LOSS PREDICTION BY GENERALIZED LEAST SQUARES

LEIGH J. HALLIWELL

Abstract

The prediction of losses, whether for ratemaking or for reserving, is the quintessential activity of the actuary. The time-honored technique of loss development is the basis for the chain ladder and Bornhuetter–Ferguson methods. These methods, particularly the chain-ladder, have been subject to a great deal of statistical analysis since the mid-1980s. It is now thought by many that development factors obtained by least squares regression are unbiased. But this paper will argue that the linear modeling and the least squares estimation found in the literature to date have overlooked an important condition of the linear model. In particular, the models for development factors regress random variables against other random variables. Stochastic regressors violate the standard linear model. Moreover, the model assumes that the errors are uncorrelated, but stochastic regressors violate this assumption as well. This paper will show that what actuaries are really seeking is found in a general linear model; i.e., a model with nonstochastic regressors but with an error matrix that allows for correlation. An example will be presented.

1. A SIMPLE ILLUSTRATION OF LOSS ESTIMATION

Consider how actuaries might approach a simple loss reserving problem. Take an exposure period now at a certain age. Based upon our best knowledge heretofore, we have believed that \$100 of losses would ultimately be paid for this period. We know that \$60 has been paid to date. We have also looked into our records and have found that, on similar exposures at the same age, 50%

of the ultimate losses have been paid. How does this new information affect our prior estimate of $100 of losses?

First, we could rely on the statistic that 50% of the losses should have been paid by this time, which implies that we should revise our estimate of ultimate loss to $120. Actuaries would normally say that the development factor from this age to ultimate is 2.0. So, our paid losses should develop from $60 dollars to $\$60 \times 2.0 = \$120$. This is often called the chain ladder (CL) method of loss development.

Second, we could rely on the prior hypothesis that the ultimate loss will be $100, and assume that the accelerated payment of $60 to date will be countered by a decelerated payment of $40 henceforth.

As is so often the case, there is a third approach which mediates between the CL and the prior hypothesis methods. The CL method disregards the prior hypothesis, and sticking to the prior hypothesis disregards the payout statistic. Why not assume that the amount yet to be paid is half of the prior hypothesis estimate, or $50? This, plus the $60 already paid, makes for an ultimate loss of $110. This appealing solution is known as the Bornhuetter–Ferguson (BF) method, and has several variants.[1]

2. THE UPWARD BIAS OF THE CHAIN LADDER METHOD UNDER PLAUSIBLE CONDITIONS

James Stanard [10] simulated thousands of loss triangles, and developed these losses according to four methods, one of which was the CL. He concluded that the CL method was biased in the direction of overestimating ultimate losses. In his Appendix

[1]See Bornhuetter and Ferguson [2]. James Stanard [10, pp. 130f.] describes four loss development methods, the second of which is a "modified" BF method. His third method, called the "Cape Cod" method, is equivalent to what Gary Patrik [9, pp. 352–354] calls the Stanard-Bühlmann (SB) method, under the assumption that all accident years have the same prior expected losses. The SB method is a variant of the BF.

A, he shows why this method should be biased, but not whether the bias should be upward (overestimation) or downward (underestimation). In this section, we will show that under reasonable conditions, the bias is upward.

We have normal random variables $X_1 \sim [\mu_1, \sigma_1^2]$ and $X_2 \sim [\mu_2, \sigma_2^2]$. The correlation coefficient between the two is ρ. Let $Y_1 = e^{X_1}$ and $Y_2 = e^{X_2}$. Y_1 will represent the losses (whether paid or incurred) as of the earlier age; Y_2 as of the later. The assumption that losses are distributed lognormally is convenient for this demonstration, as well as frequently realistic. $X_2 - X_1$ is normally distributed as $[\mu_2 - \mu_1, \sigma_2^2 + \sigma_1^2 - 2\rho\sigma_1\sigma_2]$.

The development factor is an estimate of $E[Y_2/Y_1]$. As Stanard shows in his appendix, the bias of the CL method depends on the relation between $E[Y_1] \times E[Y_2/Y_1]$ and $E[Y_2]$. Given the lognormal assumption, $Y_2/Y_1 = e^{X_2 - X_1}$ is lognormal. Therefore,

$$E[Y_2] = e^{\mu_2 + \sigma_2^2/2}, \qquad \text{and}$$

$$\begin{aligned} E[Y_1]E\left[\frac{Y_2}{Y_1}\right] &= e^{\mu_1 + \sigma_1^2/2} e^{\mu_2 - \mu_1 + \sigma_2^2/2 + \sigma_1^2/2 - \rho\sigma_1\sigma_2} \\ &= e^{\mu_2 + \sigma_2^2/2 + \sigma_1^2 - \rho\sigma_1\sigma_2} \\ &= E[Y_2]e^{\sigma_1^2 - \rho\sigma_1\sigma_2}. \end{aligned}$$

So whether the CL method is biased downward, unbiased, or biased upward depends on whether $e^{\sigma_1^2 - \rho\sigma_1\sigma_2}$ is less than, equal to, or greater than one. And this depends on whether σ_1 is less than, equal to, or greater than $\rho\sigma_2$. Since ρ is less than one ($\rho = 1$ is unrealistic), σ_1 is greater than $\rho\sigma_2$ unless σ_2 is larger than σ_1. This means that the CL method is biased upward unless σ_2 is sufficiently larger than σ_1. The closer ρ is to zero, the less likely σ_2 will be sufficiently large; it is impossible when ρ is less than or equal to zero. Therefore, the CL method works best, or has the least upward bias, when the loss at the later time is highly positively correlated with the loss at the earlier time.

Furthermore, one very plausible assumption is that as a loss ages, its standard deviation remains proportional to its mean, or equivalently, that its coefficient of variation (CV) remains constant. Given the lognormal assumption, this means that:

$$\mathrm{CV}[Y_1] = \sqrt{e^{\sigma_1^2} - 1} = \sqrt{e^{\sigma_2^2} - 1} = \mathrm{CV}[Y_2],$$

which is true if and only if $\sigma_1 = \sigma_2$. But if $\sigma_1 = \sigma_2$, then σ_2 is not sufficiently large, and the CL method will be biased upward. Thus, we have some assumptions regarding lognormality and the coefficient of variation, having verisimilitude singly and together, under which the CL method must be biased upward.

3. AN ATTEMPT TO REHABILITATE THE CHAIN LADDER METHOD

Stanard's findings have disconcerted actuaries, who are very fond of using the CL method for estimating ultimate losses. The CL logic is simple and appealing. For example, "If half the losses should have been paid by now, and \$60 have indeed been paid, then \$120 should ultimately be paid." Moreover, the CL method makes no use of a prior hypothesis, so it seems to have the benefit of parsimony.[2] As for an upward bias, many actuaries would consider this to be a windfall since, if true, it would add an extra bit of conservatism to their estimates.[3]

---

[2]Recall Ockham's razor.

[3]It is ironic that although in theory and in simulation the CL method should be biased upward, in practice it frequently seems to be biased downward. Several years ago, while employed by NCCI, the author conducted a study of how accurately losses were developed in NCCI ratemaking. He found that the development was usually underestimated by five to ten percent. Of course, this is not really an indictment against our belief that the CL method is biased upward. Rather, it is reflective of the runaway conditions of workers compensation in the late 1980s; i.e., of the worsening conditions not reflected in projections of ultimate losses. It is assumed throughout this paper that all the rows of a loss triangle are commensurate (akin to one another), and that we are cognizant of, and can adjust for, the important exogenous effects on the losses. Doing justice to this assumption involves the hardest work of the actuary, and is more actuarial art than science.

However, most actuaries desire unbiased estimates—not just because of statistical purity, but also because of competitive pressures in business. If loss estimates need to be conservative, then the conservatism should be a deliberate and measured addition to an unbiased estimate. Therefore, Stanard's findings have been one impetus in the search for a better approach.

Daniel Murphy [8] has sought to extract unbiased loss development factors from loss triangles by the application of linear regression techniques. His model is $\mathbf{Y} = \mathbf{J}\alpha + \text{Prev}(\mathbf{Y})\beta + \mathbf{e}$, where α and β are the regression coefficients to be estimated and $\mathbf{Y}$, $\mathbf{J}$, Prev($\mathbf{Y}$), and $\mathbf{e}$ are $(t \times 1)$ vectors. Prev($\mathbf{Y}$) and $\mathbf{Y}$ are adjacent matching columns in the loss triangle, and $\mathbf{J}$ is a vector of ones, or an intercept vector. As for Var[$\mathbf{e}$], a $(t \times t)$ matrix, it is assumed to be diagonal; i.e., $\text{Cov}[\mathbf{e}_i, \mathbf{e}_j] = \sigma_i^2$ for $i = j$, but 0 otherwise.

Murphy [8, p. 187] appeals to the Gauss–Markov theorem in affirming that the least-squares estimates of the regression coefficients are best linear unbiased estimates (BLUE). From there, he fills in the loss triangle with supposedly unbiased estimates, and constructs a confidence interval for the aggregate incurred loss. However, it appears that Murphy has overlooked one of the conditions of the Gauss–Markov theorem, thus invalidating his claim of unbiasedness.

First, Murphy shows in his appendix that the familiar simple-average and weighted-average development factors fall out from a regression model with no intercept ($\alpha = 0$), given appropriate assumptions as to the σ_i^2 elements. This in itself should raise doubt: if a special case of the linear regression model reduces to the CL method which is biased, then how can the regression estimates be unbiased? One might be tempted to answer that the special case is biased, whereas the model with the intercept (nonzero α) is unbiased. However, the Gauss–Markov theorem, starting from the assumption that the linear model $\mathbf{Y} = \mathbf{X}\beta + \mathbf{e}$ is well specified, where Var[$\mathbf{e}$] = $\mathbf{\Phi}$, proves $(\mathbf{X}'\mathbf{\Phi}^{-1}\mathbf{X})^{-1}(\mathbf{X}'\mathbf{\Phi}^{-1}\mathbf{Y})$

to be the best linear unbiased estimator of β, irrespective of whether the regressor matrix $\mathbf{X}$ has a column of ones to serve as an intercept.[4]

The flaw in Murphy's claim of achieving unbiasedness is that his regressor matrix, the $(t \times k)$ matrix $\mathbf{X}$ in the model $\mathbf{Y} = \mathbf{X}\beta + \mathbf{e}$, contains stochastic regressors; viz., one of its columns is Prev(**Y**), which is stochastic. George Judge [5, Ch. 13] discusses the ramifications of stochastic regressors at some length. In short, if the stochastic regressors are independent of the error vector, then the least-squares estimator is still unbiased. However, even in this case, the usual formulas for $\mathrm{Var}[\hat{\beta}]$ and for $\hat{\sigma}^2$ do not include the variation inherent in the fact that other values of the stochastic regressors could have been realized.

More to the point, Murphy's model is an example of what Judge [5, pp. 574–576] calls "partially independent stochastic regressors." Here Prev(**Y**) is not independent of all the error terms, and the most that can be said is that, under certain conditions, the least-squares estimator is consistent; i.e., asymptotically unbiased. This is the fundamental problem with the CL method. Rather than try to rehabilitate it, this paper introduces a different model that honors all the conditions of the Gauss–Markov theorem.

4. THE NECESSITY OF CONSIDERING EXPOSURE

Consider the conclusions of Stanard and Murphy as to their loss-development simulations:

> The common age-to-age factor approach (Method 1) is clearly inferior to the other three methods [Standard 10, p. 134].
>
> The performance of the incurred loss development technique based on the more general least squares

[4]For a proof of the Gauss–Markov theorem, see Appendix A.

> estimator may approach that of the Bornhuetter–Ferguson (BF) and Stanard–Bühlmann (SB) techniques in some situations [Murphy 8, p. 185].

What Murphy calls the BF and SB techniques correspond to Stanard's second and third methods. Stanard himself favors his fourth method, the additive model [10, pp. 131, 135]: "In fact, Method 4 may be completely unbiased." The SB and additive models can be considered variants of the BF.

The obvious question is this: If it is so hard to beat the BF method and its variants, then why continue to refine the CL method? Of course, loss development factors are used in some BF variants, as Murphy notes [8, p. 207]:

> The average bias of the BF and SB methods should be greater than zero as well because the LDFs on which they rely are themselves overstated more often than not.

But what is unique to the BF variants to give them a performance advantage over the CL method? The answer is that the BF variants incorporate prior knowledge, whether it be a prior estimate of incurred losses or a knowledge of exposure relativities.

I suspect that the desire to avoid relying on prior knowledge is one motive for actuaries to try to perfect the CL method, as if reliance on such knowledge would be tantamount to circular reasoning. However, what could be more axiomatic than a statement such as "Twice the exposure should produce twice the expected loss, all else being equal?" Nevertheless, this information is unknown to the CL method. But is it unknown, or just ignored? If there is enough information in the form of a loss triangle to produce development factors, then there must also be substantial knowledge of the underlying exposures. Otherwise, how would the actuary know that the rows of the triangle were commensurate, or that they represented the same process of development?

Substantial prior knowledge is implied in John Robertson's comment [11, p. 149]:

> Previous literature on reserving techniques generally has concentrated on overcoming the effects of changes in the underlying mix of business, changes in the individual claim reserving and settling policies, and changes in claims reporting systems. Most of this prior literature assumes that once the changes are accounted for and the data has been restated so as to have relatively constant underlying conditions, then any number of loss development methods can be applied to obtain valid forecasts.

Even Murphy, who has worked diligently to further the CL method, resorts to a knowledge of exposures in his argument for a non-zero intercept term [8, p. 204]:

> From Equation 2 one can see that the slope factor b_n does not depend on the exposure (N) but only on the reporting pattern, and that the constant a_n is proportional to the exposure. An increase in exposure from one accident year to the next will cause an upward, parallel shift in the development regression line.

The extent to which his simulated regression results outperform the BF method may be due not only to the extra parameter a_n, but also to a BF-like use of exposure.[5]

It is time to introduce a method that gives exposure its proper place.

---

[5]In his Section 5 [8, p. 204], Murphy considers the model $\mathbf{Y} = \mathbf{E}\alpha + \mathrm{Prev}(\mathbf{Y})\beta + \mathbf{e}$, where $\mathbf{E}$ contains exposures for each row. It is unclear to the author whether he ever used this model in his simulations. Of course, if his simulated triangles had equal exposures in all rows, as did Stanard's, then the "J" and "E" models are equivalent. In the auto liability incurred loss and ALAE example (Figure 1A), exposures are obviously unequal, and the "J" model is used to produce estimates of $\mathbf{a}_{\mathrm{LSL}}$ = \$374 and $\mathbf{b}_{\mathrm{LSL}}$ = 2.027 for the 12 : 24 development [8, p. 190].

5. LOSS COVARIANCE VERSUS LOSS DEVELOPMENT

There is a distinction between loss covariance and loss development. To an actuary, loss development connotes the estimation of a loss as of time t_{i+1}, $X(t_{i+1})$, from the loss as of earlier times. In other words, there is some estimation function, f, such that $X(t_{i+1}) = f(X(t_1), X(t_2), \ldots, X(t_{i-1}), X(t_i)) + e_{i+1}$. Actuaries also simplify the functional form to $X(t_{i+1}) = f(X(t_i)) + e_{i+1}$. This simplification assumes that the most recent value of the loss is all-determinative of its future development; i.e., that the path the loss took in getting to $X(t_i)$ is irrelevant. Thomas Mack [7, p. 108] points out that this simplification may be inappropriate; however, without it, the functional forms could easily become overspecified. In any case, $X(t_{i+1}) = f(X(t_1), X(t_2), \ldots, X(t_{i-1}), X(t_i)) + e_{i+1}$ expresses the familiar and appealing concept of loss development. It is appealing because actuaries feel that earlier values of X should affect the later values. However, as was pointed out in Section 3, this entails estimation with stochastic regressors.

Loss covariance involves the following idea: Let $\mathbf{X}$ be an $(n \times 1)$ vector,

$$\mathbf{X} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix},$$

where x_i is the incremental loss, whether paid or incurred, during the ith time interval. Through research, we believe that we have a good idea of the mean and variance of $\mathbf{X}$, which depends on our knowledge of exposure, inflation, etc. Then, as x_is become known, the x_is still unknown can be considered elements of a conditional random vector. They are affected by the known elements in a Bayesian sense, through the variance matrix.

