Unbiased Development for Individual Claims—Taming the Wild Burning Cost

Joseph Boor, FCAS, PhD, CERA

Abstract

The ultimate cost of an unpaid individual claim follows a probability distribution, and usually will not be the exact point resulting from use of a loss development factor. So, when actuaries apply loss development factors to individual claims, they often create biased estimates of excess loss costs. Methods for creating a loss development probability distribution are developed. It is shown that the results of those methods are unbiased not only in terms of developing losses overall, but also projecting the ultimate costs in any layer. Methods to adjust the probability distribution for other development maturities and different claim handling are provided.

Keywords: Excess loss development, stochastic loss development factor, large deductible reserving, excess loss reserving, unbiased development

1   Introduction

This paper presents a theory and methods for estimating ultimate costs in excess layers from immature claims data. It is not unusual for actuaries to be encouraged to develop individual claims. For example, a paper by Lowe, Jing and Lebens 2009 suggests that in some cases developing the remaining open claims may be among the more accurate methods for estimating the reserve needs of very mature years. However, as Holler and Philbrick 1996 noted, applying the loss development factors derived from the entire body of claims to the large claims to get a "burning cost" estimate of the excess losses underestimates the excess losses. Looking at things symmetrically, applying unlimited[1] loss development factors to limited claims, and then eliminating the "developed" portion above the limit will overestimate the limited losses.

So, an alternate approach to loss development is necessary. Holler and Philbrick 1996 suggest strategies such as beginning with the distribution of claims as reported at an early maturity, then adjusting the mean and variance to reflect changes in the severity distribution that are expected as losses migrate to their ultimate values. This indirectly addresses the overriding issue with developing individual claims—-that while aggregate claims may develop in a relatively predictable fashion, the ultimate cost of an individual claim is best characterized by a probability distribution[2]. It is not difficult to see that the probability distribution model explains the problems noted in Holler and Philbrick. One might expect that the claims that develop into large claims might have larger than average development. Hence, applying a standard development factor to those claims underestimates their (excess) loss. However, it would of course be absurd to apply larger

---

[1]This statement assumes the claims handling underlying the loss development factors is a good match for that employed on the book of business being analyzed.

[2]This is sometimes referred to as a "stochastic development factor".

than average development factors to all the claims. So, done properly, the implementation of a probability distribution approach to loss development has potential to create unbiased estimates of excess loss costs.

One may approach this in more than one way. The particular focus of this paper is developing a single probability distribution that applies to all claims of a given maturity, and generates unbiased estimates of the claims in any layer, including the various excess layers. There may be alternate approaches that use different probability distributions for different claims, but this method is presented because it is relatively tractable, yet it eliminates the bias in the burning cost approach.

This article contains general methodologies for computing probability distributions for loss development, together with a toolkit of practical methods to implement the methodologies. Except where noted otherwise[3], the toolkit methods employ only functionality available in the most common spreadsheet software, so the tools may be implemented by any actuary with that software who obtains the needed data. So, the tools enable broad use of probability distributions for loss development. That has potential to significantly improve actuarial reserving for unpaid losses above a large deductible, excess-of-loss reinsurance, and so forth. The discussion involves several changes of variables and other transformations, but the reader will be rewarded with methods that produce unbiased estimates of ultimate excess losses for reserving and pricing.

The basic concept of probability distributions for development of individual claims is not entirely new. That process of using a probability distribution to gain a more accurate estimate of the ultimate severity was previously articulated by Gillam and Couret 1997, and discussed in greater detail by Mahler 1998. This approach builds on that concept and presents a different approach to the distribution.

2   The Conceptual Approach

The key goal of this approach is to create a loss development factor distribution that, combined with the immature losses, creates unbiased estimates of the excess loss costs. The core principle begins with the severity distribution underlying the "raw" undeveloped losses[4], "$s_X(x)$", and the severity distribution expected at ultimate, "$s_Y(y)$". Then one must create some severity distribution for the development "$s_R(r)$" generating a variable "R" such that independent samples from the raw distribution, "X" and the development distribution "R", generate a product $X \times R$ that is distributed according to $s_Y(y)$. Thus, R represents the desired distribution of possible loss development factors. Even if X takes a single value, $X \times R$ represents possible values that X may "develop"[5] to. In other words, given a very large body of claims, combining the products of the undeveloped data and the development distribution would reproduce the ultimate severity distribution (times the number of claims).

A few comments are in order before proceeding further. First, note that even in the presence of limited data, say, only "c" claims, $\sum_{h=1}^{c} X_h \times R_h$, or for an interval $\sum_{h=1}^{c} min\{max\{X_h \times R_h -$

---

[3]All the methods may be implemented on a common spreadsheet platform, but the author found implementing the Fourier transform method using the approach in the spreadsheet program to be challenging.

[4]This paper will follow the standard convention from statistics that capital letters to the random variable as a whole, whereas lowercase letters refer to specific instances or outcomes of the random variables.

[5]The term development is a bit of a misnomer as the development distribution $f(r)$ could be determined so that it is appropriate across the entire body of claims, including closed claims. Further, it is not designed to, say, create a perfect predictor of the distribution of possible ultimate costs of a claim presently reserved at exactly $10,000.

$B, 0\}, T - B\}$, generates unbiased[6] estimates of the loss costs that will develop from those c claims in any interval such as $[B, T]$. By construction, extending a set of claims that fully represent $s_X(x)$ through the distribution of R, produces $s_Y(y)$. Therefore, it will represent the ultimate claims cost in any interval of possible values of Y. So, one would say

$$E_{distribution}[X \times R] = Y, \tag{1}$$

where the expectation is across all possible values of X and it equals the full probability distribution for Y. In other words, applying the spread induced by R to and across all sets of raw claims creates an unbiased estimate of the entire distribution, values in various intervals, etc. for Y. So, this creates a powerful tool for evaluating development into various bands of large losses.

Of note, then Mahler's expression

$$\sum_{\text{all claims } ``X_h''} X_h \int_{L/X_h}^{\infty} s_R(r) \left(r - \frac{L}{X_h} \right) dr, \tag{2}$$

for the developed loss excess (of some limit/retention "L") arising from a group of claims provides an unbiased estimate of the potential ultimate cost of those claims that is excess of L.

Basic Criteria for X and Y

It stands to reason that, unless R is a constant, it will add what may be vaguely described as "additional spread" to the distribution X. So one would strongly expect that if the variance of X is larger than that of Y, then no independent R exists so that $X \times R$ has the same severity distribution as Y. Similarly if X has a similarly vaguely defined "fatter tail" at the largest values than Y, then no independent R exists to link them.

3 The Underlying Approach to Estimating R

Considering the discussion above, it appears that, given some severity distribution for the undeveloped losses that generates X, and some other severity distribution for the ultimate losses Y, one seeks a third distribution R such that

$$X \times R \sim Y, \tag{3}$$

(where the "$\sim$" indicates that the two sides are essentially identical[7].) It is often extremely difficult to determine a distribution for such an R, but the process may be simplified somewhat by taking natural logarithms of both sides to get

$$\ln(X) + \ln(R) \sim \ln(Y). \tag{4}$$

That simplifies an extremely complex multiplication-type problem, producing a more tractable addition problem. However, it may be confusing to discuss a distribution labeled, for example, "$\ln(X)$". So, it should be helpful to use new variables to describe the logarithms, $U = \ln(X)$, $Z = \ln(R)$, and $W = \ln(Y)$. Then

---

[6]A potential issue with late reported claims does exist, but a correction will be noted later in this article.

[7]Technically, that the cumulative density functions of the two sides are identical

$$U + Z \sim W. \tag{5}$$

Note that converting the severity probability distribution s_X of x to the severity distribution s_U for $U = \ln(x)$ (and vice versa) requires a key principle from calculus. An example will illustrate the issue that must be dealt with, and the resolution. Say, for example, that U is normally distributed with mean μ and variance σ^2. Then, its density function would be

$$s_U(u) = \frac{1}{\sqrt{2\pi}\sigma} \exp -\frac{(u - \mu)^2}{2\sigma^2}. \tag{6}$$

If one were to attempt to evaluate s_X by simply substituting $\ln(x)$ for u in that formula, one would obtain

$$s_X(x) \ ? =? \ \frac{1}{\sqrt{2\pi}\sigma} \exp -\frac{(\ln(x) - \mu)^2}{2\sigma^2}. \tag{7}$$

Usually, that is not even a probability distribution (usually its total integral is not unity (1.00)). Obviously, in this case s_X should be a lognormal distribution with parameters μ and σ. So, the formula should be

$$s_X(x) = \frac{1}{x\sqrt{2\pi}\sigma} \exp -\frac{(\ln(x) - \mu)^2}{2\sigma^2}. \tag{8}$$

The additional "$1/x$" factor illustrates a key principle involved in substituting a function of a variable (in this case $u = \ln(x)$) for the main variable in a severity distribution. To keep a distribution function that integrates to unity (1.00), one must follow the rules for substituting variables in integrals. In effect, that requires multiplying the expression in (7) by the $\frac{d\ln(x)}{dx} = 1/x$ needed in equation (8). This illustrates the fact that, since $\ln(x)$ grows more slowly, and generally has much lower values that $f(x) = x$ above unity, its values tend to be more compressed relative to the x's that gave rise to the $\ln(x)$'s. Therefore, and especially for large x, without the correction the weights given to the $s_U(\ln(x))$'s would be diluted.

