Measuring Loss Reserve Uncertainty

Panning, William H.[1]

Willis Re
1 Wall Street Plaza
88 Pine Street, 4th Floor
New York, NY 10005

Office Phone: 212-820-7680
Fax: 212-344-4646
Email: bill.panning@willis.com

Abstract

An accurate and understandable measure of loss reserve uncertainty – the magnitude of potential differences between forecast and actual future loss payments – can be used to address significant issues in surplus management, in pricing and capital allocation, and in the management of uncertainty. Here I present a parametric method for measuring loss reserve uncertainty, specifically defined as the coefficient of variation of estimated future loss payments. The method is simple relative to the alternative methods described in the literature, and it avoids a number of pitfalls that can potentially distort any method of estimating reserves or loss reserve uncertainty. I validate its accuracy by using 10,000 simulated loss reserve triangles to demonstrate that it successfully estimates both loss reserves and reserve uncertainty, which includes both process and parameter risk. The measure utilized and the data employed by this method facilitate comparisons across different lines of business and different firms, and can arguably be applied to reserve estimates obtained using somewhat different data and methods. In presenting the method I have taken pains to make the exposition accessible to a wide audience, and I explain in concrete detail how the method can be implemented in Excel.

Keywords: Loss Reserve, Reserve Estimation, Reserve Uncertainty, Coefficient of Variation, Paid Loss Triangle.

---

[1] The author can be contacted at bill.panning@Willis.com. I am grateful to my colleagues at Willis Re for valuable comments and suggestions concerning this paper and the method it presents. I am especially indebted to Yves Provencher, of Willis Re, for his patient and thorough evaluation of two versions of this paper, and to Richard Goldfarb, of Ernst and Young, for his insightful comments and careful verification of key results. Given such expert advice, I can hardly avoid taking responsibility for any errors that remain. The opinions and recommendations expressed in this paper are those of the author, and do not necessarily represent the views of Willis Re.

1: What loss reserve uncertainty is, and why it matters

Property-casualty loss reserves are estimates of the total future payments that will be required to settle claims on accidents that have already occurred. Because such estimates are inherently imprecise, for reasons discussed later on, insurers may ultimately pay out more or less to claimants than is forecast in the firm's current reserve. Loss reserve uncertainty (LRU) is a measure of the magnitude of this potential difference between forecast and actual loss payments.

In this paper I propose, explain, and justify a particular method of estimating loss reserve uncertainty. This method has several important virtues. First, it is simple, and so can be implemented on a spreadsheet and applied to universally available data. Second, the method is accurate, since it addresses and avoids a number of pitfalls in statistical estimation and also meets a Monte Carlo test of its precision. Third, the resulting estimates are comparable across different lines of business and different firms. Finally, the measure of LRU is scalable, so that it is applicable to reserves that have been estimated in different ways.

A method for estimating LRU that has these characteristics is likely to be extremely useful to insurers, investors, regulators, and rating agencies, for estimating surplus adequacy, for pricing and capital allocation, and for determining the potential significance of reserve developments.

Estimating surplus adequacy. The uncertainty of an insurer's loss reserve has direct implications for its required surplus or reinsurance. The greater an insurer's LRU, the greater the surplus or reinsurance it needs to cope with potential scenarios in which ultimate losses exceed forecast losses. In the absence of an accepted measure of LRU, these various audiences have relied on indirect measures of surplus adequacy such as premium-to-surplus or reserve-to-surplus ratios relative to peer companies or to industry averages. Such relative evaluations can be quite misleading in an industry that exhibits profound swings in pricing and reserve adequacy.

The problem of estimating surplus adequacy is a fundamental issue in Enterprise Risk Management, which attempts to estimate the total capital needed by an insurer to withstand potential losses from all sources of risk. Before total enterprise risk can be managed, it must first be measured. For most property casualty firms, the principal sources of risk are loss reserve uncertainty, asset risk (principally due to equities), pricing risk (potential differences between forecast losses and actual losses), and credit risk on receivables and recoverables. Of these, LRU is typically the largest and also the most difficult to estimate.

Pricing and capital allocation. Many insurers allocate capital to different lines of business and evaluate pricing adequacy by the return on capital achieved in each line. Although firms may employ different methods for allocating capital among different lines of business, there is consensus that the capital allocated to a particular line should reflect the fact that estimated losses are uncertain. Consequently, the capital allocated to a line of business should reflect the rapidity with which its reserve runs off and the magnitude of uncertainty involved. Measuring loss reserve uncertainty can therefore inform and improve capital allocation and pricing.

Managerial feedback. An insurer's loss reserve is a forecast of all future loss payments, including those anticipated during the next calendar year. The measure I propose can be adapted to estimate the uncertainty of this calendar year estimate. What makes this important is that this estimated uncertainty provides a useful benchmark against which any difference between actual and forecast loss payments can be evaluated. For example, if calendar year paid losses are 20% higher than forecast, this is of little concern when the standard deviation of those forecast losses is 15%. But if, instead, the standard deviation is 6%, then the 20% deviation should trigger significant managerial concern.

2: Prior studies of loss reserve uncertainty

Given the potential importance of measuring LRU, it is not surprising that the number of papers on the subject has grown significantly during the past decade. Relevant papers include Ashe (1986), Barnett and Zehnwirth (2000), Braun (2004), Brehm (2002), England and Verrall (1999, 2001, 2002), Halliwell (1996), Hayne (2003), Hodes, Feldblum, and Blumsohn (1996), Holmberg (1994), Kloek (1998), Mack (1993, 1994, 1995, 1999), Murphy (1994), Taylor (1987, 2004), Taylor and Ashe (1983), and Verrall (1994). Rather than describing each paper individually, I shall comment on this body of work taken as a whole.

Chain ladder focus. First, a central assumption of much of this literature is that the chain ladder method for estimating reserves is the obligatory starting point for estimating reserve uncertainty. For example, in their excellent review of a variety of models and techniques for estimating reserves and reserve uncertainty, England and Verrall (2002) note that a principal objective of the models they review is "to give the same reserve estimates as the chain-ladder technique" (p. 448). By contrast, there are relatively few studies like Standard (1985), Narayan and Warthen (1997), Barnett and Zehnwirth (2000), and Taylor (2003) that focus on the key assumptions and comparative adequacy of the chain ladder method. Here I make no attempt to ensure that my proposed method agrees with the chain ladder method in estimating the unknown parameters of some paid loss triangle. Instead, I shall use known parameters to simulate paid loss triangles and determine whether my proposed method is able to accurately estimate these parameters, simulated reserves, and simulated reserve uncertainty. The point is to obtain estimates that are correct, whether or not they agree with a widely-used method.

Absence of estimation criteria. Second, apart from the special place accorded to the chain ladder method, much of the literature seems to assume a kind of algorithmic democracy, in which one technique for estimating reserves or LRU is considered as good as any other. (This assumption reaches its inevitable conclusion when the results obtained from different methods are averaged.) With few exceptions, there is no discussion of criteria that must be met in order for estimates of reserves or LRU to be accurate. The notable exceptions here are Ashe (1986), Barnett and Zehnwirth (2000), Halliwell (1996), Taylor (1987), and Taylor and Ashe (1983), but even here the relevant issues are typically either assumed or discussed very briefly. Here I explain at some length conditions that are crucial to accurate estimation, and show specifically what must be done to meet those conditions.