As an example, consider a two-part loss

$$\mathbf{X} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \sim N\left(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix}\right),$$

which means that $\mathbf{X}$ is distributed as a bivariate normal random variable with the mu vector as its mean and the sigma matrix as its variance. The variance matrix must be symmetric, because $\sigma_{ij} = \mathrm{Cov}(x_i,x_j) = \mathrm{Cov}(x_j,x_i) = \sigma_{ji}$. Hence, $\sigma_{21} = \sigma_{12}$. The distribution of x_2 conditional on x_1 is:[6]

$$x_2 \mid x_1 \sim N\left(\mu_2 + \frac{\sigma_{12}}{\sigma_{11}}(x_1 - \mu_1), \sigma_{22} - \frac{\sigma_{12}^2}{\sigma_{11}}\right).$$

Knowledge of x_1 affects our expectation of x_2, and even lessens the variance of x_2. Since $\sigma_{11} = \sigma_1^2$ and $\sigma_{12} = \rho\sigma_1\sigma_2$, we can rewrite the conditional expectation as:

$$\mathrm{E}[x_2 \mid x_1] = \mu_2 + \rho\sigma_2\left(\frac{x_1 - \mu_1}{\sigma_1}\right).$$

Thus,

$$\frac{\mathrm{E}[x_2 \mid x_1] - \mu_2}{\sigma_2} = \rho\left(\frac{x_1 - \mu_1}{\sigma_1}\right).$$

Consider what this means: the conditional mean $\mathrm{E}[x_2 \mid x_1]$ will differ from the unconditional mean μ_2 in terms of standard deviation units (σ_2) by some proportion (ρ) of the standardized error of x_1. This is the essence of the loss covariance approach: that the known losses affect the unknown not through their absolute levels, but rather through a combination both of their relative departure from their expected values and of the covariance of the known with the unknown. The covariance defines the persistency of this departure.

---

[6] For a general proof, see Appendix B. Johnson [4, p. 138] has another proof. The author's thinking was helped by Julien McKee's presentation at the 1994 Casualty Loss Reserve Seminar [6].

We do not yet have a linear model for loss prediction; nevertheless, we have uncovered a fundamental truth, a truth that has been missed in loss development methodology. Covariance is the link from one random variable to another. When actuaries seek to predict losses, they must consider how the known losses affect the unknown—and that involves covariance. Actuaries do not have to know how, or even whether, the known "causes" the unknown; the question of causality is academic. Prediction is a matter of covariance, which informs how one random variable is expected to differ from its mean, given the departure of another random variable from its mean. So covariance requires the estimation of means, which are functions of exogenous, known quantities such as exposures and price indices.

Let us solve for α and β in the equation $Y = \alpha + \beta X + e$, where $\mathrm{Cov}[X,e] = 0$, and $\mathrm{E}[e] = 0$. Applying the expectation operator, we have $\mathrm{E}[Y] = \alpha + \beta \mathrm{E}[X]$. Also, $\mathrm{Cov}[Y,X] = \beta \mathrm{Var}[X]$. Therefore, $\beta = \mathrm{Cov}[Y,X]/\mathrm{Var}[X]$, and $\alpha = \mathrm{E}[Y] - \beta \mathrm{E}[X]$. Moreover, $Y - \mathrm{E}[Y] = \beta(X - \mathrm{E}[X]) + e = \{\mathrm{Cov}[Y,X]/\mathrm{Var}[X]\}(X - \mathrm{E}[X]) + e$. What then is the function of the intercept α? Is it not to supply the proper combination of the mean values of the dependent and independent random variables? But the real relation is between the departures of the random variables from their means. In a well-constructed linear model, the intercept is replaced with expressions for mean values based on outside information. This amplifies the reason for the abolition of intercept terms advanced by Gregory Alff [1, p. 89], that "a constant does nothing to describe the underlying contributory causes of change in the dependent variable."

Though not a model, the example above encompasses the three approaches to loss reserving discussed in Section 1. Let us also treat σ_1 and σ_2 as equal. When ρ is positive, a greater than expected x_1 will raise $\mathrm{E}[x_2 \mid x_1]$. This is in keeping with the first approach, the CL method. When ρ is negative, a greater than expected x_1 will lower the conditional expectation. At the extreme, when $\rho = -1$, $\mathrm{E}[x_2 \mid x_1] = \mu_2 - (x_1 - \mu_1)$. In this case,

$x_1 + \mathrm{E}[x_2 \mid x_1] = \mu_1 + \mu_2$. All this keeps with the second approach, that of sticking to the prior hypothesis. When ρ is zero, $\mathrm{E}[x_2 \mid x_1]$ is unaffected, which keeps with the third approach, the BF method. A theory becomes very attractive when it unifies partial explanations. Such is the case with loss covariance. CL, prior hypothesis, or BF—which to choose? The answer will lie in a continuum dependent on the variance matrix of the incremental losses.[7]

6. A LINEAR MODEL OF LOSS COVARIANCE

The idea of loss covariance introduced in the previous section needs to be expanded before we consider a real-life example. Actuaries typically attempt to fill in a "loss rectangle" when all that is known is a triangular portion. The usual case is to have n observations of the earliest accident (or policy) time period, $n-1$ of the next, and so on until the latest time period, for which there is one observation. So there are $1+2+\cdots+n = n(n+1)/2$ known cells, and $n(n-1)/2$ unknown cells in the $(n \times n)$ rectangle. In this discussion we are not concerned about extrapolating beyond the nth interval.[8] The (ij)th cell in the rectangle will

[7] Two more points in closing this section: A suitable variance matrix can make a conditional mean dependent on more than just the latest known loss, thus recognizing Thomas Mack's caveat mentioned earlier in this section. And second, Stanard's fourth model, the additive, which he claims to have performed best in his simulations [10, pp. 131, 135], is the method closest to the covariance method.

[8] So too Murphy: The model does not attempt to predict "beyond the triangle" [8, p. 205]. At this writing, the author is expecting the publication of a paper, "Statistical and Financial Aspects of Self-Insurance Funding," in the 1996 Discussion Paper Program. In Section 3 of that paper the author estimates losses from the 84th month (seventh report) to ultimate. Since the risk treated there had no loss history beyond 84 months, bureau data was invoked, according to which ninety percent of the losses were paid by the 84th month. One might interpret this to mean that there is a development factor from 84th to ultimate of $1.00/0.90 \approx 1.111$, and that the CL method with its bias resurfaces. Even if this were true, at least the use of the CL method would be restricted to a hopefully small role. However, in the paper just mentioned, cumulative predictions as of 84 months were not multiplied by 1.111. Rather, the pure premium for payments up to 84 months, for which an estimate had been derived, was divided by nine to arrive at an estimate of the pure premium for payments after 84 months. The payments after 84 months were then predicted as the product of exposures and the latter pure premium. Assumptions

contain Y_{ij}, the incremental loss of the ith accident period during the jth interval from the beginning of that accident period. The (ij) subscript is a link to much information about the distribution of Y_{ij}; e.g., information about the premium or exposure in the ith period, or inflation trends in absolute time (which is represented by $i + j$).

Now imagine the transpose of the ith row of the rectangle. This is an $(n \times 1)$ vector, the first $(n + 1 - i)$ elements of which are known. Take the known elements of each vector, stack them into an $(n[n + 1]/2 \times 1)$ vector, and call it $\mathbf{Y}_1$. Similarly, stack the unknown elements into an $(n[n - 1]/2 \times 1)$ vector, and call it $\mathbf{Y}_2$. Finally, stack $\mathbf{Y}_1$ on top of $\mathbf{Y}_2$, creating the partitioned $(n^2 \times 1)$ vector $\mathbf{Y}$. Each element of this $\mathbf{Y}$ was originally some Y_{ij} in the rectangle.

We can form the linear model $\mathbf{Y}_{(t\times 1)} = \mathbf{X}_{(t\times k)}\boldsymbol{\beta}_{(k\times 1)} + \mathbf{e}_{(t\times 1)}$, where $t = n^2$ and $\mathrm{Var}[\mathbf{e}] = \Sigma_{(t\times t)}$. $\mathbf{X}$ is the design matrix, each row of which contains pertinent information affixed to the (ij)th location implicit in the same row of $\mathbf{Y}$. The variance, Σ, determines how errors will influence one another. There is no reason why there cannot be correlations between errors of different accident periods (e.g., calendar-year effects), although it will not be considered in the following example. Σ has to be estimated with a minimum of parameters, so it is best to start with only correlation within accident periods.

The objectives are to estimate $\boldsymbol{\beta}$ and to predict the mean and the variance of $\mathbf{Y}_2$ conditional on $\mathbf{Y}_1$, which are done by the method of generalized least squares. The formulas for these objectives are derived in Appendix C. The model outlined here and treated in Appendix C is more general than the idea of the previous section in that (1) it provides for the estimation of unknown parameters, (2) it does not require that error terms be

---

were specified as to the covariance of these payments with payments prior to 84 months, so that the payments to ultimate could be affected by the departures of the observations from their predicted values.

normally distributed, and (3) it allows for correlation between, as well as within, accident periods.

7. AN EXAMPLE

Exhibit 1 shows paid workers compensation indemnity losses for eight accident quarters at quarterly evaluations. The numbers above and to the left of the dotted line are actual observations; those below and to the right are projections based on the loss development factors in the bottom row. The development factors are weighted-averages between matched columns; e.g., 1.114 = (756,879 + 2,327,141)/(701,411 + 2,067,233). This is a typical example of the chain ladder method. Notice that the total penultimate (at 24 months) loss, 45,377,646, is obtained without any knowledge of exposures.

Exhibit 2 is an example of Stanard's additive model [10, p. 131]. Incremental losses are related to an exposure base, which in this case is on-level premium. For example, based on two observations, between eighteen and twenty-one months the indemnity payout of an accident quarter will be 0.69% of premium, (55,468 + 259,908)/(11,631,592 + 33,995,192). The incremental payments below and to the right of the dotted line can be projected, and a cumulative table can be constructed. The ultimate losses of the additive model are lower than those of the CL method for every accident quarter.

We will use the linear model $\mathbf{Y}_1 = \mathbf{X}_1\beta + \mathbf{e}_1$, where $\mathbf{Y}_1$ and $\mathbf{X}_1$ are shown in Exhibit 3. We will assume, as is frequently done, that the variance of an observation is proportional to its exposure [Venter 12, p. 445]. The values in the column entitled "Scale $\mathbf{A}_1$" are the square roots of the respective premiums in $\mathbf{X}_1$. If we diagonalize Scale $\mathbf{A}_1$ and call it $\mathbf{\Lambda}_1$, then $\mathrm{Var}[\mathbf{e}_1] = \sigma^2\mathbf{\Lambda}_1^2 = \sigma^2\mathbf{\Psi}_{11}$, where $\mathbf{\Psi}_{11}$ is defined as $\mathbf{\Lambda}_1^2$. The generalized least-squares estimator for β, $(\mathbf{X}_1'\mathbf{\Psi}_{11}^{-1}\mathbf{X}_1)^{-1}(\mathbf{X}_1'\mathbf{\Psi}_{11}^{-1}\mathbf{Y}_1)$, turns out to be:

$$\hat{\beta} = \begin{bmatrix} 0.0099 \\ 0.0196 \\ 0.0142 \\ 0.0123 \\ 0.0108 \\ 0.0096 \\ 0.0069 \\ 0.0061 \end{bmatrix}.$$

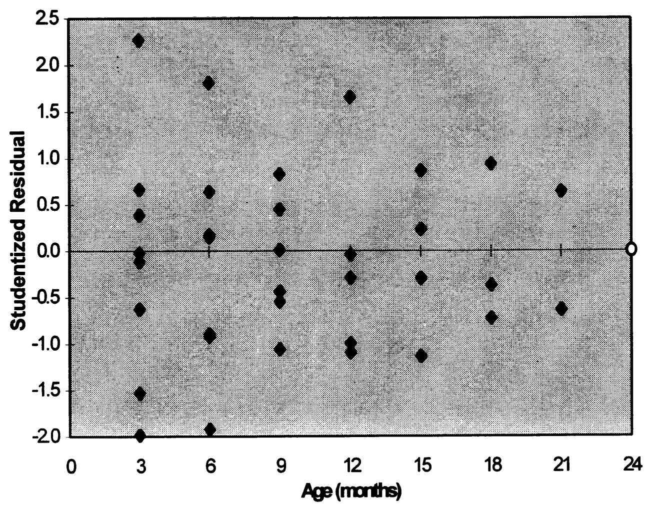
It should come as no surprise that these are the same coefficients as were obtained in Exhibit 2. The linear model with such a proportional variance produces weighted averages [7, pp. 111f]. The formula for the sample variance [5, p. 332] is: $\hat{\sigma}^2 = (\mathbf{Y} - \hat{\mathbf{Y}})'\Psi^{-1}(\mathbf{Y} - \hat{\mathbf{Y}})/(36 - 8)$, which is 176.3242.

Exhibit 4 contains results from the regression. $\mathbf{Y}$ is the same as $\mathbf{Y}_1$ in Exhibit 3; $\hat{\mathbf{Y}}$ is the fitted vector, or $\mathbf{X}_1\hat{\beta}$; and $\hat{\mathbf{e}} = \mathbf{Y} - \hat{\mathbf{Y}}$. Appendix D derives the formula for Var[$\hat{\mathbf{e}}$]. The square roots of the diagonal elements of this matrix are contained in the column Std($\hat{\mathbf{e}}$). $\hat{\mathbf{e}}$ divided by these numbers forms the column Student($\hat{\mathbf{e}}$). If the model is homoskedastic, these studentized residuals should show no increase or decrease by accident quarter age. However, it appears from the graph in Figure 1 that the studentized residuals decrease by age. This is obvious from the following table of sample variances of the studentized residuals by age:

Age	Count	Variance	Y = Ln(Var)	$\hat{Y}$	exp($\hat{Y}$)
3	8	1.759	0.565	0.297	1.345
6	7	1.495	0.402	0.181	1.198
9	6	0.482	−0.729	0.065	1.067
12	5	1.226	0.204	−0.051	0.950
15	4	0.719	−0.329	−0.167	0.846
18	3	0.767	−0.265	−0.283	0.753
21	2	0.813	−0.206	−0.399	0.671

FIGURE 1

ACCIDENT QUARTER RESIDUALS

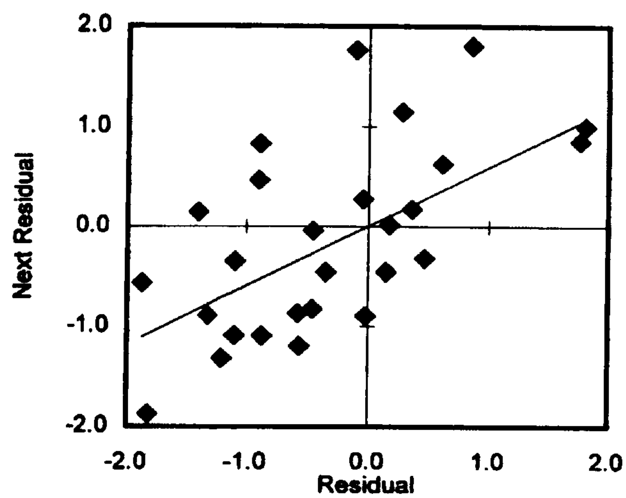


The table also shows that an exponential regression explains well the tapering off of the variance. Moreover, the variance can be predicted for age 24 months, which is 0.597. We now can remodel the variance of an observation as proportional not only to the premium, but also to the fitted or predicted sample variances. The square roots of these new variances are found in the column Scale B of Exhibit 3. For example, for AQ.Age 1.03, the standard deviation of the error is proportional to the square root of the product of 11,631,592 and 1.345, or 3,955.31.

The regression reweighted with Scale $\mathbf{B}_1$ diagonalized as $\mathbf{\Lambda}_1$, produces new standardized residuals that are homoskedastic. The estimate for β changes negligibly (no change within the first ten decimal places). The results of this regression are not shown; however, Figure 2 contains some of the studentized residuals. This exhibit shows that there is a relation between one studentized residual and the next; viz., that the next studentized resid-

FIGURE 2

CORRELATION OF RESIDUALS



ual tends to be 59.31% of the previous. Since we are dealing with studentized, homoskedastic residuals, whose variances should all be unity, the slope coefficient $\hat{\rho}$ should be a correlation coefficient [5, pp. 391f.].