Using that principle, one may note that

$$\begin{aligned}
\frac{s_U(\ln(x))}{x} &= s_X(x) = s_X(\exp(u)); s_X(\exp(u))\exp(u) = s_U(u); s_X(\exp(u)) = s_U(u)\exp(-u); \\
\frac{s_Z(\ln(r))}{r} &= s_R(r) = s_R(\exp(z)); s_R(\exp(z))\exp(z) = s_Z(z); s_R(\exp(z)) = s_Z(z)\exp(-z); \text{ and,} \\
\frac{s_W(\ln(y))}{y} &= s_Y(y) = s_Y(\exp(w)); s_Y(\exp(w))\exp(w) = s_W(w); s_Y(\exp(w)) = s_W(w)\exp(-w).
\end{aligned} \tag{9}$$

So, using the exponential and logarithm functions and their derivatives, one may translate from the original multiplicative equation for R to an additive problem, and vice versa.

3.1   A Basic Approach Using Matrices

Most modern spreadsheet packages contain the ability to perform matrix arithmetic. And the process of adding the values of U to those of Z to produce the values of W may be mirrored

numerically in a set of equations that create a matrix. The first step is to set up the discrete analogues to U, Z, and W. A set of evenly spaced points is to be used to represent each probability distribution. They will be organized so that the difference between each two adjacent points (of each distribution) is equal to the "grid size" g. So, $[\mathcal{U}]$ is defined [8] on the points $0, g, 2g, ..., lg$ indexed by $i = 0, 1, 2, ..., l$. The brackets denote a vector or matrix. The probability density function at each point, following equation (9), is

$$s_U(ig) = s_X(\exp(ig))\exp(ig). \tag{10}$$

Further, ig represents an interval $[(i - .5)g, (i + .5)g]$ of size g. So, one may estimate the probability that u falls in the interval, taking $s_U(ig)$ as representative of the probability density function in the interval. Then

$$P_U[(i - .5)g, (i + .5)g] \approx gs_U(ig) = gs_X(\exp(ig))\exp(ig) = [\mathcal{U}]_i. \tag{11}$$

$[\mathcal{U}]$ and the related vectors are used to clarify the difference between the actual distributions and their discrete analogues. $[\mathcal{U}]$ contains the $gs_U(ig) = gs_X(\exp(ig))\exp(ig)$'s for $i = 0, 1, 2, ..., l$ from equation (11); $[\mathcal{Z}]$ represents the $gs_Z(jg) = gs_R(\exp(jg))\exp(jg)$'s for $j = 0, 1, 2, ..., m$; and $[\mathcal{W}]$ contains the $gs_W(kg) = gs_Y(\exp(kg))\exp(kg)$'s for $k = 0, 1, 2, ..., n$. Note that the number of points in each vector (l, m, and n), may all be different depending on the spread of probabilities of Z desired and how many of the sums (discussed later) are to be evaluated.

Now that the setup of the vectors is complete, one may set up the matrix equation for $[\mathcal{Z}]$. This uses the probabilities of the various combinations of i and j that add to each k. In this example, the lowest value of both i and j is zero. So, the only way to sum an index of $[\mathcal{U}]_i$ and an index of $[\mathcal{Z}]_j$ to index $[\mathcal{W}]_0 = 0 \times g$ is if $i = j = 0$. So, since X and R are independent, U and Z are too. Therefore, the discrete probability $[\mathcal{W}]_0$ is approximately $[\mathcal{U}]_0 \times [\mathcal{Z}]_0$ (the probability of a zero occurring in both vectors). Similarly, with these vectors the only way to obtain $k = 1$ for $[\mathcal{W}]_1$ is if one of i and j is zero, and the other equals one. So,

$$[\mathcal{W}]_1 \approx [\mathcal{U}]_0 \times [\mathcal{Z}]_1 + [\mathcal{U}]_1 \times [\mathcal{Z}]_0, \tag{12}$$

and for the general case of k

$$[\mathcal{W}]_k \approx \sum_{i=0}^{k} [\mathcal{U}]_i \times [\mathcal{Z}]_{k-i}. \tag{13}$$

Of course, the above could possibly have different limits of summation depending on the limits imposed by l, m, and n. For example, if $l < k = n$, then the limit of summation could only go to l.

Clearly, equation (13) represents a series of linear equations in the $[\mathcal{Z}]_j$'s that target the $[\mathcal{W}]_k$'s. The various $[\mathcal{U}]_i$'s are the coefficients. So, one may represent the approximate relationships with the matrix equation

$$[\mathcal{W}] = [\mathcal{U}^*] \times [\mathcal{Z}], \tag{14}$$

---

[8]This example starts the index at zero to simplify the illustration. One could just as easily start with -1 or even -10 to reflect possible downward development by factors of $\exp(-1)$ or $\exp(-10)$.

Further $[\mathcal{U}^*]$ is defined[9], using the $[\mathcal{U}]_i$'s, as

$$[\mathcal{U}^*] = \begin{bmatrix} \mathcal{U}_0 & 0 & 0 & ... \\ \mathcal{U}_1 & \mathcal{U}_0 & 0 & ... \\ \mathcal{U}_2 & \mathcal{U}_1 & \mathcal{U}_0 & ... \\ ... & ... & ... & ... \end{bmatrix} \tag{15}$$

(omitting the brackets on the individual values inside the matrix for simplicity).

Of course, this is only a "discrete" representation of the probability distributions of $\ln(x)$, $\ln(r)$, and $\ln(y)$. Hence, it will not generate the exact, or even a continuous, distribution for R, however such a discrete approximation should reflect some of the basic characteristics of R. So it should usually be preferable to simply multiplying each loss by a single loss development factor (as is done in burning cost).

To complete the estimation, one may use the linear algebra formula for the $[\mathcal{Z}]$ generating the $[\mathcal{U}^*] \times [\mathcal{Z}]$ that has the least squared error[10] in approximating $[\mathcal{W}]$

$$[\mathcal{Z}] = \left([\mathcal{U}^*]^T \times [\mathcal{U}^*] \right)^{-1} \times \left([\mathcal{U}^*]^T \times [\mathcal{W}] \right) \tag{16}$$

where the upper T denotes transpose and the upper -1 denotes matrix inverse. Importantly, matrix multiplication, matrix transposition, and matrix inversion functions are all available in the most common spreadsheet software package (and likely other spreadsheet software as well). So, virtually any actuary, when given the distributions s_X and s_Y, can perform these calculations.

Of course, one seeks the distribution s_R, not the vector $[\mathcal{Z}]$. One way to use it is to convert $[\mathcal{Z}]$ to samples from the probability mass function s_Z (recall that $[\mathcal{Z}]$ contains approximate probabilities of intervals, not points), and then convert those to the probability function s_R at various values of R. As a first step, for each of the values in $[\mathcal{Z}]$, and indexed by j

$$s_Z(jg) = s_Z(Z = jg) = [\mathcal{Z}]_j/g. \tag{17}$$

Then, one may use equation (9) to get values of s_R at the points $\exp(0)$, $\exp(g)$, $\exp(2g)$, ...

$$s_R(\exp(jg)) = s_Z(jg)\exp(-jg) = [\mathcal{Z}]_j \exp(-jg)/g. \tag{18}$$

That provides a sample of s_R at the $\exp(jg)$ points. A curve of some kind may be fitted to them to complete the construction of s_R. Then, the Mahler excess function in equation (2) may be computed for each claim.

There is a second way that the $[\mathcal{Z}]_j$'s may be used. Each value $[\mathcal{Z}]_j$ represents the probability that Z falls in an interval between $(j - .5)g$ and $(j + .5)g$. So, they also represent the probability that R falls in an interval between $\exp\left((j - .5)g\right)$ and $\exp\left((j + .5)g\right)$. As long as the range of the corresponding values from R cover a wide enough range, one could use the set of discrete points $\exp(0)$, $\exp(g)$, $\exp(2g)$, ..., and assign each the probability of the interval they represent, $[\mathcal{Z}]_j$. The resulting discrete distribution could represent s_R. One would simply take each undeveloped claim and multiply it by $\exp(0)$, $\exp(g)$, $\exp(2g)$, ... with probabilities $[\mathcal{Z}]_0$, $[\mathcal{Z}]_1$, $[\mathcal{Z}]_2$, ... to

---

[9]Slightly different situations may arise when the lowest index of one or more of the vectors $[\mathcal{U}]$, $[\mathcal{Z}]$, or $[\mathcal{W}]$ is not zero, but the conversion should be straightforward.

[10]Generally, since $[\mathcal{Z}]$ has m potential values and $[\mathcal{U}]$ has l potential values, $[\mathcal{W}]$ potentially has $l + m$ values, so there are typically more equations to solve than entries in $[\mathcal{Z}]$. This is an overdetermined system of equations.

estimate the range of ultimate values the claim might represent. Then, the amounts excess over the deductible or attachment for each claim X, using all the $XR = X \exp(jg)$'s, with probabilities $[\mathcal{Z}]_j$ may be used to compute the excess loss

$$\sum_{\text{all claims } "X_i"} \sum_{j=0}^{m-1} \max\{X_i \exp(jg) - L, 0\}[\mathcal{Z}]_j \tag{19}$$

An example of the calculations in this section may be found in Appendix A.