Complexity. Third, the procedures used to estimate LRU are typically quite complex. Moreover, some recommended procedures, such as generalized least squares (GLS) in fact require very strong *a priori* assumptions about variances and covariances. Checking and, when necessary, appropriately modifying these assumptions is indeed feasible, but at the expense of making a complex procedure even more vulnerable to the temptation to over-fit the model, thereby “finding” what one has really assumed. Here I utilize a much simpler procedure that is less elegant but, in this respect, more robust.

3: A measure of loss reserve uncertainty and its merits

No single method of estimating loss reserve uncertainty is appropriate under all circumstances. Much depends upon the type and extent of data available for such an analysis. For example, actuaries within an insurance firm may have access to data that is far more extensive and detailed than the data available to external analysts. Given these differences in available data, internal and external analysts may appropriately utilize different methods to estimate loss reserve uncertainty. Nonetheless, I believe that the results obtained from the method presented here can be applied directly to reserve estimates obtained using other methods and more extensive data.

The procedure I propose has two steps. The first is estimating the loss reserve itself, in dollars. The second is estimating the standard deviation of that loss reserve, again in dollars. Because both of these estimates are in dollars, comparisons across lines of business or across different firms are essentially meaningless, since differences in these numbers will principally be affected by differences in the volume of business in each line or each firm. But if we instead express reserve uncertainty as a ***coefficient of variation*** (the standard deviation of the estimated reserve as a percentage of the estimated reserve), we arrive at a measure that has three important properties. First, it ***can be compared across different lines of business***. A line of business in which the coefficient of variation is 6% is clearly less risky (in this respect, at least) than one in which the coefficient of variation is, say, 15%. Second, this measure of LRU ***can be compared across different firms for the same line of business***. If the coefficient of variation for workers' compensation is smaller for one firm than for another, it is pretty clear that this line of business is less risky for the first firm than for the second. This fact has enormous implications for the measurement of capital adequacy. Third, I consider it likely that this measure of LRU ***can be applied to reserves that have been estimated by methods other than the one recommended here***. My argument here is very simple. Suppose that I utilize the method and data proposed here to forecast future loss payments (i.e., the reserve) for some insurer and obtain a value R. Suppose also that the firm's own actuaries, utilizing a different method and far more extensive data, obtain an estimated reserve value of R^*, where $R^* = aR$ (i.e., some positive constant times the value R obtained using the method and data recommended here). Under rather broad conditions it is the case that if $R^* = aR$, then the standard deviation $S^* = aS$, where S^* is the standard deviation of the R^* and S is the standard deviation of R. If this is so, then it is necessarily true that $S^*/R^* = S/R$. In other words, the coefficient of variation S/R will be (approximately) the same regardless of the method used to estimate reserves.

4: The data needed to measure loss reserve uncertainty

Comparing LRU across different lines of business and, in particular, across different firms, requires that data that is commonly available and consistently defined. The data utilized here consists of the paid loss triangles reported in Schedule P, Part 3, of the Annual Statement required by the National Association of Insurance Commissioners. This data is publicly available for all insurance companies licensed in the United States.

Table 1 is an example of such data.[2] The rows of this table are accident years: the calendar years in which accidents occurred. The columns are development years: calendar years in which claims payments for those accidents were actually made. A single accident can trigger multiple claim payments occurring in different development years. For example, an auto accident in November 1995 could trigger a payment for physical damage to the insured's vehicle in December of that same year, and an additional claim payment, for bodily injury medical costs, in 1996. Litigation, if it occurs, may delay claim payments into later years. Table 1 shows that, for all accidents occurring in 1994, $ 624 million in claims were paid that same year, a cumulative total of $ 2.1 billion had been paid by the year-end 1996, and $ 2.9 billion by year-end 2003.

Table 1: Cumulative Paid Losses (millions)

Years in Which Losses Were Incurred	**Development Year**									
	1994	**1995**	**1996**	**1997**	**1998**	**1999**	**2000**	**2001**	**2002**	**2003**
1994	624	1,595	2,066	2,366	2,559	2,685	2,765	2,818	2,860	2,895
1995		695	1,503	1,975	2,295	2,496	2,631	2,727	2,784	2,831
1996			668	1,477	1,968	2,263	2,447	2,562	2,645	2,707
1997				696	1,540	2,055	2,357	2,551	2,699	2,806
1998					770	1,670	2,225	2,583	2,822	2,985
1999						690	1,515	2,051	2,436	2,666
2000							544	1,321	1,859	2,191
2001								563	1,355	1,852
2002									593	1,416
2003										621

In Table 2, which is a reformatted version of Table 1, each row after the first has been shifted to the left, and development years have been renumbered, from zero to nine, to represent the number of years that have elapsed since the year in which the accident occurred. (I shall refer to these development years as DY0, DY1, and so on, and to accident years, also renumbered, as AY0, AY1, and so on.) The rearranged data in Table 2 more clearly shows how the claim payments for an accident year develop over time, represented by the number of development years subsequent to the year of the accident. As before, these are cumulative claim payments. Table 2 is typical of the data commonly used to estimate loss reserves for a single line of

[2] This data is real rather than artificial, but multiplied by a constant to conceal the identity of its source.

business. To estimate reserves, one must estimate, for all accident years, the difference between the amounts already paid and the ultimate amounts that will have been paid when all claims are finally settled. (This may occur well after DY9; if so, then years prior to AY0 will also have to be analyzed, using separate data. Here I will ignore all prior years.)

Table 2: Accident Year x Development Year Cumulative Paid Losses (millions)

		0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
1994	**0**	624	1,595	2,066	2,366	2,559	2,685	2,765	2,818	2,860	2,895
1995	**1**	695	1,503	1,975	2,295	2,496	2,631	2,727	2,784	2,831	
1996	**2**	668	1,477	1,968	2,263	2,447	2,562	2,645	2,707		
1997	**3**	696	1,540	2,055	2,357	2,551	2,699	2,806			
1998	**4**	770	1,670	2,225	2,583	2,822	2,985				
1999	**5**	690	1,515	2,051	2,436	2,666					
2000	**6**	544	1,321	1,859	2,191						
2001	**7**	563	1,355	1,852							
2002	**8**	593	1,416								
2003	**9**	621									

5: The sources of loss reserve uncertainty

Property-casualty loss reserves are estimates – forecasts -- of the total future payments that will be required to settle claims on accidents that have already occurred. The actual future payments may deviate from the forecast amount for several reasons.

Process risk. First, some degree of uncertainty is inherent in the process of settling claims payments. The amount actually paid in a given development year is a complex result of numerous factors – among them the uncertain outcomes and costs of medical diagnoses and treatments, and of court proceedings or settlement negotiations. None of these factors can be easily forecast. Consequently, even at an aggregate level, attempts to predict future claim payments are inescapably imprecise. The existence of such process risk is one reason why our models fit past paid losses only imperfectly, and accounts for the presence of an error term in the prediction equation discussed below.

Parameter risk. Actuarial methods necessarily use past experience to forecast future patterns. But past experience can be misleading. The culprit here is the relatively short period of time – ten years – covered by a typical paid loss triangle, so that parameter estimates are derived from a relatively small number of observations. The paid losses in the triangle are all affected by process risk, but the small number of observations creates substantial sampling error. The result is that past data may, simply by chance, reflect unusually favorable or unfavorable claims experience, and thereby affect the model parameters we are trying to estimate.