Thus, we will use as our final model one whose error variance matrix is first-order autocorrelated within accident quarters. Exhibit 6 shows partitions of the correlation matrix $\mathbf{P}$, where ρ has been estimated to be 0.5931. It is not necessary to show $\mathbf{P}_{12}$, since it is the transpose of $\mathbf{P}_{21}$. For an explanation as to how first-order correlation produces correlation matrices such as these, see Judge [5, pp. 384–388]. Letting $\mathbf{\Lambda}_1$ and $\mathbf{\Lambda}_2$ be diagonalizations of Scale $\mathbf{B}_1$ and Scale $\mathbf{B}_2$ respectively, we can express the error variance matrix as:

$$\mathbf{\Sigma} = \begin{bmatrix} \mathbf{\Sigma}_{11} & \mathbf{\Sigma}_{12} \\ \mathbf{\Sigma}_{21} & \mathbf{\Sigma}_{22} \end{bmatrix} = \sigma^2 \begin{bmatrix} \mathbf{\Lambda}_1 & \\ & \mathbf{\Lambda}_2 \end{bmatrix} \begin{bmatrix} \mathbf{P}_{11} & \mathbf{P}_{12} \\ \mathbf{P}_{21} & \mathbf{P}_{22} \end{bmatrix} \begin{bmatrix} \mathbf{\Lambda}_1 & \\ & \mathbf{\Lambda}_2 \end{bmatrix}$$

$$= \sigma^2 \begin{bmatrix} \mathbf{\Lambda}_1 \mathbf{P}_{11} \mathbf{\Lambda}_1 & \mathbf{\Lambda}_1 \mathbf{P}_{12} \mathbf{\Lambda}_2 \\ \mathbf{\Lambda}_2 \mathbf{P}_{21} \mathbf{\Lambda}_1 & \mathbf{\Lambda}_2 \mathbf{P}_{22} \mathbf{\Lambda}_2 \end{bmatrix} = \sigma^2 \begin{bmatrix} \mathbf{\Psi}_{11} & \mathbf{\Psi}_{12} \\ \mathbf{\Psi}_{21} & \mathbf{\Psi}_{22} \end{bmatrix}.$$

Therefore, we can estimate β as $(\mathbf{X}_1'\Psi_{11}^{-1}\mathbf{X}_1)^{-1}(\mathbf{X}_1'\Psi_{11}^{-1}\mathbf{Y}_1)$, or:

$$\hat{\beta} = \begin{bmatrix} 0.0099 \\ 0.0199 \\ 0.0145 \\ 0.0125 \\ 0.0108 \\ 0.0100 \\ 0.0079 \\ 0.0078 \end{bmatrix}.$$

Also, the estimate for $\hat{\sigma}^2 = (\mathbf{Y}_1 - \mathbf{X}_1\hat{\beta})'\Psi_{11}^{-1}(\mathbf{Y}_1 - \mathbf{X}_1\hat{\beta})/(36 - 8 - 2)$ is 149.9509. The denominator has two less degrees of freedom because two parameters were estimated in creating the correlation matrix; viz., the decay factor in the exponentially fitted variances and the correlation coefficient ρ. The predicted values of $\mathbf{Y}_2$ are calculated according to the formula derived in Appendix C: $E[\mathbf{Y}_2 \mid \mathbf{Y}_1] = \mathbf{X}_2\hat{\beta} + \Sigma_{21}\Sigma_{11}^{-1}(\mathbf{Y}_1 - \mathbf{X}_1\hat{\beta})$.

Exhibit 7 contains selected values from this final regression. The observed and predicted Ys are carried over to the incremental table of Exhibit 8. The cumulative table follows, and shows at age 24 the expected values of quarters 2 through 8. It can be seen that the estimates at age 24 for this method are higher than those of the additive method (Exhibit 2). They are lower than those of the CL method (Exhibit 1), except for quarters 2 and 3 (and even here the losses are only about 0.1 percent higher). Thus it seems that this "covariance method" mediates between a BF variant and the CL method.

Exhibit 9 is like Exhibit 8 except that it displays the $\mathbf{X}_1\hat{\beta}$ and $\mathbf{X}_2\hat{\beta}$ columns of Exhibit 7. These are predictions of $\mathbf{Y}_1$ and $\mathbf{Y}_2$ prior to any observation (of course, we needed observations in order to obtain $\hat{\beta}$). Exhibit 9 helps us to see that the covariance is working in Exhibit 8. For example, the prediction of incremental AQ.Age 2.24 is 261,487. Ignoring observations of 2.03 through 2.21, the prediction would have been 266,326. Why is the a posteriori prediction less than the a priori? It is because the actual 2.21 observation of 259,908 is less than the a priori prediction of 268,555. So the covariance is carrying over to the prediction. AQ 8 is observed to commence with a payment higher than expected, and this excess is perpetuated in forecasts 8.06 through 8.24. However, the excess dampens over time, as expected.

A better but more complicated model would recognize a trend in the observed payments. By comparing Exhibits 2 and 9, one can see that the model tends to overestimate the payments of the first three accident quarters, and to underestimate those of the last five. So perhaps it is no surprise that for quarters 2 and 3 the model gives higher results than does the CL method. Building trend into the model, by applying some sort of inflation index to the exposures would probably lessen the estimate of ρ, and make better use of the error variance matrix. As it is now, it seems that the variance matrix is trying to chase the trend, as well as to capture covariance.

It should be noticed that the column totals of Exhibit 8 are identical to those of Exhibit 9. The author did not expect this, and checked the programming for errors (the work was done both on an Excel spreadsheet and in a SAS program,[9] and the results were the same). It is also of interest that the column rates are identical to the estimate of $\hat{\beta}$. The author thinks of these

[9] After the body of this paper was written, the author learned of a procedure in SAS, Proc Mixed, which has the ability to estimate simultaneously, by the method of maximum likelihood, both the regression coefficients and any parameters in the variance matrix. SAS users who will be using generalized least squares would do well to study the following SAS publications on Proc Mixed: SAS Institute Inc., SAS® Technical Report P-229, *SAS/STAT® Software: Changes and Enhancements, Release 6.07*, Cary, NC: SAS Institute Inc., 1992, ch. 16, and SAS Institute Inc., Introduction to the MIXED Procedure Course Notes, Cary, NC: SAS Institute Inc., 1995.

At the end of Chapter 16 of the Technical Report is an extensive bibliography of the literature devoted to the subject of mixed models, of which generalized least squares is a subset.

somewhat appealing qualities as "balance properties." Appendix E gives a demonstration of these properties, a demonstration that relies on the peculiarities of this example and so cannot be generalized.

Another interesting property, perhaps related to the column balance just described, is that within an accident quarter, the predictions depend only on the last observation of that quarter. Recall the prediction formula: $E[\mathbf{Y}_2 \mid \mathbf{Y}_1] = \mathbf{X}_2\hat{\beta} + \Sigma_{21}\Sigma_{11}^{-1}(\mathbf{Y}_1 - \mathbf{X}_1\hat{\beta})$. The matrix $\Sigma_{21}\Sigma_{11}^{-1}$ is zero only where $\mathbf{P}_{21}\mathbf{P}_{11}^{-1}$ is zero, since the Σs are $\mathbf{P}$s times diagonal scaling matrices. Exhibit 6 shows $\mathbf{P}_{21}\mathbf{P}_{11}^{-1}$. (For details about the inverse of a first-order autocorrelation matrix see Judge [5, p. 389].) From this it can be seen that the prediction adjustment of any AQ.Age is proportional to the proper power of ρ times the error of the last observation for that AQ. The errors of earlier ages, though correlated with the predictions, in the first-order autocorrelation model are impounded within, or built into, the error of the latest observation. This has a bearing on Mack's remark about path dependence [7, p. 108], discussed earlier in Section 5. If there is path independence in the first-order autocorrelation model, which is the most basic of generalized linear models, then perhaps actuaries have not been too remiss in developing losses from the last observation only.

The last column of the cumulative table of Exhibit 8 contains the standard deviations of the cumulative paid predictions at 24 months. Appendix C derives the formula for the variance of the predictions:

$$\begin{aligned}\mathrm{Var}[\mathbf{Y}_2 \mid \mathbf{Y}_1] = {} & (\mathbf{X}_2 - \Sigma_{21}\Sigma_{11}^{-1}\mathbf{X}_1)\mathrm{Var}[\hat{\beta}](\mathbf{X}_2 - \Sigma_{21}\Sigma_{11}^{-1}\mathbf{X}_1)' \\ & + (\Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}).\end{aligned}$$

This is a (28×28) matrix, too large to print, but whose row and column headings would be the same as those of $\mathbf{P}_{22}$ in Exhibit 6. The variance of the prediction of AQ 4 at 24 months, for example, is the variance of the sum of the predictions of 4.18, 4.21, and 4.24. This would be the sum of the nine variances and covariances which occupy the square whose diagonal is from (4.18,4.18) to (4.24,4.24). The square root of this number, 293,083, is found in the last column of Exhibit 8. The Total row contains the square root of the sum of all 784 elements of $\mathrm{Var}[\mathbf{Y}_2 \mid \mathbf{Y}_1]$, which is the standard deviation of the sum of all the predictions.

If the errors were normally distributed, then the total predicted would be t-distributed with twenty-six degrees of freedom, with a mean of 41,778,516 and a standard deviation of 1,598,047. So, for example, the 95% upper bound for the total predicted would be $41{,}778{,}516 + 1{,}598{,}047 \times 1.706 = 44{,}504{,}784$. However, as stated in the appendices, the errors need not be normally distributed. We could just as easily assume that the total predicted is lognormal (41,778,516, 1,598,047). This is equivalent to $e^{N[\mu=17.54716,\sigma^2=0.001462]}$. The normal random variable has a 95th percentile at $17.54716 + (0.001462)^{1/2} \times 1.645 = 17.61006$. Therefore, the 95% upper bound with a lognormal distribution is 44,457,985.

8. CONCLUSION

Generalized least squares is a better method of loss prediction than the chain ladder method and the other loss development methods. Even when linear models are imposed on loss development methods, they incorporate stochastic regressors, and the estimates are not guaranteed to be either best or unbiased. The confidence intervals derived therefrom are not trustworthy. The fault lies in trying to make the level of one variable affect the level of the next, whereas the statistical idea is that the departure of one variable from its mean affects the departure of the next from its mean. This is the idea of covariance, and it is accommodated in the general linear model and generalized least squares estimation. It may not be easy to determine a good structure for the error variance matrix; but then again, the prediction of losses in itself is no easy feat.

REFERENCES

[1] Alff, Gregory N., "A Note Regarding Evaluation of Multiple Regression Models," *PCAS* LXXI, 1984, pp. 84–95.

[2] Bornhuetter, R. L. and R. E. Ferguson, "The Actuary and IBNR," *PCAS* LIX, 1972, pp. 181–195.

[3] Healy, M. J. R., *Matrices for Statistics*, Oxford, Clarendon Press, 1986.

[4] Johnson, R. A. and D. W. Wichern, *Applied Multivariate Statistical Analysis*, Third Edition, Englewood Cliffs, NJ, Prentice Hall, 1992.

[5] Judge, G. G., R. C. Hill, W. E. Griffiths, H. Lütkepohl and T.-C. Lee, *Introduction to the Theory and Practice of Econometrics*, Second Edition, New York, John Wiley & Sons, 1988.

[6] McKee, J. D., "Traditional Regression Methods in Loss Reserving," unpublished presentation, 1994 Casualty Loss Reserve Seminar.

[7] Mack, T., "Measuring the Variability of Chain Ladder Reserve Estimates," *Casualty Actuarial Society Forum*, Spring 1994, pp. 101–182.

[8] Murphy, D. M., "Unbiased Loss Development Factors," *PCAS* LXXXI, 1994, pp. 154–222.

[9] Patrik, G. S., "Reinsurance," *Foundations of Casualty Actuarial Science*, Casualty Actuarial Society, 1990, pp. 277–374.

[10] Stanard, J. N., "A Simulation Test of Prediction Errors of Loss Reserve Estimation Techniques," *PCAS* LXXII, 1985, pp. 124–148.

[11] Robertson, J. P., Discussion of Stanard: "A Simulation Test of Prediction Errors of Loss Reserve Estimation Techniques," *PCAS* LXXII, 1985, pp. 149–153.

[12] Venter, G. G., "Credibility," *Foundations of Casualty Actuarial Science*, Casualty Actuarial Society, 1990, pp. 375–483.

EXHIBIT 1

CHAIN LADDER METHOD

Cumulative Workers Compensation Indemnity Losses Paid

AQ	@3	@6	@9	@12	@15	@18	@21	@24
Qtr 1	87,248	275,126	393,511	492,620	605,497	701,411	756,879	827,621
Qtr 2	189,320	712,837	1,157,102	1,496,943	1,786,132	2,067,233	2,327,141	2,544,648
Qtr 3	392,599	1,457,421	2,182,953	2,840,352	3,440,079	4,017,704	4,475,360	4,893,652
Qtr 4	675,634	2,214,303	3,282,110	4,289,915	5,113,465	5,950,530	6,628,354	7,247,876
Qtr 5	720,152	2,052,734	3,134,652	4,004,034	4,805,445	5,592,088	6,229,082	6,811,286
Qtr 6	746,772	2,160,682	3,177,136	4,107,877	4,930,073	5,737,117	6,390,631	6,987,934
Qtr 7	769,063	2,218,298	3,331,899	4,307,979	5,170,225	6,016,581	6,701,929	7,328,328
Qtr 8	853,758	2,644,494	3,972,049	5,135,660	6,163,567	7,172,532	7,989,554	8,736,301
Total	4,434,546	13,735,895	20,631,412	26,675,379	32,014,483	37,255,195	41,498,932	45,377,646
Development Factor		3.097	1.502	1.293	1.200	1.164	1.114	1.093

EXHIBIT 2

ADDITIVE METHOD

Incremental Workers Compensation Indemnity Losses Paid

AQ	@3	@6	@9	@12	@15	@18	@21	@24	Premium
Qtr 1	87,248	187,878	118,385	99,109	112,877	95,914	55,468	70,742	11,631,592
Qtr 2	189,320	523,517	444,265	339,841	289,189	281,101	259,908	206,755	33,995,192
Qtr 3	392,599	1,064,822	725,532	657,399	599,727	577,625	371,000	326,440	53,674,098
Qtr 4	675,634	1,538,669	1,067,807	1,007,805	823,550	666,271	479,042	421,505	69,305,023
Qtr 5	720,152	1,332,582	1,081,918	869,382	788,639	700,314	503,518	443,042	72,846,114
Qtr 6	746,772	1,413,910	1,016,454	877,931	771,773	685,337	492,750	433,567	71,288,246
Qtr 7	769,063	1,449,235	1,006,824	870,551	765,286	679,577	488,608	429,922	70,689,000
Qtr 8	853,758	1,246,933	906,681	783,963	689,167	611,983	440,009	387,161	63,658,000
Total	4,434,546	8,757,546	6,367,866	5,505,980	4,840,208	4,298,123	3,090,303	2,719,133	447,087,265
Rate	0.0099	0.0196	0.0142	0.0123	0.0108	0.0096	0.0069	0.0061	

Cumulative Workers Compensation Indemnity Losses Paid

AQ	@3	@6	@9	@12	@15	@18	@21	@24
Qtr 1	87,248	275,126	393,511	492,620	605,497	701,411	756,879	827,621
Qtr 2	189,320	712,837	1,157,102	1,496,943	1,786,132	2,067,233	2,327,141	2,533,896
Qtr 3	392,599	1,457,421	2,182,953	2,840,352	3,440,079	4,017,704	4,388,704	4,715,143
Qtr 4	675,634	2,214,303	3,282,110	4,289,915	5,113,465	5,779,736	6,258,778	6,680,284
Qtr 5	720,152	2,052,734	3,134,652	4,004,034	4,792,673	5,492,987	5,996,505	6,439,547
Qtr 6	746,772	2,160,682	3,177,136	4,055,067	4,826,840	5,512,177	6,004,927	6,438,494
Qtr 7	769,063	2,218,298	3,225,122	4,095,673	4,860,958	5,540,535	6,029,143	6,459,065
Qtr 8	853,758	2,100,691	3,007,372	3,791,335	4,480,502	5,092,485	5,532,495	5,919,655
Total	4,434,546	13,192,092	19,559,958	25,065,938	29,906,146	34,204,269	37,294,572	40,013,705