A Brief Note on Discrete/Numerical Analysis Approximations

Two campers are in the woods and see a hungry bear...

The first camper says 'I hope I can run faster than the bear.',
The second one says 'I don't need to run faster than the bear, I just need to run faster than you.'
-Old joke

As one may see in Appendix A, the discrete approximations above (and below) do not generate exact or perfect solutions. That is of course a general characteristic of numerical methods. However, note that actuarial science is not itself a method for exactly computing exact losses, or the exact cost of the claims covered by a loss reserve. The key questions to ask are "Does it generate a better approximation than burning cost?" and "Do I nevertheless have a method to make my estimate more accurate?".

3.2   An Improved Matrix Approach—Covering the Upper Range (or Another Wider Range) of $\ln(Y) = W$

Experimentation with matrix calculations suggests that the accuracy may be improved by letting the number of entries (rows) in $[\mathcal{W}]$ be much larger[11] than the number of rows in $[\mathcal{U}]$. This makes sense, since equation (4) says that $\ln(X) + \ln(R) \sim \ln(Y)$. Therefore, $\ln(Y) = W$ should tend toward larger values than $\ln(X) = U$. In fact, if X and R were constrained to be less than $\max(X)$ and $\max(R)$, $\ln(Y)$ would range up to $\max(\ln(X)) + \max(\ln(R))$. Further, numerical analysts know well that when the data and curves may contain mild distortions, so-called "overdetermined" systems of equations, with more equations than variables to be solved for, often perform better than those using square matrices. Therefore, one might replace equation (13) with (for each k indexing $[\mathcal{W}]$) between $kmin$ and $kmax$, inclusive,

$$[\mathcal{W}]_k = \sum_{j=\max(jmin,k-jmax)}^{\min(k,jmax)} [\mathcal{U}]_{k-j}[\mathcal{Z}]_j. \tag{20}$$

So, one might have a matrix $[\mathcal{U}^*]$ that looks like

---

[11]Note, though, that using a large multiple of the number of elements in $[\mathcal{Z}]$ will increase the complexity, and thus the likely error, in the matrix arithmetic.

$$[\mathcal{U}^*] = \begin{bmatrix} \mathcal{U}_0 & 0 & 0 & ... \\ \mathcal{U}_1 & \mathcal{U}_0 & 0 & ... \\ \mathcal{U}_2 & \mathcal{U}_1 & \mathcal{U}_0 & ... \\ ... & ... & ... & ... \\ ... & ... & ... & ... \\ ... & 0 & \mathcal{U}_l & \mathcal{U}_{l-1} \\ ... & 0 & 0 & \mathcal{U}_l \end{bmatrix}, \tag{21}$$

but the earlier equation (16) for computing $[\mathcal{Z}]$ still applies. An example of the calculations may be found in Appendix B.

Another approach would be to focus on the upper tail of the distribution. One might seek only use versions of equation (21) that are above some value "$kmin$". Then, $[\mathcal{U}^*]$ might look like

$$[\mathcal{U}^*] = \begin{bmatrix} \mathcal{U}_{kmin} & \mathcal{U}_{kmin-1} & \mathcal{U}_{kmin-2} & ... \\ \mathcal{U}_{kmin+1} & \mathcal{U}_{kmin} & \mathcal{U}_{kmin-1} & ... \\ ... & ... & ... & ... \\ ... & ... & ... & ... \\ ... & 0 & \mathcal{U}_l & \mathcal{U}_{l-1} \\ ... & 0 & 0 & \mathcal{U}_l \end{bmatrix}. \tag{22}$$

Some variations on that format exist depending on items such as whether $kmin$ is larger than the largest index of $[\mathcal{U}]$, specifically l.

Experimentation with this approach suggests that it generally seems to be better conditioned than the situations where $[\mathcal{Z}]$ and $[\mathcal{W}]$ are about the same size. Further, when the scopes of $[\mathcal{U}]$ and $[\mathcal{W}]$ (specifically, the ranges of the ig's and kg's) are limited to areas where the probability distributions have substantial weight, the calculations seem to be better conditioned. Unfortunately, if the development pattern under Z features very large development, but only on a very small number of claims, then the truncated problem may not provide a proper estimate of Z. However, there are some methods following that may be of assistance.

3.3 Enhancing the Results of the Matrix Calculations

There are a couple of types of error that were often observed in the calculation of $[\mathcal{Z}]$ using equation (16). First, one should understand that, especially when very low values of $[\mathcal{U}]_i$ or $[\mathcal{W}]_k$ are used in a matrix approach, the calculations may generate significant numerical error. The first type of error that might be corrected tends to arise near the outlying values of jg where the actual values of $s_Z(jg)$ might be expected to be very low. The matrix calculations might have difficulty estimating these very low values in conjunction with the calculation of the larger values near the main body of the distribution s_Z. So, one might see output points here with negative probability, or very large probability at one point followed immediately by very low probability. One could likely enhance the quality of the final estimate of $[\mathcal{Z}]_j$ at these points by fitting curves to extend the information in the main body of the distribution, by limiting the fit to the upper end of the distribution as in Appendix C, or, if the exact values are less material, simply by using judgment.

Next, in practice sometimes the probabilities in $[\mathcal{Z}]$ do not quite sum to unity. This may be resolved by simply multiplying all the values $[\mathcal{Z}]_j$ by a common correction factor $\frac{1.00}{\sum[\mathcal{Z}]_j}$

A last pair of corrections involve finding inaccuracies in the estimated mean and variance. For background, given distributions for U and W, and the assumed independence of U and Z, then the means and variances of three distributions satisfy

$$E[U] + E[Z] = E[W], \text{ or } E[Z] = E[W] - E[U], \text{ and} \tag{23}$$
$$Var[U] + Var[Z] = Var[W], \text{ or } Var[Z] = Var[W] - Var[U].$$

So, since the distributions of U and W are known, the mean and variance of Z may be computed before the matrix calculations begin. However, sometimes the errors in the matrix calculations lead to a slightly different mean and variance for $[\mathcal{Z}]$. First, one may consider the case where the mean of the result of the matrix calculations is $\mu_Z + e$ and the correct value is, not surprisingly, μ_Z. The straightforward solution is to move the index associated with each probability down by e/g. Or one may replace the probability associated with each index j with the probability at $j + e/g$. So, the revised probability assigned in $[\mathcal{Z}]_j$ would be the current value in $[\mathcal{Z}]_{j+e/g}$. Recognizing that e/g may not be an integer, this may require interpolation. Straight linear interpolation may be used, or the improved interpolation along the curve in Boor 2014 may generate better results.

The remaining correctable issue arises when the variance of $[\mathcal{Z}]$ does not match the value projected in equation (23). Then, the values must all be moved uniformly closer to the mean, or uniformly further away from the mean, in order to fulfill the mathematical identity. One may simply employ the correction formula

$$Revised\left([\mathcal{Z}]_j\right) = [\mathcal{Z}]_{\left[\mu_Z + (jg - \mu_Z)\sqrt{(Var[W] - Var[U])/Var[\mathcal{Z}])}\right]/g}. \tag{24}$$

The formula is somewhat lengthy, but one may see that it simply replaces the distance from the mean with a value scaled to produce the proper variance. As with the correction of the mean, it will likely require interpolation. Further, if the result generates more or fewer points, the distribution may need to be rebalanced to sum to unity overall.

Using these techniques, one may improve the utility of the calculated $[\mathcal{Z}]$ and consequently obtain a better estimation of R. Examples of these methods may be found in Appendix C.

3.4   Fitting a Distribution for Z by Mean and Variance Matching

As an alternative, one could consider determining the optimum choice from some family of distributions. For example, if one had reason to believe that the underlying distribution of R was a lognormal distribution, then one could assume that $\ln(R)$ has a normal distribution. And all that would be needed is to determine the μ_Z and σ_Z. But, as stated previously, since $\ln(R) = Z$ is constructed to be independent of $\ln(X) = U$,

$$\begin{aligned} E[W] &= E[U] + E[Z] \\ Var[W] &= Var[U] + Var[Z] \end{aligned}$$

$$\tag{25}$$

Thus, since the distributions of X and Y are known inputs, so are those of U and W. As noted earlier, one could compute the normal parameters of Z, and hence the lognormal parameters of R.

$$\begin{aligned} \mu_Z &= E[W] - E[U] \\ \sigma_Z^2 &= Var[W] - Var[U]. \end{aligned}$$

(26)

Of course, one often assumes that the distribution of R comes from some other family (i.e., is not lognormal). However, if the first and second moments are to be matched to choose the specific distribution, then equations (26) may still be used to pick a specific s_Z from a family of distributions corresponding to possible distributions of $\ln(R)$. Then, an exponential change of variable (following equation (9)) will provide the distribution for R.

