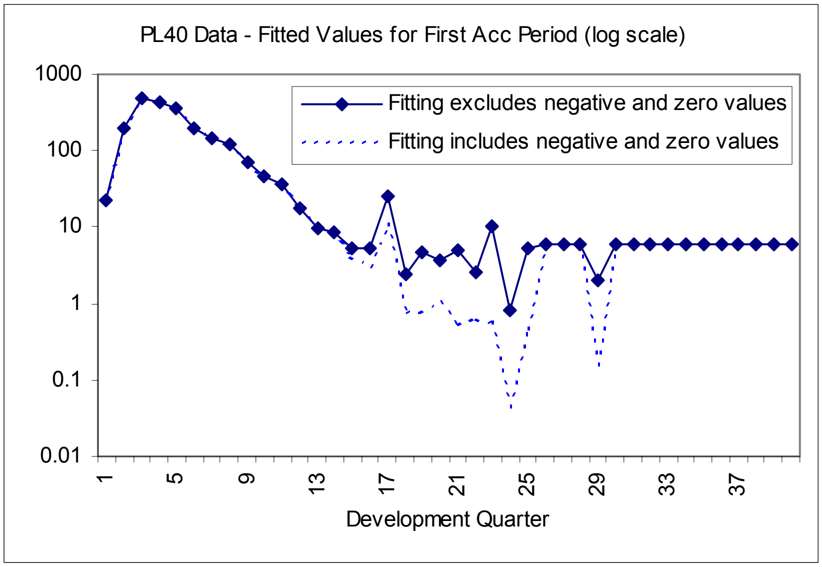


Figure 5.5 Fitted values (log scale) for the first accident quarter versus development quarter for the over-dispersed Poisson model applied to the PL40 data

5.2.7 Criterion 11: Consistency of Simulated Data with Actual Data

Without a model, it is not possible to simulate data, so this check can only be done on the *ER* method if a corresponding model is specified. The way the variance of the loss development factors is calculated in the *ER* method suggests the following underlying model: for any given pair of development periods, the individual accident year factors are randomly chosen from some distribution with a variance that does not depend on the accident year. This is precisely Murphy's model with $i = 2$. The usual assumption is that the distribution is normal. With these assumptions, it is possible to create simulated data, although it is possible that negative ratios will occasionally be generated if the variance is large.

Three triangles were simulated as follows:

1. Ratios were generated for each pair of development periods and each accident period, using the mean and standard deviation of the ratios fitted to the last 10 accident quarters of the PL40 data (PL10).

2. These ratios were multiplied into the first development period data for PL10.

Then a simple model was fitted to each of these triangles and the original PL10 data. This model was fitted to the logs of the data, had one parameter for each development period, and a single trend parameter in the calendar direction. The residuals from this model are plotted against accident period in Figure 5.6. It is clear that the original data (bottom left) has different properties than the simulated data. In the simulated data, random variation in the second development period is propagated and amplified in the later development periods. In the real data, this does not happen. It appears that the chain ladder assumptions are not appropriate for this data.

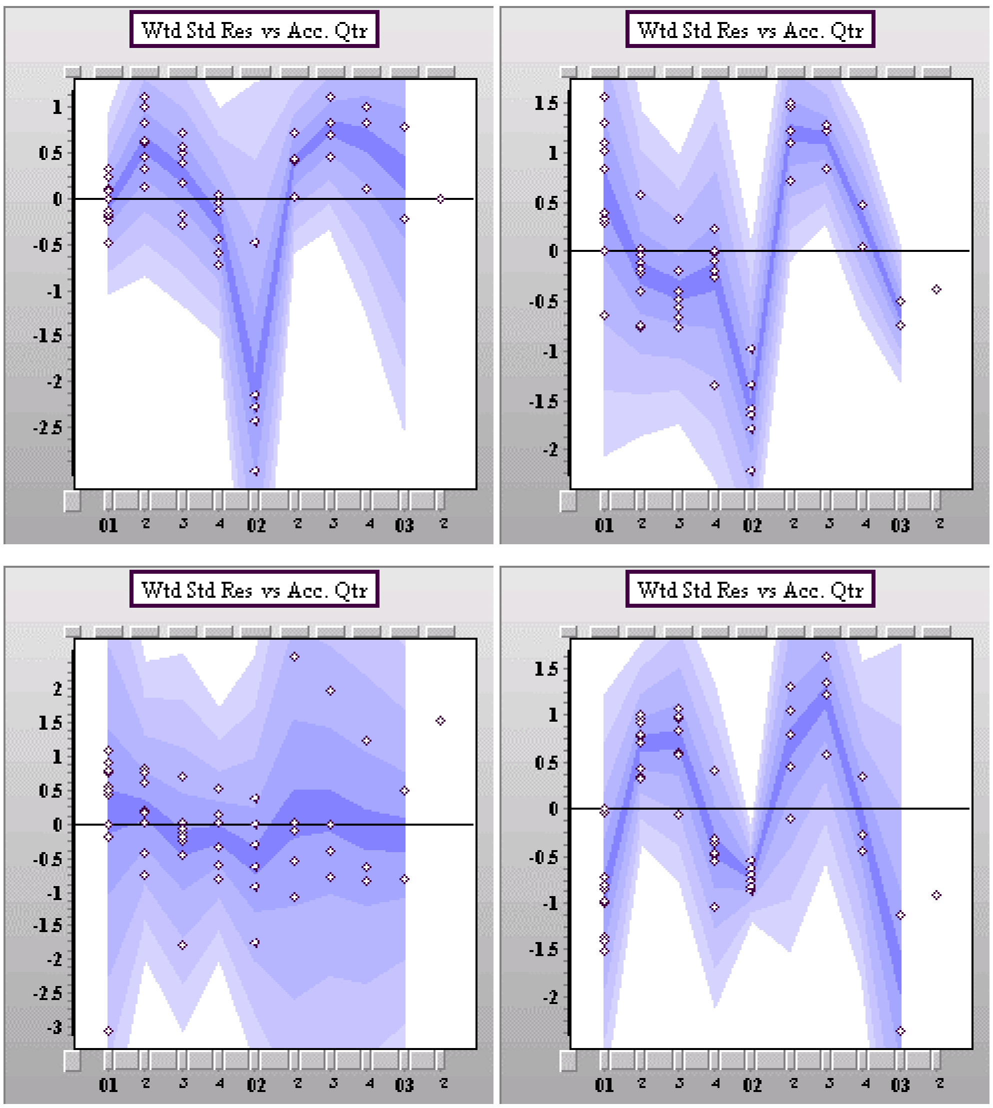


Figure 5.6 Residuals plotted against accident quarter for three sets of simulated data and one of real data (PL10)

There is a difficulty with testing this criterion on the *ODP* model – we need to assume some distributional form for the errors to perform the simulation. Either normal or negative binomial would be reasonable choices, but it is possible that the data might follow some other over-dispersed Poisson distribution. Some differences between the real and simulated data could be due to differences in this error distribution.

5.2.8 Criterion 12: Model Completeness and Consistency

If additional information is incorporated, the model prediction may be improved. For example, if there is a relevant index of inflation available, the plots of residuals versus calendar period, suggested under Criterion 15, could be compared with and without inflation adjustment, to see if one or the other appears more like a random sample. This could be useful for models like the *ODP* that have no parameters in the calendar direction. Similarly, an exposure measure, such as number of policies, may be used to normalize the data. Plots of residuals versus accident period may indicate whether this gives any improvement.

Prior information may be useful when there is a large amount of uncertainty in any aspect of the model parameter estimates. For example, if the triangle must be projected to future development periods, information from other similar triangles may be used in conjunction with information within the triangle. However, the prior information should always be appropriately weighted by its uncertainty compared to the uncertainty of the estimate from the triangle.

There is no obvious way to integrate other information into the results of the *ER* method. Consistency of assumptions for future development trends with the trends in the data could be difficult to determine for both the *ER* method and the *ODP* model, as neither explicitly estimate trends. However, it would be possible to use the estimated levels for later development periods to fit a parametric curve and to project that curve into future development periods. It is less clear how to estimate the uncertainty associated with that projection.

5.3 Model Goodness-of-Fit and Prediction Error Evaluation

5.3.1 Criterion 13: Validity of Link Ratios

If the *ER* method is used in conjunction with link ratios, it would be advisable to apply tests for the validity of link ratios, such as those in Barnett and Zehnwirth[83] and Venter[84]. As an illustration, we will apply some of these tests to the Murphy model with $i = 2$ applied to the IL10 data.

Venter's second test compares this model with some alternative models. The plot on the left in Figure 5.7 shows that there appears to be a linear relationship between the incremental in the second development period and the cumulative in the first development period, as expected under Murphy's model. However, the plot on the right shows that there is also a linear relationship between the incremental in the second development period and the accident period. Statistically, the second relationship provides a better fit to this data. This suggests that there may be a better alternative model to link ratios, at least for this development period.

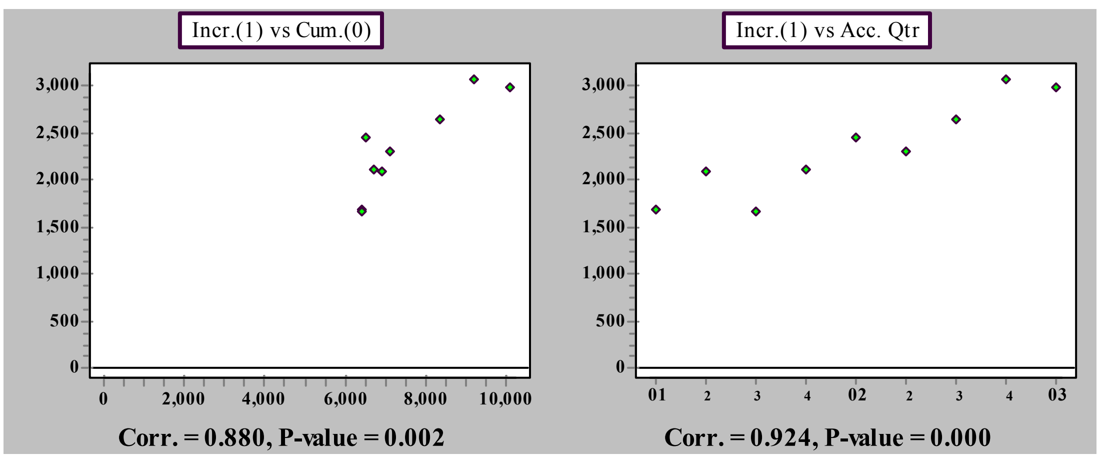


Figure 5.7 Incrementals in the second development period of the IL10 data plotted against the cumulative in the first development period (left) and the accident quarter (right)

Venter's fourth and sixth tests relate to patterns in the residuals against accident period and calendar period respectively. The residuals do not appear to be random (see Figure 5.8), particularly when plotted against accident period – the residuals are mostly greater than zero in the latest six accident quarters, so the actual data is nearly always higher than the fitted values from the model. This suggests that this is not a good model for this data.

[83] See Barnett and Zehnwirth [1].
[84] See Venter [71].

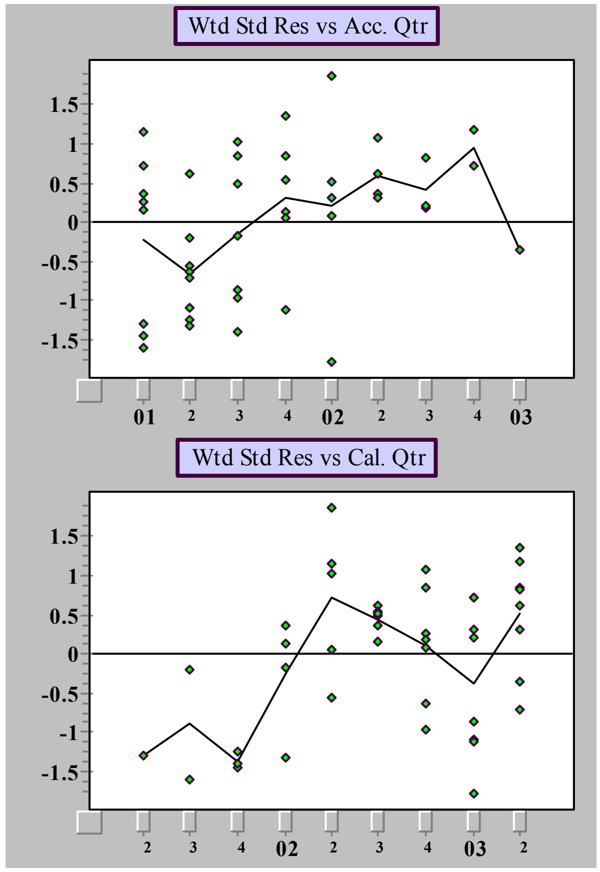


Figure 5.8 Residuals from the Murphy model with i=2 applied to the IL10 data plotted against the accident quarter (left) and the calendar quarter (right)

5.3.2 Criterion 14: Standardization of Residuals

Q-Q plots are used to assess whether data has a particular distribution. The sorted data is plotted against the distribution values at the corresponding percentiles. If the data follows the selected distribution, then the plot will be approximately a straight line (the extreme points are expected to have more variability than points toward the center). A plot that is bent down on the left and bent up on the right means that the data have a longer tail than the distribution.

This criterion cannot be applied to the *ER* method, unless we use some corresponding model such as Murphy's model with $i = 2$. When this model is applied to the last ten quarters of the IL40 data, the normality of the residuals is reasonable. However, the full IL40 data gives the Q-Q plot in Figure 5.9, indicating that the tails of the distribution of the residuals are much heavier than would be expected if they were normally distributed. This may be indicating a lack of fit of the model – the residual plots should be checked carefully for patterns.