EXHIBIT 3

PART 1
X_1, Y_1 MATRICES

AQ.Age	Y_1	X_1								Scale A_1	Scale B_1
1.03	87,248	11,631,592	0	0	0	0	0	0	0	3,410.51	3,955.81
1.06	187,878	0	11,631,592	0	0	0	0	0	0	3,410.51	3,732.86
1.09	118,385	0	0	11,631,592	0	0	0	0	0	3,410.51	3,522.47
1.12	99,109	0	0	0	11,631,592	0	0	0	0	3,410.51	3,323.94
1.15	112,877	0	0	0	0	11,631,592	0	0	0	3,410.51	3,136.59
1.18	95,914	0	0	0	0	0	11,631,592	0	0	3,410.51	2,959.81
1.21	55,468	0	0	0	0	0	0	11,631,592	0	3,410.51	2,792.99
1.24	70,742	0	0	0	0	0	0	0	11,631,592	3,410.51	2,635.58
2.03	189,320	33,995,192	0	0	0	0	0	0	0	5,830.54	6,762.77
2.06	523,517	0	33,995,192	0	0	0	0	0	0	5,830.54	6,381.61
2.09	444,265	0	0	33,995,192	0	0	0	0	0	5,830.54	6,021.94
2.12	339,841	0	0	0	33,995,192	0	0	0	0	5,830.54	5,682.53
2.15	289,189	0	0	0	0	33,995,192	0	0	0	5,830.54	5,362.26
2.18	281,101	0	0	0	0	0	33,995,192	0	0	5,830.54	5,060.03
2.21	259,908	0	0	0	0	0	0	33,995,192	0	5,830.54	4,774.84
3.03	392,599	53,674,098	0	0	0	0	0	0	0	7,326.26	8,497.64
3.06	1,064,822	0	53,674,098	0	0	0	0	0	0	7,326.26	8,018.70
3.09	725,532	0	0	53,674,098	0	0	0	0	0	7,326.26	7,566.76
3.12	657,399	0	0	0	53,674,098	0	0	0	0	7,326.26	7,140.28
3.15	599,727	0	0	0	0	53,674,098	0	0	0	7,326.26	6,737.85
3.18	577,625	0	0	0	0	0	53,674,098	0	0	7,326.26	6,358.09

EXHIBIT 3

PART 1—PAGE 2

AQ.Age	Y_1	X_1								Scale A_1	Scale B_1
4.03	675,634	69,305,023	0	0	0	0	0	0	0	8,324.96	9,656.02
4.06	1,538,669	0	69,305,023	0	0	0	0	0	0	8,324.96	9,111.79
4.09	1,067,807	0	0	69,305,023	0	0	0	0	0	8,324.96	8,598.24
4.12	1,007,805	0	0	0	69,305,023	0	0	0	0	8,324.96	8,113.63
4.15	823,550	0	0	0	0	69,305,023	0	0	0	8,324.96	7,656.34
5.03	720,152	72,846,114	0	0	0	0	0	0	0	8,534.99	9,899.63
5.06	1,332,582	0	72,846,114	0	0	0	0	0	0	8,534.99	9,341.68
5.09	1,081,918	0	0	72,846,114	0	0	0	0	0	8,534.99	8,815.17
5.12	869,382	0	0	0	72,846,114	0	0	0	0	8,534.99	8,318.33
6.03	746,772	71,288,246	0	0	0	0	0	0	0	8,443.24	9,793.20
6.06	1,413,910	0	71,288,246	0	0	0	0	0	0	8,443.24	9,241.25
6.09	1,016,454	0	0	71,288,246	0	0	0	0	0	8,443.24	8,720.40
7.03	769,063	70,689,000	0	0	0	0	0	0	0	8,407.68	9,751.96
7.06	1,449,235	0	70,689,000	0	0	0	0	0	0	8,407.68	9,202.32
8.03	853,758	63,658,000	0	0	0	0	0	0	0	7,978.60	9,254.27

EXHIBIT 3

PART 2

X_2, Y_2 MATRICES

AQ.Age	Y_2	X_2							Scale A_2	Scale B_2
2.24		0	0	0	0	0	0	033,995,192	5,830.54	4,505.73
3.21		0	0	0	0	0	053,674,098	0	7,326.26	5,999.74
3.24		0	0	0	0	0	0	053,674,098	7,326.26	5,661.59
4.18		0	0	0	0	069,305,023	0	0	8,324.96	7,224.82
4.21		0	0	0	0	0	069,305,023	0	8,324.96	6,817.62
4.24		0	0	0	0	0	0	069,305,023	8,324.96	6,433.37
5.15		0	0	0	072,846,114	0	0	0	8,534.99	7,849.50
5.18		0	0	0	0	072,846,114	0	0	8,534.99	7,407.09
5.21		0	0	0	0	0	072,846,114	0	8,534.99	6,989.62
5.24		0	0	0	0	0	0	072,846,114	8,534.99	6,595.68
6.12		0	0	071,288,246	0	0	0	0	8,443.24	8,228.90
6.15		0	0	0	071,288,246	0	0	0	8,443.24	7,765.11
6.18		0	0	0	0	071,288,246	0	0	8,443.24	7,327.46
6.21		0	0	0	0	0	071,288,246	0	8,443.24	6,914.48
6.24		0	0	0	0	0	0	071,288,246	8,443.24	6,524.77

EXHIBIT 3

PART 2—PAGE 2

AQ.Age	Y_2	X_2								Scale A_2	Scale B_2
7.09		0	0	70,689,000	0	0	0	0	0	8,407.68	8,683.67
7.12		0	0	0	70,689,000	0	0	0	0	8,407.68	8,194.25
7.15		0	0	0	0	70,689,000	0	0	0	8,407.68	7,732.41
7.18		0	0	0	0	0	70,689,000	0	0	8,407.68	7,296.60
7.21		0	0	0	0	0	0	70,689,000	0	8,407.68	6,885.35
7.24		0	0	0	0	0	0	0	70,689,000	8,407.68	6,497.29
8.06		0	63,658,000	0	0	0	0	0	0	7,978.60	8,732.69
8.09		0	0	63,658,000	0	0	0	0	0	7,978.60	8,240.50
8.12		0	0	0	63,658,000	0	0	0	0	7,978.60	7,776.06
8.15		0	0	0	0	63,658,000	0	0	0	7,978.60	7,337.79
8.18		0	0	0	0	0	63,658,000	0	0	7,978.60	6,924.22
8.21		0	0	0	0	0	0	63,658,000	0	7,978.60	6,533.97
8.24		0	0	0	0	0	0	0	63,658,000	7,978.60	6,165.70

EXHIBIT 4

FIRST REGRESSION

AQ	Age	Y	$\hat{Y}$	$\hat{e}$	Std($\hat{e}$)	Student($\hat{e}$)
Qtr 1	3	87,248	115,371	−28,123	44,694	−0.629
Qtr 1	6	187,878	227,840	−39,962	44,595	−0.896
Qtr 1	9	118,385	165,669	−47,284	44,437	−1.064
Qtr 1	12	99,109	143,246	−44,137	44,183	−0.999
Qtr 1	15	112,877	125,925	−13,048	43,697	−0.299
Qtr 1	18	95,914	111,822	−15,908	42,552	−0.374
Qtr 1	21	55,468	80,398	−24,930	39,091	−0.638
Qtr 1	24	70,742	70,742	0	0	0
Qtr 2	3	189,320	337,190	−147,870	74,420	−1.987
Qtr 2	6	523,517	665,898	−142,381	73,910	−1.926
Qtr 2	9	444,265	484,194	−39,929	73,093	−0.546
Qtr 2	12	339,841	418,658	−78,817	71,765	−1.098
Qtr 2	15	289,189	368,035	−78,846	69,178	−1.140
Qtr 2	18	281,101	326,817	−45,716	62,786	−0.728
Qtr 2	21	259,908	234,978	24,930	39,091	0.638
Qtr 3	3	392,599	532,380	−139,781	91,257	−1.532
Qtr 3	6	1,064,822	1,051,368	13,454	90,218	0.149
Qtr 3	9	725,532	764,480	−38,948	88,543	−0.440
Qtr 3	12	657,399	661,009	−3,610	85,792	−0.042
Qtr 3	15	599,727	581,081	18,646	80,320	0.232
Qtr 3	18	577,625	516,002	61,623	65,943	0.934
Qtr 4	3	675,634	687,419	−11,785	101,616	−0.116
Qtr 4	6	1,538,669	1,357,547	181,122	100,057	1.810
Qtr 4	9	1,067,807	987,112	80,695	97,530	0.827
Qtr 4	12	1,007,805	853,507	154,298	93,341	1.653
Qtr 4	15	823,550	750,303	73,247	84,836	0.863
Qtr 5	3	720,152	722,542	−2,390	103,690	−0.023
Qtr 5	6	1,332,582	1,426,910	−94,328	102,001	−0.925
Qtr 5	9	1,081,918	1,037,548	44,370	99,261	0.447
Qtr 5	12	869,382	897,116	−27,734	94,707	−0.293
Qtr 6	3	746,772	707,090	39,682	102,789	0.386
Qtr 6	6	1,413,910	1,396,394	17,516	101,157	0.173
Qtr 6	9	1,016,454	1,015,359	1,095	98,512	0.011
Qtr 7	3	769,063	701,146	67,917	102,438	0.663
Qtr 7	6	1,449,235	1,384,656	64,579	100,828	0.640
Qtr 8	3	853,758	631,408	222,350	98,114	2.266

EXHIBIT 5

AUTOCORRELATION (ρ)

AQ	Age	X = Student($\hat{e}$)	Y = Next($\hat{e}$)	$\hat{Y} = X\hat{\rho}$
Qtr 1	3	−0.577	−0.871	−0.342
Qtr 1	6	−0.871	−1.097	−0.517
Qtr 1	9	−1.097	−1.091	−0.650
Qtr 1	12	−1.091	−0.346	−0.647
Qtr 1	15	−0.346	−0.458	−0.205
Qtr 1	18	−0.458	−0.829	−0.272
Qtr 2	3	−1.823	−1.873	−1.081
Qtr 2	6	−1.873	−0.563	−1.111
Qtr 2	9	−0.563	−1.199	−0.334
Qtr 2	12	−1.199	−1.319	−0.711
Qtr 2	15	−1.319	−0.893	−0.782
Qtr 2	18	−0.893	0.829	−0.530
Qtr 3	3	−1.406	0.145	−0.834
Qtr 3	6	0.145	−0.453	0.086
Qtr 3	9	−0.453	−0.046	−0.269
Qtr 3	12	−0.046	0.269	−0.027
Qtr 3	15	0.269	1.146	0.159
Qtr 4	3	−0.106	1.760	−0.063
Qtr 4	6	1.760	0.853	1.044
Qtr 4	9	0.853	1.805	0.506
Qtr 4	12	1.805	0.999	1.071
Qtr 5	3	−0.021	−0.899	−0.013
Qtr 5	6	−0.899	0.461	−0.533
Qtr 5	9	0.461	−0.320	0.273
Qtr 6	3	0.354	0.168	0.210
Qtr 6	6	0.168	0.011	0.100
Qtr 7	3	0.608	0.623	0.361

EXHIBIT 6

PART 1

P_{11} CORRELATION MATRICES

P_{11}	1.03	1.06	1.09	1.12	1.15	1.18	1.21	1.24	2.03	2.06	2.09	2.12	2.15	2.18	2.21	3.03	3.06	3.09	3.12	3.15	3.18	4.03	4.06	4.09	4.12	4.15	5.03	5.06	5.09	5.12	6.03	6.06	6.09	7.03	7.06	8.03
1.03	ρ^0	ρ^1	ρ^2	ρ^3	ρ^4	ρ^5	ρ^6	ρ^7																												
1.06	ρ^1	ρ^0	ρ^1	ρ^2	ρ^3	ρ^4	ρ^5	ρ^6																												
1.09	ρ^2	ρ^1	ρ^0	ρ^1	ρ^2	ρ^3	ρ^4	ρ^5																												
1.12	ρ^3	ρ^2	ρ^1	ρ^0	ρ^1	ρ^2	ρ^3	ρ^4																												
1.15	ρ^4	ρ^3	ρ^2	ρ^1	ρ^0	ρ^1	ρ^2	ρ^3																												
1.18	ρ^5	ρ^4	ρ^3	ρ^2	ρ^1	ρ^0	ρ^1	ρ^2																												
1.21	ρ^6	ρ^5	ρ^4	ρ^3	ρ^2	ρ^1	ρ^0	ρ^1																												
1.24	ρ^7	ρ^6	ρ^5	ρ^4	ρ^3	ρ^2	ρ^1	ρ^0																												
2.03									ρ^0	ρ^1	ρ^2	ρ^3	ρ^4	ρ^5	ρ^6																					
2.06									ρ^1	ρ^0	ρ^1	ρ^2	ρ^3	ρ^4	ρ^5																					
2.09									ρ^2	ρ^1	ρ^0	ρ^1	ρ^2	ρ^3	ρ^4																					
2.12									ρ^3	ρ^2	ρ^1	ρ^0	ρ^1	ρ^2	ρ^3																					
2.15									ρ^4	ρ^3	ρ^2	ρ^1	ρ^0	ρ^1	ρ^2																					
2.18									ρ^5	ρ^4	ρ^3	ρ^2	ρ^1	ρ^0	ρ^1																					
2.21									ρ^6	ρ^5	ρ^4	ρ^3	ρ^2	ρ^1	ρ^0																					
3.03																ρ^0	ρ^1	ρ^2	ρ^3	ρ^4	ρ^5															
3.06																ρ^1	ρ^0	ρ^1	ρ^2	ρ^3	ρ^4															
3.09																ρ^2	ρ^1	ρ^0	ρ^1	ρ^2	ρ^3															
3.12																ρ^3	ρ^2	ρ^1	ρ^0	ρ^1	ρ^2															
3.15																ρ^4	ρ^3	ρ^2	ρ^1	ρ^0	ρ^1															
3.18																ρ^5	ρ^4	ρ^3	ρ^2	ρ^1	ρ^0															
4.03																						ρ^0	ρ^1	ρ^2	ρ^3	ρ^4										
4.06																						ρ^1	ρ^0	ρ^1	ρ^2	ρ^3										
4.09																						ρ^2	ρ^1	ρ^0	ρ^1	ρ^2										
4.12																						ρ^3	ρ^2	ρ^1	ρ^0	ρ^1										
4.15																						ρ^4	ρ^3	ρ^2	ρ^1	ρ^0										
5.03																											ρ^0	ρ^1	ρ^2	ρ^3						
5.06																											ρ^1	ρ^0	ρ^1	ρ^2						
5.09																											ρ^2	ρ^1	ρ^0	ρ^1						
5.12																											ρ^3	ρ^2	ρ^1	ρ^0						
6.03																															ρ^0	ρ^1	ρ^2			
6.06																															ρ^1	ρ^0	ρ^1			
6.09																															ρ^2	ρ^1	ρ^0			
7.03																																		ρ^0	ρ^1	
7.06																																		ρ^1	ρ^0	
8.03																																				ρ^0

EXHIBIT 6

PART 2

P_{21} CORRELATION MATRICES

P_{21}	1.03	1.06	1.09	1.12	1.15	1.18	1.21	1.24	2.03	2.06	2.09	2.12	2.15	2.18	2.21	3.03	3.06	3.09	3.12	3.15	3.18	4.03	4.06	4.09	4.12	4.15	5.03	5.06	5.09	5.12	6.03	6.06	6.09	7.03	7.06	8.03
2.24									ρ^7	ρ^6	ρ^5	ρ^4	ρ^3	ρ^2	ρ^1																					
3.21																ρ^6	ρ^5	ρ^4	ρ^3	ρ^2	ρ^1															
3.24																ρ^7	ρ^6	ρ^5	ρ^4	ρ^3	ρ^2															
4.18																						ρ^5	ρ^4	ρ^3	ρ^2	ρ^1										
4.21																						ρ^6	ρ^5	ρ^4	ρ^3	ρ^2										
4.24																						ρ^7	ρ^6	ρ^5	ρ^4	ρ^3										
5.15																											ρ^4	ρ^3	ρ^2	ρ^1						
5.18																											ρ^5	ρ^4	ρ^3	ρ^2						
5.21																											ρ^6	ρ^5	ρ^4	ρ^3						
5.24																											ρ^7	ρ^6	ρ^5	ρ^4						
6.12																															ρ^3	ρ^2	ρ^1			
6.15																															ρ^4	ρ^3	ρ^2			
6.18																															ρ^5	ρ^4	ρ^3			
6.21																															ρ^6	ρ^5	ρ^4			
6.24																															ρ^7	ρ^6	ρ^5			
7.09																																		ρ^2	ρ^1	
7.12																																		ρ^3	ρ^2	
7.15																																		ρ^4	ρ^3	
7.18																																		ρ^5	ρ^4	
7.21																																		ρ^6	ρ^5	
7.24																																		ρ^7	ρ^6	
8.06																																				[illegible]
8.09																																				[illegible]
8.12																																				[illegible]
8.15																																				[illegible]
8.18																																				[illegible]
8.21																																				[illegible]
8.24																																				