3.5 Fitting a Distribution for Z by Matrix-Based Parameter Estimation

An alternative method for selecting the parameters of a specific distribution for Z from a curve family involves a numerical method. In this case, one would select the best estimate of s_Z using the solution finder routines available in spreadsheet software packages. For example, one may desire that s_Z come from some class of distributions such as the Pareto, lognormal, etc. Then one could fit the parameters of the distribution by choosing the parameters that result in the best estimates of key values of $[\mathcal{W}]$ when the discrete probability distribution corresponding to those parameters is combined with the proper $[\mathcal{U}^*]$. Effectively one would feed the matrix $[\mathcal{Z}]$ of the Pareto distribution through the $[\mathcal{U}^*] \times [\mathcal{Z}]$ process and compute the corresponding error in this estimate of $[\mathcal{Z}]$. Then, one would vary the parameters (as standard spreadsheet software does) to minimize the error estimating $[\mathcal{W}]$ and find the best choice of parameters. Such a calculation is shown in Appendix D. Admittedly, this is essentially another matrix approach. However, since the type of distribution is specified, one should need fewer points, and less computational complexity, to simply estimate the parameters of the distribution.

3.6 Matching the Pareto Alpha Parameter

When developing individual claims for excess layers, it may often be more important to just know the upper end of the s_R distribution. Further, the upper end of the s_Y distribution may itself result from an extrapolation, often using a Pareto distribution. Therefore, it may be more important to find an R that creates the Pareto character of Y than to find an R that generates a good match across the entire domain of Y. This approach skips the intermediate step of taking logarithms U and W of X and Y and estimates characteristics of R directly from X and Y. To do so, one may begin with an assumption that Y's Pareto parameters are α and y_M, generating

$$s_Y(y) = \frac{\alpha \times y_M^\alpha}{y^{\alpha+1}},$$

(27)

and similarly that

$$s_X(x) = \frac{\beta \times x_M^\beta}{x^{\beta+1}},$$

(28)

Of course, it must be that $\alpha \le \beta$, since R adds "spread" to X. Then, a paper by Pederzoli and Rathie 1980 states that if X and Y are independent, with $\alpha < \beta$, then R, the quotient of Y over X, will follow the Pareto distribution

$$s_R(r) = \frac{\alpha\beta}{\alpha + \beta} \times \frac{(y_M/x_M)^\alpha}{r^{\alpha+1}}, \tag{29}$$

which is itself a Pareto distribution. However, Y is actually dependent on X. So the above represents more of a general guideline.

However, the distribution of the product of any Pareto distributions X and R generally has a Pareto character at the upper severities, matching the lowest (thickest tail) of β and the R's power parameter "γ". Specifically, if X and R are Pareto distributions, then Y will have Pareto-type character in the upper tail, corresponding to the lowest power parameter. To fully specify R, one would assume that it has the parameters γ and r_M, and $\gamma < \beta$. Then, if X has the Pareto parameters listed earlier,

$$s_Y(y) = \frac{\beta\gamma}{\beta - \gamma} \left\{ \frac{(x_M r_M)^\gamma}{y^{\gamma+1}} - \frac{(x_M r_M)^\beta}{y^{\beta+1}} \right\}. \tag{30}$$

One may see that as y gets large, the first term, which has Pareto power γ (from R) will dominate the final result.[12] This indicates suggests that the Pareto power of R will generally match the Pareto power of Y.

So, R may be estimated by simply matching the α parameter of the distribution of Y. Note that even if X has too thin a tail[13] to be a Pareto distribution, using a Pareto distribution with parameter α for R will still mean that the power parameter of $X \times R$ will be α. For example, consider the case where X is a constant.

Then, the primary goal is to find an R which, in conjunction with X will generate a distribution $X \times R$ with a Pareto shape parameter of α. So, it is necessary to estimate α from the values of Y. One may use the various upper percentiles of F_Y, the cumulative distribution function associated with the severity mass function s_Y. If one chooses p_1 and p_2 to be probabilities fairly close to unity, and their corresponding $100p_1\%$ and $100p_2\%$ percentiles are $F_Y^{-1}(p_1)$ and $F_Y^{-1}(p_2)$, then

$$\alpha = \frac{\ln\left(\frac{1-p_1}{1-p_2}\right)}{\ln\left(\frac{F_Y^{-1}(p_2)}{F_Y^{-1}(p_1)}\right)}. \tag{31}$$

Alternately, one may use any other standard method to determine the Pareto power parameter in the upper tail.

One may also correct the absolute value of the result. If the final distribution of $X \times R$ results in probabilities for large values that are $d\%$ of the probabilities Y generates for the same values, one need only alter the truncation point of R by $(1/d)^{1/\alpha}$ to create the desired absolute level of the distribution of Y at the upper amounts. So, using this approach, a distribution for R that provides the key Pareto-type characteristics at the upper limits of the distribution of Y may be computed.

3.7 Fourier Analysis—A Heavily Mathematical Approach

Fourier analysis represents an advanced mathematical approach, using complex variables, to (among other things) compute the distributions of sums and differences of random variables.[14] For example,

---

[12]The calculations underlying this formula, which is likely not new, involve using $u = \ln(x)$, etc. logarithmic substitutions and standard calculus.

[13]When $\beta = \alpha$, however, this process may not work.

[14]The reader that is not familiar with complex-valued functions, characteristic functions or Fourier transforms should not let this digression become a deterrent. The use of those is limited to this section.

when U and W are instances of well-known types of probability distributions, it might be possible to find the distribution Z using characteristic functions (Fourier transforms). Such a transform of some random variable, for example "T" would be

$$\varphi_T(\omega) = E[\exp(i\omega X)]. \tag{32}$$

It is well known that when $U + Z \sim W$ that

$$\varphi_U(\omega) \times \varphi_Z(\omega) = \varphi_W(\omega) \quad \text{(for all } \omega\text{).} \tag{33}$$

So

$$\varphi_Z(\omega) = \varphi_W(\omega) \div \varphi_U(\omega), \tag{34}$$

and using the inverse of the characteristic function, (so as to simplify the formula)

$$s_Z(z) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp(-i\omega z) \frac{\varphi_W(\omega)}{\varphi_U(\omega)} d\omega, \tag{35}$$

which may then be used to construct s_R using equation (9). It is not difficult to see, though, that even computing the Fourier transform is complicated (and uses complex numbers to boot). So, at least insofar as a paper and pencil approach is involved, this Fourier transform approach may be difficult to use on the less structured loss distributions associated with insurance claims. Possibly, there may be packaged software that can reliably estimate whole complex functions and, therefore, compute the needed transforms. The author is not aware of such software, but the author rarely uses complex valued functions at work. A discussion of the discrete estimates of Fourier transforms, and some discussion of when they are effective and ineffective may be found in Halliwell 2014. However, a significant concern exists with any such approach. Since complex analysis is not included on the Casualty Actuarial Society's syllabus of examinations, this approach may be too technical for many actuaries to use or understand.

3.8 Testing the Results

Noting that the different methods have different strengths and weaknesses, it may be prudent, after estimating R, to run a Monte Carlo sample of $X \times R$ and see whether or not the result mirrors Y. Usually, special attention should be placed on the larger values in the range of Y. Further, it may be preferable to use more than one method and test each method to determine which one performs best.

4 Finding Data and Using Data Effectively

Implementing the various procedures above requires having a severity distribution at the same maturity as the claims to be developed, and the ultimate severity distribution (or least suitable approximations to the two). Of course, if one has a benchmark ultimate loss distribution from some source, and the s_X severity distribution is from the actual claims to be developed, $X \times R$, up to approximations in the process, has to simply regurgitate the benchmark ultimate loss distribution one began with. So, it would often be necessary to use external data to derive s_R. It may be appropriate, though, to modify s_R to reflect differences between the characteristics of the data being developed and the characteristics of the benchmark. Further, considering the difficulty inherent in

obtaining severity distributions for immature data, it may be helpful to use adjustment processes to adjust an s_R distribution for data at 12 months maturity to develop s_R's suitable for data at, say, 24 or 36 months maturity.

4.1 Using the Severity Distributions of a Larger Body of Claims

Using the severity distributions of a broader group that contains or matches the body of claims to be developed provides a good alternative to datasets with limited credibility. Note that while the 12 months and ultimate severity distributions of some line of business in a single state may contain too few claims to reliably compute an s_R for that state, the data from the entire country for that line of business may well provide a good benchmark. So, it may be helpful to develop the s_R distribution from a broader dataset. Similarly, instead of using the data from a program[15] or class, it may be helpful to use an s_R for the entire line of business. Lastly, although the discussion of this approach is short, it should not diminish the relevance of the method. Not only can it solve the problem of constructing s_R for small programs and classes, it also eliminates the problem of obtaining separate s_X and s_Y distributions for each class.

4.2 Using the Severity Distributions Underlying Advisory/Rating Organization Loss Costs/Rates

Another approach is to develop s_X and s_Y from material provided by advisory or rating organizations. For example, one might request (typically, at some cost) a sample of all the individual claims at twelve months for the class underlying some increased limit or excess charge table. Then, one might use that twelve month's data to generate s_X, and the organization's loss distribution[16] underlying the increased limit or excess ratio table as s_Y. As long as the scope of the data in such an s_X is a good match for the scope of the data in the s_Y, those would provide the raw materials needed to generate an s_R. Further, it would appear to be fairly important that the s_X distribution and the s_Y distribution come from the same block of business, since the resulting s_R is computed using the differences between them. Consider, though, a small block of business that is a subset of the original, presumably fairly homogeneous, block[17]. Using the resulting s_R from a slightly different, but larger block of business[18] would not be expected to generate significant error, especially if the adjustments in the following section are used. Therefore, one might expect the resulting s_R to often have relatively small errors in estimating the loss development distribution for the modified block. So, this could be a good source of data for computing s_R for the slightly modified block of business.