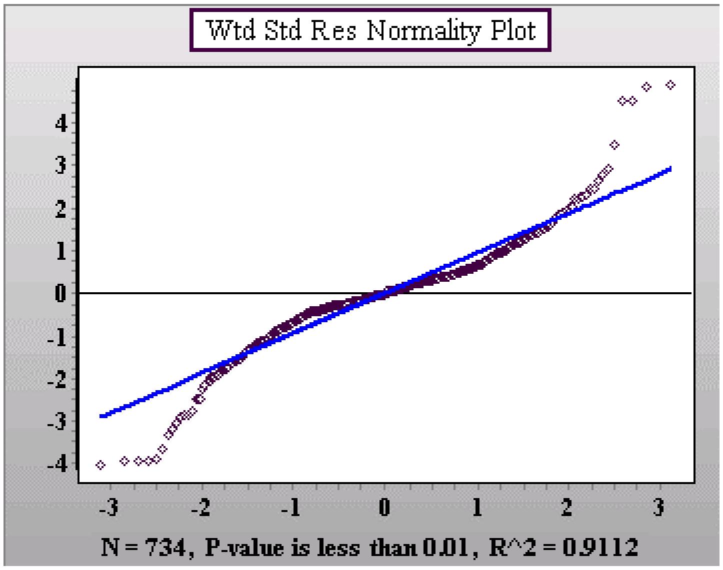


Figure 5.9 Normal Q-Q plot for the residuals when Murphy's model with i=2 is applied to the IL40 data

It is also problematic to apply this criterion to the *ODP* model, as there is no explicit distributional assumption for the residuals. Some possible choices are the normal distribution or the negative binomial. The negative binomial presents difficulties in doing the usual Q-Q plot as the shape of the distribution can change with the mean, so the percentile corresponding to a "standardized" residual depends on the mean. The normal Q-Q plot in Figure 5.10 for the *ODP* model applied to the PL40 data indicates that the residuals are far from normality. The highest point corresponds to accident quarter 1Q1998, development quarter 17. It has a standardized residual of 12, having an actual value of 560 and a fitted distribution with mean 51 and standard error of 43. Under any reasonable distribution this will be an outlier. The model should be refitted with this point removed.

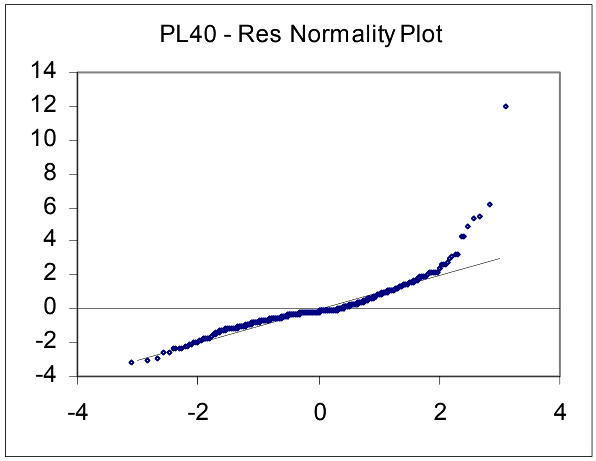


Figure 5.10 Normal Q-Q plot for the residuals from the over-dispersed Poisson model applied to the PL40 data

A lack of normality may be an indication that the variance assumption in the model does not hold for this data. A plot of residuals against fitted values is a good test of this assumption. Figure 5.11 shows this plot, with an estimate of the standard deviation of the residuals shown as dashed lines. This estimate is about one for the lower fitted values, as would be expected of standardized residuals, but increases to about 1.5 at the larger values. This suggests that the assumption of the *ODP* model that the variance is proportional to the mean may not be satisfied by this data.

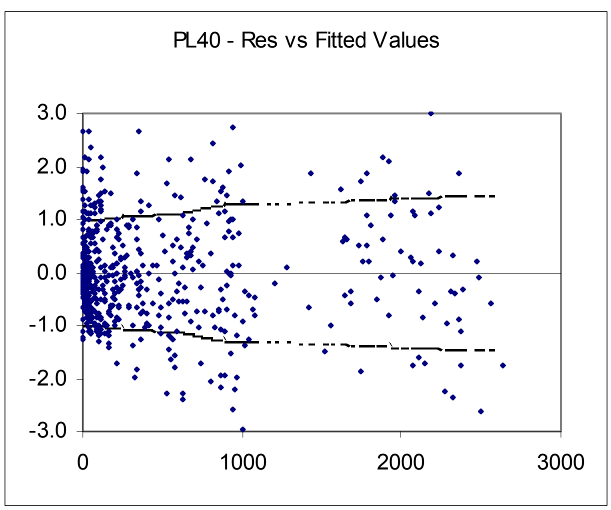


Figure 5.11 Residuals *vs.* fitted values from the over-dispersed Poisson model applied to the PL40 data (residuals larger than 3 not displayed); dashed lines indicate estimated standard deviation

5.3.3 Criterion 15: Analysis of Residual Patterns

This test cannot be applied to the *ER* method, unless we use some corresponding model such as Murphy's model with $i = 2$. When this model is applied to the last 10 quarters of the IL40 data, there is an interesting pattern in the residuals plotted against accident quarter. On average, the fitted values are higher than the actual in the early accident quarters, but lower than the actual in the later accident quarters (see Figure 5.12). This pattern suggests that the assumption that the development pattern is the same in all accident periods may be incorrect, and the model may under-forecast in the later accident quarters, where most of the remaining IBNR is found.

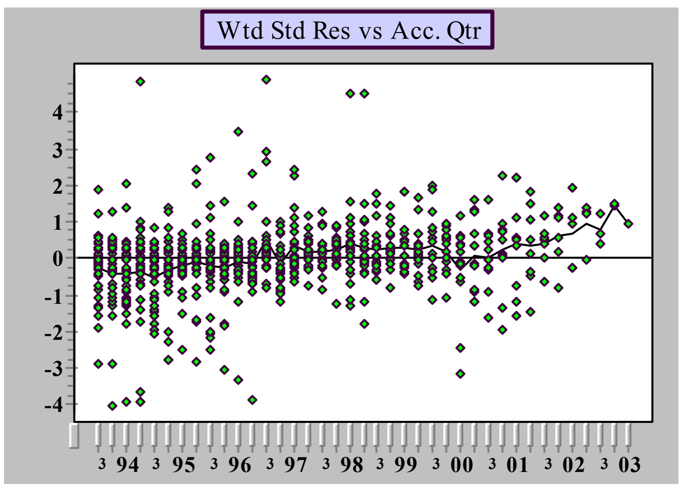


Figure 5.12 Residuals *vs.* accident quarter from the Murphy model with i=2 applied to the IL40 data; the black line indicates the average of the residuals in each period

Plots of residuals against calendar quarters should be examined to see if there is evidence for any calendar period effects. When the *ODP* model is applied to the PL40 data, there is some suggestion of changing calendar period trends in the residual plot in Figure 5.13. If there is relevant economic inflation data available, the triangle should be adjusted by that inflation and the adjusted triangle's residuals should be plotted again. If there are still patterns in the residuals, some further tests are needed. Either the patterns could be statistically tested directly, or a model could be fitted that accounts for those patterns and tested to see if it has statistically a better fit than the original model (see Criterion 17).

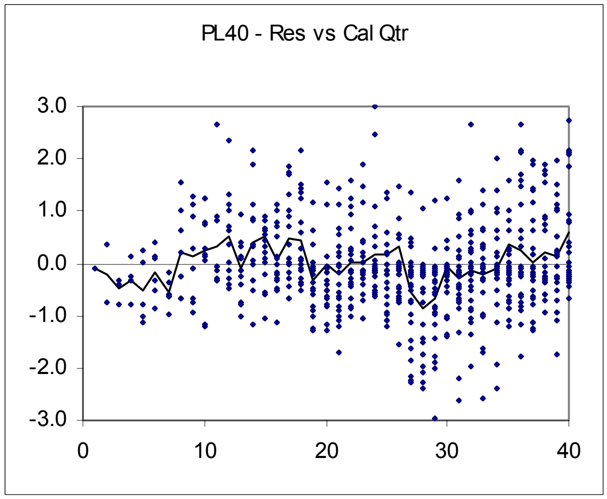
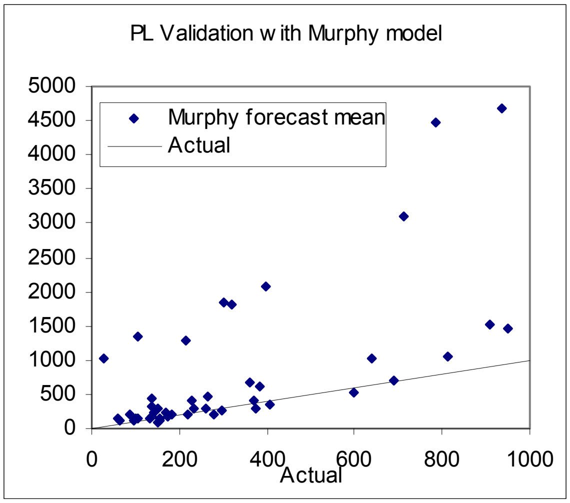


Figure 5.13 Residuals *vs.* calendar quarter from the over-dispersed Poisson model applied to the PL40 data; the black line indicates the average of the residuals in each period; residuals with magnitude greater than 3 are not displayed

5.3.4 Criterion 16: Prediction Error and Out-of-Sample Data

This is the "gold standard" of the criteria – evaluating a model against data not used in the model selection and fitting. A model that satisfies all the other criteria may still fail this test, as the past is not always a good predictor of the future.

The first 10 accident quarters of the PL40 data was used to validate the *ER* method using Murphy's model with $i = 2$ and the *ODP* model. The forecast means of the two models for each future development and accident period are plotted against the actual values in Figure 5.14. The very high values for the *ER* method are all in the last accident quarter and are due to a very large estimated ratio between the first two development periods. The *ODP* model is less sensitive to individual high ratios, which in this case are due to two small values in the first development period. In all other accident quarters, there is very little difference between the forecast means of the two models.



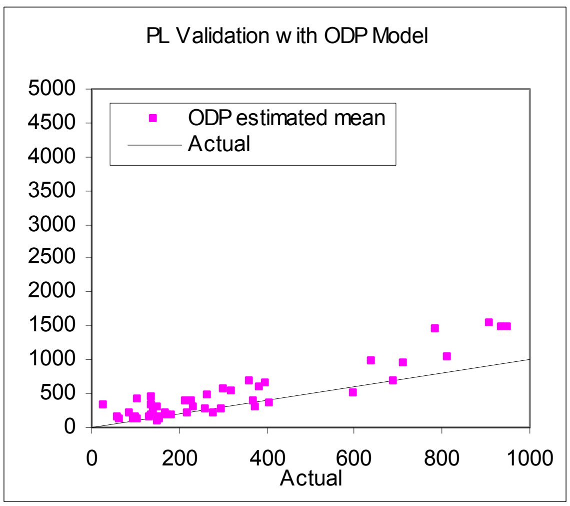


Figure 5.14 Forecast *vs*. actual values for the Murphy model with i=2 (left) and the over-dispersed Poisson model (right), applied to the first 10 quarters of the PL40 data

Even if the model continues to hold in the future, the actual values will not match the forecast mean due to process and parameter uncertainty, so we should take this uncertainty into account when comparing forecast and actual values. For the *ODP* model, all the actual values lie within two standard errors of the forecast mean.

We can also compare the actual total with the forecast distribution to see how plausible the actual value is as a sample from the distribution. For this example, the actual total is 14,400. The forecast distribution for the *ER* method is lognormal with mean 35,851 and standard error 38,474, so the actual value is 0.6 standard errors below the mean, a plausible 27^{th} percentile of the distribution. The forecast distribution for the *ODP* model has a mean of 20,786 and standard error of 4,445, so the actual value is 1.4 standard errors below the mean. The type of the distribution is unspecified, but if we assume normality, the actual value is the 8th percentile of the forecast distribution.

Other tests can be done on the prediction errors. Do they follow the expected probability distribution, for example, does a Q-Q plot indicate they are normally distributed? Do they have any structure when plotted against development period, accident period, calendar period or fitted value?

5.3.5 Criterion 17: Goodness-of-Fit Measures

To use this criterion, we need a sufficiently flexible family of models to compare. Clearly it is pointless comparing models that are not a good fit to the data, so the residual plots of Criterion 15 should appear reasonably random for each of the models. Residuals of models with a small number of parameters (for example, a single parameter for the accident direction, or a two or three parameter curve for the development direction) should be examined very closely, and compared with more generously parameterized models.

It should be noted that the *AIC* and *BIC* measures are intended to be used to compare models in the same "family". If models have different variances on the same observation, their *AIC/BIC* are not comparable. In particular, you should not use the *AIC/BIC* to compare models that have had different outliers removed, or have significantly different assumptions about the variances of the error terms.

The different measures of goodness-of-fit will often choose different models. The other criteria should also be applied and may suggest that one is better than the others on grounds other than strict goodness-of-fit. However, particularly when process variability is low, there may be several models that all qualify as "good". The range of forecasts from these good models gives some measure of model uncertainty. Allowance should be made for this uncertainty in the spread of the forecast distribution, either informally, or formally by model averaging techniques based on the statistical and common sense likelihood of the various models.