EXHIBIT 6

PART 3

P_{22} CORRELATION MATRICES

P_{22}	2.24	3.21	3.24	4.18	4.21	4.24	5.15	5.18	5.21	5.24	6.12	6.15	6.18	6.21	6.24	7.09	7.12	7.15	7.18	7.21	7.24	8.06	8.09	8.12	8.15	8.18	8.21	8.24
2.24	ρ^0																											
3.21		ρ^0	ρ^1																									
3.24		ρ^1	ρ^0																									
4.18				ρ^0	ρ^1	ρ^2																						
4.21				ρ^1	ρ^0	ρ^1																						
4.24				ρ^2	ρ^1	ρ^0																						
5.15							ρ^0	ρ^1	ρ^2	ρ^3																		
5.18							ρ^1	ρ^0	ρ^1	ρ^2																		
5.21							ρ^2	ρ^1	ρ^0	ρ^1																		
5.24							ρ^3	ρ^2	ρ^1	ρ^0																		
6.12											ρ^0	ρ^1	ρ^2	ρ^3	ρ^4													
6.15											ρ^1	ρ^0	ρ^1	ρ^2	ρ^3													
6.18											ρ^2	ρ^1	ρ^0	ρ^1	ρ^2													
6.21											ρ^3	ρ^2	ρ^1	ρ^0	ρ^1													
6.24											ρ^4	ρ^3	ρ^2	ρ^1	ρ^0													
7.09																ρ^0	ρ^1	ρ^2	ρ^3	ρ^4	ρ^5							
7.12																ρ^1	ρ^0	ρ^1	ρ^2	ρ^3	ρ^4							
7.15																ρ^2	ρ^1	ρ^0	ρ^1	ρ^2	ρ^3							
7.18																ρ^3	ρ^2	ρ^1	ρ^0	ρ^1	ρ^2							
7.21																ρ^4	ρ^3	ρ^2	ρ^1	ρ^0	ρ^1							
7.24																ρ^5	ρ^4	ρ^3	ρ^2	ρ^1	ρ^0							
8.06																						ρ^0	ρ^1	ρ^2	ρ^3	ρ^4	ρ^5	ρ^6
8.09																						ρ^1	ρ^0	ρ^1	ρ^2	ρ^3	ρ^4	ρ^5
8.12																						ρ^2	ρ^1	ρ^0	ρ^1	ρ^2	ρ^3	ρ^4
8.15																						ρ^3	ρ^2	ρ^1	ρ^0	ρ^1	ρ^2	ρ^3
8.18																						ρ^4	ρ^3	ρ^2	ρ^1	ρ^0	ρ^1	ρ^2
8.21																						ρ^5	ρ^4	ρ^3	ρ^2	ρ^1	ρ^0	ρ^1
8.24																						ρ^6	ρ^5	ρ^4	ρ^3	ρ^2	ρ^1	ρ^0

EXHIBIT 6

PART 4

$P_{21}P_{11}^{-1}$ CORRELATION MATRICES

$P_{21}P_{11}^{-1}$	1.03	1.06	1.09	1.12	1.15	1.18	1.21	1.24	2.03	2.06	2.09	2.12	2.15	2.18	2.21	3.03	3.06	3.09	3.12	3.15	3.18	4.03	4.06	4.09	4.12	4.15	5.03	5.06	5.09	5.12	6.03	6.06	6.09	7.03	7.06	8.03
2.24															ρ^1																					
3.21																					ρ^1															
3.24																					ρ^2															
4.18																										ρ^1										
4.21																										ρ^2										
4.24																										ρ^3										
5.15																														ρ^1						
5.18																														ρ^2						
5.21																														ρ^3						
5.24																														ρ^4						
6.12																																	ρ^1			
6.15																																	ρ^2			
6.18																																	ρ^3			
6.21																																	ρ^4			
6.24																																	ρ^5			
7.09																																			ρ^1	
7.12																																			ρ^2	
7.15																																			ρ^3	
7.18																																			ρ^4	
7.21																																			ρ^5	
7.24																																			ρ^6	
8.06																																				ρ^1
8.09																																				ρ^2
8.12																																				ρ^3
8.15																																				ρ^4
8.18																																				ρ^5
8.21																																				ρ^6
8.24																																				ρ^7

EXHIBIT 7

SELECTED VALUES

AQ.Age	$X_1\hat{\beta}$	Y_1	AQ.Age	$X_2\hat{\beta}$	$E[Y_2 \mid Y_1]$
1.03	115,371	87,248	2.24	266,326	261,487
1.06	231,615	187,878	3.21	424,014	446,060
1.09	169,126	118,385	3.24	420,496	432,834
1.12	145,210	99,109	4.18	694,978	735,877
1.15	125,953	112,877	4.21	547,495	570,385
1.18	116,639	95,914	4.24	542,952	555,763
1.21	91,887	55,468	5.15	788,818	766,410
1.24	91,125	70,742	5.18	730,487	717,947
2.03	337,190	189,320	5.21	575,468	568,450
2.06	676,931	523,517	5.24	570,694	566,766
2.09	494,297	444,265	6.12	889,970	878,725
2.12	424,399	339,841	6.15	771,948	765,655
2.15	368,119	289,189	6.18	714,865	711,343
2.18	340,897	281,101	6.21	563,162	561,190
2.21	268,555	259,908	6.24	558,489	557,386
3.03	532,380	392,599	7.09	1,027,833	1,051,136
3.06	1,068,788	1,064,822	7.12	882,489	895,531
3.09	780,433	725,532	7.15	765,459	772,758
3.12	670,073	657,399	7.18	708,856	712,941
3.15	581,212	599,727	7.21	558,428	560,714
3.18	538,234	577,625	7.24	553,794	555,074
4.03	687,419	675,634	8.06	1,267,593	1,392,036
4.06	1,380,040	1,538,669	8.09	925,601	995,248
4.09	1,007,710	1,067,807	8.12	794,713	833,692
4.12	865,211	1,007,805	8.15	689,324	711,139
4.15	750,473	823,550	8.18	638,351	650,560
5.03	722,542	720,152	8.21	502,884	509,718
5.06	1,450,552	1,332,582	8.24	498,712	502,536
5.09	1,059,198	1,081,918			
5.12	909,418	869,382			
6.03	707,090	746,772			
6.06	1,419,531	1,413,910			
6.09	1,036,546	1,016,454			
7.03	701,146	769,063			
7.06	1,407,598	1,449,235			
8.03	631,408	853,758			

EXHIBIT 8

COVARIANCE METHOD

Incremental Workers Compensation Indemnity Losses Paid

AQ	@3	@6	@9	@12	@15	@18	@21	@24
Qtr 1	87,248	187,878	118,385	99,109	112,877	95,914	55,468	70,742
Qtr 2	189,320	523,517	444,265	339,841	289,189	281,101	259,908	261,487
Qtr 3	392,599	1,064,822	725,532	657,399	599,727	577,625	446,060	432,834
Qtr 4	675,634	1,538,669	1,067,807	1,007,805	823,550	735,877	570,385	555,763
Qtr 5	720,152	1,332,582	1,081,918	869,382	766,410	717,947	568,450	566,766
Qtr 6	746,772	1,413,910	1,016,454	878,725	765,655	711,343	561,190	557,386
Qtr 7	769,063	1,449,235	1,051,136	895,531	772,758	712,941	560,714	555,074
Qtr 8	853,758	1,392,036	995,248	833,692	711,139	650,560	509,718	502,536
Total	4,434,546	8,902,649	6,500,745	5,581,484	4,841,305	4,483,308	3,531,892	3,502,587
Rate	0.0099	0.0199	0.0145	0.0125	0.0108	0.0100	0.0079	0.0078

Cumulative Workers Compensation Indemnity Losses Paid

AQ	@3	@6	@9	@12	@15	@18	@21	@24	Standard Deviation of 24 Month Prediction
Qtr 1	87,248	275,126	393,511	492,620	605,497	701,411	756,879	827,621	
Qtr 2	189,320	712,837	1,157,102	1,496,943	1,786,132	2,067,233	2,327,141	2,588,628	87,982
Qtr 3	392,599	1,457,421	2,182,953	2,840,352	3,440,079	4,017,704	4,463,764	4,896,598	189,783
Qtr 4	675,634	2,214,303	3,282,110	4,289,915	5,113,465	5,849,342	6,419,727	6,975,489	293,083
Qtr 5	720,152	2,052,734	3,134,652	4,004,034	4,770,444	5,488,391	6,056,841	6,623,606	359,330
Qtr 6	746,772	2,160,682	3,177,136	4,055,861	4,821,515	5,532,858	6,094,049	6,651,434	405,187
Qtr 7	769,063	2,218,298	3,269,434	4,164,964	4,937,723	5,650,664	6,211,378	6,766,452	453,553
Qtr 8	853,758	2,245,794	3,241,042	4,074,734	4,785,873	5,436,433	5,946,151	6,448,687	470,040
Total	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

EXHIBIT 9

EXPECTED LOSSES PRIOR TO ANY OBSERVATIONS

Incremental Workers Compensation Indemnity Losses Paid

AQ	@3	@6	@9	@12	@15	@18	@21	@24
Qtr 1	115,371	231,615	169,126	145,210	125,953	116,639	91,887	91,125
Qtr 2	337,190	676,931	494,297	424,399	368,119	340,897	268,555	266,326
Qtr 3	532,380	1,068,788	780,433	670,073	581,212	538,234	424,014	420,496
Qtr 4	687,419	1,380,040	1,007,710	865,211	750,473	694,978	547,495	542,952
Qtr 5	722,542	1,450,552	1,059,198	909,418	788,818	730,487	575,468	570,694
Qtr 6	707,090	1,419,531	1,036,546	889,970	771,948	714,865	563,162	558,489
Qtr 7	701,146	1,407,598	1,027,833	882,489	765,459	708,856	558,428	553,794
Qtr 8	631,408	1,267,593	925,601	794,713	689,324	638,351	502,884	498,712
Total	4,434,546	8,902,649	6,500,745	5,581,484	4,841,305	4,483,308	3,531,892	3,502,587
Rate	0.0099	0.0199	0.0145	0.0125	0.0108	0.0100	0.0079	0.0078

Cumulative Workers Compensation Indemnity Losses Paid

AQ	@3	@6	@9	@12	@15	@18	@21	@24
Qtr 1	115,371	346,986	516,111	661,321	787,275	903,914	995,801	1,086,926
Qtr 2	337,190	1,014,121	1,508,418	1,932,818	2,300,936	2,641,834	2,910,388	3,176,715
Qtr 3	532,380	1,601,168	2,381,601	3,051,674	3,632,887	4,171,120	4,595,134	5,015,630
Qtr 4	687,419	2,067,459	3,075,169	3,940,380	4,690,853	5,385,831	5,933,325	6,476,277
Qtr 5	722,542	2,173,094	3,232,293	4,141,711	4,930,528	5,661,016	6,236,484	6,807,178
Qtr 6	707,090	2,126,621	3,163,168	4,053,137	4,825,085	5,539,951	6,103,112	6,661,601
Qtr 7	701,146	2,108,745	3,136,578	4,019,067	4,784,526	5,493,382	6,051,810	6,605,604
Qtr 8	631,408	1,899,001	2,824,602	3,619,315	4,308,639	4,946,989	5,449,874	5,948,585
Total	4,434,546	13,337,195	19,837,940	25,419,423	30,260,728	34,744,037	38,275,929	41,778,516

APPENDIX A

THE GAUSS–MARKOV THEOREM

This proof is an extension of the proof found in Judge [5, pp. 202–205]. Some matrix theory assumed here can be studied from that text, especially from its Appendix A. One principle merely stated here is that if $\mathbf{Z}$ is an $(n \times 1)$ random vector, distributed as $[\mu, \Sigma]$, and $\mathbf{A}$ is a nonstochastic $(m \times n)$ matrix, then $\mathbf{AZ} \sim [\mathbf{A}\mu, \mathbf{A}\Sigma\mathbf{A}']$. The distribution does not have to be normal.

We have a model $\mathbf{Y} = \mathbf{X}\beta + \mathbf{e}$, where $\mathbf{e} \sim [\mathbf{0}, \mathbf{\Phi}]$, $\mathbf{Y}$ and $\mathbf{e}$ are $(t \times 1)$, $\mathbf{X}$ is $(t \times k)$, and β is $(k \times 1)$. $\mathbf{e}$ does not have to be normally distributed. The rank of $\mathbf{X}$ is k, and $\mathbf{\Phi}$ is positive definite. These are standard and nonrestrictive conditions. The last two conditions guarantee that $(\mathbf{X}'\mathbf{\Phi}^{-1}\mathbf{X})^{-1}$ exists, and that there is a nonsingular $(t \times t)$ matrix $\mathbf{W}$ such that $\mathbf{WW}' = \mathbf{\Phi}$. (See Appendix C regarding the Cholesky procedure.)

The generalized least-squares estimator is:

$$\begin{aligned}
\hat{\beta} &= (\mathbf{X}'\mathbf{\Phi}^{-1}\mathbf{X})^{-1}(\mathbf{X}'\mathbf{\Phi}^{-1}\mathbf{Y}) \\
&= (\mathbf{X}'\mathbf{\Phi}^{-1}\mathbf{X})^{-1}(\mathbf{X}'\mathbf{\Phi}^{-1}\mathbf{X}\beta) + (\mathbf{X}'\mathbf{\Phi}^{-1}\mathbf{X})^{-1}(\mathbf{X}'\mathbf{\Phi}^{-1}\mathbf{e}) \\
&= \beta + (\mathbf{X}'\mathbf{\Phi}^{-1}\mathbf{X})^{-1}(\mathbf{X}'\mathbf{\Phi}^{-1}\mathbf{e}) \\
&\sim [\beta, (\mathbf{X}'\mathbf{\Phi}^{-1}\mathbf{X})^{-1}(\mathbf{X}'\mathbf{\Phi}^{-1})\mathrm{Var}[\mathbf{e}](\mathbf{\Phi}^{-1}\mathbf{X})(\mathbf{X}'\mathbf{\Phi}^{-1}\mathbf{X})^{-1}] \\
&\sim [\beta, (\mathbf{X}'\mathbf{\Phi}^{-1}\mathbf{X})^{-1}(\mathbf{X}'\mathbf{\Phi}^{-1})\mathbf{\Phi}(\mathbf{\Phi}^{-1}\mathbf{X})(\mathbf{X}'\mathbf{\Phi}^{-1}\mathbf{X})^{-1}] \\
&\sim [\beta, (\mathbf{X}'\mathbf{\Phi}^{-1}\mathbf{X})^{-1}(\mathbf{X}'\mathbf{\Phi}^{-1}\mathbf{X})(\mathbf{X}'\mathbf{\Phi}^{-1}\mathbf{X})^{-1}] \\
&\sim [\beta, (\mathbf{X}'\mathbf{\Phi}^{-1}\mathbf{X})^{-1}].
\end{aligned}$$

Now consider some alternative estimator:

$$\begin{aligned}
\tilde{\beta} &= \mathbf{AY} \\
&= \mathbf{AX}\beta + \mathbf{Ae} \\
&\sim [\mathbf{AX}\beta, \mathbf{A}\,\mathrm{Var}[\mathbf{e}]\mathbf{A}'] \\
&\sim [\mathbf{AX}\beta, \mathbf{A\Phi A}'].
\end{aligned}$$

Both estimators are linear functions of $\mathbf{Y}$. The first is unbiased; the second is unbiased, whatever β may be, if and only if $\mathbf{AX}$ is the $(k \times k)$ identity matrix $\mathbf{I}_k$. Hence, $\mathbf{AX} = \mathbf{I}_k$.