As an example, consider the step in the chain ladder method in which one of several weighting methods is used to produce a ratio of cumulative losses in DY5 to cumulative losses in DY4. This and other similar ratios are key parameters in the chain ladder model. But note that in DY5 there are only five cumulative AY losses from which ratios can be formed. If one or more of these five cumulative loss numbers is affected by an unusually large, or unusually small, claim payment in DY5 or in any preceding DY, then the resulting ratio will be atypically large or small. As this example suggests, this problem of sampling error is more acute for firms and lines of business that have few claims involving large payments than for firms with many small claims. Sampling error is likely to be less relevant to private passenger auto than to, say, product liability or D&O.

Especially in low-frequency high-severity lines of business, then, sampling error can lead to distorted estimates of key loss reserve parameters. This important consequence of sampling error can be called parameter risk, since it pertains to the accuracy with which we can use past data to estimate key parameters in our model of reserves or reserve uncertainty. Unlike process risk, which is inherent in the claims settlement process, parameter risk reflects our ignorance of the true parameters that characterize that process and the consequent need for us to use imperfect data to estimate them.

Model risk. All reserve estimates require extrapolation from the past to the future. We use data from the past to create a model of the evolution of claims payments, and we then use this model to forecast future payments. Implicit in this process are two crucial assumptions. One is that we have correctly modeled the past: that we have included all the relevant variables and specified the correct functional form of the model. A second implicit assumption is that the pattern of future claims payments will continue to conform to this model. That is, the way claims are settled in the future will closely resemble the way they have been settled in the past. This implicit assumption may become misleading if there are fundamental changes – known as regime changes – in the claims settlement process at a particular firm (perhaps as a consequence of regulatory or judicial decisions), or in the types of claims being settled (which may change over time due to changes in business mix). These two components of reserve uncertainty can be called model risk, since they pertain to the capability of a model to correctly extrapolate from the parameters of past experience to estimates of future payments.

Model risk can often, although not invariably, be avoided by means of a thorough analysis of the differences (residuals) between the fitted values of past paid losses obtained from a model and the actual paid losses that have been observed. (The paper by Barnett and Zehnwirth (2000) provides an excellent example of the analysis of residuals.) If these residuals exhibit a trend or a sudden temporal shift, then there is good reason to suspect that a regime change has occurred. This possibility can be confirmed by testing a more advanced model that incorporates temporal changes in the value of key parameters. Unfortunately, however, there are tradeoffs in introducing additional variables that reflect temporal changes in key parameters, since doing so is likely to increase our estimate of LRU.

Summary. Process risk essentially reflects the fact that some aspects of the claims settlement process are inherently unpredictable. Parameter risk reflects the fact that, even if we have a correct model of the evolution of paid losses, our estimates of the parameters of this correct model will necessarily be somewhat imprecise. Model risk reflects the possibility that the model we are using may itself be incorrect, so that our ability to predict future loss payments from past paid losses may be impaired. A satisfactory approach to estimating LRU should address all three of its sources. In particular, it should provide systematic ways to avoid, minimize, or detect model risk, and it should quantify both process and parameter risk.

6: Criteria for accurately estimating reserves and reserve uncertainty

The method presented here uses linear regression to fit past loss payments, forecast future loss payments, and estimate LRU. But the use of linear regression will not produce accurate estimates of reserves and LRU unless certain crucial problems are avoided or corrected. Here I describe these problems, their relevance, and what can be done about them. [3]

Linear regression. In linear regression we initially assume a simple relationship between some dependent variable **Y** (here specified as paid losses) and one or more independent variables **X.** The relationship between the two is represented by the equation $\mathbf{Y} = \boldsymbol{\beta}\mathbf{X} + \boldsymbol{\varepsilon}$, where $\boldsymbol{\beta}$ is one or more parameters to be estimated, and $\boldsymbol{\varepsilon}$ is an error term that represents random disturbances or deviations from the predicted relationship between **Y** and **X**. In the simplest possible model, **X** consists of a single independent variable. I shall refer to this as model 1.

In this simple model, process risk is represented by $\boldsymbol{\varepsilon}$, which consists of disturbances that are assumed to have an expected value of zero and a standard deviation that is constant across all observations. Parameter risk is reflected in the fact that the resulting estimated value of $\boldsymbol{\beta}$, represented by **b**, is assumed to be correct, so that $\mathbf{b} = \boldsymbol{\beta}$, which may not be true. Finally, model risk is represented in several ways. For example, model 1 directly assumes that the relationship between **Y** and **X** is indeed linear, that all variables pertinent to **Y** are included in **X**, and that $\boldsymbol{\beta}$ is constant. All of these may in fact be false, but can typically be checked by thorough examining the residuals from the model – the deviations between actual and fitted paid losses.

Bias. If important variables affecting Y are omitted from model 1, the error term is likely to have a nonzero mean, the fitted and forecast values from the model will be biased – their estimated values will systematically deviate from their true values. In the absence of specific data concerning the omitted variables, we can take their influence into account by adding an intercept term to the original model, which now becomes $\mathbf{Y} = \boldsymbol{\alpha} + \boldsymbol{\beta}\mathbf{X} + \boldsymbol{\varepsilon}$. I shall refer to this model as model 2. Since the unnecessary use of an intercept term affects our estimate of LRU, we should use model 2 only when there is convincing evidence that the error terms from model 1 have a mean that significantly differs from zero.

Varying parameters. Another source of model risk is change over time in the value of $\boldsymbol{\beta}$. This may occur due to changes in (a) the firm's claims settlement process, (b) judicial decisions or regulatory requirements, (c) the composition of the firm's policyholders in that line of business, or (d) the structure of a firm's reinsurance program (since paid losses are reported net of

[3] In this section I rely very heavily on Kennedy (2003), a superb elementary presentation of the essentials of econometrics, and on Greene (2000), one of the most widely used advanced texts.

reinsurance recoverable). These and other possible changes may produce sudden or gradual changes over time in the true value **β**, but these changes that will not be reflected in its estimated value **b**. Fortunately, situations of this sort exhibit a characteristic pattern of residuals, and can be corrected by using a slightly more complex model in which **β** is assumed to change linearly over time, so that $\boldsymbol{\beta} = \boldsymbol{\beta_0} + \boldsymbol{\beta_1}t$, where t = 0, 1, . . . n is a time index. When this is substituted into the original model we have a new model, which I shall refer to as model 3: $\mathbf{Y} = \boldsymbol{\beta_0}\mathbf{X} + \boldsymbol{\beta_1}\mathbf{tX} + \boldsymbol{\varepsilon}$. If $\boldsymbol{\beta_1} = 0$, then this model collapses into the original, simpler model 1.

Correlated disturbances. Linear regression models assume that the disturbance terms for each observation are uncorrelated with one another. For the data in Table 2, it sensible to assume – as many others have -- that the disturbances in different accident years are uncorrelated. The important question is whether the disturbance terms within the same accident year are correlated across development years. I will show that they are in cumulative data.