The application of goodness-of-fit measures is illustrated under Criterion 18 where the *ODP* model is compared with variations of that model that use fewer parameters. This test cannot be applied to the *ER* method, unless we assume some underlying model.

5.3.6 Criterion 18: Ockham's Razor and the Principle of Parsimony

Applying the principle of parsimony also requires a sufficiently flexible family of models, but in this case the flexibility must extend to allowing models with a small number of parameters. One approach is to use smoothing, as in Verrall[85] , where a smoothing parameter controls the effective number of parameters. Another approach is to use a flexible family of parameterized curves, such as the piecewise linear or constant curves used by Zehnwirth[86]. This approach has the advantage that it is easy to add parameters in the calendar direction as well as in the development and accident direction.

A simple first check of whether a model may be over-parameterized is to look at the ratio between the parameter estimates and their standard errors (we will refer to this as the t-value of the parameter). A formal statistical test can be done (such as an F-test), but as a rough rule-of-thumb, if this ratio is less than two, the parameter is not significantly different from zero and can probably be omitted. For example, in Table 3 of Appendix B, some of the accident parameters and most of the development parameters are not significantly different from zero.

Often models can be parameterized in many different ways, and some ways will make it easier to spot the "redundant" parameters. The tests for non-significance should be based on what we know about the way the loss process behaves. For example, Figure 5.4 shows that the fitted values in the accident direction for this data tend to be more or less constant for a number of periods with occasional jumps up. It would make sense to test whether the level had changed between adjacent periods, and, if not, to use the same parameter for those periods. On the other hand, Figure 5.5 shows that many of the fitted values in the development direction lie more or less on a single trend line. It would make sense to test whether this trend had changed between adjacent periods, and, if not, to use the same trend for those periods. This kind of parameterization, with the addition of trends in the calendar direction, is described in Zehnwirth[87], for the linear regression model applied to the logs of the data. The associated design matrix is described in Barnett and Zehnwirth[88] .

This parameterization was applied to the *ODP* model on the IL10 data. When all 18 parameters are fitted, the standard errors are large (of a similar size to those for the

[85] See Verrall [73].
[86] See Zehnwirth [81].
[87] See Zehnwirth [81].
[88] See Barnett and Zehnwirth [1].

parameter estimates in Appendix B, Table 3, 10 Accident Quarters, which use the same data but an alternative parameterization). As a result, many of the t-values are less than two – for nine of the ten parameters in the accident direction and for four of the eight parameters in the development direction. A more parsimonious model can be obtained by removing parameters until some goodness-of-fit measure is minimized.

A simple way of choosing the next parameter to remove is to choose the one with the smallest absolute t-value. Following this procedure, slightly different models are obtained using different goodness-of-fit measures – the Adjusted SSE gives a model with eight parameters, the "maximum absolute t-value > 2" gives a model with seven parameters and the "F-test p-value > 0.05" gives six parameters. The AIC and BIC do not seem to be useful – they continue to decrease until there are only two parameters left, at which stage the fit of the model is clearly poor.

The various models can be compared visually by plotting the fitted values in the accident and development directions. Figure 5.15 shows the fitted values in the accident direction (for development quarter 1), for three models. The original *ODP* model has ten accident parameters (one for each accident period). The minimum Adjusted SSE model has four accident parameters (one for the first five quarters, one for the next two, one for the next two and one for the last quarter). The F-test model has three accident parameters (one for the first five quarters, one for the next two, one for the last three quarters). Figure 5.16 shows the fitted values in the development direction (for the first accident quarter) for the same models.

The accident parameter estimates from the parsimonious models are very similar to the *ODP* model accident parameter estimates, except for the last accident parameter, where the *ODP* model estimates this value from a single observation. The development parameter estimates from the parsimonious models differ more from the *ODP* model estimates, particularly in the later development periods, where the *ODP* model has to estimate four parameters from just eight observations.

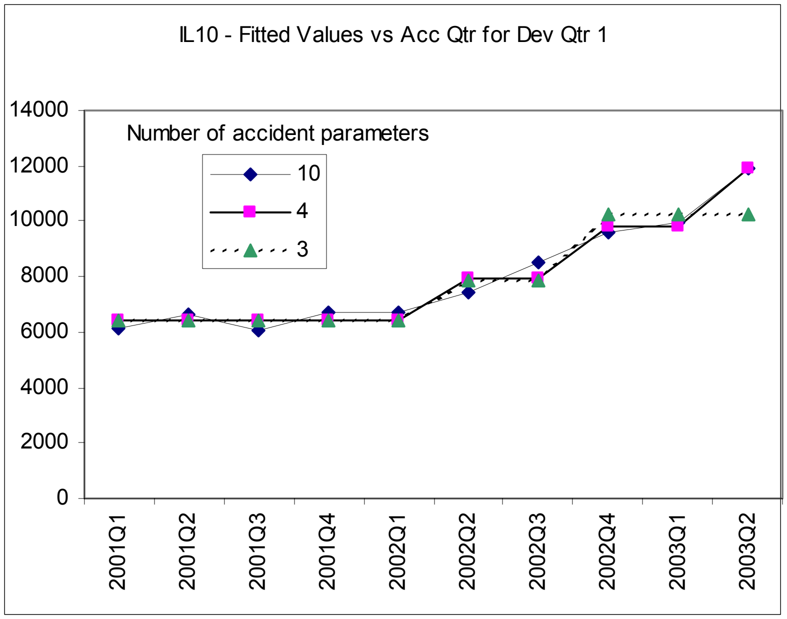


Figure 5.15 Fitted values in the accident direction from the variants of the over-dispersed Poisson model, applied to the IL10 data

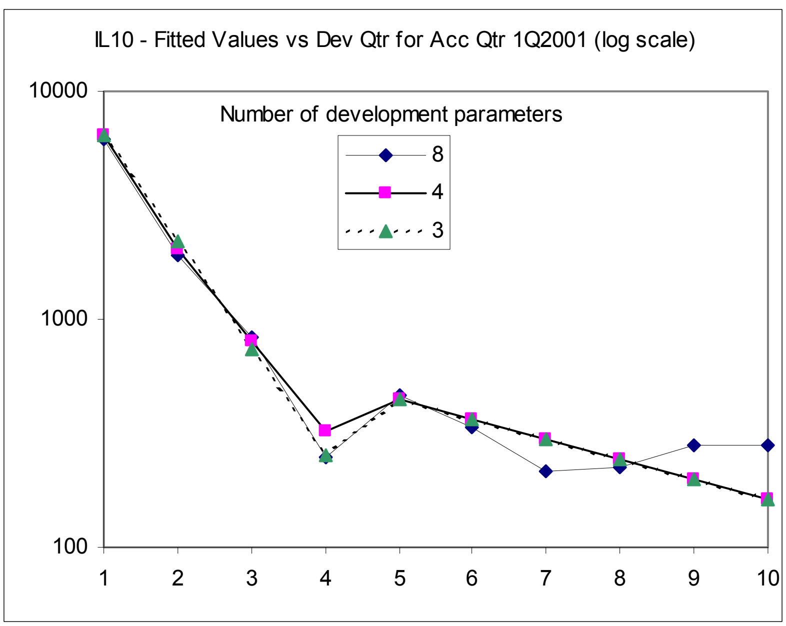


Figure 5.16 Fitted values in the development direction from the variants of the over-dispersed Poisson model, applied to the IL10 data

The standard errors of the fitted values are lower on the parsimonious models. For the original *ODP* model, the forecast standard error of the total is 2,996. For the Adjusted SSE model, it is 2,446. The mean of the total may also be significantly different on the parsimonious models to the over-parameterized model. On average, they should give more reliable forecasts. In this case, for the original *ODP* model, the forecast mean of the total is 24,053. For the Adjusted *SSE* model, it is 21,997, and for the *F*-test model, it is 20,948.

There are two main reasons why these forecasts are lower. These models fit an exponential decline to the data from development quarter five to nine, which gives lower forecasts in development periods nine and ten. This would seem to be more in accord with common sense. The F -test model fits a single parameter to the last three accident quarters, which gives a lower forecast in the last accident quarter. As this forecast is based on a number of observations, instead of a single observation in the original *ODP* model, it is less sensitive to random process variation.

The plot of residuals against calendar period appears to have some trend structure. With the parsimonious models, it is possible to add parameters in the calendar direction and test whether they improve the model according to the goodness-of-fit criteria. In fact, it seems that a better model for this data may be one that has a single accident level, three development parameters and one calendar parameter. Under this model, the calendar trend is zero in the first four quarters, then is 9.5% $\pm$ 1.1% after 4Q2001. Actuarial knowledge may indicate whether or not this is plausible.

Murphy's model does not have a parameterization that lends itself to significance tests between development periods. Often the incremental payments increase or decrease by an approximately constant percentage for some development periods, so a "natural" parameter might be this percentage change. Then this parameter could be tested for changes between pairs of development periods. Similarly, inflationary effects act as an approximately constant percentage change on incremental payments in the calendar direction. There is no obvious way to test this under Murphy's model as cumulating the payments disguises the changes in inflation. There is no parameter corresponding to accident periods, so there is no way to test for the difference between accident periods.

5.3.7 Criterion 19: Predictive Variability

The probability distribution of the future payment estimate is provided by the *ER* method, although it is not based on any model, so its validity cannot be tested. The *ODP* model does not provide a probability distribution, unless some assumption is made about the distribution of the errors. Both provide an estimate of the standard deviation of the future payment estimate, although the *ER* method may not make an adequate allowance for parameter uncertainty in many cases. Care must also be taken with the *ODP* model when a statistical package is used to estimate the standard deviation of forecast values, as the "built-in" estimates may not include process variability.

5.3.8 Criterion 20: Model Validation

The process of validation using within-sample data is similar to the process in Criterion 16 of using out-of-sample data. It is particularly useful in determining if the most recent data is indicating changes – perhaps a flattening of the tail of the development pattern or a recent increase in superimposed inflation. These changes might suggest that the future could be more uncertain than the model indicates. The absence of such changes in the most recent data will increase our confidence in the model forecast.

5.4 Summary

The process of determining forecast distributions consists of a number of steps:

1. Choose a family of models that is suitable for your purpose and sufficiently flexible to model all the features in the data (criteria 1-4).
2. Identify the members of that family that provide an adequate fit to the data (criteria 14-15).
3. Select the "best" models. Are the models reasonable (criteria 5-8, 10)? Do they validate well (criteria 16, 20)? Are simulated datasets similar to the real data (criterion 11)? Are the models parsimonious (criteria 13, 17-18)?
4. Utilize any other information that would improve the model estimates (criterion 12).
5. Decide what assumptions are reasonable for the future, bearing in mind what the data says about the past (criterion 12).
6. Produce forecasts that incorporate model uncertainty, parameter uncertainty and process variability (criterion 19).

6. FUTURE RESEARCH

The CAS Working Party on Quantifying Variability in Reserve Estimates has identified a number of areas in which the reserving actuary would benefit from future research.

These areas are described below:

- Latent Liabilities;
- Correlation of Multiple Segments;
- Making Use of External Information;
- Adjusting Data for Operational Changes; and
- Making Use of Individual Claim Detail.

These five topics are described individually in the sections below. It is hoped that each of these topics can be viewed as a Request for Proposal (RFP) for research papers.

6.1 Latent Liabilities

6.1.1 Statement of the Problem

The nature of US casualty insurance creates exposure to types of claims that do not fit into standard loss development techniques. These exposures may be for coverages not seen in historical loss experience, and are subject to lengthy litigation. These include:

- Asbestos;
- Environmental Pollution;
- Mass Tort Events; and
- Toxic Mold and Construction Defect.

The exposures are often described as "latent" because the insured and insurer may have been unaware of the potential for losses at the time that the original policy was issued. A common element of these latent exposures is that recognition of the loss is more likely to take place on a calendar year basis. The assumption underlying most development triangle methods is that each accident year will show a similar pattern of loss emergence; this assumption is patently untrue for latent exposures.

6.1.2 Estimation Techniques Used

Because traditional casualty loss development techniques are not applicable for latent exposures, other approaches are needed. In broad terms, two types of approaches are taken to estimating future loss emergence: “bottom-up” approaches and “top-down” approaches.

6.1.2.1 Bottom-Up Approaches

A bottom-up approach begins with a detailed review of individual contracts that the insurance company has written historically. For example, for asbestos liability estimation this begins with a listing of all the policies that either have had or can have claims made against them, along with their historical experience. The historical policies can be grouped into categories or “tiers” according to the relative likelihood of claims being made. For the most exposed (tier 1) policies, the insurance company may simply set a reserve equal to the available aggregate limit. For policies with less likelihood of claims, a reserve is set judgmentally at some percent of the available limits.

This bottom-up approach may be applied on a sample of policies, with the results extrapolated to the total population of policies written historically. A rigorous description for the asbestos example is given in Cross and Doucette[89] .

6.1.2.2 Top-Down Approaches

Rather than working with a sample of detailed policy information and extrapolating to the company level, a “top-down” approach instead begins with an industry-wide estimate and attempts to determine the insurance company’s share of the total.

The most naïve top-down approach is “survival analysis”, which simply calculates the number of years until the carried reserve would be exhausted if losses were paid at the current rate. The survival ratio is typically the total carried reserve divided by the average annual payment of the latest three-year period. Survival analysis is not strictly an estimation method, but rather a key benchmark statistic for comparing relative adequacy against peer companies.

[89] See Cross and Doucette [11].

More sophisticated methods involve taking an industry estimate for the type of latent claim, and allocating it to the insurance company based on market share of premium for the years of maximum exposure. The market share may include different percents based on such things as state mix, mono-line versus package business, and reinsurance versus direct insurance.