So far we have two linear unbiased estimators of β; we have the "LUE" of "BLUE." We show that the first estimator is better (or best of all) by showing that the difference of the variance matrices is nonnegative definite:

$$\begin{aligned}
&\mathrm{Var}(\tilde{\beta}) - \mathrm{Var}(\hat{\beta}) \\
&\quad = \mathbf{A\Phi A}' - (\mathbf{X}'\mathbf{\Phi}^{-1}\mathbf{X})^{-1} \\
&\quad = \mathbf{A\Phi A}' - (\mathbf{X}'\mathbf{\Phi}^{-1}\mathbf{X})^{-1} - (\mathbf{X}'\mathbf{\Phi}^{-1}\mathbf{X})^{-1} + (\mathbf{X}'\mathbf{\Phi}^{-1}\mathbf{X})^{-1} \\
&\quad = \mathbf{A\Phi A}' - \mathbf{AX}(\mathbf{X}'\mathbf{\Phi}^{-1}\mathbf{X})^{-1} - (\mathbf{X}'\mathbf{\Phi}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{A}' \\
&\qquad + (\mathbf{X}'\mathbf{\Phi}^{-1}\mathbf{X})^{-1}(\mathbf{X}'\mathbf{\Phi}^{-1}\mathbf{X})(\mathbf{X}'\mathbf{\Phi}^{-1}\mathbf{X})^{-1} \\
&\quad = \mathbf{AWW}'\mathbf{A}' - \mathbf{AWW}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{\Phi}^{-1}\mathbf{X})^{-1} \\
&\qquad - (\mathbf{X}'\mathbf{\Phi}^{-1}\mathbf{X})^{-1}\mathbf{X}'(\mathbf{W}^{-1})'\mathbf{W}'\mathbf{A}' \\
&\qquad + (\mathbf{X}'\mathbf{\Phi}^{-1}\mathbf{X})^{-1}\mathbf{X}'(\mathbf{W}^{-1})'\mathbf{W}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{\Phi}^{-1}\mathbf{X})^{-1} \\
&\quad = \{\mathbf{AW} - (\mathbf{X}'\mathbf{\Phi}^{-1}\mathbf{X})^{-1}\mathbf{X}'(\mathbf{W}^{-1})'\} \\
&\qquad \times \{\mathbf{W}'\mathbf{A}' - \mathbf{W}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{\Phi}^{-1}\mathbf{X})^{-1}\} \\
&\quad = \{\mathbf{AW} - (\mathbf{X}'\mathbf{\Phi}^{-1}\mathbf{X})^{-1}\mathbf{X}'(\mathbf{W}^{-1})'\} \\
&\qquad \times \{\mathbf{AW} - (\mathbf{X}'\mathbf{\Phi}^{-1}\mathbf{X})^{-1}\mathbf{X}'(\mathbf{W}^{-1})'\}' \\
&\quad \geq \mathbf{0}.
\end{aligned}$$

The last line means that the matrix in the previous line is non-negative definite, which indeed it is since it is the product of a matrix and its transpose. Therefore, $\hat{\beta} = (\mathbf{X}'\mathbf{\Phi}^{-1}\mathbf{X})^{-1}(\mathbf{X}'\mathbf{\Phi}^{-1}\mathbf{Y})$ is BLUE.

APPENDIX B

THE CONDITIONAL MULTIVARIATE NORMAL DISTRIBUTION

We start with $\mathbf{X} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $\mathbf{X}$ and $\boldsymbol{\mu}$ are $(n \times 1)$, and $\boldsymbol{\Sigma}(n \times n)$ is symmetric and positive definite. Then we partition $\mathbf{X}$ into p known elements and q unknown $(p + q = n)$:

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{bmatrix} \sim N\left(\begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix} \right),$$

where $\mathbf{X}_1$ and $\boldsymbol{\mu}_1$ are $(p \times 1)$, $\mathbf{X}_2$ and $\boldsymbol{\mu}_2$ are $(q \times 1)$, $\boldsymbol{\Sigma}_{11}$ is $(p \times p)$, $\boldsymbol{\Sigma}_{12}$ is $(p \times q)$, $\boldsymbol{\Sigma}_{21}$ is $(q \times p)$, and $\boldsymbol{\Sigma}_{22}$ is $(q \times q)$. Because $\boldsymbol{\Sigma}$ is symmetric, $\boldsymbol{\Sigma}_{11}$ and $\boldsymbol{\Sigma}_{22}$ are symmetric, and $\boldsymbol{\Sigma}_{21} = \boldsymbol{\Sigma}_{12}'$. Moreover, because $\boldsymbol{\Sigma}$ is positive definite, $\boldsymbol{\Sigma}_{11}$ and $\boldsymbol{\Sigma}_{22}$ are positive definite. Furthermore, $\boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}$ and $\boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}$ are symmetric and positive definite. Every positive definite matrix has an inverse. The probability density function for $\mathbf{X}$ is [Johnson 4, p. 128]:

$$f_{\mathbf{X}}(\mathbf{X}) = \frac{1}{(2\pi)^{n/2}|\boldsymbol{\Sigma}|^{1/2}} e^{-(1/2)(\mathbf{X}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{X}-\boldsymbol{\mu})}$$

$$\propto e^{-(1/2)(\mathbf{X}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{X}-\boldsymbol{\mu})}.$$

Let

$$\mathbf{A}_{(p \times p)} = (\boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21})^{-1}$$

and

$$\mathbf{D}_{(q \times q)} = (\boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12})^{-1}.$$

$\mathbf{A}$ and $\mathbf{D}$ must exist because they are inverses of positive definite matrices. They are also symmetric. An important equation is $\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}\mathbf{D} = \mathbf{A}\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}$, which is proven as follows:

$$
\begin{aligned}
&\Sigma_{11}^{-1}\Sigma_{12}\mathbf{D} - \mathbf{A}\Sigma_{12}\Sigma_{22}^{-1} \\
&\quad = \mathbf{A}(\mathbf{A}^{-1}\Sigma_{11}^{-1}\Sigma_{12} - \Sigma_{12}\Sigma_{22}^{-1}\mathbf{D}^{-1})\mathbf{D} \\
&\quad = \mathbf{A}([\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}]\Sigma_{11}^{-1}\Sigma_{12} \\
&\qquad\quad - \Sigma_{12}\Sigma_{22}^{-1}[\Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}])\mathbf{D} \\
&\quad = \mathbf{A}([\Sigma_{12} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}] \\
&\qquad\quad - [\Sigma_{12} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}])\mathbf{D} \\
&\quad = \mathbf{A}(\mathbf{0})\mathbf{D} \\
&\quad = \mathbf{0}.
\end{aligned}
$$

And let $\mathbf{B}_{(p\times q)} = -\Sigma_{11}^{-1}\Sigma_{12}\mathbf{D} = -\mathbf{A}\Sigma_{12}\Sigma_{22}^{-1}$, so $\mathbf{B}' = -\Sigma_{22}^{-1}\Sigma_{21}\mathbf{A}$. It can be shown, from multiplying Σ by the following matrix and obtaining the identity matrix, that:

$$
\Sigma^{-1} = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B}' & \mathbf{D} \end{bmatrix}.
$$

Therefore,

$$
\begin{aligned}
&(\mathbf{X} - \mu)'\Sigma^{-1}(\mathbf{X} - \mu) \\
&\quad = (\mathbf{X} - \mu)' \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B}' & \mathbf{D} \end{bmatrix} (\mathbf{X} - \mu) \\
&\quad = [(\mathbf{X}_1 - \mu_1)' \;\; (\mathbf{X}_2 - \mu_2)'] \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B}' & \mathbf{D} \end{bmatrix} \begin{bmatrix} (\mathbf{X}_1 - \mu_1) \\ (\mathbf{X}_2 - \mu_2) \end{bmatrix} \\
&\quad = (\mathbf{X}_1 - \mu_1)'\mathbf{A}(\mathbf{X}_1 - \mu_1) + (\mathbf{X}_1 - \mu_1)'\mathbf{B}(\mathbf{X}_2 - \mu_2) \\
&\qquad + (\mathbf{X}_2 - \mu_2)'\mathbf{B}'(\mathbf{X}_1 - \mu_1) + (\mathbf{X}_2 - \mu_2)'\mathbf{D}(\mathbf{X}_2 - \mu_2).
\end{aligned}
$$

If $\mathbf{X}_1$ is given, then $(\mathbf{X}_1 - \mu_1)'\mathbf{A}(\mathbf{X}_1 - \mu_1)$ is constant, so:

$$
\begin{aligned}
&f_{\mathbf{X}_2|\mathbf{X}_1}(\mathbf{X}_2) \\
&\qquad \propto e^{-(1/2)[(\mathbf{X}_2-\mu_2)'\mathbf{D}(\mathbf{X}_2-\mu_2)+(\mathbf{X}_2-\mu_2)'\mathbf{B}'(\mathbf{X}_1-\mu_1)+(\mathbf{X}_1-\mu_1)'\mathbf{B}(\mathbf{X}_2-\mu_2)]}.
\end{aligned}
$$

Now, because $\mathbf{D}$ is symmetric and positive definite, there exists a nonsingular $\mathbf{W}_{(q\times q)}$ such that $\mathbf{W}'\mathbf{W} = \mathbf{D}$. Therefore,

$$
\begin{aligned}
&(\mathbf{X}_2-\boldsymbol{\mu}_2)'\mathbf{D}(\mathbf{X}_2-\boldsymbol{\mu}_2)+(\mathbf{X}_2-\boldsymbol{\mu}_2)'\mathbf{B}'(\mathbf{X}_1-\boldsymbol{\mu}_1)\\
&\qquad+(\mathbf{X}_1-\boldsymbol{\mu}_1)'\mathbf{B}(\mathbf{X}_2-\boldsymbol{\mu}_2)\\
&\quad=(\mathbf{X}_2-\boldsymbol{\mu}_2)'\mathbf{W}'\mathbf{W}(\mathbf{X}_2-\boldsymbol{\mu}_2)\\
&\qquad+(\mathbf{X}_2-\boldsymbol{\mu}_2)'\mathbf{W}'(\mathbf{W}')^{-1}\mathbf{B}'(\mathbf{X}_1-\boldsymbol{\mu}_1)\\
&\qquad+(\mathbf{X}_1-\boldsymbol{\mu}_1)'\mathbf{B}\mathbf{W}^{-1}\mathbf{W}(\mathbf{X}_2-\boldsymbol{\mu}_2)\\
&\quad=[(\mathbf{X}_2-\boldsymbol{\mu}_2)'\mathbf{W}'+(\mathbf{X}_1-\boldsymbol{\mu}_1)'\mathbf{B}\mathbf{W}^{-1}]\\
&\qquad\times[\mathbf{W}(\mathbf{X}_2-\boldsymbol{\mu}_2)+(\mathbf{W}')^{-1}\mathbf{B}'(\text{''}X_1-\boldsymbol{\mu}_1)]\\
&\qquad-(\mathbf{X}_1-\boldsymbol{\mu}_1)'\mathbf{B}\mathbf{W}^{-1}(\mathbf{W}')^{-1}\mathbf{B}'(\mathbf{X}_1-\boldsymbol{\mu}_1).
\end{aligned}
$$

This expression goes into the exponent of the probability distribution. When $\mathbf{X}_1$ is given, the term after the minus sign in the last equation is constant. Therefore, dropping this term does not change the proportionality of the conditional distribution. So we continue:

$$
\begin{aligned}
&[(\mathbf{X}_2-\boldsymbol{\mu}_2)'\mathbf{W}'+(\mathbf{X}_1-\boldsymbol{\mu}_1)'\mathbf{B}\mathbf{W}^{-1}]\\
&\qquad\times[\mathbf{W}(\mathbf{X}_2-\boldsymbol{\mu}_2)+(\mathbf{W}')^{-1}\mathbf{B}'(\mathbf{X}_1-\boldsymbol{\mu}_1)]\\
&\quad=[(\mathbf{X}_2-\boldsymbol{\mu}_2)'+(\mathbf{X}_1-\boldsymbol{\mu}_1)'\mathbf{B}\mathbf{W}^{-1}(\mathbf{W}')^{-1}]\mathbf{W}'\mathbf{W}\\
&\qquad\times[(\mathbf{X}_2-\boldsymbol{\mu}_2)+\mathbf{W}^{-1}(\mathbf{W}')^{-1}\mathbf{B}'(\mathbf{X}_1-\boldsymbol{\mu}_1)]\\
&\quad=[(\mathbf{X}_2-\boldsymbol{\mu}_2)'+(\mathbf{X}_1-\boldsymbol{\mu}_1)'\mathbf{B}(\mathbf{W}'\mathbf{W})^{-1}]\mathbf{W}'\mathbf{W}\\
&\qquad\times[(\mathbf{X}_2-\boldsymbol{\mu}_2)+(\mathbf{W}'\mathbf{W})^{-1}\mathbf{B}'(\mathbf{X}_1-\boldsymbol{\mu}_1)]\\
&\quad=[(\mathbf{X}_2-\boldsymbol{\mu}_2)'+(\mathbf{X}_1-\boldsymbol{\mu}_1)'\mathbf{B}\mathbf{D}^{-1}]\mathbf{D}\\
&\qquad\times[(\mathbf{X}_2-\boldsymbol{\mu}_2)+\mathbf{D}^{-1}\mathbf{B}'(\mathbf{X}_1-\boldsymbol{\mu}_1)]\\
&\quad=[(\mathbf{X}_2-\boldsymbol{\mu}_2)+\mathbf{D}^{-1}\mathbf{B}'(\mathbf{X}_1-\boldsymbol{\mu}_1)]'(\mathbf{D}^{-1})^{-1}\\
&\qquad\times[(\mathbf{X}_2-\boldsymbol{\mu}_2)+\mathbf{D}^{-1}\mathbf{B}'(\mathbf{X}_1-\boldsymbol{\mu}_1)].
\end{aligned}
$$

Therefore,

$$f_{\mathbf{X}_2|\mathbf{X}_1}(\mathbf{X}_2)$$
$$\propto e^{-(1/2)[(\mathbf{X}_2-\mu_2)+\mathbf{D}^{-1}\mathbf{B}'(\mathbf{X}_1-\mu_1)]'(\mathbf{D}^{-1})^{-1}[(\mathbf{X}_2-\mu_2)+\mathbf{D}^{-1}\mathbf{B}'(\mathbf{X}_1-\mu_1)]}.$$

This form is multivariate normal with the following characteristics:

$$\mathbf{X}_2 \mid \mathbf{X}_1 \sim N(\mu_2 - \mathbf{D}^{-1}\mathbf{B}'(\mathbf{X}_1 - \mu_1), \mathbf{D}^{-1}).$$

But $\mathbf{D}^{-1} = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}$ and $\mathbf{D}^{-1}\mathbf{B}' = \mathbf{D}^{-1}(-\mathbf{D}\Sigma_{21}\Sigma_{11}^{-1}) = -\Sigma_{21}\Sigma_{11}^{-1}$. Therefore,

$$\mathbf{X}_2 \mid \mathbf{X}_1 \sim N(\mu_2 + \Sigma_{21}\Sigma_{11}^{-1}(\mathbf{X}_1 - \mu_1), \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}).$$

Finally, notice that:

$$\begin{aligned}\mathrm{Var}[\mathbf{X}_2] - \mathrm{Var}[\mathbf{X}_2 \mid \mathbf{X}_1] &= \Sigma_{22} - (\Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12})\\ &= \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}\\ &\geq \mathbf{0},\end{aligned}$$

because $\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}$ is nonnegative definite. Therefore, the conditional variance of $\mathbf{X}_2$ is less than or equal to the unconditional variance.

APPENDIX C

THE LEAST SQUARES PREDICTOR

This appendix relies much upon Appendix B, and is similar to the derivation by Judge [5, pp. 343–346]. We start with the standard linear model $\mathbf{Y} = \mathbf{X}\beta + \mathbf{e}$, where $\mathbf{Y}$ and $\mathbf{e}$ are $(t \times 1)$, $\mathbf{X}$ is $(t \times k)$, and β is $(k \times 1)$. $\mathbf{e} \sim [\mathbf{0}, \Sigma]$, not necessarily normal, where Σ is symmetric and positive definite. However, the first p rows of $\mathbf{Y}$ have been observed; the last $q = t - p$ rows are to be predicted. So the partitioned model is:

$$\begin{bmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{bmatrix} \beta + \begin{bmatrix} \mathbf{e}_1 \\ \mathbf{e}_2 \end{bmatrix},$$

where

$$\operatorname{Var} \begin{bmatrix} \mathbf{e}_1 \\ \mathbf{e}_2 \end{bmatrix} = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}.$$

The four submatrices of Σ have all the properties described in Appendix B. Note that the known matrices are $\mathbf{Y}_1, \mathbf{X}_1, \mathbf{X}_2$, and Σ. $\mathbf{Y}_1$ is known by observation, and Σ usually has to be estimated.