Suppose that, for a given line of business and a given accident year, the expected paid losses are \$40, \$30, \$20, and \$10 in development years zero through three. However, in any given development year the actual paid losses will deviate from these expected paid losses due to a variety of random factors whose net effect in those development years is ε_0, ε_1, ε_2, and ε_3, respectively. There are good reasons for assuming that these four random terms are independent of one another and of all other similar random terms affecting other accident years and development years. However, if we create a table of cumulative paid losses, as in Table 2, we will create correlations among these random terms, since the new disturbance term for AY1, for example, is now $\varepsilon_0 + \varepsilon_1$, which is clearly positively correlated with ε_0, the disturbance term for AY0. In cumulative data, then, an unusually large disturbance in any development year will be reflected in all subsequent cumulative paid losses for that accident year.

The typical consequence of correlated disturbances, explained in both Kennedy and Greene, is that a given model will appear to fit the historical data better than it actually does, so that process error will be underestimated. This, in turn will result in LRU being underestimated as well. Fortunately, the remedy for the correlated disturbances in cumulative paid loss triangles is simple: we should use incremental paid losses rather than cumulative ones. Consequently, the data we will utilize to estimate reserves and reserve uncertainty will be incremental, like that shown in Table 3, which is derived from Table 2.[4] (The boxes in Table 3 are explained later.)

Heteroskedasticity (disturbances with non-constant variances). Linear regression assumes that the disturbance terms for past observations are homoskedastic – i.e., have a constant variance or standard deviation. This assumption is clearly violated in paid loss triangles like Table 3, for the variability of disturbances typically decreases from one development year to the next. Heteroskedastic disturbances reduce the precision of reserve estimates and can also dramatically affect estimates of LRU.

---

[4] See Halliwell (1996) for an alternative solution using Generalized Least Squares, which I discuss below.)

Table 3: Accident Year x Development Year Paid Losses (incremental)

		0	1	2	3	4	5	6	7	8	9
1994	**0**	**624**	971	**471**	300	193	126	80	53	42	35
1995	**1**	**695**	808	**473**	319	201	135	96	57	47	
1996	**2**	**668**	809	**491**	295	184	115	83	63		
1997	**3**	**696**	844	**515**	302	194	148	107			
1998	**4**	**770**	900	**555**	358	239	162				
1999	**5**	**690**	825	**536**	384	231					
2000	**6**	**544**	777	**537**	332						
2001	**7**	**563**	792	**497**							
2002	**8**	**593**	823								
2003	**9**	**621**									

There are two remedies for heteroskedasticity that are relevant to the problem at hand. One is to use a procedure known as Generalized Least Squares (GLS), which is a variation of linear regression that incorporates the use of an assumed or estimated variance-covariance matrix of disturbances (Halliwell, 1996; Taylor and Ashe, 1983). One typical assumption, for example, is that the standard deviation of disturbances is proportional to the observed losses themselves. Besides its complexity, there is a fundamental problem with the use of GLS for estimating reserves and LRU. Whether the variance-covariance matrix is assumed or estimated, the use of GLS introduces additional parameter risk that is not taken into account in the estimate of LRU. Moreover, however useful GLS may be in increasing the accuracy of reserves estimates, when it is applied to the problem of estimating LRU it comes dangerously close to assuming precisely what we are trying to estimate.

A second and far simpler remedy is to assume – quite plausibly – that the standard deviation of disturbance terms is constant within the same development year. What this implies, in practice, is the need to perform separate regressions on each development year. While this procedure may be less elegant than performing a single comprehensive regression for the whole paid loss triangle, it avoids the need to make problematic assumptions about variances and covariances.

Zero correlation between disturbances and independent variables. Halliwell (1996) correctly points out that the classical linear regression model required that the independent variables be non-stochastic. However, both Greene (2000) and Kmenta (1977, pp. 297ff.) demonstrate that this stringent and seldom-met condition can be replaced by one that is far less demanding, namely, that the disturbance terms be independent of the values of the independent variables. However, even if the correlation is slightly positive rather than zero, the effect on the resulting estimates of reserves and LRU is imperceptible, as I shall demonstrate.

Implications and summary. In using linear regression to estimate reserves and LRU, it is essential to avoid the various pitfalls just described. If one or more of these problems do occur, then estimates of reserves and LRU may be seriously affected. It should be noted that this conclusion applies not only to the use of linear regression, but to the use of other estimation procedures as well.

The immediate implications for modeling reserves and LRU can be summarized as follows: (a) if bias appears to be a problem, use model 2 rather than model 1; (b) if the model parameters appear to change over time, use model 3; (c) to avoid correlated errors, use incremental paid loss triangles; (d) to avoid heteroskedasticity, analyze different development years separately; (e) the use of non-stochastic independent variables, as advocated by Halliwell (1996) is unnecessary provided that there is no correlation between disturbances and the independent variables.

7: Estimating loss reserves

I will present the full procedure for estimating reserves in this section, and for estimating LRU in the next one. In both, the presentation will focus initially on DY0 through DY7 and subsequently on DY8 and DY9, where data is minimal and extrapolation from the results for preceding DY's becomes necessary.

Fitting and forecasting losses for DY1 to DY7. The procedure I use here is linear regression. As explained in the previous section, we will analyze each DY separately. The independent variable **X** used to fit each DY is the column of paid losses in DY0, shown in the left box in Table 3. We will illustrate the procedure by fitting DY2, the right box in Table 3, as the dependent variable. In model 1, these are the only two variables.

In the absence of specialized statistical software, one would typically perform the linear regression in Excel to obtain the regression coefficients and fitted values. Then one would obtain forecast values and, finally, calculate their forecast standard deviations. This final step can be especially complex. Here I introduce a method first suggested by Salkever (1976), later recommended by Kennedy, and described briefly but clearly by Greene (pp. 308-310), that makes it possible to do all three steps simultaneously.

Table 4 shows the key steps in the Salkever algorithm. First, augment the dependent variable **Y** with zeros so that it is the same length as DY0. Second, for each of the zeros added to **Y**, create additional columns in the independent variable **X**, in each of which there is a single entry, -1, corresponding to one of the zeros in Y. These additional variables are known as "dummy" variables, and so I have labeled them as D9 and D8, since their nonzero entries correspond to AY9 and AY8. In this example X now consists of three variables. Third, perform the linear regression (LINEST in Excel, with no intercept). The results are shown at the bottom of Table 4, with one slight difference from those obtained in Excel: I have reversed the left-to-right order of the first two rows of regression results, so that they now appear in the same order as the three variables in **X**, thereby doing what Microsoft should have done.

Table 4: Fitting and Forecasting Development Year 2

Y	X		
DY2	**DY0**	**D9**	**D8**
471	624	0	0
473	695	0	0
491	668	0	0
515	696	0	0
555	770	0	0
536	690	0	0
537	544	0	0
497	563	0	0
0	593	0	-1
0	621	-1	0
b	0.77	477	456
se_b	0.03	62	62
R^2, se_{est}	0.99	**59.0**	
t-statistic	**24.3**	**7.7**	**7.4**

<u>Results</u>. The first column of results is identical to what one would have obtained by simply regressing Y against X. The estimated regression coefficient **b** is 0.77, which indicates that the losses in DY2 are about 77% of those in DY0. The standard error of **b**, in the second row, tells us that **b** has an estimated standard deviation of 0.03. The t-statistic, in the fourth row, is the ratio of **b** to its standard error. As a general rule of thumb, t-statistics with absolute values greater than 2.0 are considered significantly different from zero. The two numbers in the third are R^2 and the standard error of the estimate, which is the estimated standard deviation of the error terms, the differences between fitted and actual values of **Y**. In the absence of an intercept R^2 is typically high, so the standard error of the estimate is a better measure of goodness of fit.