---

[15] As will be seen later, it is possible to adjust s_R for differences in claim handling, and adjustments of that sort may be required.

[16] It should be noted that whatever loss cost "trend" adjustments that are needed to match the occurrence, etc. periods of the two distributions will be done.

[17] Or one that has development and reserving patterns that are otherwise similar to those of the original block of business

[18] As stated previously, the key questions to ask are 'Does it generate a better approximation than burning cost?' and 'Do I nevertheless have a method to make my estimate more accurate?'.

4.3 Modifying the Key Parameters of s_R or X to Mirror Different Loss Development Patterns

There are some fairly straightforward adjustments one might make to s_R to make it useful for blocks of business that have significant, but tractable, differences. For example, the time and expense of computing an s_R for data at 12 months may necessitate that the results be adjusted for use at 24, 36, etc. months. It would not be unusual for an insurance enterprise to have a large general liability, commercial auto liability, or workers compensation program handled by in-house claims adjusters, yet also have a similar program where losses are adjusted by a third party claims administrator ("TPA"). It is also possible that the company might have some block of business that is administered in-house, but has some claims with different features than the base program. Volume considerations might necessitate that s_R be computed from the main or "benchmark" program and reflect the loss development distribution of the main program. But, quite often the basic loss development factors for the "alternate" (for example, the TPA-administered) program[19] will be different than those of the main program. It is possible to correct s_R for the overall development difference. For example, assuming s_R is created so that the mean of R is $LDF_{benchmark}$, and that the alternate program has a loss development factor of $LDF_{alternate}$. Then, one may use the formula

$$s_{R,alternate}(r) = s_R \left(\frac{LDF_{alternate}}{LDF_{benchmark}} r \right) \frac{LDF_{alternate}}{LDF_{benchmark}}. \tag{36}$$

As one may see, this is similar to the calculations in Appendix C, Table 6 that were discussed in subsection 3.3, when one considers the expansion of the intervals between the discrete "r" points induced in Table 6. In fact, one could replace the formula in equation (36) with the calculations on Table 6 in discrete cases[20]. So, adjusting the random development distribution s_R for differences in overall development is fairly tractable.

On the other hand, it is often more preferable to leave R as computed and modify the X. In that event, one could simply compute

$$x_{adjusted} = x_{alternate} \frac{LDF_{alternate}}{LDF_{benchmark}} \tag{37}$$

for each $x_{alternate}$ in the dataset of claims. Then, one would apply the range of potential development factors s_R to the resulting $x_{adjusted}$'s in accordance with equation (2). In some respects, this is a more straightforward process.

There is a more subtle aspect of reserving that may be considered as well. The differences in loss development factors do reflect the relative levels of overall development. But, one should also consider the accuracy of reserving. For example, suppose a TPA uses a strict formula approach to reserving, where every claim is reserved at the average cost at the given maturity of the reserves set by the benchmark adjusters. The loss development factors needed for the TPA will equal that of the reference (in this case in-house) data. However, it should be clear that since the TPA's reserves feature no claims of above average size, combining the TPA's data with s_R will predict too few excess losses. So, in that case, one may regard the benchmark data as being more accurately reserved. In fact, to the extent that the variance in loss sizes in the X distribution mirrors the

---

[19]This is not intended to provide commentary on TPA-administered programs. It is merely used to provide an example where the claims profile may be different than that of the standard program.

[20]Excepting that Table 6 is more oriented toward matching the logarithm of R to the logarithms of X and Y than adjusting s_R.

differences in the losses at ultimate Y, one may regard greater variance as greater accuracy in reserving. Of course, it is possible that some differences in reserves set by a claims staff may be due to factors that do not relate to ultimate costs. So, variance should be regarded as an indicator, not a perfect measure of the reserving accuracy in X.

An important aspect of the variance of the immature, undeveloped loss should be considered. Even when there is too little data to construct a reliable loss distribution for the undeveloped losses, there may nonetheless be enough to provide an acceptable estimate of the variance of the undeveloped losses.

The key factor to match to the reference data is not the variance, though. It is the coefficient of variation, c_v. That is because correcting $X_{alternate}$, for example, for a higher development factor raises the variance by $\frac{LDF_{alternate}}{LDF_{benchmark}}$ along with the values of $X_{alternate}$. So it is important to either use LDF-adjusted data or strictly use the coefficient of variation to compute the variance of the alternate claims regime/alternate block of business.

So, at the start, a logical correction might be to simply expand or contract[21] the variance of the x's. For instance, if the variance[22] is presently α^2 and the R was developed using a distribution with a variance of β^2, and μ_X is the mean, then one may use

$$x_{transformed} = \mu_X + \frac{\beta}{\alpha}(x - \mu_X) \tag{38}$$

to get the desired variance. Then, one might apply R to each revised claim amount $x_{transformed}$.

An alternate method is also needed. Although the method guarantees the needed variance of $X_{transformed}$, it has a substantial weakness. When X contains values that are small enough, and $\frac{\beta}{\alpha} > 1$, one may obtain negative values of $x_{transformed}$. Of course, often such values will be irrelevant to the pricing of the excess layers. But, an alternative is needed for the cases where the smaller losses affect the expected excess losses. Such an approach involves looking at the logarithms of the values of X and using a power function (to some γ^{th} power)[23] to approximately correct the variance. The first steps involve computing the geometric mean $\mu_{geometric}$ and adjustment factor γ

$$\mu_{geometric} = E[ln(X)]$$
$$\gamma = \sqrt{\frac{\ln(c_v^2[benchmark]+1)}{\ln(c_v^2[alternate]+1)}}.$$

Then, one may compute the distribution with the "power" variance correction, by computing each $x_{transformed}$ via

$$x_{transformed} = \left(\frac{x}{\mu_{geometric}}\right)^{\gamma} \mu_{geometric} \tag{39}$$

This only technically promises the correct variance when X is a lognormal distribution. However, it could be expected to produce reasonable results as long as X is not too different from a lognormal distribution. Further, when it does not appropriately mirror the desired variance, the γ parameter could always be revised.

---

[21]This implicitly provides a rationale for adjusting the x's rather than the distribution of R. The variance of the x's may be corrected up or down as needed, but computing a R with lower variance, since it is a distribution rather than a set of values, is more complicated.

[22]α and β are reused here from subsection 3.6 to avoid using unfamiliar symbols, but of course they have a completely different meaning in this context.

[23]The variable γ is reused here to avoid using unusual variable names, just as with α and β.

So, with these changes, especially when the mean adjustment and variance adjustment are made in tandem, it is possible to modify the claims data to make them suitable for combination with a R from some not-too-different severity distribution.

4.4   Non-Matching, but Usable, Benchmarks

Subsection 4.3 opens up a large range of opportunities to use benchmark data. For example, insurance company data might be used for claims severity distributions reported by TPA's. Conceivably, as long as the general curve is reasonably predictive of the range of development, it might be used for claims severity data from other maturities. Of course, there is no reason to believe that similar-developing subsets of some benchmark could not use the s_R of the benchmark. So, once s_R is established for some benchmark, one might expect it to have wide applicability.

5   Pure IBNR Claims

A criticism that may be raised about the processes in this paper is that they do not recognize true late reported or IBNR claims. One must also recognize the potential that the late reported claims may be more severe than the remaining claims. If the late reported claims were part of the distribution of the x's, they would definitely affect the $X \times R$ combination. But they are not. They are, however, part of the y's. So, even though the severities of the late reported claims are not included in X, $X \times R$ generates all the ultimate severities of Y. Therefore, the expected distribution of possible x's is sufficient for the distribution $X \times R$ to generate the full severity distribution Y. By construction, R implicitly recognizes the missing, potentially larger, severities that would be generated by the late reported claims.

However, the developed excess loss will be missing some claims counts across-the-board. For example, if 10% of all the ultimate claim counts in some product line are as yet unreported at 24 months, and the x's used in Mahler's equation (2) for the excess loss estimate are the loss amounts of the claims reported at 24 months, then the result will be deficient by 10% every time. That is both bad news and good news. All the excess loss estimates must be adjusted, but the predicted costs in each and every layer need only be corrected via a division by .9. That simple correction (multiplying the end result of Mahler's equation by the reported count development factor) suffices to create fully unbiased predictions of the costs in the various layers.

6   Those Pesky Closed Claims

A common criticism from users of loss reserve reports covering excess programs is that some of the claims projected to develop into the excess layer are already closed. In the context of this paper it should be clear that if some large closed claims are expected in the dataset, removing them would bias the resulting estimate of the ultimate excess loss. Thus, an alternative that does not combine closed claims with the standard construction of R is desirable. The solution involves simply removing the closed claim distribution within X from both: the severity distribution of the claims reported to-date (s_X); and, that of the ultimate claims (s_Y). To do so, one needs the total number of expected claims to-date in X, "a", the expected number of closed claims to-date in X ,"b", the expected number of ultimate claims "c", and the special distribution of the closed claims

to-date in X, "$s_{X,C}$". Then, one may compute the distributions of the open claims in X, "$s_{X,O}$" and the ultimate loss distribution excluding the claims closed in X, "$s_{Y,O}$", as

$$s_{X,O}(x) = \frac{as_X(x) - bs_{X,C}(x)}{a - b}, \text{ for all x, and} \tag{40}$$

$$s_{Y,O}(y) = \frac{cs_Y(y) - bs_{X,C}(y)}{c - b} \text{ for all y.} \tag{41}$$

The resulting $s_{X,O}$ and $s_{Y,O}$ may then be used with all the techniques presented previously to develop a random development distribution R for just the open claims.