It is a theorem of matrix algebra that given a symmetric and positive definite Σ, there exists a nonsingular, lower-triangular matrix $\mathbf{W}$, such that $\mathbf{W}\Sigma\mathbf{W}' = \mathbf{I}$. Equivalently, $\Sigma^{-1} = \mathbf{W}'\mathbf{W}$. A suitable matrix $\mathbf{W}$ can be found by the Cholesky procedure [Healy, 3, pp. 54f]. $\mathbf{W}$ can be partitioned as:

$$\begin{bmatrix} \mathbf{A}_{(p \times p)} & \mathbf{0}_{(p \times q)} \\ \mathbf{C}_{(q \times p)} & \mathbf{D}_{(q \times q)} \end{bmatrix},$$

where $\mathbf{A}$ and $\mathbf{D}$ are nonsingular and lower-triangular. Choose $\mathbf{A}$ such that $\mathbf{A}\Sigma_{11}\mathbf{A}' = \mathbf{I}_p$, and $\mathbf{D}$ such that $\mathbf{D}(\Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12})\mathbf{D}' = \mathbf{I}_q$. The existence of suitable matrices $\mathbf{A}$ and $\mathbf{D}$ is guaranteed, since Σ_{11} and $\Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}$ are both symmetric and positive definite. And let $\mathbf{C} = -\mathbf{D}\Sigma_{21}\Sigma_{11}^{-1}$. Partitioned matrix multiplication will show:

$$\mathbf{W}\Sigma\mathbf{W}' = \begin{bmatrix} \mathbf{A}_{(p\times p)} & \mathbf{0}_{(p\times q)} \\ \mathbf{C}_{(q\times p)} & \mathbf{D}_{(q\times q)} \end{bmatrix} \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \begin{bmatrix} \mathbf{A}'_{(p\times p)} & \mathbf{C}'_{(p\times q)} \\ \mathbf{0}_{(q\times p)} & \mathbf{D}'_{(q\times q)} \end{bmatrix}$$

$$= \begin{bmatrix} \mathbf{I}_p & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_q \end{bmatrix} = \mathbf{I}_t.$$

The matrix $\mathbf{W}$ provides a convenient linear one-to-one transformation of the original model $\mathbf{WY} = \mathbf{WX}\beta + \mathbf{We}$, where $\mathbf{We} \sim [\mathbf{0}, \mathbf{W}\Sigma\mathbf{W}'] \sim [\mathbf{0}, \mathbf{I}_t]$. In partitioned matrices, we have:

$$\begin{bmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{C} & \mathbf{D} \end{bmatrix} \begin{bmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{C} & \mathbf{D} \end{bmatrix} \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{bmatrix} \beta + \begin{bmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{C} & \mathbf{D} \end{bmatrix} \begin{bmatrix} \mathbf{e}_1 \\ \mathbf{e}_2 \end{bmatrix}, \quad \text{and}$$

$$\begin{bmatrix} \mathbf{AY}_1 \\ \mathbf{CY}_1 + \mathbf{DY}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{AX}_1\beta \\ (\mathbf{CX}_1 + \mathbf{DX}_2)\beta \end{bmatrix} + \begin{bmatrix} \mathbf{Ae}_1 \\ \mathbf{Ce}_1 + \mathbf{De}_2 \end{bmatrix}.$$

Since $\mathbf{Y}_2$ is unknown, β must be estimated from the first p rows. And since $\mathbf{A}$ is nonsingular, the model for estimating β may be reduced to $\mathbf{Y}_1 = \mathbf{X}_1\beta + \mathbf{e}_1$, where $\mathbf{e}_1 \sim [\mathbf{0}, \Sigma_{11}]$. Therefore, the best linear unbiased estimator is (see Appendix A):

$$\begin{aligned} \hat{\beta} &= (\mathbf{X}_1'\Sigma_{11}^{-1}\mathbf{X}_1)^{-1}(\mathbf{X}_1'\Sigma_{11}^{-1}\mathbf{Y}_1) \\ &= \beta + (\mathbf{X}_1'\Sigma_{11}^{-1}\mathbf{X}_1)^{-1}(\mathbf{X}_1'\Sigma_{11}^{-1}\mathbf{e}_1) \\ &\sim [\beta, (\mathbf{X}_1'\Sigma_{11}^{-1}\mathbf{X}_1)^{-1}]. \end{aligned}$$

Now, instead of considering the predictor of $\mathbf{Y}_2$, let us consider the predictor of $\mathbf{CY}_1 + \mathbf{DY}_2$. This is easier because its error term, $\mathbf{Ce}_1 + \mathbf{De}_2$, is uncorrelated with the error term of the $\mathbf{AY}_1$, which is $\mathbf{Ae}_1$. This is the reason for finding $\mathbf{W}$ such that $\mathbf{W}\Sigma\mathbf{W}' = \mathbf{I}$. Therefore, $\mathrm{Cov}[\mathbf{Ae}_1, \mathbf{Ce}_1 + \mathbf{De}_2] = \mathbf{0}_{(p\times q)}$. Moreover, premultiplying both sides of this equation by $\mathbf{A}^{-1}$ yields $\mathrm{Cov}[\mathbf{A}^{-1}\mathbf{Ae}_1, \mathbf{Ce}_1 + \mathbf{De}_2] = \mathrm{Cov}[\mathbf{e}_1, \mathbf{Ce}_1 + \mathbf{De}_2] = \mathbf{0}_{(p\times q)}$. And since $\hat{\beta}$ is a linear combination of $\mathbf{e}_1$, this implies that

$\mathrm{Cov}[\hat{\boldsymbol{\beta}}, \mathbf{Ce}_1 + \mathbf{De}_2] = \mathbf{0}$. Hence:

$$\begin{aligned}
\mathbf{CY}_1 + \mathbf{DY}_2 &= (\mathbf{CX}_1 + \mathbf{DX}_2)\boldsymbol{\beta} + (\mathbf{Ce}_1 + \mathbf{De}_2) \\
&= (\mathbf{CX}_1 + \mathbf{DX}_2)\hat{\boldsymbol{\beta}} - (\mathbf{CX}_1 + \mathbf{DX}_2)(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) + (\mathbf{Ce}_1 + \mathbf{De}_2) \\
&= (\mathbf{CX}_1 + \mathbf{DX}_2)\hat{\boldsymbol{\beta}} + \mathbf{h},
\end{aligned}$$

where $\mathbf{h} = -(\mathbf{CX}_1 + \mathbf{DX}_2)(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) + (\mathbf{Ce}_1 + \mathbf{De}_2)$ is the total error term. $E[\mathbf{h}] = \mathbf{0}_{(q\times 1)}$. And, because $\mathrm{Cov}[\hat{\boldsymbol{\beta}}, \mathbf{Ce}_1 + \mathbf{De}_2] = \mathbf{0}$,

$$\begin{aligned}
\mathrm{Var}[\mathbf{h}] &= \mathrm{Var}[(\mathbf{CX}_1 + \mathbf{DX}_2)\hat{\boldsymbol{\beta}}] + \mathrm{Var}[\mathbf{Ce}_1 + \mathbf{De}_2] \\
&= (\mathbf{CX}_1 + \mathbf{DX}_2)\mathrm{Var}[\hat{\boldsymbol{\beta}}](\mathbf{CX}_1 + \mathbf{DX}_2)' + \mathbf{I}_q,
\end{aligned}$$

where, of course, $\mathrm{Var}[\hat{\boldsymbol{\beta}}] = (\mathbf{X}_1'\boldsymbol{\Sigma}_{11}^{-1}\mathbf{X}_1)^{-1}$.

When $\mathbf{Y}_1$ is observed, $\hat{\boldsymbol{\beta}}$ is determined; and we have:

$$\begin{aligned}
\mathbf{DY}_2 \mid \mathbf{Y}_1 &= (\mathbf{CX}_1 + \mathbf{DX}_2)\hat{\boldsymbol{\beta}} - \mathbf{CY}_1 + \mathbf{h} \\
\mathbf{Y}_2 \mid \mathbf{Y}_1 &= \mathbf{D}^{-1}(\mathbf{CX}_1 + \mathbf{DX}_2)\hat{\boldsymbol{\beta}} - \mathbf{D}^{-1}\mathbf{CY}_1 + \mathbf{D}^{-1}\mathbf{h} \\
&= \mathbf{D}^{-1}\mathbf{CX}_1\hat{\boldsymbol{\beta}} + \mathbf{X}_2\hat{\boldsymbol{\beta}} - \mathbf{D}^{-1}\mathbf{CY}_1 + \mathbf{D}^{-1}\mathbf{h} \\
&= \mathbf{X}_2\hat{\boldsymbol{\beta}} - \mathbf{D}^{-1}\mathbf{C}(\mathbf{Y}_1 - \mathbf{X}_1\hat{\boldsymbol{\beta}}) + \mathbf{D}^{-1}\mathbf{h} \\
&\sim [\mathbf{X}_2\hat{\boldsymbol{\beta}} - \mathbf{D}^{-1}\mathbf{C}(\mathbf{Y}_1 - \mathbf{X}_1\hat{\boldsymbol{\beta}}), \mathbf{D}^{-1}\mathrm{Var}[\mathbf{h}]\mathbf{D}'^{-1}].
\end{aligned}$$

But $\mathbf{D}^{-1}\mathbf{C} = \mathbf{D}^{-1}(-\mathbf{D}\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}) = -\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}$. And

$$\begin{aligned}
&\mathbf{D}^{-1}\mathrm{Var}[\mathbf{h}]\mathbf{D}'^{-1} \\
&= \mathbf{D}^{-1}\{(\mathbf{CX}_1 + \mathbf{DX}_2)\mathrm{Var}[\hat{\boldsymbol{\beta}}](\mathbf{CX}_1 + \mathbf{DX}_2)' + \mathbf{I}_q\}\mathbf{D}'^{-1} \\
&= \mathbf{D}^{-1}(\mathbf{CX}_1 + \mathbf{DX}_2)\mathrm{Var}[\hat{\boldsymbol{\beta}}](\mathbf{CX}_1 + \mathbf{DX}_2)'\mathbf{D}'^{-1} + \mathbf{D}^{-1}\mathbf{D}'^{-1} \\
&= (\mathbf{D}^{-1}\mathbf{CX}_1 + \mathbf{X}_2)\mathrm{Var}[\hat{\boldsymbol{\beta}}](\mathbf{D}^{-1}\mathbf{CX}_1 + \mathbf{X}_2)' + (\mathbf{D}'\mathbf{D})^{-1} \\
&= (-\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\mathbf{X}_1 + \mathbf{X}_2)\mathrm{Var}[\hat{\boldsymbol{\beta}}](-\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\mathbf{X}_1 + \mathbf{X}_2)' + (\mathbf{D}'\mathbf{D})^{-1} \\
&= (\mathbf{X}_2 - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\mathbf{X}_1)\mathrm{Var}[\hat{\boldsymbol{\beta}}](\mathbf{X}_2 - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\mathbf{X}_1)' \\
&\quad + (\boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}).
\end{aligned}$$

Therefore, per the Gauss–Markov theorem, $E[\mathbf{Y}_2 \mid \mathbf{Y}_1] = \mathbf{X}_2\hat{\beta} + \Sigma_{21}\Sigma_{11}^{-1}(\mathbf{Y}_1 - \mathbf{X}_1\hat{\beta})$ is the best linear unbiased predictor. Leaving aside the meaning of the square root of a variance matrix (i.e., a standard deviation matrix), we will write this as:

$$\Sigma_{22}^{-0.5}(E[\mathbf{Y}_2 \mid \mathbf{Y}_1] - \mathbf{X}_2\hat{\beta}) = \{\Sigma_{22}^{-0.5}\Sigma_{21}\Sigma_{11}^{-0.5}\}\{\Sigma_{11}^{-0.5}(\mathbf{Y}_1 - \mathbf{X}_1\hat{\beta})\}.$$

The terms on the ends of the equation look like standardized random vectors, and the middle term looks like a correlation matrix. This is a matrix generalization of the bivariate conditional expectation of Section 5 and Appendix B.

APPENDIX D

THE VARIANCE OF THE RESIDUALS

In Section 7, residuals were studentized; i.e., divided by the square root of the diagonal elements of a variance matrix. In this appendix, the expression for the variance of the residuals is derived.

We have the usual model $\mathbf{Y} = \mathbf{X}\beta + \mathbf{e}$, where $\mathbf{e} \sim [\mathbf{0}, \mathbf{\Phi}]$. $\mathbf{\Phi}$ is symmetric and positive definite, and the rank of $\mathbf{X}_{(t\times k)}$ is k. These conditions guarantee that $(\mathbf{X}'\mathbf{\Phi}^{-1}\mathbf{X})^{-1}$ exists. We have shown in Appendix A:

$$\hat{\beta} = \beta + (\mathbf{X}'\mathbf{\Phi}^{-1}\mathbf{X})^{-1}(\mathbf{X}'\mathbf{\Phi}^{-1}\mathbf{e}) \sim [\beta, (\mathbf{X}'\mathbf{\Phi}^{-1}\mathbf{X})^{-1}].$$

By definition,

$$\begin{aligned}
\hat{\mathbf{e}} &= \mathbf{Y} - \hat{\mathbf{Y}} \\
&= \mathbf{X}\beta + \mathbf{e} - \mathbf{X}\hat{\beta} \\
&= \mathbf{e} - \mathbf{X}(\hat{\beta} - \beta) \\
&= \mathbf{e} - \mathbf{X}(\mathbf{X}'\mathbf{\Phi}^{-1}\mathbf{X})^{-1}(\mathbf{X}'\mathbf{\Phi}^{-1}\mathbf{e}) \\
&= (\mathbf{I}_t - \mathbf{X}(\mathbf{X}'\mathbf{\Phi}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{\Phi}^{-1})\mathbf{e}.
\end{aligned}$$

Therefore,

$$\begin{aligned}
&\mathrm{Var}[\hat{\mathbf{e}}] \\
&= (\mathbf{I}_t - \mathbf{X}(\mathbf{X}'\mathbf{\Phi}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{\Phi}^{-1})\mathrm{Var}[\mathbf{e}](\mathbf{I}_t - \mathbf{X}(\mathbf{X}'\mathbf{\Phi}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{\Phi}^{-1})' \\
&= (\mathbf{I}_t - \mathbf{X}(\mathbf{X}'\mathbf{\Phi}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{\Phi}^{-1})\mathbf{\Phi}(\mathbf{I}_t - \mathbf{\Phi}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{\Phi}^{-1}\mathbf{X})^{-1}\mathbf{X}') \\
&= (\mathbf{\Phi} - \mathbf{X}(\mathbf{X}'\mathbf{\Phi}^{-1}\mathbf{X})^{-1}\mathbf{X}')(\mathbf{I}_t - \mathbf{\Phi}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{\Phi}^{-1}\mathbf{X})^{-1}\mathbf{X}') \\
&= (\mathbf{\Phi} - \mathbf{X}(\mathbf{X}'\mathbf{\Phi}^{-1}\mathbf{X})^{-1}\mathbf{X}') - (\mathbf{X}(\mathbf{X}'\mathbf{\Phi}^{-1}\mathbf{X})^{-1}\mathbf{X}' - \mathbf{X}(\mathbf{X}'\mathbf{\Phi}^{-1}\mathbf{X})^{-1}\mathbf{X}') \\
&= \mathbf{\Phi} - \mathbf{X}(\mathbf{X}'\mathbf{\Phi}^{-1}\mathbf{X})^{-1}\mathbf{X}'.
\end{aligned}$$

If instead of $\mathbf{\Phi}$ we have $\sigma^2\mathbf{\Phi}$, with σ^2 unknown, we use the estimate for σ^2.

APPENDIX E

THE BALANCE PROPERTIES OF SECTION 7

The a priori predictions are expressed as a (64×1) partitioned vector

$$\begin{bmatrix} \mathbf{X}_1\hat{\boldsymbol{\beta}} \\ \mathbf{X}_2\hat{\boldsymbol{\beta}} \end{bmatrix}.$$

Let $\mathbf{E}_1$ be a (36×36) diagonal matrix whose diagonal elements are the exposures (or premiums) in $\mathbf{X}_1$. In other words, $\mathbf{E}_1$ is Scale $\mathbf{A}_1$ (Exhibit 3) diagonalized and squared. Since the exposure must be positive, $\mathbf{E}_1$ is nonsingular. Let $\mathbf{J}_1 = \mathbf{E}_1^{-1}\mathbf{X}_1$. $\mathbf{J}_1$ has ones where $\mathbf{X}_1$ has positive numbers, and like $\mathbf{X}_1$ is zero everywhere else. Let $\mathbf{E}_2$ (28×28) be similarly defined, but with respect to $\mathbf{X}_2$. And let $\mathbf{J}_2 = \mathbf{E}_2^{-1}\mathbf{X}_2$.