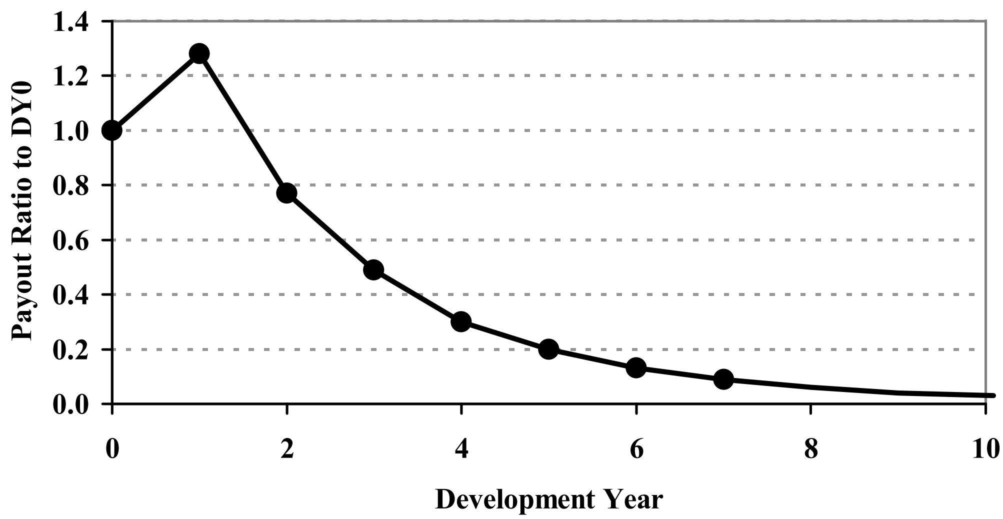
The real value of the Salkever procedure lies in the remaining two columns of regression results. The regression coefficients in the first row are <u>the forecast paid losses</u> for AY9 and AY8, respectively, and the values in the second row are the corresponding <u>forecast standard errors</u>. Note that the standard errors for the forecast values, which include parameter risk, are appropriately higher (62) than the standard error of the estimate (59), which does not.

The results from applying this procedure to DY1 through DY7 are summarized in Table 5. The top part shows the fitted values of past loss payments; the lower part shows the regression coefficients and other summary measures of goodness of fit. (In all cases R^2 was 0.99.) For DY1 through DY7, all estimated coefficients **b** were relatively precise, as indicated by their small standard errors. In each DY goodness of fit, as measured by the standard error of the estimate, is likewise small relative to the average paid loss. The overall pattern of these coefficients is shown in Figure 1.

Table 5: Fitted Values and Regression Coefficients for each Development Year

Accident Year	t	0	1	2	3	4	5	6	7	8	9
1994	0	624	800	480	303	187	124	85	54		
1995	1	695	890	534	337	208	138	95	61		
1996	2	668	855	513	324	200	133	91	58		
1997	3	696	891	535	338	209	139	95			
1998	4	770	986	592	374	231	153				
1999	5	690	884	530	335	207					
2000	6	544	697	418	264						
2001	7	563	722	433							
2002	8	593	760								
2003	9	621									
	b	1.00	1.28	0.77	0.49	0.30	0.20	0.14	0.09		
	se_b		0.05	0.03	0.02	0.01	0.01	0.01	0.00		
	se_{est}		91	59	40	15	12	9	4		

Figure 1: Actual and Estimated Model 1 Payout Pattern



These regression coefficients can be used in a fashion similar to chain ladder link ratios. The regression coefficient for any given DY is the estimated incremental dollars paid in that DY relative to the dollars paid in DY0. In DY1, for example, one can anticipate paying, on average, about 28% more than was paid out DY0. For a given AY, then, the remaining payments can be estimated by adding up the coefficients for the remaining DY's and then multiplying by the amount paid in DY0. In this example, the sum of the coefficients is 3.43 when one includes the tail. For AY9 the estimated remaining paid losses are 3.43 times the $621 million loss in DY0, or $2.130 billion, so that the estimated ultimate AY total is $2,751 billion.

Analysis of residuals. Table 6 shows the residuals -- the difference between actual and fitted values – for the data analyzed here. Two questions are central to the analysis of these residuals. First, do they exhibit patterns that may alert us to variables or unusual conditions not reflected in Model 1. Second, are the magnitudes of any particular residuals significant or noteworthy. Table 6 provides a basis for answering the first question, and Table 7, which shows standardized residuals (residuals divided by the DY standard error), facilitates answering the second. Table 7 shows, for example, that only one standardized residual has an absolute value greater than 2, which can be expected to occur about five percent of the time, or in about two instances of the 42 values shown in the table.

Table 6: Residuals from Fitted Values for Model 1

Accident		Development Year									
Year	t	0	1	2	3	4	5	6	7	8	9
1994	0		171	-9	-3	5	2	-5	-1		
1995	1		-82	-62	-18	-7	-4	1	-3		
1996	2		-47	-22	-29	-16	-18	-8	4		
1997	3		-47	-19	-36	-15	10	12			
1998	4		-86	-37	-16	8	9				
1999	5		-59	5	49	24					
2000	6		80	119	68						
2001	7		70	64							
2002	8		63								
2003	9										

The fact that an extensive discussion of the art of residual analysis is beyond the scope of this paper should by no means obscure its fundamental importance. The estimation of loss reserves and LRU should not be a mechanical application of a standard algorithm to standard data. As experienced actuaries and analysts know, the scientific model-building that lies at the core of actuarial science must necessarily be accompanied by skillful judgment in determining how those models are applied and interpreted for particular firms and lines of business.

Table 7: Standardized Residuals: Residuals divided by Standard Deviation

Accident Year	t	0	1	2	3	4	5	6	7	8	9
1994	**0**		1.88	-0.15	-0.08	0.35	0.15	-0.61	-0.27		
1995	**1**		-0.90	-1.04	-0.45	-0.46	-0.33	0.14	-0.83		
1996	**2**		-0.51	-0.37	-0.72	-1.05	-1.59	-0.95	1.11		
1997	**3**		-0.51	-0.33	-0.89	-0.96	0.85	1.31			
1998	**4**		-0.94	-0.62	-0.39	0.54	0.79				
1999	**5**		-0.64	0.09	1.22	1.53					
2000	**6**		0.88	2.02	1.69						
2001	**7**		0.77	1.08							
2002	**8**		0.69								
2003	**9**										

Forecasting the tails. Table 8 shows the forecast future paid losses obtained from applying model 1 to the data in Table 3 as well as the estimated payments for the tails, DY8 and beyond. The procedures used to obtain these tail estimates makes two important assumptions. The first is that the regression coefficients from DY4 and beyond decrease exponentially. Figure 1 already demonstrated that this assumption does not hold for earlier DYs. The second assumption is that the rate of exponential decay can be estimated from the coefficients already obtained for DY4 through DY7. I now describe the two steps needed to derive forecasts from these assumptions.