7   Summary

Developing individual claims with standard, single-valued[24] loss development factors generally understates the cost of the excess layers. Although eliminating the bias in that process is a challenging problem, an overall framework for unbiased development of excess layers based on stochastic development (probabilistic loss development factors) using a single development distribution was shown. A number of methods for estimating the values of that probability distribution were developed. Some involve a matrix approach, some involve differing approaches to curve-fitting, and some involve the complexities of Fourier transforms. So, one may use more than one method to develop the probability distribution for loss development, and mirror the use of multiple reserving methods in underlying layers with multiple reserve indications for the losses in the excess layer(s). Further, a number of methods for adjusting the distribution when data that exactly matches the reserving situation is not available were discussed.

Hopefully, these methods will be broadly used. It is also hoped that other papers will expand the tools available. As promised, though, the methods provided for computing this random development factor offer a means to provide an unbiased development approach. Specifically, these methods offer an opportunity to prepare unbiased estimates of claim groups such as losses excess of large deductibles or losses excess of a specific retention. Those results may in turn be used in both reserving and pricing. Consequently, this approach has the potential to improve the accuracy of actuarial projections.

References

J. Boor. Interpolation along a curve. *Variance*, 08:9–22, 2014.

W. Gillam and J. Couret. Retrospective rating: 1997 excess loss factors . *Proceedings of the Casualty Actuarial Society*, 84:450–481, 1997.

L. Halliwell. The discrete fourier transform and cyclical overflow . *Variance*, 08:73–79, 2014.

K. Holler and S.Philbrick. A survey of methods used to reflect loss development in ratemaking. *Casualty Actuarial Society Forum,Winter Edition*, pages 243–296, 1996.

---

[24]Specifically, a single value for each development maturity.

S. Lowe, Y. Jing and J. Lebens. Claim reserving: Performance testing and the control cycle. *Variance*, 03:161–193, 2009.

H. Mahler. Retrospective rating: 1997 excess loss factors [discussion]. *Proceedings of the Casualty Actuarial Society*, 85:316–344, 1998.

G. Perzoli and P. Rathie. Distribution of product and quotient of pareto variates. *Metrika*, 27: 165–169, 1980.

A   Example of the Matrix Method for Estimating Z and R

Suppose one begins by computing $[\mathcal{U}]$ from X and s_X(whose values are given below), starting from values of u that start at zero and are .3 apart.

Table 1: Calculation of $[\mathcal{U}]$ from Values of X

u (or .3i)	$x = \exp(u)$	$s_X(x)$	$s_U(u) = x s_X(x)$	$[\mathcal{U}]_i = .3 s_U(u)$
0	1.000	.333	.333	.100
.3	1.350	.494	.667	.200
.6	1.822	.549	1.000	.300
.9	2.460	.339	.833	.250
1.2	3.320	.151	.500	.150

Therefore

$$[\mathcal{U}] = \begin{bmatrix} .10 \\ .20 \\ .30 \\ .25 \\ .15 \end{bmatrix}. \tag{42}$$

Similar calculations for $[\mathcal{W}]$,this time beginning with some given values of Y, yields the values tabulated below.

So

$$[\mathcal{W}] = \begin{bmatrix} .010 \\ .040 \\ .100 \\ .185 \\ .235 \end{bmatrix}. \tag{43}$$

Then the index j of $[Z]_j$ is set to cover 0, .3,.6, and .9. Following equation (14) the corresponding matrix equation for $[\mathcal{Z}]$ using $[\mathcal{U}^*]$ and $[\mathcal{W}]$ is

Table 2: Calculation of $[\mathcal{W}]$ from Values of Y

w (or $.3k$)	$y = \exp(w)$	$s_Y(y)$	$s_W(w) = ys_Y(y)$	$[\mathcal{W}]_k = .3s_W(w)$
0	1.000	.033	.033	.010
.3	1.350	.099	.133	.040
.6	1.822	.183	.333	.100
.9	2.460	.251	.617	.185
1.2	3.320	.236	.783	.235

$$
\begin{bmatrix}
.10 & 0 & 0 & 0 \\
.20 & .10 & 0 & 0 \\
.30 & .20 & .10 & 0 \\
.25 & .30 & .20 & .10 \\
.15 & .25 & .30 & .20
\end{bmatrix} \times [\mathcal{Z}] =
\begin{bmatrix}
.010 \\
.040 \\
.100 \\
.185 \\
.235
\end{bmatrix}
\tag{44}
$$

That illustration shows the structure of the matrix equation (and $[\mathcal{U}^*]$) . The columns of $[\mathcal{U}^*]$ contain equal values from $[\mathcal{U}]$, but they are offset, and have successively more and more preceding zeros. Per the equation, that information can provide an estimate for $[Z]$. Even though $[\mathcal{Z}]$ has four entries and $[\mathcal{W}]$ has five, standard linear algebra provides a way to find the four dimensional array $[\mathcal{Z}]$ for which $[\mathcal{U}^*] \times [\mathcal{Z}]$ best estimates[25] $[\mathcal{W}]$. To provide a view of the calculations and the characteristics of the various matrices, calculations for this numerical example follow.

The best estimate process generates a four by four system of equations by multiplying both sides of equation (44) by $[\mathcal{U}^*]^T$, the matrix transpose of $[\mathcal{U}^*]$. That gives the initial matrix equation

$$
[\mathcal{U}^*]^T \times [\mathcal{U}]^* \times [\mathcal{Z}] = [\mathcal{U}^*]^T \times [\mathcal{W}],
\tag{45}
$$

or

$$
\begin{bmatrix}
.10 & .20 & .30 & .25 & .15 \\
0 & .10 & .20 & .30 & .25 \\
0 & 0 & .10 & .20 & .30 \\
0 & 0 & 0 & .10 & .20
\end{bmatrix} \times
\begin{bmatrix}
.10 & 0 & 0 & 0 \\
.20 & .10 & 0 & 0 \\
.30 & .20 & .10 & 0 \\
.25 & .30 & .20 & .10 \\
.15 & .25 & .30 & .20
\end{bmatrix} \times [\mathcal{Z}]
\tag{46}
$$

$$
=
\begin{bmatrix}
.10 & .20 & .30 & .25 & .15 \\
0 & .10 & .20 & .30 & .25 \\
0 & 0 & .10 & .20 & .30 \\
0 & 0 & 0 & .10 & .20
\end{bmatrix} \times
\begin{bmatrix}
.010 \\
.040 \\
.100 \\
.185 \\
.235
\end{bmatrix} ,
\tag{47}
$$

---

[25]Least sum of squared errors in estimating the various elements of $[\mathcal{W}]$.

which may be reduced, using a popular spreadsheet program, to

$$
\begin{bmatrix}
.2250 & .1925 & .1250 & .0550 \\
.1925 & .2025 & .1550 & .0800 \\
.1250 & .1550 & .1400 & .0800 \\
.0550 & .0800 & .0800 & .0500
\end{bmatrix}
\times [\mathcal{Z}] =
\begin{bmatrix}
.12050 \\
.13825 \\
.11750 \\
.06550
\end{bmatrix}.
\tag{48}
$$

Note that at this stage, the values still look tractable. Then, since $[\mathcal{U}^*]^T \times [\mathcal{U}^*]$ is a four by four matrix, it potentially has an inverse. Assuming such an inverse $\left([\mathcal{U}^*]^T \times [\mathcal{U}^*]\right)^{-1}$ exists, then one may solve for $[\mathcal{Z}]$. In fact, $[\mathcal{U}^*]^T \times [\mathcal{U}]^*$ does have an inverse (as determined using the same popular spreadsheet program) and it is

$$
\left([\mathcal{U}^*]^T \times [\mathcal{U}^*]\right)^{-1} =
\begin{bmatrix}
56.896 & -87.931 & 22.413 & 42.241 \\
-87.931 & 208.620 & -198.275 & 80.172 \\
22.413 & -198.275 & 460.344 & -443.965 \\
42.241 & 80.172 & -443.965 & 555.603
\end{bmatrix}.
\tag{49}
$$

Note that this matrix, although an inverse of a matrix with strictly positive elements, all between 5% and 23%, has very large elements and some very large negative elements. So, one might expect such matrices to magnify small errors or be "ill-conditioned". In this case, though, in conjunction with equation (48) it produces an accurate solution[26] for the matrix equation of

$$
[\mathcal{Z}] =
\begin{bmatrix}
.1 \\
.2 \\
.3 \\
.4
\end{bmatrix}.
\tag{50}
$$

Having $[\mathcal{Z}]$ should allow us to estimate the excess losses from undeveloped losses. For example, suppose one has only three losses reported so far, at \$5,000, \$50,000, and \$75,000, and one seeks the expected ultimate loss excess of \$100,000. The first step is to follow the protocols of equation (19). Those essentially say that we may take the probabilities generally represented by the $[\mathcal{Z}]_j$'s as representing the probability in intervals around the $r = \exp(jg)$'s. So, each $\exp(jg)$ may be taken as representative of the r-values in its representative interval with the probability of r lying in the interval at $[\mathcal{Z}]_j$. So, Table 3 shows how the ultimate excess loss may be estimated using the discrete values associated with the $[\mathcal{Z}]_j$'s.