The column totals of the a priori predictions are represented by the (8×1) matrix

$$[\mathbf{J}_1' \quad \mathbf{J}_2'] \begin{bmatrix} \mathbf{X}_1\hat{\boldsymbol{\beta}} \\ \mathbf{X}_2\hat{\boldsymbol{\beta}} \end{bmatrix} = \mathbf{J}_1'\mathbf{X}_1\hat{\boldsymbol{\beta}} + \mathbf{J}_2'\mathbf{X}_2\hat{\boldsymbol{\beta}}.$$

Similarly, the column totals of the a posteriori predictions are

$$[\mathbf{J}_1' \quad \mathbf{J}_2'] \begin{bmatrix} \mathbf{Y}_1 \\ \mathrm{E}[\mathbf{Y}_2 \mid \mathbf{Y}_1] \end{bmatrix} = \mathbf{J}_1'\mathbf{Y}_1 + \mathbf{J}_2'\mathrm{E}[\mathbf{Y}_2 \mid \mathbf{Y}_1].$$

The first balance property to be demonstrated is that these two vectors are equal, or that their difference is $\mathbf{0}_{(8\times 1)}$:

$$\begin{aligned}
\mathbf{0} &= \mathbf{J}_1'\mathbf{Y}_1 + \mathbf{J}_2'\mathrm{E}[\mathbf{Y}_2 \mid \mathbf{Y}_1] - \mathbf{J}_1'\mathbf{X}_1\hat{\boldsymbol{\beta}} - \mathbf{J}_2'\mathbf{X}_2\hat{\boldsymbol{\beta}} \\
&= \mathbf{J}_1'(\mathbf{Y}_1 - \mathbf{X}_1\hat{\boldsymbol{\beta}}) + \mathbf{J}_2'(\mathrm{E}[\mathbf{Y}_2 \mid \mathbf{Y}_1] - \mathbf{X}_2\hat{\boldsymbol{\beta}}) \\
&= \mathbf{J}_1'(\mathbf{Y}_1 - \mathbf{X}_1\hat{\boldsymbol{\beta}}) + \mathbf{J}_2'(\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}(\mathbf{Y}_1 - \mathbf{X}_1\hat{\boldsymbol{\beta}})) \\
&= (\mathbf{J}_1' + \mathbf{J}_2'\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1})(\mathbf{Y}_1 - \mathbf{X}_1\hat{\boldsymbol{\beta}}) \\
&= (\mathbf{J}_1'\boldsymbol{\Sigma}_{11} + \mathbf{J}_2'\boldsymbol{\Sigma}_{21})\boldsymbol{\Sigma}_{11}^{-1}(\mathbf{Y}_1 - \mathbf{X}_1\hat{\boldsymbol{\beta}}).
\end{aligned}$$

But, since $\hat{\beta} = (\mathbf{X}_1'\boldsymbol{\Sigma}_{11}^{-1}\mathbf{X}_1)^{-1}(\mathbf{X}_1'\boldsymbol{\Sigma}_{11}^{-1}\mathbf{Y}_1)$,

$$(\mathbf{X}_1'\boldsymbol{\Sigma}_{11}^{-1}\mathbf{X}_1)\hat{\beta} = (\mathbf{X}_1'\boldsymbol{\Sigma}_{11}^{-1}\mathbf{Y}_1).$$

Hence,

$$\mathbf{0} = \mathbf{X}_1'\boldsymbol{\Sigma}_{11}^{-1}(\mathbf{Y}_1 - \mathbf{X}_1\hat{\beta}).$$

Therefore, if we can show that $(\mathbf{J}_1'\boldsymbol{\Sigma}_{11} + \mathbf{J}_2'\boldsymbol{\Sigma}_{21})$ can be factored as the product of some matrix and $\mathbf{X}_1'$, then we will have proved the first balance property.

We need to define some more notation. Let $\mathbf{L}_1$ be the (36×36) diagonal matrix of the variance relativities that were introduced into the weighted regression to remove heteroskedasticity. In other words, $\mathbf{E}_1\mathbf{L}_1$ is Scale $\mathbf{B}_1$ (Exhibit 3) diagonalized and squared. Of course, since both $\mathbf{E}_1$ and $\mathbf{L}_1$ are diagonal, $\mathbf{E}_1\mathbf{L}_1 = \mathbf{L}_1\mathbf{E}_1$.

Therefore, $\boldsymbol{\Sigma}_{11} = \mathbf{L}_1^{0.5}\mathbf{E}_1^{0.5}\mathbf{P}_{11}\mathbf{E}_1^{0.5}\mathbf{L}_1^{0.5}$. (To be more accurate, we should introduce the proportionality constant σ^2. However, the same constant would apply to $\boldsymbol{\Sigma}_{21}$, and the balance property is unaffected.) $\mathbf{P}_{11}$ is shown in Exhibit 6, and it is easy to see that $\mathbf{P}_{11}$ and $\mathbf{E}_1$ commute. Therefore, $\boldsymbol{\Sigma}_{11} = \mathbf{L}_1^{0.5}\mathbf{P}_{11}\mathbf{L}_1^{0.5}$ times $\mathbf{E}_1$, where $\mathbf{E}_1$ can be inserted anywhere after the equals sign. $\mathbf{P}_{11}$ and $\mathbf{L}_1$ do *not* commute, so the two factors of $\mathbf{L}_1^{0.5}$ cannot be combined. So, $\mathbf{J}_1'\boldsymbol{\Sigma}_{11} = \mathbf{J}_1'\mathbf{E}_1\mathbf{L}_1^{0.5}\mathbf{P}_{11}\mathbf{L}_1^{0.5} = \mathbf{X}_1'\mathbf{L}_1^{0.5}\mathbf{P}_{11}\mathbf{L}_1^{0.5}$.

Moreover, $\boldsymbol{\Sigma}_{21} = \mathbf{L}_2^{0.5}\mathbf{E}_2^{0.5}\mathbf{P}_{21}\mathbf{E}_1^{0.5}\mathbf{L}_1^{0.5}$, where $\mathbf{L}_2$ (28×28) is similar to $\mathbf{L}_1$ in that $\mathbf{E}_2\mathbf{L}_2$ is Scale $\mathbf{B}_2$ (Exhibit 3) diagonalized and squared. $\mathbf{P}_{21}$ is also shown in Exhibit 6, and it is not hard to see that $\mathbf{E}_2\mathbf{P}_{21} = \mathbf{E}_2^{0.5}\mathbf{P}_{21}\mathbf{E}_1^{0.5} = \mathbf{P}_{21}\mathbf{E}_1$. Of course, the Es and the Ls commute. Therefore, $\mathbf{J}_2'\boldsymbol{\Sigma}_{21} = \mathbf{J}_2'\mathbf{E}_2\mathbf{L}_2^{0.5}\mathbf{P}_{21}\mathbf{L}_1^{0.5} = \mathbf{X}_2'\mathbf{L}_2^{0.5}\mathbf{P}_{21}\mathbf{L}_1^{0.5}$.

Let $\mathbf{L}$ without any subscript be the (8×8) diagonal matrix whose elements are the eight homoskedasticizing relativities. Again, it is not too hard to see that $\mathbf{L}\mathbf{X}_1' = \mathbf{X}_1'\mathbf{L}_1$, and that

$\mathbf{LX}_2' = \mathbf{X}_2'\mathbf{L}_1$. Therefore,

$$\begin{aligned}
(\mathbf{J}_1'\boldsymbol{\Sigma}_{11} + \mathbf{J}_2'\boldsymbol{\Sigma}_{21}) &= \mathbf{X}_1'\mathbf{L}_1^{0.5}\mathbf{P}_{11}\mathbf{L}_1^{0.5} + \mathbf{X}_2'\mathbf{L}_2^{0.5}\mathbf{P}_{21}\mathbf{L}_1^{0.5} \\
&= \mathbf{L}^{0.5}\mathbf{X}_1'\mathbf{P}_{11}\mathbf{L}_1^{0.5} + \mathbf{L}^{0.5}\mathbf{X}_2'\mathbf{P}_{21}\mathbf{L}_1^{0.5} \\
&= \mathbf{L}^{0.5}(\mathbf{X}_1'\mathbf{P}_{11} + \mathbf{X}_2'\mathbf{P}_{21})\mathbf{L}_1^{0.5} \\
&= \mathbf{L}^{0.5}(\mathbf{J}_1'\mathbf{E}_1\mathbf{P}_{11} + \mathbf{J}_2'\mathbf{E}_2\mathbf{P}_{21})\mathbf{L}_1^{0.5} \\
&= \mathbf{L}^{0.5}(\mathbf{J}_1'\mathbf{P}_{11}\mathbf{E}_1 + \mathbf{J}_2'\mathbf{P}_{21}\mathbf{E}_1)\mathbf{L}_1^{0.5} \\
&= \mathbf{L}^{0.5}(\mathbf{J}_1'\mathbf{P}_{11} + \mathbf{J}_2'\mathbf{P}_{21})\mathbf{E}_1\mathbf{L}_1^{0.5}.
\end{aligned}$$

We will show that $(\mathbf{J}_1'\mathbf{P}_{11} + \mathbf{J}_2'\mathbf{P}_{21})$ can be factored as $\mathbf{QJ}_1'$ for some $\mathbf{Q}$. Then

$$\begin{aligned}
(\mathbf{J}_1'\boldsymbol{\Sigma}_{11} + \mathbf{J}_2'\boldsymbol{\Sigma}_{21}) &= \mathbf{L}^{0.5}\mathbf{QJ}_1'\mathbf{E}_1\mathbf{L}_1^{0.5} \\
&= \mathbf{L}^{0.5}\mathbf{QX}_1'\mathbf{L}_1^{0.5} \\
&= \mathbf{L}^{0.5}\mathbf{QL}^{0.5}\mathbf{X}_1'.
\end{aligned}$$

As we saw earlier, this will amount to a proof of the first balance property. The considerable maneuvering to this point is to show that the result is independent of exposure and variance relativity, as long as exposure is constant by accident period and variance relativity is constant by age.

The remainder of the proof relies on the fact that the error correlation matrix is first-order autoregressive by accident period. The (36×8) matrix $\mathbf{J}_1$ can be considered as a left-justified stack of identity matrices:

$$\mathbf{J}_1 = \begin{bmatrix} \mathbf{I}_8 \\ \mathbf{I}_7 \\ \vdots \\ \mathbf{I}_1 \end{bmatrix} = \text{Left}(\mathbf{I}_8, \mathbf{I}_7, \ldots, \mathbf{I}_1).$$

This notation can be made formal. The gaps to the right side caused by the decreasing dimensions of the identity matrices are filled with zeroes. Therefore,

$$\mathbf{J}_1' = [\mathbf{I}_8 \quad \mathbf{I}_7 \quad \cdots \quad \mathbf{I}_1] = \mathrm{Top}(\mathbf{I}_8, \mathbf{I}_7, \ldots, \mathbf{I}_1).$$

We can write $\mathbf{P}_{11}$ as $\mathrm{Diag}(\mathbf{V}_8, \mathbf{V}_7, \ldots, \mathbf{V}_1)$, where $\mathbf{V}_i$ is $(i \times i)$. As an example,

$$\mathbf{V}_4 = \begin{bmatrix} \rho^0 & \rho^1 & \rho^2 & \rho^3 \\ \rho^1 & \rho^0 & \rho^1 & \rho^2 \\ \rho^2 & \rho^1 & \rho^0 & \rho^1 \\ \rho^3 & \rho^2 & \rho^1 & \rho^0 \end{bmatrix}.$$

According to the rules of multiplying partitioned matrices, $\mathbf{J}_1'\mathbf{P}_{11} = \mathrm{Top}(\mathbf{V}_8, \mathbf{V}_7, \ldots, \mathbf{V}_1)$.

Similarly, $\mathbf{J}_2 = \mathrm{Right}(\mathbf{I}_1, \mathbf{I}_2, \ldots, \mathbf{I}_7)$. However, this is a (28×7) matrix. But we need to make it a (28×8) matrix by padding it with an extra leftmost column of zeroes. So we will say that $\mathbf{J}_2 = \mathbf{0}_{(28\times 1)} \,\|\, \mathrm{Right}(\mathbf{I}_1, \mathbf{I}_2, \ldots, \mathbf{I}_7)$. And $\mathbf{J}_2' = \mathbf{0}_{(1\times 28)} // \mathrm{Bottom}(\mathbf{I}_1, \mathbf{I}_2, \ldots, \mathbf{I}_7)$.

It is not so easy to see how $\mathbf{J}_2'\mathbf{P}_{21}$ works. However, just as premultiplying $\mathbf{P}_{11}$ by $\mathbf{J}_1'$ had the effect of elevating the submatrices of $\mathbf{P}_{11}$ to the top of an (8×36) matrix, so too a little thought will convince the reader that premultiplying $\mathbf{P}_{21}$ by $\mathbf{J}_2'$ has the effect of dropping the submatrices of $\mathbf{P}_{21}$ to the bottom of an (8×36) matrix.

The sparse, or zero, areas of $\mathbf{J}_1'\mathbf{P}_{11}$ and $\mathbf{J}_2'\mathbf{P}_{21}$ are complements of each other. The first eight columns of $\mathbf{J}_1'\mathbf{P}_{11} + \mathbf{J}_2'\mathbf{P}_{21}$ contain $\mathbf{V}_8$. The next seven columns contain $\mathbf{V}_8$ without its last column. The next six columns contain $\mathbf{V}_8$ without its last two columns. The pattern continues down to the last column, which contains the first column of $\mathbf{V}_8$. But this is also the result of multiplying $\mathbf{V}_8$ by $\mathrm{Top}(\mathbf{I}_8, \mathbf{I}_7, \ldots, \mathbf{I}_1)$. Therefore,

$$\begin{aligned} \mathbf{J}_1'\mathbf{P}_{11} + \mathbf{J}_2'\mathbf{P}_{21} &= \mathbf{V}_8 \,\mathrm{Top}(\mathbf{I}_8, \mathbf{I}_7, \ldots, \mathbf{I}_1) \\ &= \mathbf{Q}\mathbf{J}_1'. \end{aligned}$$

This is the factorization that we sought, so the proof is complete. Notice that even though we proved the property for eight accident periods, the proof can easily be generalized to any number of periods greater than or equal to two.

The second property can be treated succinctly.

$$[\mathbf{J}_1' \quad \mathbf{J}_2'] \begin{bmatrix} \mathbf{Y}_1 \\ \mathrm{E}[\mathbf{Y}_2 \mid \mathbf{Y}_1] \end{bmatrix}$$

represents the column totals of the a posteriori predictions. Each column total needs to be divided by the total exposure. As an (8×8) matrix, this is

$$[\mathbf{J}_1' \quad \mathbf{J}_2'] \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{bmatrix} = \mathbf{J}_1'\mathbf{X}_1 + \mathbf{J}_2'\mathbf{X}_2.$$

One can verify that this is an (8×8) diagonal matrix, each diagonal element of which is the sum of all accident period exposures. So the column weighted averages are

$$\begin{aligned}
&\left([\mathbf{J}_1' \quad \mathbf{J}_2'] \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{bmatrix}\right)^{-1} \left([\mathbf{J}_1' \quad \mathbf{J}_2'] \begin{bmatrix} \mathbf{Y}_1 \\ \mathrm{E}[\mathbf{Y}_2 \mid \mathbf{Y}_1] \end{bmatrix}\right) \\
&= (\mathbf{J}_1'\mathbf{X}_1 + \mathbf{J}_2'\mathbf{X}_2)^{-1}(\mathbf{J}_1'\mathbf{Y}_1 + \mathbf{J}_2'\mathrm{E}[\mathbf{Y}_2 \mid \mathbf{Y}_1]) \\
&= (\mathbf{J}_1'\mathbf{X}_1 + \mathbf{J}_2'\mathbf{X}_2)^{-1}(\mathbf{J}_1'\mathbf{X}_1\hat{\beta} + \mathbf{J}_2'\mathbf{X}_2\hat{\beta}) \\
&= (\mathbf{J}_1'\mathbf{X}_1 + \mathbf{J}_2'\mathbf{X}_2)^{-1}(\mathbf{J}_1'\mathbf{X}_1 + \mathbf{J}_2'\mathbf{X}_2)\hat{\beta} \\
&= \hat{\beta}.
\end{aligned}$$