Table 8: Forecast Future Paid Losses

Accident Year	t	0	1	2	3	4	5	6	7	8	9	Tail
1994	**0**											48
1995	**1**										27	53
1996	**2**									35	24	47
1997	**3**								61	43	29	57
1998	**4**							105	67	47	31	62
1999	**5**						137	94	60	42	28	56
2000	**6**					163	108	74	47	32	21	42
2001	**7**				274	169	112	77	49	33	22	44
2002	**8**			456	288	178	118	81	52	35	23	46
2003	**9**		796	477	302	186	124	85	54	37	24	48
Development Year Total			796	933	863	696	600	517	390	305	230	504

In step one we extrapolate the regression coefficients already obtained to DYs beyond DY7. Because we have assumed that the coefficients decrease exponentially, it is appropriate to use logarithmic regression. We create a variable **W** that consists of the regression coefficients for DY4 through DY7, shown previously in Table 5. We also create a variable **V** consisting simply of the numbers 4, 5, 6, and 7. Then we obtain estimates **a** and **b** of the coefficients **α** and **β** in the logarithmic regression **lnW** = **α** + **βV** + **ε**. From this we obtain **b**, the estimated value of **β**, which is the logarithm of the rate of exponential decay. In this analysis I use d = exp(**b**) = 0.66 as the estimated rate at which the coefficients decrease from one year to the next in the tail.[5]

In step two we create robust forecasts of the paid losses in subsequent development years by using an average of three separate forecasts. For each accident year, the forecast paid loss for DY8 is forecast as $P_8 = (P_4d^4 + P_5d^3 + P_6d^2)/3$. The three terms in parentheses are three different forecasts of P_8 created from the actual or forecast paid losses in DY4, DY5, and DY6. The forecasts for P_9 are made in the same way, except that the exponents of d are each increased by one. Finally, the forecast value for the tail, consisting of paid losses for all development years after nine, is calculated as the forecast for P_{10} multiplied by 1/(1-d), the formula for the sum of the infinite exponentially decreasing series $(1 + d + d^2 + \ldots)$. The results of this procedure have already been shown in Table 8. The estimated reserve for these data, based on Model 1, is $5,835 million.

8: Estimating loss reserve uncertainty

As in estimating loss reserves, here we deal first with DY1 through DY7, and then tackle DY8 and beyond. We will first estimate the uncertainty of the total forecast payments for each DY. Then we will estimate total LRU by appropriately aggregating the uncertainties obtained for each DY.

Estimating the uncertainty of DY forecast totals. For DY1 there is only one future payment to be forecast, and we can obtain that forecast and its forecast standard deviation directly from the Salkever method. Here again it is noteworthy that the standard error (standard deviation of disturbances) for past observations, which include only process risk, is 91 (as shown in Table 5), whereas the standard deviation of the forecast future paid loss, which is subject to both process and parameter risk, is accordingly higher, at 96, as shown later in Tables 10 and 11.

For subsequent DY the problem is more complex, since forecast future payments within a DY share the same parameters and are therefore correlated since they share common parameter risk. The estimation procedure for DY2 and subsequent DY is shown in Table 9. The input data are shown at the top of the table. One is the column of paid losses in DY0. Recall that in Table 4 the first eight entries of DY0 were used to fit the eight paid losses already observed in DY2, and the remaining to entries were used to forecast future paid losses. Here we need to split DY0 into two separate parts, which we label **X** and $\mathbf{X_0}$ to correspond to the notation used by Greene (2000,

[5] Technically, unless the logarithmic regression perfectly fits the data, one should include a slight adjustment for the error term in order to obtain the mean estimated value of **b**. By deliberately failing to include this adjustment I instead obtain the median value of **b**, which is presumably more robust. In most cases the difference is miniscule and difficult to explain to a non-technical audience.

p. 309). Another input, **s**, is the standard error of the estimate for DY2, already reported in Tables 4 and 5. Finally, we need an identity matrix **I**, a square matrix of size n, where n is the number of the DY, with one's on its main diagonal and zero's elsewhere. From these inputs we obtain VCV, the variance-covariance matrix for forecast errors, as VCV = $s^2[\mathbf{I} + \mathbf{X_0}(\mathbf{X'X})^{-1}\mathbf{X_0'}]$. (The Excel formula for this calculation is shown at the bottom of Table 9.) We then sum these entries and take the square root of that result to obtain the standard deviation of the DY2 sum of forecast paid losses, which in this case is 92.

Table 9: Calculating the standard deviation of forecast paid losses for DY2

		DY0				
Input data:		624				
		695				
		668	**s=**	59		**= the standard error of**
		696				**the estimate, se_{est},**
	X=	770				**shown in Tables 4 and 5.**
		690				
		544				
		563				
	X_0=	593	**I=**	1	0	**The identity matrix**
		621		0	1	**(For DYn it is n x n)**

The variance-covariance matrix $s^2[\mathbf{I} + \mathbf{X_0}(\mathbf{X'X})^{-1}\mathbf{X_0'}]$ =

VCV=	3,838	369
	369	3,872

The variance of the sum of forecast paid losses for DY2 = the sum of the entries in the variance-covariance matrix

Σ VCV= 8,447

The standard deviation of the sum of forecast paid losses for DY2 = the square root of Σ VCV

= 92

Note: the following is the Excel array formula for VCV, where range names are shown in boldface type:
=**S***(**I.2**+MMULT(MMULT(**XZERO**,MINVERSE(MMULT(TRANSPOSE(**X**),**X**))),TRANSPOSE(**XZERO**)))

The standard deviations of the sum of forecast paid losses for DY3 to DY7 are calculated in the same way. Note that as we move from one DY to the next we must increase the number of entries in $\mathbf{X_0}$ by one, correspondingly decrease the number in **X** by one, and increase the dimension of **I** by one. The results are reported in Table 11, to which we shall return after we first obtain standard deviations for paid losses in DY8 and beyond.

Estimating the standard deviations of forecast tail paid losses. Salkever's method, applied to DY1 through DY7, provided forecasts of future paid losses (shown in Table 8) as well as standard errors (standard deviations) for these forecast values, are shown in Table 10. The table also shows the estimated standard errors of forecast losses for DY8 and beyond, which we calculated as follows.

Table 10: Standard Errors of Forecast Future Paid Losses

Accident		**Development Year**										
Year	**t**	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**Tail**
1994	**0**											5
1995	**1**										2	5
1996	**2**									3	2	5
1997	**3**								5	3	2	5
1998	**4**							10	5	4	2	5
1999	**5**						13	10	5	4	3	5
2000	**6**					16	12	10	4	4	2	5
2001	**7**				42	16	12	10	4	4	2	5
2002	**8**			62	42	16	12	10	5	4	2	5
2003	**9**		96	62	43	16	12	10	5	4	3	5

As with regression coefficients, the assumption is that the standard errors decrease exponentially in the tail. As before, we use logarithmic regression to estimate the rate of decrease. Here, however, the dependent variable **U** consists of the average standard error for each DY from DY1 to DY7, and the independent variable **T** consists of the numbers from 1 to 7. For the regression $\ln\mathbf{U} = \boldsymbol{\alpha} + \boldsymbol{\beta}\mathbf{T} + \boldsymbol{\varepsilon}$, we obtain an estimate **b** such that the rate of decrease $g = \exp(\mathbf{b}) = 0.61$. In a manner identical to the one used for paid losses, we forecast the standard deviations for DY8 in each AY as $E_8 = (E_4g^4 + E_5g^3 + E_6g^2)/3$, an average of three forecasts. Here each E within parentheses is the standard error of the forecast (for cells with forecast values) or the standard error of the estimate (for cells with observed values).