Of note, using the $[\mathcal{Z}]$-weighted average of the loss development factors (in other words, the single loss development factor from, say, the chain-ladder method) would only produce an "estimated" excess loss of \$42,533. Note also that the \$42,533 is unusually close to the excess value of \$54,226 because of the large weight on the highest development factor. In conclusion, though, the results of Table 3 are directionally correct.

Another alternative is to compute the s_R values using the $[\mathcal{Z}]_j$'s at the $z = \ln(r)$ points that correspond to the ig's, then fit a curve to them. Such an analysis is shown in Table 4. Note that the increase in probability at higher limits in this example is unusual (one would expect s_R to decrease slowly till r reaches either some large value or infinity). The fact that a uniform distribution was used emanates from that situation, and the choice of a uniform distribution should not be

---

[26]It should be stated that ill-conditioned matrices do tend to magnify errors. The quality of the result above appears to result from the fact that $[\mathcal{U}^*]^T \times [\mathcal{U}^*]$ is only a four by four matrix.

Table 3: Sample Calculation of Estimated Excess Loss Using $[\mathcal{Z}]$

| | Index "j" | | | | $[\mathcal{Z}]_j$ Wtd. |
	0	1	2	3	Average
$[\mathcal{Z}]_j$	0.1	0.2	0.3	0.4	
$\exp(.3j) = r = LDF$	1.000	1.350	1.822	2.460	
Loss 1	$5,000	$5,000	$5,000	$5,000	
Developed	$5,000	$6,749	$ 9,111	$12,298	
Excess $100,000	$0	$0	$0	$0	$0
Loss 2	$50,000	$50,000	$50,000	$50,000	
Developed	$50,000	$67,493	$91,106	$122,980	
Excess $100,000	$0	$0	$0	$22,980	$9,192
Loss 3	$75,000	$75,000	$75,000	$75,000	
Developed	$75,000	$101,239	$136,659	$184,470	
Excess $100,000	$0	$1,239	$36,659	$84,470	$45,034
Total Est. Excess					$54,226

considered normal. So, one should simply take this as an illustration of curve-fitting using $[\mathcal{Z}]$. The poor quality fit that results, and the fact that the mean of the fitted distribution is less than the average loss development factor (at least that per the point estimate approach) yields the very unusual result that the fitted curve excess loss is less than that derived by burning cost. This should be taken to reinforce the importance of matching the mean when the distribution of R is determined. Fitting a curve of course differs radically from curve family to curve family[27].

So, by estimating $[\mathcal{Z}]$ representing s_Z, then dividing out the exponent of the matrix indices (along with the scale factor) gives the severity distribution s_R as noted in equation (9). So, when it is effective, this discrete approximation may produce a useful estimation of the transition probability density (and a consequently useful measure of the excess losses).

However, it must be noted that there are two concerns with the use of this specific method. The first is that when one begins with 0 as an index of $[\mathcal{Z}]$, that represents a lower bound for R of 1.00. So, one is beginning with a cap from below for r of unity. Thus, one is assuming that no claims will develop downward[28] under any circumstances. In most circumstances, such an assumption might be thought of as unreasonable. A second concern arises from the setup of the matrix equations. When using this method with a fairly large number (perhaps twenty or so) of values of z, or for which values along long stretches of $[\mathcal{U}]$ or $[\mathcal{W}]$ are near zero, computational error may cause the numerical solution may be so far off as to be completely unreasonable. So, there is significant motivation for improving this method.

---

[27]Other than the two stalwarts of method of moments and the use of a solution routine to solve a least squares problem (per the example in Table 9).

[28]Of course, this may be resolved by accepting some negative indices, perhaps even only one negative index, for the values of Z.

Table 4: Sample Severity Curve Fitting and Calculation of Estimated Excess Loss Using $[\mathcal{Z}]$

	Index "j"			
	0	1	2	3
$[\mathcal{Z}]_j$	0.1	0.2	0.3	0.4
$\exp(.3j) = r$	1.000	1.350	1.822	2.460
$s_R(r) = \mathcal{Z}]_j/(.3\exp(.3j))$	0.33333	0.49387	0.54881	0.54209
Uniform distribution of Best Fit:				
Avg. Value s = .47953; Inverse=Interval Length = 2.0837 (Use 2.0) s_R-Wtd. Avg. of Points = Center of Interval=1.7378 (Use 1.7) Selected Uniform Distribution with Mass .5 on $[.7, 2.7)$				
Mahler Excess Function = $\int_{100,000/C}^{2.7} .5(rC - 100,000)dr$, for each claim amount C such that $2.7C \geq 100,000$				
Loss 1	5,000	Excess =	0	
Loss 2	50,000	Excess =	6,125	
Loss 2	75,000	Excess =	35,021	
Total			41,146	

B Example of How to Set Up the Key Calculations for an Overdetermined Matrix

An overdetermined matrix system brings many more data points (values in $[\mathcal{W}]$) that one would attempt to match, but can create more complex matrix arithmetic. Per equation (14), though, the calculations are specified and use functions available in at least one standard spreadsheet program.

An example will illustrate how to set up the matrix calculations. In it, the introductory setup of $[\mathcal{U}]$ and $[\mathcal{W}]$ (as illustrated in the Appendix A example, with different values) will be presumed to have already been completed. Further, the final conversion from $[\mathcal{Z}]$ to R (also illustrated in Appendix A, again with different values) will be omitted as well. The focus will simply be on the setup of the matrix arithmetic.

If one has the following vectors:

$$[\mathcal{W}] = \begin{bmatrix} .080 \\ .100 \\ .140 \\ .180 \\ .180 \\ .140 \\ .100 \\ .080 \end{bmatrix} \tag{51}$$

for $-2g,\ -1g,\ 0g,\ 1g,\ 2g,\ 3g,\ 4g,\ 5g$;

$$[\mathcal{U}] = \begin{bmatrix} .10 \\ .20 \\ .30 \\ .25 \\ .15 \end{bmatrix} \tag{52}$$

for $-1g,\ 0g,\ 1g,\ 2g,\ 3g,$; and

$[\mathcal{Z}]$ is set to cover $-1g,\ 0g,\ 1g,\ 2g$;

then the matrix equation using $[\mathcal{U}^*] \times [\mathcal{Z}] = [\mathcal{W}]$ is

$$\begin{bmatrix} .10 & 0 & 0 & 0 \\ .20 & .10 & 0 & 0 \\ .30 & .20 & .10 & 0 \\ .25 & .30 & .20 & .10 \\ .15 & .25 & .30 & .20 \\ 0 & .15 & .25 & .30 \\ 0 & 0 & .15 & .25 \\ 0 & 0 & 0 & .15 \end{bmatrix} \times [\mathcal{Z}] = \begin{bmatrix} .080 \\ .100 \\ .140 \\ .180 \\ .180 \\ .140 \\ .100 \\ .080 \end{bmatrix}. \tag{53}$$

Clearly, one may then solve for $[\mathcal{Z}]$ using the methods of subsection 3.1, specifically, using equation (16).

C An Example of How the Matrix Equation Output May be Corrected

As noted in subsection 3.5 sometimes computer-based matrix arithmetic may distort the mean and variance of Z somewhat. However, the author's experience suggests it is likely to be possible to remove much of the distortion. An example of the correction process follows.

One may numerically solve the discretized problem specified as:

- W is a normal distribution with mean 5.5 and variance .74 ($W \sim N(5.5, .74)$);

- $U \sim N(3, .49)$;

- both are discretized to $[\mathcal{U}]$ and $[\mathcal{W}]$ with increments of .1;

- the matrix equation covers W between 3.2 and 7.6 and U between 1.2 and 4.8; and

- a discretized solution for Z ($[\mathcal{Z}]$)in increments of .1 between 1.2 and 3.6 is sought.

The results of the calculation of $[\mathcal{Z}]$ are shown below:

$$[\mathcal{Z}] = \begin{bmatrix} 0.06384 \\ -0.02399 \\ -0.01848 \\ -0.00732 \\ 0.00756 \\ 0.02386 \\ 0.03937 \\ 0.05234 \\ 0.07509 \\ 0.05579 \\ 0.06094 \\ 0.06446 \\ 0.06644 \\ 0.06686 \\ 0.06556 \\ 0.06228 \\ 0.11377 \\ 0.07348 \\ 0.05943 \\ 0.04086 \\ 0.01971 \\ -0.00130 \\ -0.01903 \\ -0.03044 \\ 0.09551 \end{bmatrix}. \tag{54}$$

One may clearly see that the values between 0.06384 and -0.00732 are dominated by error, while 0.00756 is suspect. At the bottom of the matrix 0.01971 is suspect, whereas -0.00130 through 0.09551 are in error. The results in the middle are initially presumed to be relatively accurate.

Of course, the problem in U and W has a known solution. The sum of two independent normal distributions is itself a normal distribution, with a mean equal to the sum of the means and a variance equal to the sum of the variances[29]. So, $R \sim N(2.5, .25)$. So, the matrix results are compared to the actual results in Table 5.