6.1.3 The Challenge for Estimating Variability

Given this brief description of the challenge in making a point estimate for the future payments for latent liabilities, the challenges for estimating variability are apparent.

Estimating latent liabilities relies on judgment at many steps, rather than on a pure statistical model. High and low selections can be made based on using more or less optimism in the selection process, but it is not at all clear how the resulting numbers correspond to a statistical distribution of outcomes, or to the "confidence level" associated with the assumptions.

6.1.4 Papers Describing these Techniques

Current papers describing these techniques include: Bhagavatula, Brown and Murphy[90] ; Bouska[91]; Cross and Doucette[92]; Diamantoukos[93] ; and Madigan and Metzner[94].

6.2 Correlation of Multiple Segments

6.2.1 Statement of the Problem

Techniques discussed in this paper provide various means for estimating the distribution of ultimate unpaid losses, which correspond to an individual segment of relatively homogenous claims. One can estimate the marginal distributions for all segments that comprise the complete set for a given insurance company. What remains for the actuary performing this analysis is a means to combine the various analyses into a single aggregate

90 See Bhagavataula, Brown and Murphy [3].
91 See Bouska [8].
92 See Cross and Doucette [11].
93 See Diamantoukos [13].
94 See Madigan and Metzner [44].

future payment distribution that provides management with a statistical picture of future payment variability in total. The question may be posed as such: how do outcomes for individual segments relate to each other? For example, if personal auto liability payments run off requiring more money than was anticipated, would we expect commercial auto liability payments to be more likely to exhibit the same "adverse" development?

There are really two parts to this question. First, we would like to measure the strength of correlation between pairs of segments. The result would be a matrix of correlation coefficients. Second, the correlation must be incorporated into a structure that defines the aggregate future payment distribution.

6.2.2 Description of Approaches Published and In Use

We will discuss various published approaches to both parts of the question.

6.2.2.1 Single Triangle Approach

The simplest approach bypasses the first step. We may combine the loss development data for individual segments into a single set of data for all claims. Then the same variability estimation techniques can be applied to the aggregate data. This approach assumes that the mix of business is constant over the historical period. To illustrate the potential problem of this naïve analysis, imagine a company that had written primarily long-tailed insurance until five years ago, at which time the company shifted its emphasis toward short-tailed lines. All of the development history for ages at which no further development is expected for the short-tailed business is drawn from long-tailed business. Just as a future payment indication itself is not selected this way, a variability analysis will be similarly distorted by non-constant business mix.

6.2.2.2 Pair-wise Correlation Approach

A second method also ignores the question of explicitly measuring correlation coefficients. An assumption that correlation between two triangles is fully exhibited in the matched accident/development year increments can be readily incorporated into the bootstrap methodology.[95] One defines a model for the loss development triangle of each segment, and resamples repeatedly from the residual history of each triangle. For each

[95] See Kirschner, Kerley and Isaacs [33].

resample, the residuals corresponding to the same accident/development increments in each triangle are selected, so that the correlation contained in the pair-wise realization of the triangles' development is captured. The sum of ultimate unpaid losses can be compiled over the trials of the bootstrap process to compute a range of estimated future payments. Correlation between triangles could be measured explicitly, but is not necessary because the aggregate future payment distribution is created in the simulation.

6.2.2.3 An Approach Based on Common Trend Factors

The construction of log-linear decay models also provides the actuary with information useful for measuring correlation and aggregating future payment distributions.[96] The model created for each segment includes additive parameters in log-space which capture trends of an incremental paid loss triangle over up to three directions: accident year, development year, and calendar year. Where parameters are fitted for the same purpose in the models of more than one segment, the fitting errors of those parameters may be used to induce correlation in a simulation of simultaneous future payment runoff. Brehm[93] has suggested that the vectors of calendar year parameters themselves can be used to measure correlation, since inflation is a common effect on all segments. The actuary in this case believes that how future payment runoff responds to inflation in the future is the primary source of multi-line correlation.

6.2.2.4 Hindsight Approach

An ad hoc method for measuring correlation coefficients involves inspecting hindsight future payment estimates by accident year in a development triangle[97]. One begins with future payment indications that have been deduced from a survey of the results of multiple estimation methods, then finds the implied future payment estimates which would have been estimated at every point of the development history had the current information been available. The result is a collection of alternate future payment indications that incorporate all available information, but are sensitive to the unique payout of individual accident years. Correlation coefficients can be measured across lines by condensing the triangles of hindsight future payment indications into vectors of alternate viewpoints of overall liabilities for each line that are sensitive to the loss development history.

[96] See Brehm [6].
[97] See Hayne [31].

6.2.2.5 Underwriting Cycle Approach

Another method to measure future payment correlation coefficients that has been suggested relies on the underwriting cycle, or prospective correlation. Because many lines of business are affected by the same market pressures on underwriting and case-reserving, one can measure the correlation in ultimate loss ratios by year and make the leap that those coefficients also apply to reserves. This method has not been explored in detail in the literature.

6.2.3 Final Step in Combining Lines of Business

Once the actuary has designed a correlation matrix between marginal future payment distributions, a model of the aggregate future payment distribution is required that is consistent with both the correlation structure and the marginal probabilities. Two well documented methods available are the normal copula algorithm[98] and the Iman-Conover method. Mildenhall[99] has compared the two methods thoroughly.

While many suggestions have been made regarding reasonable approaches to the question of measuring correlation between segments, little has been written regarding implementation of the suggestions and testing them with actual or simulated data. For example, some actuaries have indicated that the correlation coefficients measured from a pair-wise bootstrapping approach are slight, but we have yet to see hard evidence in the literature. There is room both for the testing of the proposed approaches and for exploration of new or revised approaches. With an understanding of copulas and the Iman-Conover method, the actuary is equipped with the tools to aggregate future payment risks, but the quantification of correlation that drives the aggregation still demands study and innovation.

6.2.4 Papers Describing these Approaches

Current papers describing these approaches include Brehm[100]; Gillet and Serra[101]; Hayne[102]; and Kirschner, Kerley and Isaacs[103].

---

[98] See Wang [76].
[99] See Mildenhall [45].

[100] See Brehm[6].
[101] See Gillet and Serra [25].
[102] See Hayne [31].
[103] See Kirschner, Kerley and Isaacs [33].

6.3 Making Use of External Information

6.3.1 Statement of the Problem

The reserving actuary is often faced with the problem that the data set available for setting reserves is not sufficiently robust to estimate a distribution of future payments. There may be a number of reasons for this.

One reason may be that the subject business has only been in place for one or two years, and there is no historical pattern for use in deriving a development pattern. In this case, a development pattern from some external source is used. The external data may be from industry sources such as consolidated Schedule P, or rating bureaus such as ISO or NCCI; it may also come from competitor companies or other business segments that are judged to be similar.

If the business is very immature, a loss development factor approach may not be reliable, and so the reserve is set based on an expected loss ratio (*ELR*) from a rate filing or from industry averages.

A second reason for using external data may be that the data has too small a volume of loss experience historically, even though the business has been written for a long period. Excess and Umbrella books may fall into this category. Again, this may require the use of external sources for development patterns or expected loss ratios.

6.3.2 The Challenge for Estimating Variability

The essence of the problem is that no single model, in combination with the available data, is viewed as sufficient to set a reserve. The carried reserve is judgmentally selected after a review of multiple data sources and models. How can we estimate a variance for a reserve that we cannot assume results from a statistical model?

6.3.3 Papers Describing these Approaches

Current papers describing these approaches include: Halliwell [104]; Robbin[105]; and Verrall [106].

[104] See Halliwell [29].
[105] See Robbin [54].
[106] See Verrall [75].

6.4 Adjusting for Operational Changes

6.4.1 Statement of the Problem

Loss development techniques based on triangles are typically used with an assumption that the same patterns in the past will be repeated in the future, or that at least the historical pattern is changing in a predictable manner. This assumption is often violated in reality. Some cases in which the relevance of a historical pattern is questionable include:

- Changes in settlement practices, claims-handling, etc., including:
    - Improvements in cellular and mobile technology allowing for faster recognition of claims as well as allowing adjusters to evaluate the settlement value of claims more quickly;
    - Improvements in fraud detection; and
    - Other claims initiatives.
- Changes in operations due to a merger or acquisition
- Retroactive changes to workers' compensation benefits
- Changes in tort law and/or a company's willingness to litigate certain claims, often driven by "size of loss" or "type of claim" criteria

This problem may be addressed by adjusting the historical data "as if" the new conditions had been in place. Other authors suggest using historical data up to some critical point in the past and using that to restate the more recent diagonals. Claims initiatives also present some dilemmas for the reserving actuary. These cannot always be summarized to changes in settlement patterns or changes in case reserve adequacy. Often more sophisticated methods may be called for in the adjusting of data triangles.

When these changes are made, how does it affect the variance structure? Articles have been written to address how to determine when incurred and paid projections would be improved by adjusting the data triangles and also address how to make those adjustments. However, few authors address how this will affect the estimate of the range of results. We would encourage research both into adjusting for operational changes and assessing how this could affect the variance of the resulting distribution of results.

6.4.2 Papers Describing Techniques for Adjusting Historical Data

Some of the basic methods for adjusting the data triangles or the development factors that they generate and current papers include:

- Use of re-stated historical results; [107] [108] [109] [110]
- Adjusting historical results for factors other than those addressed in the Berquist-Sherman type papers (including various claims initiatives);[111]:
- Preserving the historical results but restating the most recent diagonals and using a frequency-severity approach; [112].
- Using regression techniques to restate both paid and incurred chain ladder factors simultaneously;[113] and
- If operational changes have led to speed up or slow down in claims payment patterns, then the mean claim amounts can be modeled as a function of operational time (percentage of claims closed) using generalized linear models. [114], [115]

6.5 Making Use of Individual Claim Detail

6.5.1 Statement of the Problem

Loss development models typically work with data in a "triangle" format, or perhaps in multiple triangles including loss dollars and counts. The methods for calculating development factors from triangles were designed to be simple enough to be accomplished with pencil and paper. The chain ladder technique is the most widely used technique in estimating future payments. This method is based on very restrictive model conditions which are quite commonly breached in practice. The underlying data need to be corrected for multiple trends, superimposed inflation, seasonal effects and many other factors. It is very difficult to quantify these factors within the chain-ladder paradigm. Should we instead be

[107] See Berquist and Sherman [5].
[108] See Thorne [69].
[109] See Fleming & Mayer [20].
[110] See Duvall [14].
[111] See Halpert, Weinstein and Gonwa [30].
[112] See Ghezzi [24].
[113] See Quarg and Mack [52].
[114] See Wright [78].
[115] See Wright [79].

looking at techniques that look at transaction histories at an individual claim level?

With advances in computer power, it is now possible to analyze individual claim level transactions to estimate future payments. Working with individual claims, we can take care of multiple trends, inflation, seasonal effects, accident quarter effects and other factors in a more direct way. We also have additional flexibility in using interaction terms and choice of error distribution. The stochastic framework also allows us to objectively compare candidate models and to validate the model that was selected. This method also provides the actuaries enhanced understanding of their data.

Future research questions that are of interest are:

- What level of data is best used for this analysis? Should we use individual claim level data? Or should we summarize the data to a more manageable size?
- What are the best ways to quantify multiple trends, inflation, seasonal effects and other such effects?
- What predictive variables are best used for this purpose? What interaction terms are most useful? Should individual claim characteristics play a role here?
- How do we explain the models to regulators? Would a model based on individual claim data be too difficult to explain?

6.5.2 Papers Describing these Techniques

Current papers describing these techniques include: England and Verrall [18]; Mack and Venter [43]; Taylor [64]; Taylor and McGuire [68]; and Weissner [77].

7. CAVEATS AND LIMITATIONS

7.1 Understanding the Nature of the Problem

Recent commentary by rating agencies on reserving actuaries make it important that we clearly define what this paper provides and what is does not provide so that there is no misunderstanding.

7.1.1 Future Payments are Uncertain

As such, the probability that the actual ultimate amount will agree with any single estimate is zero. We know that the estimate will differ from the actual amount and the question is, "what is the degree of variability present in the estimate at this time?"

7.1.2 The Estimate is at a Point in Time

Users of liability estimates need to understand that every estimate is an estimate of future payments (and thus current liabilities) using the information available at a given point in time. For a given block of historical exposures at a given point in time, the actual value of the liabilities will emerge in the future as actual payments are made. As those future payments are made, future estimates of those liabilities will become more certain as less of the future payments remain to be estimated.

7.1.3 The Actual Future Payments are Currently Unknown

Given that we know that ultimate future payments will differ from any prior point estimate of them, we as actuaries would like to provide the users of our product a quantification of the variability of the estimate; that is, potentially how much could it differ and what are the different probabilities at different levels of variability?

7.1.4 State of The Art in 2004

Given this intent, this paper represents a depiction of the "state of the art" circa 2004, of the means of producing the quantification of variability. It is not all inclusive and what we are doing is changing almost daily as new methods are being worked on, written about and evaluated. This paper likely could be updated at periodic intervals, perhaps once every five years.

7.2 General Items of Future Payment Uncertainty

The limitations mentioned in this paper do not specifically address the standard elements of the estimation process that create uncertainty in the future payment estimate. Future payment estimates and estimates of their distributions are forecasts of the future and are generally based on specific assumptions regarding the future which are often based on past performance. There are no guarantees that future events will correspond to these assumptions. Some specific assumptions about the future that future payment models often make include the following:

- Data quality, availability, homogeneity and credibility;
- Emergence patterns, settlement patterns, development patterns;
- Frequency and severity of claims;
- Limits or reinsurance;
- Policy form or deductible levels;
- Salvage and subrogation or collateral sources; and
- Company operations.