The last step requires that we obtain the standard errors of the sum of forecast paid losses for development years eight, nine, and the tail. To do this requires that we estimate what the variance-covariance matrices for those years might look like. We can in fact do this by examining the matrices already calculated for earlier development years.

The function of the variance-covariance matrix is to reflect interrelationships among the forecast errors. These interrelationships exist because the various forecast values all depend upon a common underlying parameter, β, whose estimate, b, may incorporate error. Any error in b will simultaneously affect all the forecast values. Moreover, as the number of observations on which estimates of β is based decreases, the interrelationships among forecast errors increase.

Were it not for these interrelationships among forecast errors, we could very easily calculate the standard deviation of total forecast paid losses by assuming that these forecasts and their errors were independent. In this case, the standard deviation of total forecast paid losses for development year n, which has n forecast values, would be $\sigma^* = (n\sigma_i^2)^{1/2} = \sigma_i n^{1/2}$, where σ^* is the standard deviation of total forecast paid losses assuming independence, and σ_i is the standard error of individual forecasts, here assumed to be equal (which is approximately true). In fact, however, we need to take into account the fact that the off-diagonal elements in the variance-covariance matrix are non-zero. Here we assume that these elements are identical in value (again, approximately true) and equal to $k\sigma_i^2$, where k is some constant to be estimated. In this case, the correct standard deviation of total forecast paid losses, σ, is $\sigma_i(n+kn(n-1))^{1/2}$. If we now calculate the ratio of σ to σ* we obtain the quantity $(1+k(n-1))^{1/2}$, which is a multiplier: it is the amount by which σ*, which assumes independence, must be multiplied to obtain σ, which does not. This approximation, when applied development years one through seven, produces results that are nearly an exact match to those obtained by having the actual variance-covariance matrix.

The key to applying this method is having a value for k, without which the multiplier cannot be calculated. The procedure used here was, first, to obtain the value of k from the variance-covariance matrices calculated for development years one through seven, second, to use linear regression to fit these values to an independent variable consisting of the number one through seven, and third, to forecast values of k for development years eight and nine and for the tail, for which the independent variable was nine plus the tail's weighted average length in years.[6]

<u>Results</u>. Table 11 shows the combined results of applying these procedures. Line A shows the sum of the forecast paid losses for each development year, as previously reported in Table 8. Line B shows the standard deviations of the values in line A. These differ from the values shown in Table 10, which are the standard deviations of the individual components of the sums in line A. Both, however, reflect parameter risk as well as process risk. The Total in line B is obtained by taking the square root of the summed squares of the values in that row. This assumes independence, which is appropriate since by using incremental paid losses we have eliminated correlations across development years.

Line C shows the coefficients of variation, the standard deviations divided by the forecast paid losses. For the total estimated reserve of \$5.8 billion, the standard deviation of \$175 million is

---

[6] Recall that d is the estimated ratio, in the tail, of the paid loss in one development year to the paid loss in the prior development year, so that $d < 1$. The average length of the tail, L, is calculated as a ratio in which the numerator is the infinite series $1+2d+3d^2+4d^3+ \ldots$, and the denominator is the infinite series $1+d+d^2+d^3+ \ldots$. The numbers 1, 2, and so on are the number of years subsequent to development year 9 in which payments occur, and each year is weighted by the percentage of total tail payments occurring in that year. The denominator is total tail payments. The value of the numerator is $1/(1-d)^2$, and the value of the denominator is $1/(1-d)$, so that the value of their ratio, L, is $1/(1-d)$. Consequently, for purposes of estimating k to calculate the multiplier for the tail, the number of the tail development year is $9 + 1/(1-d)$, which in this case is 12.

approximately 3.0% of the reserve. The fact that a consistent methodology was here used to estimate both the reserve and its standard deviation underscores a point made earlier: even if other methods or information are used to obtain a different estimated reserve, this estimate of loss reserve uncertainty, the coefficient of variation, should nonetheless remain valid**.**

Table 11 also shows, in line D, the sum of the forecast paid losses for calendar year 2004, which consists of the sum of the forecast losses immediately below the diagonal solid line in Table 8. The standard deviation of this value, shown in line E, is $124 million, or about 6% of the estimated calendar year total forecast payments of $2.07 billion. This calendar year measure of LRU can be especially important for helping managers to determine whether actual calendar year paid losses (for AY1 to AY9) deviate significantly from their forecast total.

Table 11: Standard Errors of Forecast Paid Losses by Development Year, Total Reserve, and Calendar Year

Development Year	1	2	3	4	5	6	7	8	9	tail	Total
A. Sum of Forecast Paid Losses	796	933	863	696	600	517	390	305	230	504	5,835
B. Standard Deviation of Forecast	96	92	81	37	34	33	17	18	15	45	175
C. Coefficient of Variation (=B/A)	12%	10%	9%	5%	6%	6%	4%	6%	6%	9%	3.0%
D. Calendar Year 2004 Forecast Paid Losses											2,070
E. Standard Deviation of CY Forecast Sum											124
F. Coefficient of Variation (=E/D)											6.0%

9: Validating the results

Here I validate the results just obtained by demonstrating that the same methods accurately estimate the future paid losses and LRU's of 10,000 simulated paid loss triangles with known parameters and outcomes.

To create simulated paid loss triangles I begin with an underlying deterministic payout pattern in which paid losses decrease exponentially from an initial value in DY0 that is identical for all accident years. (In this particular simulation, paid losses in each DY are half those in the preceding one.) I then add to each of these expected payments a random deviation drawn from a normal distribution, with a mean of zero and a standard deviation that increases linearly from 10% of the expected paid loss in DY0 to 100% in DY9 and 110% in DY10 and beyond. The simulations in fact generate the entire path of paid losses to the point where they become miniscule. Consequently, the ultimate paid losses can, in principle and in fact, depart considerably from the expected values established by the underlying pattern, and the standard deviations of these simulated variations can be calculated.

The results of the simulation are shown in Table 12. The first half of the table reports the accuracy of the method used here in forecasting DY sums of future loss payments. Line A shows the DY sums of expected future loss payments before random disturbances are added. Line B shows the average, over the 10,000 scenarios, of the simulated DY sums of future loss payments. Section C reports the results of using the procedure used in this paper to estimate DY sums of forecast future loss payments. When the independent variable is stochastic, and consists of the simulated loss payments in development year zero, the results are only trivially different from those obtained by using as the independent variable the expected (i.e., deterministic) loss payments in DY0 as if they were in fact known. This confirms the assertion in section 6 that the use of a stochastic independent variable is not a problem if its disturbances are independent of those that affect the dependent variable**.** The admittedly ad hoc procedure used here to calculate the tail values overestimates them somewhat. This is undoubtedly due to the fact that using the exponential decay function to project tail payments rules out negative payments, while the simulation does not. Nonetheless, forecasts of paid losses in the next calendar year, shown in the last column in Table 12, are remarkably accurate.