Interestingly, Table 5 essentially confirms the judgment-based assessment of which points are valid and which are not. The "correct" points are also imperfect, but they are somewhat close to the distribution. This illustrates how one might extract a useful approximation from a matrix approximation that generates some amount of approximation error.

The results may be enhanced further. First, one may limit the computed points to those between 1.7 and 3.1. Then, the probabilities only sum to roughly 92%. So, the probabilities should be rescaled to total unity. Also, assuming that one did not know (as one generally would not) that the distribution was a normal distribution, one would still know the mean and variance of $\ln(X) = U$ (3 and .49) and the mean and variance of $\ln(Y) = W$ (5.5 and .74), since those

---

[29]Note that the last two sum formulas are not unique to normal distributions, but apply to any two independent distributions.

Table 5: Comparison of Matrix-Estimated $[\mathcal{Z}]$ Distribution and Actual $N(2.5, .25)$ Distribution

$\ln(r) = z$	Estimated $.1 \times N(2.5, .25)$ Value $s_Z(z)$	True $.1 \times N(2.5, .25)$ Value $s_Z(z)$	Estimation Error%
1.2	0.06384	0.00272	2247 %
1.3	-0.02399	0.00448	-635 %
1.4	-0.01848	0.00709	-361 %
1.5	-0.00732	0.01080	-168 %
1.6	0.00756	0.01579	-52 %
1.7	0.02386	0.02218	8 %
1.8	0.03937	0.02995	31 %
1.9	0.05234	0.03884	35 %
2.0	0.07509	0.04839	55 %
2.1	0.05579	0.05794	-4 %
2.2	0.06094	0.06664	-9 %
2.3	0.06446	0.07365	-12 %
2.4	0.06644	0.07821	-15 %
2.5	0.06686	0.07979	-16 %
2.6	0.06556	0.07821	-16 %
2.7	0.06228	0.07365	-15 %
2.8	0.11377	0.06664	71 %
2.9	0.07348	0.05794	27 %
3.0	0.05943	0.04839	23 %
3.1	0.04086	0.03884	5 %
3.2	0.01971	0.02995	-34 %
3.3	-0.00130	0.02218	-106 %
3.4	-0.01903	0.01579	-221 %
3.5	-0.03044	0.01080	-382 %
3.6	0.09551	0.00709	1247 %

distributions are known starting points. So, by subtraction, the mean of $\ln(R) = Z$ would have to be 2.5 and its variance would have to be .25. For the adjusted distribution, the mean is 2.46 and the variance is .155. So, although the mean is relatively close, the variance contains serious error.

To correct this, two steps are needed. First, since the mean should be moved up by $.04/.1 = 40\%$ of the grid size. So, each point is assigned 60% of its current value and 40% of the value below. The results are shown below

However, the problems with the variance remain. To understand the problem better, it is helpful to compute the ratio of the standard deviation desired, $\sqrt{.25} = .5$, and the standard deviation presently[30] in the estimated[31] distribution of $[\mathcal{Z}]$, $\sqrt{.155} \approx .39$. That ratio is approximately 1.25.

---

[30]After correcting the mean.

[31]Note that since the distribution is discretized, some values may add up to something slightly different than unity when they should add to unity and modest effects on the mean and variance may occur.

Table 6: $[\mathcal{Z}]$ with Mean Adjusted to 2.5

$\ln(r) = j$	Adjusted $[\mathcal{Z}]_j$
1.7	0.01555
1.8	0.03603
1.9	0.05122
2.0	0.07169
2.1	0.06899
2.2	0.06396
2.3	0.06850
2.4	0.07132
2.5	0.07245
2.6	0.07178
2.7	0.06908
2.8	0.10122
2.9	0.09733
3.0	0.07067
3.1	0.05246
3.2	0.01775

Then, if maintaining the grid size is not needed[32], one may simply expand the grid as $j = 2.5 + 1.25 * (j - 2.5)$ to get the values in Table 7.

Alternately, linear interpolation using the reference values in Table 7, plus an approximate $1/1.25$ correction for the grid size difference (essentially, there are now 23 points rather than 17) produces the mean/variance adjusted distribution[33] from the original grid in Table 5. Table 8 compares the results to the true distribution.

In this case, both the mean/variance adjusted[34] distribution and the result of the original matrix arithmetic have significantly more weight towards larger and smaller development. So, they would tend to overestimate excess losses. However, that does not imply that the methodology will always do so, and it should still be more accurate than applying the same loss development factor to all claims.

Although the process is not presented in detail here, one could assign greater weight to the large values in $[\mathcal{W}]$ by multiplying each of the equations (one for each "k') represented by equation (13) by an individual weight w_k, and then creating $[\mathcal{U}^*]$ and $[\mathcal{W}]$ from the weighted values. One must be careful, though to avoid placing too much weight, say, on the equations of the largest five k's, or the problem $[\mathcal{U}^*] \times [\mathcal{Z}] = [\mathcal{W}]$ will essentially become only a five dimensional problem.

---

[32]For example, if the goal is simply to provide a distribution between a discrete set of points (per equation (19)) to use as a proxy for the continuous distribution of R.

[33]Including a needed off-balance correction, so the probabilities sum to unity

[34]To a mean of 2.529 and variance of .2634

Table 7: $[\mathcal{Z}]$ with Variance Adjusted to .25 by Expanding Indices

Adjusted $\ln(r) = j$	$[\mathcal{Z}]_j$
1.5	0.01248
1.6	0.02512
1.7	0.03603
1.8	0.04742
1.9	0.05717
2.0	0.05537
2.1	0.05226
2.2	0.05346
2.3	0.05584
2.4	0.05739
2.5	0.05815
2.6	0.05773
2.7	0.05635
2.8	0.06536
2.9	0.08071
3.0	0.07811
3.1	0.06165
3.2	0.04819
3.3	0.03139
3.4	0.00996

D   Example of Fitting a Distribution Using a Matrix-Type Approach.

For example, if one desired that s_Z follow a shifted (to negative one—to easily accommodate the indices of $[\mathcal{U}]$ and $[\mathcal{W}]$ in the example beginning at (43) and (42)) Pareto distribution[35], then one could fit the optimum Pareto parameters[36] by minimizing a weighted sum of the squared errors between the Pareto-generated discrete probability distribution at 0, 1, 2, and 3.

The light gray Pareto parameters at the top of the Table 9 were varied to obtain the lowest value of the weighted sum at the bottom of the table. In most practical situations, numerical solution software present in common spreadsheet programs may be used to find the optimum weighted sum of squares in dark gray by varying the input Pareto parameters in light gray. In this case, the numerical solution process in the software failed, but an approximate solution (and of course all numerical solutions are approximate anyway) was found by varying the parameters in light gray until changes in the parameters ceased to reduce the weighted sum.

---

[35]Use of the shifted Pareto is not intended to endorse its utility for modeling. It is merely used because its calculation is simple.

[36]x_M is the shift, α is the power.

Table 8: Comparison of Approximate Mean/Variance Adjusted Matrix-Derived $[\mathcal{Z}]$ Distribution on Original Grid and Actual Distribution

$\ln(r) = j$	Estimated $.1 \times N(2.5, .25)$ Value $[\mathcal{Z}]_j$	True $.1 \times N(2.5, .25)$ Value $[\mathcal{Z}]_j$	Estimation Error%
1.5	0.01248	0.01080	16 %
1.6	0.02512	0.01579	59 %
1.7	0.03603	0.02218	62 %
1.8	0.04742	0.02995	58 %
1.9	0.05717	0.03884	47 %
2.0	0.05537	0.04389	26 %
2.1	0.05226	0.05794	-10 %
2.2	0.05346	0.06664	-20 %
2.3	0.05584	0.07365	-24 %
2.4	0.05739	0.07821	-27 %
2.5	0.05815	0.07979	-27 %
2.6	0.05773	0.07821	-26 %
2.7	0.05635	0.07365	-23 %
2.8	0.06536	0.06664	-2 %
2.9	0.08071	0.05794	39 %
3.0	0.07811	0.04389	78 %
3.1	0.06165	0.03884	59 %
3.2	0.04819	0.02995	61 %
3.3	0.03139	0.02218	42 %
3.4	0.00996	0.01579	-37 %

Table 9: Illustration of Minimum Weighted Sum of Squares Approach with Shifted Pareto Distribution and Data from Equation (43)

Optimal Pareto Parameters				$x_M =$	2.79		$\alpha =$	1.64	

Index	$[\mathcal{U}^*]$				Pareto Values $[\mathcal{Z}]$	$[\mathcal{U}^*] \times [\mathcal{Z}]$	$[\mathcal{W}]$	Squared Error	Weight
0	0.1	0.0	0.0	0.0	0.2618	0.0262	0.010	0.00026	4
1	0.2	0.1	0.0	0.0	0.1411	0.0665	0.040	0.00070	5
2	0.3	0.2	0.1	0.0	0.0855	0.1153	0.100	0.00023	6
3	0.4	0.3	0.2	0.1	0.0562	0.1698	0.185	0.00023	7
4	0.5	0.4	0.3	0.2		0.2243	0.235	0.00003	8

| | | | | | Weighted Sum = | .0078 | |