7.3 Reference to Specific Papers

The limitations mentioned in this paper do not specifically include any limitations mentioned in the papers that have been surveyed, but these are implicitly present. As such, all limitations and conditions included in the original papers are implicitly carried forward into this paper.

7.4 Predictive Value of the Past

As with the nature of most actuarial work, one of our biggest limitations is using past history to predict the future. We are only as good as our assumptions of the future state and our ability to estimate the likelihood of that future state (Bayesian approach). This includes the distribution of future payments as well as the point estimate.

7.5 Model Uncertainty

Our biggest source of uncertainty is the model uncertainty:

- Do we have the right model or models?
- Have we parameterized the model correctly?
- How sensitive is the model and its variables and what does the sensitivity of each variable imply for the distribution of future payments.

7.6 Defining the Asymptotic Value

Who on September 10th could have imagined the events of September 11th ? The fact of the matter is that is impossible to quantify the entire width of the distribution and to account for extreme and unimaginable events.

7.7 Quantification is Not Elimination

The fact that we can measure and quantify uncertainty does not eliminate it. Therefore, management must employ the estimates we provide with other tools to mitigate this uncertainty. These tools include but are not limited to:

- Insurance;
- Reinsurance; and
- Hedging, etc.

7.8 What to Book

What to do with the estimate of variability is beyond the scope of this paper. We are opening a lot of doors by creating the ability to estimate the distribution of future payments. That being said, we are not stating an opinion as to what level within that distribution should be booked. Assuming a reasonable distribution can be estimated, what to book becomes an issue for various professional organizations concerned with financial statements such as the AAA, AICPA, SEC, IRS, etc. It is possible that different professional organizations might reach different conclusions as to the question of what to book, but the actuarial profession should provide leadership and wisdom to the debate.

REFERENCES

[1] Barnett, G. and B. Zehnwirth. 1998. Best Estimates for Reserves. CAS *Forum* (Fall): 2-54.

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

[3] Bhagavatula, Raja, Brian Brown, and Kevin Murphy. 1994. Estimation of Liabilities Due to Inactive Hazardous Waste Sites. CAS *Forum* (Summer): 301-65.

[4] Benktander, G. 1976. An Approach to Credibility in Calculating IBNR for Casualty Excess Reinsurance. *The Actuarial Review* 3, no. 2:7.

[5] Berquist, J.R. and R.E. Sherman. 1977. Loss Reserve Adequacy Testing: A Comprehensive, Systematic Approach. *PCAS* LXIV:123-84.

[6] Brehm, Paul J. 2002. Correlation and the Aggregation of Unpaid Loss Distributions. CAS *Forum* (Fall): 1-23.

[7] Bornhuetter, R. and R. Ferguson. 1972. The Actuary and IBNR. *PCAS* LIX:181-95.

[8] Bouska, Amy S. and Thomas McIntyre. 1994. Measurement of U.S. Pollution Liabilities. CAS *Forum* (Summer): 75-160.

[9] Bühlmann, Hans. 1970. *Mathematical Models in Risk Theory*. New York: Springer Verlag.

[10] Clark, D. R. 2003. LDF Curve-Fitting and Stochastic Reserving: A Maximum Likelihood Approach. CAS *Forum* (Fall): 41-92.

[11] Cross, Susan and John Doucette. 1997. Measurement of Asbestos Bodily Injury Liabilities. *PCAS* LXXXIV:187-300.

[12] De Alba, E. 2002. Bayesian Estimation of Outstanding Claims Reserves. *North American Actuarial Journal* 6, no. 4:1-20.

[13] Diamantoukos, Christopher. 2000. Estimating U.S. Environmental Pollution Liabilities by Simulation. *PCAS* LXXXVII:79-161.

[14] Duvall, Richard, M. 1992. Testing for Shifts in Reserve Adequacy. *PCAS* LXXIX:1-20.

[15] Efron, B. and R. J. Tibshirani. 1993. *An Introduction to the Bootstrap*. San Francisco: Chapman & Hall.

[16] England, P. D. and R. J. Verrall. 1999. Analytic and Bootstrap Estimates of Prediction Errors in Claims Reserving. *Insurance: Mathematics and Economics* 25:281-93.

[17] England, P. D. and R. J. Verrall. 2001. A Flexible Framework for Stochastic Claims Reserving. *PCAS* LXXXVIII:1-38.

[18] England, P. D. and R. J. Verrall. 2002. Stochastic Claims Reserving in General Insurance. *British Actuarial Journal* 8:443-544.

[19] Fisher, W. and J. Lange. 1973. Loss Reserve Testing: A Report Year Approach. *PCAS* LX:189-207.

[20] Fleming, Kirk G, and Jeffrey H. Mayer. 1988. Adjusting Incurred Losses for Simultaneous Shifts in Payment Patterns and Case Reserve Adequacy Levels. CAS *Discussion Paper Program, Evaluating Insurance Company Liabilities:* 189-214.

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

[22] Gerber, H. 1979. An Introduction to Mathematical Risk Theory. Huebner Foundation Monograph 8, Wharton School of the University of Pennsylvania, Richard D. Irwin Inc.

[23] Gerber, H. and D. Jones. 1975. Credibility Formulas of the Updating Type. *Transactions of the Society of Actuaries* 27:31-52.

[24] Ghezzi, Thomas L. 2001. Loss Reserving Without Loss Development Patterns—Beyond Berquist-Sherman. CAS *Forum* (Fall): 43-104.

[25] Gillet, Antonin and Benjamin Serra. (2003) Effets de la Dépendance Entre Différentes Branches Sur le Calcul des Provisions. Paper presented at the 2002 ASTIN Colloquium, Cancún. www.casact.org/research/summaries/links.htm.

[26] Gogol, D. 1993. Using Expected Loss Ratios in Reserving. *Insurance: Mathematics and Economics* 12:297-9.

[27] Hachemeister, C., and J. Stanard. 1975. IBNR Claims Count Estimation With Static Lag Functions. Paper presented at the 1975 ASTIN Colloquium, Portimao, Portugal and the 1976 CAS Spring Meeting Greenbrier, Va.

[28] Hachemeister, C. 1980. A Stochastic Model for Loss Reserving. *Proceedings of the ICA*: 185-94.

[29] Halliwell, Leigh J. 1997. Conjoint Prediction of Paid and Incurred Losses. CAS *Forum* 1 (Summer): 241-380.

[30] Halpert, Aaron, Scott Weinstein, and Christopher Gonwa. 2001. Evaluating Reserves in a Changing Claims Environment. CAS *Forum* (Fall): 193-238.

[31] Hayne, Roger. 2004. Estimating and Incorporating Correlation in Reserve Variability. CAS *Forum* (Fall): 53-72.

[32] Hesselager, O. 1994. A Markov Model For Loss Reserving. *ASTIN Bulletin* 24, 2:183-93.

[33] Kirschner, Gerald S., Colin Kerley, and Belinda Isaacs. Two Approaches to Calculating Correlated Reserve Indications Across Multiple Lines of Business. CAS *Forum* (Fall): 211-46.

[34] Klugman, S., H. Panjer, and G. Willmot. 2004. *Loss Models: From Data to Decisions*, 2nd ed. New York: John Wiley and Sons, Inc.

[35] Kremer, E. 1982. IBNR Claims and the Two Way Model of ANOVA. *Scandinavian Actuarial Journal* 1:47-55.

[36] Mack, T. 1990. Improved Estimation of IBNR Claims by Credibility Theory. *Insurance: Mathematics and Economics* 9, no. 1:51-7.

[37] Mack, T. 1993. Distribution-Free Calculation of the Standard Error of Chain Ladder Reserve Estimates. *ASTIN Bulletin* 23, no. 2 :213-25.

[38] Mack, Thomas. 1994. Measuring the Variability of Chain Ladder Reserve Estimates. CAS *Forum* 1 (Spring): 101-82.

[39] Mack, Thomas. 1994. Which Stochastic Model Is Underlying the Chain Ladder Method? *Insurance: Mathematics and Economics* 15, no. 2-3:133-8.

[40] Mack, Thomas. 1995. Which Stochastic Model is Underlying the Chain Ladder Method? CAS *Forum* (Fall): 229-40.

[41] Mack, Thomas. 1999. The Standard Error of Chain Ladder Reserve Estimates: Recursive Calculation and Inclusion of a Tail Factor. *ASTIN Bulletin* 29, no. 2:361-6.

[42] Mack, Thomas. 2000. Credible Claims Reserves: The Benktander Method. *ASTIN Bulletin* 30, no. 2:333-47.

[43] Mack, Thomas and Gary Venter. 2000. A Comparison of Stochastic Models That Reproduce Chain Ladder Reserve Estimates. *Insurance: Mathematics and Economics* 26, no. 1:101-7.

[44] Madigan, Kevin M. and Claus S. Metzner. 2003. Reserving for Asbestos Liabilities. CAS *Forum* (Fall): 173-212.

[45] Mildenhall, Stephen. Forthcoming. Correlation and Aggregate Loss Distributions With An Emphasis On The Iman-Conover Method. Working paper chapter contribution to the Casualty Actuarial Society Research Working Party on Correlations and Dependencies Among All Risk Sources. http://www.mynl.com/wp/ic.pdf.

[46] Murphy, D. 1994. Unbiased Loss Development Factors. *PCAS* LXXXI:154-222.

[47] Neuhaus, W. 1992. Another Pragmatic Loss Reserving Method or Bornhuetter/Ferguson Revisited. *Scandinavian Actuarial Journal*: 151-162

[48] Norberg, R. 1993. Prediction of Outstanding Liabilities in Non-Life Insurance. *ASTIN Bulletin* 23, no. 1:95-115; Norberg, R. 1999. Prediction of Outstanding Liabilities II: Model Variations and Extensions. *ASTIN Bulletin* 29, no. 1:5-26.

[49] Ntzoufras, I. and Dellaportas, P. 2002. Bayesian Modelling of Outstanding Liabilities Incorporating Claim Count Uncertainty (with discussion). *North American Actuarial Journal* 6:113-128.

[50] Patel, C. and A. Raws III. 1998. Statistical Modeling Techniques for Reserve Ranges: A Simulation Approach. CAS *Forum* (Fall): 229-55.

[51] Quarg, Gerhard. 2003. The Munich Chain Ladder—A Claims Reserving Technique That Closes the Gap Between Paid and Incurred Based IBNR-Estimates. Paper presented at the 2003 ASTIN Colloquium, Berlin. http://www.astin2003.de/img/papers/quarg.pdf.

[52] Quarg, Gerhard and Thomas Mack. Forthcoming. Munich Chain Ladder: A Reserving Method That Reduces the Gap Between IBNR Projections Based on Paid and Incurred Losses.

[53] Renshaw, A. and R. J. Verrall. 1998. A Stochastic Model Underlying the Chain Ladder Technique. *British Actuarial Journal* 4:903-23.

[54] Robbin, Ira. 1986. A Bayesian Credibility Formula for IBNR Counts. *PCAS* LXXIII:129-64.

[55] *S-Plus 6 for Windows Guide to Statistics, Volume 2*. Seattle, Washington: Insightful Corporation.. http://www.insightful.com/support/splus60win/statman2.pdf

[56] Scollnik, D.P.M. 2001. Actuarial Modeling with MCMC and BUGS. *North American Actuarial Journal* 5, no. 2:96-124.

[57] Seal, Hilary L. 1980. Survival Probabilities Based On Pareto Claim Distributions. *ASTIN Bulletin* 11: 61-71

[58] Shapland, Mark R. 2003. Loss Reserve Estimates: A Statistical Approach for Determining 'Reasonableness'. CAS *Forum* (Fall): 321-60.

[59] Tarbell, Thomas F. 1933. Incurred But Not Reported Claim Reserves. *PCAS* XX:275-80.

[60] Taylor, G. 1977. Separation of Inflation and Other Effects From the Distribution of Non-Life Insurance Claim Delays. *ASTIN Bulletin* 9, no. 1-2:219-30.

[61] Taylor, G. 1978. Testing Goodness of Fit of an Estimated Runoff Triangle. *ASTIN Bulletin* 10, no. 1: 78-86.

[62] Taylor, G. 1981. Speed of Finalization of Claims and Claims Runoff Analysis. *ASTIN Bulletin* 12, no. 2:81-100.

[63] Taylor, G. C. and F. R. Ashe. 1983. Second Moments of Estimates of Outstanding Claims. *Journal of Econometrics* 23:37-61.

[64] Taylor, G. C. 1987. Regression Models In Claim Analysis I: Theory. *PCAS* LXXIV:354-83.

[65] Taylor, G. C. 2000. *Loss Reserving: An Actuarial Perspective.* Dordrecht, Netherlands: Kluwer Academic Publishers.

[66] Taylor, G. C. 2003. Chain Ladder Bias. *ASTIN Bulletin* 33, no. 2:313-30.

[67] Taylor, Greg, Gráinne McGuire, and A. Greenfield. 2003. Loss Reserving: Past, Present and Future. Working Paper 109, Research Paper Series, Centre for Actuarial Studies, University of Melbourne, Australia. http://www.economics.unimelb.edu.au/actwww/wps2003.html.

[68] Taylor, Greg and Gráinne McGuire. 2004. Loss Reserving with GLMs: A Case Study. *CAS Discussion Paper Program*, *Applying and Evaluating Generalized Linear Models:* 327-92.