Table 12: Monte Carlo results for estimating reserves and reserve uncertainty

DY:	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10+**	**Total**	**CY**
DY Sum of Future Loss Payments												
A. True (deterministic) values	400	400	300	200	125	75	44	25	14	16	1,598	800
B. Average simulated values	399	401	301	200	124	75	44	25	14	16	1,599	799
C. Average forecast values												
-- using stochastic X	397	397	298	198	124	75	43	34	22	33	1,622	796
-- using fixed X	400	400	300	200	125	75	44	34	22	33	1,633	802
SD of DY Sum of Future Loss Payments												
D. True SD from parameters	80	85	69	50	34	21	13	8	5	5	151	112
E. SD of simulated payments	80	85	71	50	34	21	13	8	5	5	152	112
F. Parameter risk multipliers		1.1	1.1	1.2	1.3	1.4	1.6					
G. True SD plus parameter risk		**89**	**77**	**60**	**43**	**30**	**21**					
H. Estimated SD												
-- using stochastic X	91	**97**	**82**	**62**	**45**	**32**	**19**	18	13	31	189	127
-- using fixed X	82	**92**	**79**	**61**	**44**	**31**	**19**	17	13	31	180	118

The second part of Table 12 verifies the accuracy of the procedure for estimating the standard deviations of forecast future loss payments. Line D shows the actual standard deviations used in the simulation, and line E shows the standard deviation of the simulated losses. As one would hope from a properly conducted simulation, the two are virtually identical. Line F shows the multipliers for parameter risk obtained from the modeled variance-covariance matrices, and line G shows the true standard deviations in line D multiplies by the corresponding values in line F. These values in line G are the values one would hope to obtain in estimating LRU. The actual estimates obtain, both with a stochastic X and a fixed X, are shown in section H. The two sets of estimates in this section agree closely with each other and with the target values in line G. However, it appears that using a fixed X, as recommended by Halliwell (1996) improves the estimates for DY1 and DY2. For the total reserve, both stochastic and fixed X's produce a similar result, and substituting one for the other would have an imperceptible effect on the coefficient of variation (CV). For stochastic X the CV is 11.65%, and for fixed X it is 11.02%.

10: Summary and conclusions

The method I have presented here for estimating loss reserve uncertainty – the coefficient of variation of estimated future loss payments -- that has a number of merits. First, it can be used to address significant issues in surplus management, in pricing and capital allocation, and in the management of uncertainty. Second, it uses a measure of loss reserve uncertainty that facilitates comparison across different lines of business and can be applied to reserve estimates obtained through alternative methods. Third, it uses a publicly-available source of data that facilitates comparison across different firms. Fourth, the method avoids a number of serious pitfalls that can distort estimates of reserves or LRU. Fifth, the method is simple, at least as compared to some of the alternative methods advocated in the relevant literature. In particular, its use of Salkever's method provides an extremely useful shortcut for obtaining results. And sixth, the method accurately captures the key parameters of simulated paid loss trajectories. The reserve estimates are extremely accurate, and the estimates of reserve uncertainty, which include parameter risk, agree closely with benchmark calculations.

I hope that I have convinced readers that the method presented here for estimating loss reserve uncertainty that is both accurate and reasonably simple to implement. I also hope that my presentation of it is accessible to a large number of professional colleagues, who are invited to apply it in their own work and to extend it to novel uses.

11: References

ASHE, F. (1986) An Essay at Measuring the Variance of Estimates of Outstanding Claim Payments. *ASTIN Bulletin* **16**, S99-S113.

BARNETT, G. and ZEHNWIRTH, B. (2000) Best Estimates for Reserves. *Proceedings of the Casualty Actuarial Society* **87**, 245-321.

BRAUN, C.(2004) The Prediction Error of the Chain Ladder Method Applied to Correlated Run-off Triangles. *ASTIN Colloquium.*

BREHM, P. (2002) Correlation and the Aggregation of Unpaid Loss Distributions. *CAS Loss Reserve Seminar.*

ENGLAND, P. D. and VERRALL, R.J. (1999) Analytic and bootstrap estimates of prediction errors in claims. *Insurance: Mathematics and Economics* **25**, 281-293.

ENGLAND, P. D. and VERRALL, R.J. (2001) A Flexible Framework for Stochastic Claims Reserving. *Proceedings of the Casualty Actuarial Society* **88**, 1-38.

ENGLAND, P. D. and VERRALL, R.J. (2002) Stochastic Claims Reserving in General Insurance. *British Actuarial Journal* **8**, 443-544.

GREENE, W. H. (2000) *Econometric Analysis*, Prentice Hall, Upper Saddle River, New Jersey. Fourth Edition.

HALLIWELL, L. J. (1996) Loss Prediction by Generalized Least Squares. *Proceedings of the Casualty Actuarial Society* **88**, 436-489.

HAYNE, R. M. (2003) Measurement of Reserve Variability. *Casualty Actuarial Society Forum*, 141-171.

HODES, D. M., FELDBLUM, S., and BLUMSOHN, G. (1996) *Casualty Actuarial Society Forum*, 61-150.

HOLMBERG, R. D. (1994) Correlation and the Measurement of Loss Reserve Variability. *Casualty Actuarial Society Forum* **1**, 247-277.

KENNEDY, P. (2003) *A Guide to Econometrics*, MIT Press, Cambridge, Massachusetts. 5th edition.

KLOEK,T. (1998) Loss development forecasting models: an econometrician's view. *Insurance: Mathematics and Economics* **23**, 251-261.

KMENTA, J. (1971) *Elements of Econometrics*. Macmillan, New York, New York.

MACK, T. (1993) Distribution-Free Calculation of the Standard Error of Chain Ladder Reserve Estimates. *ASTIN Bulletin* **23**, 213-225.

MACK, T. (1994) Measuring the Variability of Chain Ladder Reserve Estimates. *Casualty Actuarial Society Forum*, 101-182.

MACK, T. (1995) Which Stochastic Model is Underlying the Chain Ladder Method? *Casualty Actuarial Society Forum*, 229-240.

MACK, T.. (1999) The Standard Error of Chain Ladder Reserve Estimates: Recursive Calculation and Inclusion of a Tail Factor. *ASTIN Bull*etin **29**, 361-366.

MADDALA, G. S. (1977) *Econometrics*. McGraw-Hill, New York, New York.

MURPHY, D. M. (1994) Unbiased Loss Development Factors. *Proceedings of the Casualty Actuarial Society* **81**, 154-222.

NARAYAN, P. and WARTHEN, T. (1997) *Casualty Actuarial Society Forum*, **1**, 175-196.

SALKEVER, D. (1976) The Use of Dummy Variables to Compute Predictions, Prediction Errors, and Confidence Intervals. *Journal of Econometrics* 4: 393-7.

STANARD, J. N. (1985) A Simulation Test of Prediction Errors of Loss Reserve Estimation Techniques. *Proceedings of the Casualty Actuarial Society* **72**, 124-148.

TAYLOR, G. (2003) Chain Ladder Bias. *ASTIN Bulletin* **33**, 313-330.

TAYLOR, G. (2004) Loss Reserving with GLM's: A case study. Centre for Actuarial Studies. University of Melbourne.

TAYLOR, G. (1987) Regression Models in Claims Analysis I: Theory. *Proceedings of the Casualty Actuarial Society* **87**, 354-383.

TAYLOR, G. and ASHE, F. R. (1983) Second Moments of Estimates of Outstanding Claims. *Journal of Econometrics* **23**, 37-61.

VERRALL, R. J. (1994) Statistical Methods for the Chain Ladder Technique. *Casualty Actuarial Society Forum*, 393-446.