[69] Thorne, Joseph O. 1978. Discussion of Loss Reserve Adequacy Testing: A Comprehensive, Systematic Approach by James R. Berquist and Richard E. Sherman. *PCAS* LXV: 10-33.

[70] Venter, Gary G. 1989. A Three-Way Credibility Approach To Loss Reserving. *Insurance: Mathematics and Economics* 8, no. 1:63-9.

[71] Venter, Gary G. 1998. Testing the Assumptions of Age-to-Age Factors. *PCAS* LXXXV:807-47.

[72] Venter, Gary G. 2000. Testing Reserve Models. Slide Presentation, Guy Carpenter Instrat, New York.

[73] Verrall, R. J. Claims Reserving and Generalised Additive Models. *Insurance: Mathematics and Economics* 19 (1996):31-43.

[74] Verrall, R. J. 2004. A Bayesian Generalized Linear Model for the Bornhuetter-Ferguson Method of Claims Reserving. *North American Actuarial Journal* 8, no. 3:67-89.

[75] Verrall, R. J. 2004. Obtaining Predictive Distributions for Reserves Which Incorporate Expert Opinion. CAS *Forum* (Fall): 283-316.

[76] Wang, Shaun. 1999. Aggregation of Correlated Risk Portfolios: Models and Algorithms. *PCAS* LXXXVI: 848-939.

[77] Weissner, Edward W. 1978. Estimation of the Distribution of Report Lags by the Method of Maximum Likelihood. *PCAS* LXV:1-9.

[78] Wright, Thomas S. 1990. A Stochastic Method for Claims Reserving in General Insurance. *The Journal of the Institute of Actuaries* 117, Part III, No. 468: 677-731. http://www.actuaries.org.uk/files/pdf/library/JIA-117/0677-0731.pdf

[79] Wright, Thomas S. 1992. Stochastic Claims Reserving When Past Claim Numbers are Known. *PCAS* LXXIX:255-361.

[80] Ye, J. On Measuring and Correcting The Effects of Data Mining and Model Selection. *Journal of the American Statistical Association* 93 (1998):120-31.

[81] Zehnwirth, Ben. 1994. Probabilistic Development Factor Models with Applications to Loss Reserve Variability, Prediction Intervals and Risk Based Capital. CAS *Forum* 2 (Spring): 447-605.

Abbreviations and notations

The abbreviations and notations used in the paper are as follows:
AIC, Akaike Information Criteria
APD, automobile physical damage
BIC, Bayesian Information Criteria
BF, Bornhuetter-Ferguson
BUGS, Bayesian Inference Using Gibbs Sampling
CL, Chain Ladder
CV, coefficient of variation
ELR, Expected Loss Ratio
EPV, Expected Process Variance
GB, Gunnar-Benktander
GLM, generalized linear models
MCMC, Markov Chain Monte Carlo
MSEP, Mean Squared Error of Prediction
ODP, Over-Dispersed Poisson
OLS, Ordinary Least Squares
VHM, Variance of Hypothetical Mean

Biographies of Working Party Members

Co-Chairs:

Roger Hayne is a Principal and Consulting Actuary in Milliman, Inc.'s Pasadena, California office. His practice includes work in extended warranties, reinsurance, as well as reserve analyses in most casualty coverages. He holds a Ph.D. in mathematics from the University of California; he is a Fellow of the CAS and a Member of the American Academy of Actuaries. He is an active volunteer in the CAS and has spoken at many CAS meetings and seminars on the issue of uncertainty in loss reserve estimates. In addition to four papers that have appeared in the *Proceedings of the Casualty Actuarial Society*, one of which was awarded the 1995 Dorweiler Prize, he has authored several papers that appeared in the *CAS Forum* as well as other publications.

James Leise is Assistant Vice President, Operational Loss Reserving at United Services Automobile Association. He graduated from the United States Military Academy and is a Fellow of the Casualty Actuarial Society and a Member of the American Academy of Actuaries. Leise has served on the CAS Committee on Reserves.

Section 2 Group:

Sandie Cagley is a Senior Financial Analyst, having responsibility for reserving at State Farm Insurance Companies. She is an ACAS and MAAA. She received a Bachelor of Arts degree in Mathematics from Southern Illinois University.

Joseph O. Marker is principal and founder of Marker Actuarial Services LLC. This firm provides actuarial and management consulting services in areas such as pricing, product development, marketing, alternative risk, forecasting and model, information systems, reinsurance, strategic planning, loss reserving, and actuarial staffing.

Prior to consulting, Marker worked for four organizations in various actuarial capacities for 28 years, the last fifteen years as Chief Actuary at two Midwestern United States regional insurance companies. Two of the companies wrote large volumes of medical malpractice insurance.

He is an FCAS and MAAA and a past president of the Midwestern Actuarial Forum. Marker is a frequent speaker at regional meetings and the Casualty Loss Reserve Seminar. He published the paper, "Studying

Policy Retention using Markov Chains," in the 1998 *PCAS*. Marker also co-authored the paper "Rating Claims-Made Insurance Policies" for the CAS Call Paper Program in 1980. This paper is part of the exam syllabus for the CAS.

His education includes Master of Science, Mathematics, from the University of Minnesota, and a B.A. from the University of Michigan, with Highest Honors in Mathematics.

David E. Sanders is consultant at Milliman's in London, concentrating mainly on London Market issues. He is a Fellow of the Institute of Actuaries and member of the American Academy of Actuaries He has been chairman of the London Market Actuaries Group and has been on the General Insurance Board of Institute of Actuaries. He has written a number of published papers on subjects such as catastrophes, extreme events and pricing.

Section 3 Group:

Mark R. Shapland is an Actuary in Milliman Inc.'s Milwaukee office. He is responsible for various reserving projects including modeling of asbestos liabilities. 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 is the leader of Section 3 of this Working Party, the current Chair of the Dynamic Risk Modeling Committee of the CAS and a member of the CAS Committee on Reserves.

Yisheng Bu is a Senior Actuarial Analyst with Liberty Mutual Insurance Group in Boston, Massachusetts. He received his Ph.D. in economics and finance from Indiana University-Bloomington. He has written a research paper on optimal reinsurance arrangement for the 2005 CAS Seminar on Reinsurance.

James Christopher Guszcza is a Senior Manager in the Actuarial and Insurance Solutions practice of Deloitte Consulting, where he specializes in data mining and predictive modeling projects. He has a B.A. degree in liberal arts from St. John's College, and a Ph.D. in Philosophy from the University of Chicago. He is a Fellow of the Casualty Actuarial Society and Member of the American Academy of Actuaries.

Greg Taylor is a consultant with Taylor Fry Consulting Actuaries in Sydney, Australia. He is also an extra-mural Professor of Actuarial Studies at the University of Melbourne, and the University of New South Wales in Sydney.

Section 4 Group:

Gary G. Venter, FCAS, MAAA, is the Managing Director in the Instrat group of Guy Carpenter, and has put in over thirty years in the insurance and reinsurance industry. He joined Instrat in 1994, directly from the Workers Compensation Reinsurance Bureau, where he served as President. His other actuarial positions include Fireman's Fund, Prudential Reinsurance, and the National Council on Compensation Insurance. Venter has an undergraduate degree in mathematics and philosophy from University of California-Berkeley and a Masters in mathematics from Stanford University. He has served on numerous committees of the Casualty Actuarial Society and American Academy of Actuaries, including Chairmanships of the International Research Committee, the Committee on Theory of Risk, and the Advisory Committee on Asset/Liability Management and Investment Policy.

Venter is well known in the actuarial field as author of research articles, some of them prize winners, and as a frequent speaker at actuarial seminars, including international forums such as the East Asia Actuarial Conference and the ASTIN Colloquium. He teaches a course on actuarial statistics at Columbia University and has taught actuarial science training programs worldwide, including the U.K., Brazil, Portugal, Paris, and China. His research focus is on applying advanced actuarial methods to practical business problems, particularly those involving risk transfer and capital management.

John T. Bonsignore is currently an Associate Actuary with Milliman and has been with the firm since 2001. He has over 13 years of property and casualty experience in both commercial and personal lines, with a concentration in reserving and financial analysis.

Atul Malhotra is a consulting actuary with the Tillinghast business of Towers Perrin in the firm's San Francisco office. He is a Fellow of the Casualty Actuarial Society and a Member of the American Academy of Actuaries

Since joining Tillinghast, Malhotra has been involved with both ratemaking and reserving for most lines of business, including medical professional liability, workers compensation, auto, homeowners, commercial multi-peril, general liability, surety, extended warranties, and asbestos-related exposures.

Prior to joining the firm, Malhotra spent several years with Fireman's Fund. During this time, he worked in pricing and reserving functions covering workers compensation, general liability, auto liability, and auto warranty.

Malhotra's responsibilities have included planning, development of pricing models for commercial and personal lines of business using the economic value-added methodology, loss portfolio transfer pricing, and reserving for retrospective rated policies.

Malhotra has co-authored a professional paper on the impacts of layoffs, plant closures, and downsizing in reserving workers compensation liabilities.

Section 5 Group:

David Ruhm is an Assistant Vice President in Corporate Actuarial Research at The Hartford. His work is primarily focused on developing systems that support and advance The Hartford's financial discipline. Ruhm graduated from the University of California at San Diego and is a Fellow of the Casualty Actuarial Society. He currently serves on the Casualty Actuarial Society's Committee on Theory of Risk as well as other industry committees, and has authored several published papers.

Julie Sims is a Senior Statistician at Insureware Pty. Ltd. in Melbourne, Australia. She contributes to program design, research, testing, documentation, and user support for ICRFS-PLUS™, software used worldwide for modelling the distribution of future payments. She has previously worked on a wide range of modelling projects in the areas of petroleum, minerals, and steel. Her qualifications include a Master of Science in Statistics from Monash University and a Doctor of Philosophy in Mathematics from the University of Oxford.

Section 6 Group:

David R. Clark works for American Re-Insurance supporting pricing and model development functions. He was awarded the 2003 Reserving Call Paper prize.

Ravi Kumar works as a Data Analytics Consultant in designing risk selection and pricing strategies for P&C insurance companies.

Gary V. Nickerson is Vice President and Corporate Actuary with Universal Underwriters Group, and has previously been a member of the CAS Committee on Reserves.

Dianne M. Phelps is currently working in Personal Lines reserving at The Hartford in Hartford, Connecticut and has been in the actuarial field for six years.

Micah Grant Woolstenhulme is an Assistant Actuary in the Corporate Actuarial and Risk Management department of Safeco Insurance Companies in Seattle, Washington. His duties include dynamic financial analysis, stochastic reserving, and reinsurance pricing. Prior to Safeco, Woolstenhulme worked as a Workers Compensation Pricing Analyst for Kemper Insurance Companies in Long Grove, Illinois, and as a

Pension Consultant for PricewaterhouseCoopers in Teaneck, New Jersey. Woolstenhulme became a Fellow of the Casualty Actuarial Society in the spring of 2004.

Section 7 Group:

C.K. "Stan" Khury, FCAS, MAAA, CLU, is a principal and founder of Bass & Khury, an independent actuarial consulting firm focusing primarily on actuarial litigation. He graduated with a BA degree from Ohio Wesleyan University and a MS degree from Ohio State University, both in mathematics. He has been active in the CAS for over thirty years, including service as president in 1984.

Bruce E. Ollodart is a consulting actuary with the firm American Actuarial LLC, located in Wallingford, Connecticut. Ollodart is a Fellow of the Casualty Actuarial Society, a member of the American Academy of Actuaries, and a member of the International Actuarial Association. He received a Bachelor of Science degree from the George Washington University with majors in Physics and Applied Mathematics. Ollodart's range of experience is broad and extensive, encompassing many of the principal property and casualty consulting areas with over 23 years experience in the actuarial profession.

R. Stephen Pulis is a Consulting Actuary at Actuarial Research Group in Houston, Texas. His current responsibilities are reserve evaluations of medical malpractice business through analyzing individual claim experience and modeling future results. During his 33 years in the actuarial profession, he has worked on pricing or reserving of most of the personal and commercial property-casualty lines of business, from high-frequency/low-severity to high-severity/low-frequency books of business. He has BS in Mathematics from Michigan State University. He is an Associate of the CAS, a member of the American Academy of Actuaries, and a member of the International Association of Actuaries. He is also a past president of the Southwest Actuarial Forum, and participated for several years as a committee member for an industry trade organization.

Reviewer

Rodney Kreps, FCAS, MAAA, is a Managing Director of Guy Carpenter, and a past Chair of the CAS Committee on the Theory of Risk. He holds a BS from Stanford and a Ph.D. from Princeton in theoretical physics. He worked as an academic for fifteen years, acquiring tenure as an Associate Professor of physics. After working in construction for seven years he went to Fireman's Fund and while there worked in workers compensation, reserving, database design, and reinsurance. He moved to Sedgwick Re (now Guy Carpenter) in 1989 and has actively pursued theoretical and practical reinsurance models, contract designs, and financial modeling. He has written papers for the *PCAS* and spoken frequently at CAS and other events.

Appendix A. Calculation of Variances of Future Payments Through Approximation

This appendix describes the calculation for each type of variance of future payments ($\mathrm{Var}_f[\hat{q}(w,d)]$, $\mathrm{Var}_f[\hat{q}(w,*)]$ and $\mathrm{Var}_f[\hat{q}(*,*)]$) through the delta method. Specifically, the discussion will first present the calculation formulas for the parameter variance and their simplified versions for the generalized linear models with logarithmic link function, then the process variance calculation for the over-dispersed Poisson model, and summarize at the end.

1. The calculation of $\mathrm{Var}_f[\hat{q}(w,d)]$ for each future incremental payment. By the first-order approximation or the delta method, the parameter variance can be estimated, and specifically, in matrix form, is:

$$\mathrm{Var}[\hat{q}(w,d)] = \underbrace{\left[\frac{\partial \hat{q}(w,d)}{\partial p_k}\right]}_{1\times K} \cdot \underbrace{\left[\mathrm{Cov}(p_{k_1}, p_{k_2})\right]}_{K\times K} \cdot \underbrace{\left[\frac{\partial \hat{q}(w.d)}{\partial p_k}\right]^T}_{K\times 1}, \tag{A.1}$$

where p_k is the model parameter, and K is the number of parameters. Notice that $\hat{q}(w,d)$ is the incremental loss function for development period d and accident year w. For the generalized linear model with logarithmic link function, or $m_{wd} = \ln(c + \alpha_w + \beta_d)$ in which α is an accident year-specific parameter and β is a development year-specific parameter, (A.1) can be simplified to

$$\mathrm{Var}[\hat{q}(w,d)] = \hat{m}_{wd}^2 \cdot \mathrm{Var}[\hat{\eta}_{wd}], \tag{A.2}$$

where, $\hat{\eta}_{wd} = c + \alpha_w + \beta_d$. The value of $\mathrm{Var}[\hat{\eta}_{wd}]$ is calculated by using the variance-covariance matrix of the model parameters, which can be directly obtained from the computer output. For the nonlinear model, $\partial \hat{q}(w,d) / \partial p_k$ can be complicated to compute, as $\hat{q}(w,d)$ may take complicated functional forms.

For the over-dispersed Poisson model, the process variance is simply

$$\mathrm{Var}[q(w,d)] = \phi m_{wd}. \tag{A.3}$$

The formula for calculating $\mathrm{Var}_f[\hat{q}(w,d)]$ is summarized as

$$\mathrm{Var}_f[\hat{q}(w,d)] = \mathrm{Var}[q(w,d)] + \left[\frac{\partial \hat{q}(w,d)}{\partial p_k}\right] \cdot \left[\mathrm{Cov}(p_{k_1}, p_{k_2})\right] \cdot \left[\frac{\partial \hat{q}(w,d)}{\partial p_k}\right]^T. \tag{A.4}$$

2. The calculation of $\mathrm{Var}_{\mathrm{f}}[\hat{q}(w,*)]$ for future payments in a particular accident year. The parameter variance of the future payment estimate for accident year w is calculated as:

$$\mathrm{Var}[\hat{q}(w,*)] = \sum_{d_1,d_2 \in \Delta_w} \mathrm{Cov}[\hat{q}(w,d_1),\hat{q}(w,d_2)] = \sum_{d \in \Delta_w} \mathrm{Var}[\hat{q}(w,d)] + 2\sum_{\substack{d_1,d_2 \in \Delta_w \\ d_1 \neq d_2}} \mathrm{Cov}[\hat{q}(w,d_1),\hat{q}(w,d_2)]. \tag{A.5}$$

The first term on the right hand side of the above equation, $\sum_{d \in \Delta_w} \mathrm{Var}[\hat{q}(w,d)]$, can be obtained from the calculation of $\mathrm{Var}_{\mathrm{f}}[\hat{q}(w,d)]$; the computation of the second term, however, is not straightforward. Again using the delta method, the second term is approximated as:

$$\sum_{\substack{d_1,d_2 \in \Delta_w \\ d_1 \neq d_2}} \mathrm{Cov}[\hat{q}(w,d_1),\hat{q}(w,d_2)] = \sum_{\substack{d_1,d_2 \in \Delta_w \\ d_1 \neq d_2}} \sum_k \left(\frac{\partial \hat{q}(w,d_1)}{\partial p_{k_1}}\right)\left(\frac{\partial \hat{q}(w,d_2)}{\partial p_{k_2}}\right) \mathrm{Cov}(p_{k_1}, p_{k_2}), \tag{A.6}$$

or in matrix form,

$$\sum_{\substack{d_1,d_2 \in \Delta_w \\ d_1 \neq d_2}} \mathrm{Cov}[\hat{q}(w,d_1),\hat{q}(w,d_2)] = \sum_{\substack{d_1,d_2 \in \Delta_w \\ d_1 \neq d_2}} \underbrace{\left[\frac{\partial \hat{q}(w,d_1)}{\partial p_{k_1}}\right]}_{1\times K} \cdot \underbrace{\left[\mathrm{Cov}(p_{k_1}, p_{k_2})\right]}_{K\times K} \cdot \underbrace{\left[\frac{\partial \hat{q}(w,d_2)}{\partial p_{k_2}}\right]^T}_{K\times 1}. \tag{A.7}$$

In the case of generalized linear models with logarithmic link function, the formula for parameter variance is simplified to:

$$\mathrm{Var}[\hat{q}(w,*)] = \underbrace{\left[m_{wd}\right]}_{1\times J^w} \cdot \underbrace{\left[\mathrm{Cov}(\eta_{wd_1}, \eta_{wd_2})\right]}_{J^w \times J^w} \cdot \underbrace{\left[m_{wd}\right]^T}_{J^w \times 1}, \tag{A.8}$$

where J^w is the number of development years left for accident year w, and $\mathrm{Cov}(\eta_{wd_1}, \eta_{wd_2})$ is the variance-covariance matrix of η_{wd}, the elements of which are computed from the variance-covariance matrix of the model parameters.

For the over-dispersed model, the process variance for the accident year future payment estimate is the product of the accident year future payment estimate and the scale parameter which tends to capture over-dispersion, or mathematically,

$$\mathrm{Var}[q(w,*)] = \sum_{d \in \Delta_w} \phi\, \hat{m}_{wd}. \tag{A.9}$$

In summary,

$$\mathrm{Var_f}[\hat{q}(w,*)] = \sum_{d\in\Delta_w}\mathrm{Var_f}[\hat{q}(w,d)] + 2\cdot \sum_{\substack{d_1,d_2\in\Delta_w\\ d_1\neq d_2}}\left[\frac{\partial\hat{q}(w,d_1)}{\partial p_k}\right]\cdot[\mathrm{Cov}(p_{k_1},p_{k_2})]\cdot\left[\frac{\partial\hat{q}(w,d_2)}{\partial p_k}\right]^T . \tag{A.10}$$

In plain English, the calculation of the variance of the distribution of accident year future payments should take into account any correlations between the predicted values for different development periods of the same accident year, besides the variance of each of the individual predicted values.

3. The calculation of $\mathrm{Var_f}[\hat{q}(*,*)]$ for total future payments for all accident years combined. The parameter variance of the total future payment estimate for all accident years combined should add in the covariance terms that account for the correlation between the predicted values of different accident years. Mathematically, it is

$$\mathrm{Var}[\hat{q}(*,*)] = \sum_{w\in\Delta}\mathrm{Var}[\hat{q}(w,*)] + 2\sum_{\substack{d_1,d_2\in\Delta\\ w_1,w_2\in\Delta\\ w_1\neq w_2}}\mathrm{Cov}[\hat{q}(w_1,d_1),\hat{q}(w_2,d_2)]. \tag{A.11}$$

Similar to the results in (A.6), the calculation of the second term is approximated as:

$$\sum_{\substack{d_1,d_2\in\Delta\\ w_1,w_2\in\Delta\\ w_1\neq w_2}}\mathrm{Cov}[\hat{q}(w_1,d_1),\hat{q}(w_2,d_2)] = \sum_{\substack{d_1,d_2\in\Delta\\ w_1,w_2\in\Delta\\ w_1\neq w_2}}\sum_k\left(\frac{\partial\hat{q}(w_1,d_1)}{\partial p_{k_1}}\right)\left(\frac{\partial\hat{q}(w_2,d_2)}{\partial p_{k_2}}\right)\mathrm{Cov}(p_{k_1},p_{k_2}), \tag{A.12}$$

or,

$$\sum_{\substack{d_1,d_2\in\Delta\\ w_1,w_2\in\Delta\\ w_1\neq w_2}}\mathrm{Cov}[\hat{q}(w_1,d_1),\hat{q}(w_2,d_2)] = \sum_{\substack{d_1,d_2\in\Delta\\ w_1,w_2\in\Delta\\ w_1\neq w_2}}\underbrace{\left[\frac{\partial\hat{q}(w_1,d_1)}{\partial p_k}\right]}_{1\times K}\cdot\underbrace{\left[\mathrm{Cov}(p_{k_1},p_{k_2})\right]}_{K\times K}\cdot\underbrace{\left[\frac{\partial\hat{q}(w_2,d_2)}{\partial p_k}\right]^T}_{K\times 1} . \tag{A.13}$$

For the generalized linear model, a simplified formula for calculating the parameter variance is:

$$\mathrm{Var}[\hat{q}(*,*)] = \underbrace{\left[m_{w_1d_1}\right]}_{1\times\sum_w J^w}\cdot\underbrace{\left[\mathrm{Cov}(\eta_{w_1d_1},\eta_{w_2d_2})\right]}_{\sum_w J^w\times\sum_w J^2}\cdot\underbrace{\left[m_{w_2d_2}\right]^T}_{\sum_w J^w\times 1}, \tag{A.14}$$

where $\sum_w J^w$ is the total number of incremental losses in the future loss triangle (the lower triangle).

Again for the over-dispersed model, the process variance for the total future payment estimate for all accident years combined is simply

$$\text{Var}[q(*,*)] = \sum_{w,d \in \Delta} \phi\, m_{wd} \; . \tag{A.15}$$

In summary, the formula for calculating $\text{Var}_{\text{f}}[\hat{q}(*,*)]$ is:

$$\text{Var}_{\text{f}}[\hat{q}(*,*)] = \sum_{w} \text{Var}_{\text{f}}[\hat{q}(w,*)] + 2 \cdot \sum_{\substack{d_1, d_2 \in \Delta \\ w_1, w_2 \in \Delta \\ w_1 \neq w_2}} \left[\frac{\partial \hat{q}(w_1, d_1)}{\partial p_k} \right] \cdot \left[\text{Cov}(p_{k_1}, p_{k_2}) \right] \cdot \left[\frac{\partial \hat{q}(w_2, d_2)}{\partial p_{k_2}} \right]^T . \tag{A.16}$$

In words, the variance of the distribution of the total future payments is the sum of the variances for each accident year future payment estimate and the covariances between accident year future payment estimates.

Appendix B. An Example of Estimating Future Payment Variability

This appendix provides an illustrative example for the estimation of future payment variability. A sample of triangle data set contains gross incurred losses for 40 accident quarters (AQ) and 40 development quarters. The data relate to bodily injury coverage in auto insurance.

Future payment is modeled in both generalized linear and non-linear models. Here are some specific assumptions made in the estimation. First, for the nonlinear model,

$$m_{wd} = U_w \cdot [G(t_{d+1} \mid \theta, \omega) - G(t_d \mid \theta, \omega)], \tag{B.1}$$

where U_w is the ultimate loss for accident year w, and $G(t_d \mid \theta, \omega)$ is the loss emergence function, which is assumed to be a loglogistic function and has the following form

$$G(t_d \mid \theta, \omega) = \frac{t_d^{\ \omega}}{t_d^{\ \omega} + \theta^{\omega}}. \tag{B.2}$$

Also in model estimation, the chain ladder estimation method has been used[116], which implies that the loss emergence pattern has been assumed to be constant over the years. This could be problematic, considering the changes in case reserve adequacy and in the rate of settlement of claims over time.[117] Considering the possibility of time varying of risk parameters, the 40-AQ triangle is divided into two smaller triangles, each of which only contains the most recent 18 and 10 accident quarters, respectively. In the estimation, for this particular data set, it is reasonable to assume that all incurred losses have fully developed after 40 calendar quarters, or tail factors are ignored after 40 quarters' development. For the two smaller triangles, the tail factors are also assumed to be zero here; that is, no loss development occurs after 18 and 10 accident quarters, respectively. In practice, appropriate tail factors should be chosen if losses in the oldest accident period have not fully developed. Second, the estimation of the *GLM* model assumes logarithmic link function and Poisson distribution form.

The estimation and model evaluation results presented below are obtained for each of the three data sets. All the calculations are programmed in *SAS/IML*. As the example is used only for illustrative purposes, exemplifying the estimation results and the procedures for

---

[116]For the nonlinear model, there are 42 parameters in total (the ultimate loss for each of the 40 accident quarters, θ and ω) that have been estimated.

[117] See Berquist and Sherman [5].

model evaluation, the tables and figures will not be discussed in detail.

Table 1 shows the estimates of the means and standard errors of the distribution of future payments for each accident quarter. The standard error for each accident quarter's future payment distribution, which is composed of process variance and parameter variance, is calculated using the approach described in Appendix A. Several observations are worthy of note. First, the standard errors are larger for more recent accident quarters, since a smaller percentage of losses has emerged. Thus, more uncertainty is associated with these quarters. However, the standard error as a percentage of the mean increases as the accident quarter ages. This may show that the company's bodily injury line is short-tailed, and outstanding liabilities after several quarters' development become very small. Second, for the recent accident quarters, the percentage standard errors calculated from the nonlinear model are larger than the corresponding ones from the *GLM* model. For instance, if using the 40-AQ gross incurred loss triangle to estimate the most recent 10 accident quarters, the nonlinear model yields the percentage standard errors that are generally 10%-20% larger when compared with the *GLM* model. This can be explained by considering that the *GLM* model is virtually a specific and simplified version of the nonlinear model. Taking logarithms on both sides of $m_{wd} = x_w y_d$ would essentially give the *GLM*, except that in the estimation of the nonlinear model, the loss emergence pattern is specifically modeled as a random variable, while for that of the *GLM*, it is treated as parameters to be directly estimated. Third, for a shorter development period (or equivalently, for the cases where fewer accident and development quarters are used in the estimation), the point estimates for future losses are much higher for each accident quarter. This is due to the fact that many of the incremental payments become negative for higher development quarters in the data set.

Table 1. Estimated Future Losses and Prediction Errors (PEs): GLM and Nonlinear Models

	40 Accident Quarters (1993Q3 - 2003Q2)			18 Accident Quarters (1999Q1 - 2003Q2)			10 Accident Quarters (2001Q1 - 2003Q2)		
AQ	**Est. Future Loss**	**PE**	**PE %**	**Est. Future Loss**	**PE**	**PE %**	**Est. Future Loss**	**PE**	**PE %**
GLM:									
1999Q3	125	122	98%	10	34	328%			
1999Q4	146	133	91%	29	58	201%			
2000Q1	151	133	88%	39	65	168%			
2000Q2	182	145	80%	58	84	146%			
2000Q3	219	154	70%	102	108	106%			
2000Q4	268	165	62%	103	108	106%			
2001Q1	356	177	50%	121	114	94%			
2001Q2	439	196	45%	213	149	70%	304	175	57%
2001Q3	501	201	40%	329	176	54%	553	271	49%
2001Q4	736	241	33%	545	226	41%	855	336	39%
2002Q1	968	272	28%	833	278	33%	1099	371	34%
2002Q2	1381	325	24%	1236	341	28%	1611	445	28%
2002Q3	2042	400	20%	1897	429	23%	2476	555	22%
2002Q4	2835	479	17%	2580	511	20%	3191	642	20%
2003Q1	4113	592	14%	4007	660	16%	4680	780	17%
2003Q2	7383	867	12%	8018	1049	13%	9285	1208	13%
Nonlinear Model:									
1999Q3	184	168	91%	233	214	92%			
1999Q4	215	182	85%	272	232	85%			
2000Q1	217	183	84%	275	233	85%			
2000Q2	246	196	79%	311	248	80%			
2000Q3	289	213	73%	364	270	74%			
2000Q4	308	220	71%	388	279	72%			
2001Q1	324	226	70%	406	287	71%			
2001Q2	393	250	64%	493	317	64%	977	404	41%
2001Q3	427	261	61%	533	332	62%	1032	418	40%
2001Q4	545	297	55%	679	378	56%	1280	470	37%
2002Q1	650	326	50%	806	416	52%	1473	512	35%
2002Q2	864	380	44%	1068	486	45%	1879	593	32%
2002Q3	1238	462	37%	1523	592	39%	2560	717	28%
2002Q4	1945	592	30%	2377	763	32%	3768	915	24%
2003Q1	3013	764	25%	3649	987	27%	5325	1147	22%
2003Q2	6439	1252	19%	7683	1626	21%	9731	1749	18%

Table 2. Comparison of Estimated AQ Ultimate Losses: GLM and Nonlinear Model

	40 Accident Quarters (1993Q3 - 2003Q2)		18 Accident Quarters (1999Q1 - 2003Q2)		10 Accident Quarters (2001Q1 - 2003Q2)	
AQ	GLM Estimates	Nonlinear Estimates	GLM Estimates	Nonlinear Estimates	GLM Estimates	Nonlinear Estimates
1999Q3	10434	10493	10319	10542		
1999Q4	11245	11314	11128	11371		
2000Q1	10480	10546	10367	10604		
2000Q2	10846	10910	10722	10975		
2000Q3	11579	11649	11462	11724		
2000Q4	11136	11176	10971	11256		
2001Q1	10526	10494	10291	10576	10170	10994
2001Q2	11280	11234	11054	11334	11145	11818
2001Q3	10684	10610	10512	10716	10736	11215
2001Q4	11778	11587	11587	11721	11897	12322
2002Q1	11856	11538	11721	11694	11986	12361
2002Q2	12955	12439	12811	12643	13185	13454
2002Q3	14645	13841	14500	14126	15079	15163
2002Q4	16737	15847	16482	16279	17093	17670
2003Q1	17195	16095	17090	16731	17763	18407
2003Q2	19316	18372	19951	19616	21217	21664
Correlation	0.998		0.999		0.998	

Table 2 compares the ultimate loss estimates for each accident quarter using *GLM* with those from the nonlinear model. The correlation between these two sets of estimates is 0.998, showing that the two models yield very similar ultimate loss estimates. Table 3 reports the parameter estimates, their standard errors, and *p*-values for the *GLM* model.

The AIC and BIC criteria are calculated as follows, respectively,

$$\mathrm{AIC}(K) = \ln\left(\frac{\sum_w \sum_d [q(w,d) - m_{wd}]^2}{n}\right) + \frac{2 \cdot K}{n}, \quad \text{and} \tag{B.3}$$

$$\mathrm{BIC}(K) = \ln\left(\frac{\sum_w \sum_d [q(w,d) - m_{wd}]^2}{n}\right) + \frac{K \cdot \ln(n)}{n}. \tag{B.4}$$

Note that K is the number of parameters that are estimated and n is the total number of incremental losses. Table 4 gives the calculation results for AIC and BIC for the *GLM* and nonlinear models.

Table 3. Parameter Estimates and Standard Errors: GLM

	40 Accident Quarters (1993Q3 - 2003Q2)			18 Accident Quarters (1999Q1 - 2003Q2)			10 Accident Quarters (2001Q1 - 2003Q2)		
Parameters (AQ)	Parameter Estimate	Standard Error	Pr > Chi Sq	Parameter Estimate	Standard Error	Pr > Chi Sq	Parameter Estimate	Standard Error	Pr > Chi Sq
Intercept	3.886	3.083	0.208	0.086	12.123	0.994	6.304	0.348	<.0001
1999Q3	-0.598	0.102	<.0001	-0.649	0.116	<.0001			
1999Q4	-0.517	0.100	<.0001	-0.561	0.113	<.0001			
2000Q1	-0.535	0.101	<.0001	-0.586	0.114	<.0001			
2000Q2	-0.468	0.100	<.0001	-0.525	0.114	<.0001			
2000Q3	-0.447	0.099	<.0001	-0.493	0.112	<.0001			
2000Q4	-0.442	0.100	<.0001	-0.499	0.113	<.0001			
2001Q1	-0.549	0.102	<.0001	-0.606	0.116	<.0001	-0.663	0.111	<.0001
2001Q2	-0.477	0.100	<.0001	-0.534	0.114	<.0001	-0.586	0.109	<.0001
2001Q3	-0.592	0.102	<.0001	-0.641	0.116	<.0001	-0.681	0.112	<.0001
2001Q4	-0.495	0.100	<.0001	-0.543	0.114	<.0001	-0.579	0.109	<.0001
2002Q1	-0.488	0.100	<.0001	-0.532	0.114	<.0001	-0.571	0.109	<.0001
2002Q2	-0.399	0.099	<.0001	-0.443	0.112	<.0001	-0.476	0.107	<.0001
2002Q3	-0.277	0.097	0.004	-0.319	0.110	0.004	-0.342	0.105	0.001
2002Q4	-0.143	0.094	0.129	-0.191	0.107	0.075	-0.216	0.103	0.035
2003Q1	-0.116	0.096	0.224	-0.155	0.109	0.154	-0.178	0.104	0.086
2003Q2	0.000	0.000	.	0.000	0.000	.	0.000	0.000	.

Parameters (DQ)	Parameter Estimate	Standard Error	Pr > Chi Sq	Parameter Estimate	Standard Error	Pr > Chi Sq	Parameter Estimate	Standard Error	Pr > Chi Sq
0-3	5.501	3.083	0.074	9.302	12.123	0.443	3.084	0.340	<.0001
3-6	4.038	3.083	0.190	8.028	12.123	0.508	1.911	0.343	<.0001
6-9	3.321	3.083	0.281	7.264	12.123	0.549	1.093	0.350	0.002
9-12	2.475	3.084	0.422	6.155	12.124	0.612	-0.136	0.403	0.736
12-15	2.567	3.084	0.405	6.443	12.123	0.595	0.490	0.369	0.184
15-18	2.291	3.084	0.458	6.142	12.124	0.612	0.169	0.389	0.663
18-21	2.025	3.085	0.512	6.089	12.124	0.616	-0.264	0.432	0.541
21-24	1.822	3.086	0.555	5.665	12.125	0.640	-0.236	0.513	0.646
24-27	1.404	3.088	0.649	5.480	12.125	0.651	0.000	0.000	.
27-30	0.626	3.100	0.840	4.863	12.133	0.689	0.000	0.000	.
30-33	1.408	3.096	0.649	3.881	12.158	0.750			
33-36	0.425	3.118	0.892	0.606	14.719	0.967			
36-39	0.078	3.118	0.980	4.153	12.149	0.733			
39-42	-0.433	3.153	0.891	3.265	12.311	0.791			
42-45	-1.328	3.230	0.681	2.856	12.197	0.815			
45-48	-0.953	3.205	0.766	3.322	12.224	0.786			

Table 4. Criteria for Assessing Goodness of Fit: GLM and Nonlinear Model

	40 Accident Quarters (1993Q3 - 2003Q2)			18 Accident Quarters (1999Q1 - 2003Q2)			10 Accident Quarters (2001Q1 - 2003Q2)		
Criterion	**DF**	**Value**	**Value/DF**	**DF**	**Value**	**Value/DF**	**DF**	**Value**	**Value/DF**
GLM:									
Scaled Deviance	536	442.10	0.82	103	106.51	1.03	33	37.97	1.15
Scaled Pearson Chi-Square	536	536.00	1.00	103	103.00	1.00	33	33.00	1.00
AIC	.	10.61	.	.	11.42	.	.	11.25	.
BIC/SIC	.	11.18	.	.	12.16	.	.	11.97	.
Nonlinear Model:									
AIC	.	10.44	.	.	11.28	.	.	11.59	.
BIC/SIC	.	10.68	.	.	11.64	.	.	12.03	.

Figures 1 and 2 plot out the scaled Pearson residuals against incremental age for each of the three data sets. For the loss triangle data used in this example, most of the residual points are randomly scattered around the zero line for both models, and as a result, neither of them should be rejected based on the validity of the model assumptions.

Figure 1a. Nonlinear Models: Scaled Pearson Residuals - 40 Accident Quarters

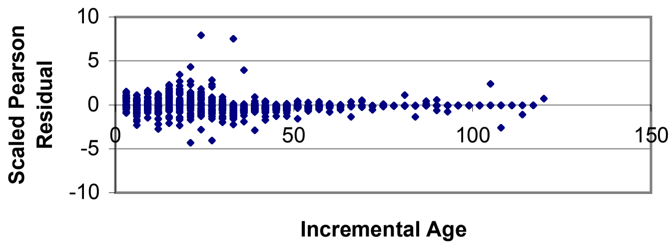


Figure 1c. Nonlinear Model: Scaled Pearson Residuals - 10 Accident Quarters

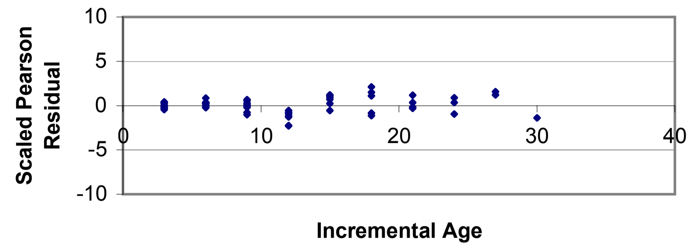


Figure 2b. GLM: Scaled Pearson Residuals - 18 Accident Quarters

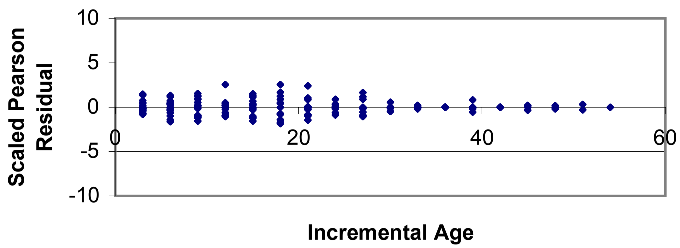


Figure 2a. GLM: Scaled Pearson Residuals - 40 Accident Quarters

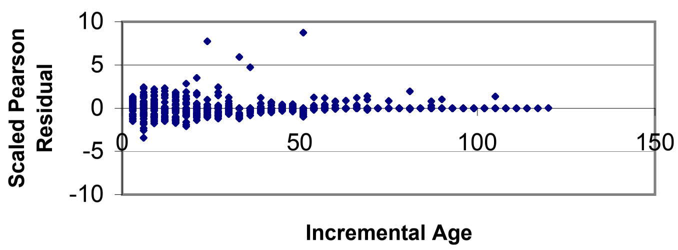


Figure 2c. GLM: Scaled Pearson Residuals - 10 Accident Quarters

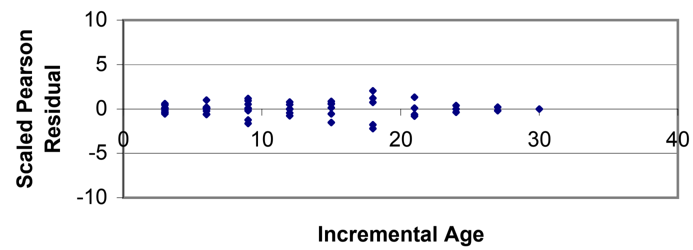


Figure 2b. GLM: Scaled Pearson Residuals - 18 Accident Quarters