Conjoint Prediction of Paid and Incurred Losses

by Leigh J. Halliwell, ACAS

Conjoint Prediction of Paid and Incurred Losses

Abstract

Actuaries use paid and incurred methods to predict losses. Then one method is selected, and usually the information contained in the other(s) is jettisoned. Sometimes methods are weighted together, but without statistical justification. And even when an incurred method is deemed sufficient, present valuing will require predictions of loss payments. This paper will present a statistical model whereby paid and incurred losses are predicted together, or interdependently. As a result, the predictions of both paid and incurred losses will attain to the same ultimate amounts.

The model will be developed from statistical theory, using a simple example. Then a realistic example will be treated, and results therefrom will be present valued. One of the appendices will treat the financial theory of valuing stochastic cash flows, in which accepted theory regarding risk-adjusted rates of return will be challenged.

1. Introduction

When forecasting insured losses, whether for ratemaking or for reserving, actuaries will often make use of several data sources and methods. The idea of arriving at an opinion by several different approaches makes practical sense and is philosophically attractive.[1] When different approaches yield essentially the same forecast, i.e., when there is a consensus, the actuary is truly happy (and relieved). But even then a fear lingers in the back of his mind that something important might have been left out. However, all too often, different approaches yield different forecasts; and the actuary must blend them into a compromise, or reject at least one approach.

Frequently ratemaking and reserving involve the development of paid and incurred losses. Incurred losses are like an onion with several layers; from innermost to outermost they are paid losses, case reserves, IBNR reserves, and sometimes bulk and/or contingency reserves. It is possible to combine these forms of losses with several projection methods to arrive at dozens of forecasts. But usually there are less than ten substantially different approaches. If the actuary favors some and not others, he must explain why the case at hand should warrant the selection. But this is just a special instance of weighting the approaches, because rejecting an approach is equivalent to giving it zero weight. The actuary often uses the term 'credibility' when weighting alternative approaches, but in truth credibility theory is more of an art than a science when it comes to blending discordant forecasts.

This paper offers something different. Instead of forecasting paid and incurred losses separately and seeking a blend, why not forecast them simultaneously? Why not let the left hand know what the right hand is doing? Conjoint prediction of paid and incurred losses can be achieved within a statistical model. The data is allowed to speak for itself, which should obviate the charge of special pleading on the part of the actuary. Moreover, in addition to the conjoint feature, statistical models offer two possible advantages over traditional actuarial methods. First, a properly constructed statistical model yields unbiased predictions, unlike many actuarial methods. And second, a statistical model yields variance measures of the prediction errors. Suppose that an actuary says something like "Method A predicts $100,000 of losses with a standard deviation of $10,000, and Method B predicts $150,000 with a standard deviation of $20,000." Suppose also that a 50/50 credibility is reasonable to all. Then a weighted estimate of $125,000 would follow. However, the standard deviation is lost without gratuitous assumptions as to how independent the methods are. But a statistical model can handle the varieties of losses and methods simultaneously, all the while preserving the first and second moments.

2. A Simple Example of Conjoint Prediction

Exhibit 1 contains paid and incurred losses for a simple example: three accident years with all losses paid within three years. We will assume that the accident years have the same exposure (by definition, one unit); and that the expected incremental loss at each cell, or at each combination of AY and age, is the product of exposure and one of six factors. The factors, to be estimated in the model, depend on whether the loss is paid or incurred and on the age (1, 2, or 3). For simplicity we will also assume that the cells are homoskedastic, which means that the losses are all of the same variance and do not covary.

Exhibit 2 shows a first attempt to model the incremental losses. In order to avoid clutter, the zeroes in the matrices X_1 and X_2 are not printed. The first three columns of X correspond to incurred losses, the last three to paid losses. Appendix C derives the formulas which are shown in the exhibits. According to that appendix, in the linear statistical model:

$$\begin{matrix} \mathbf{y}_1 = X_1\beta + \mathbf{e}_1 \\ \mathbf{y}_2 = X_2\beta + \mathbf{e}_2 \end{matrix} \quad \text{, where } \mathrm{Var}\begin{bmatrix} \mathbf{e}_1 \\ \mathbf{e}_2 \end{bmatrix} = \Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$$

the best linear unbiased estimate of $\mathbf{y}_2$ is:

$$\begin{aligned} \hat{\mathbf{y}}_2 &= X_2\hat{\beta} + \Sigma_{21}\Sigma_{11}^{-1}(\mathbf{y}_1 - X_1\hat{\beta}) \\ \mathrm{Var}[\mathbf{y}_2 - \hat{\mathbf{y}}_2] &= \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12} + (X_2 - \Sigma_{21}\Sigma_{11}^{-1}X_1)\mathrm{Var}[\hat{\beta}](X_2 - \Sigma_{21}\Sigma_{11}^{-1}X_1)', \text{ where} \\ \hat{\beta} &= (X_1'\Sigma_{11}^{-1}X_1)^{-1}X_1'\Sigma_{11}^{-1}\mathbf{y}_1 \text{ and} \\ \mathrm{Var}[\hat{\beta}] &= (X_1'\Sigma_{11}^{-1}X_1)^{-1} \end{aligned}$$

The model of Exhibit 2 has the simplest variance matrix: Σ_{11} is σ^2 times an (18×18) identity matrix, Σ_{22} is σ^2 times a (6×6) identity matrix, and Σ_{12} and Σ_{21} are zero. Therefore, the formulas of Exhibit 3 simplify. $\mathbf{y}_2$ having been estimated, Exhibit 4 shows the completed loss rectangles.

In Model 1 the paid and the incurred losses are estimated separately.[2] One could complete the incurred rectangle without knowing the paid losses, and *vice versa*. As a result of this separation, the ultimate paid losses for AY2 and AY3 differ from the ultimate incurred losses. Moreover, if one were to make a rate for some future period, AY4, according to the incurred betas the rate would be 66.667 + 20 + 10 = 96.667; whereas according to the paid betas it would be 51.667 + 27.5 + 20 = 99.167 (Exhibit 3). The paid rate is 2.6 percent higher than the incurred rate.

This suggests a more sophisticated model. Why don't we constrain β such that the sum of the first three elements equals the sum of the last three? Hence, Model 2 of Exhibit 5 is Model 1 with the constraint that $[\,1 \;\; 1 \;\; 1 \; -1 \; -1 \; -1\,]\beta = [0]$. Appendices B and C show how one can determine that β satisfies this constraint if and only if there is a (5×1) γ such that:

$$\beta = \begin{bmatrix} 0.88808 & 0. & 0. & 0.19954 & 0.06949 \\ -0.26202 & 0.80783 & -0.11865 & 0.30139 & 0.08472 \\ -0.26202 & -0.30116 & 0.75893 & 0.30139 & 0.08472 \\ 0.26202 & 0.50667 & 0.64028 & -0.30139 & -0.08472 \\ 0.05958 & 0. & 0. & 0.75008 & -0.51688 \\ 0.04244 & 0. & 0. & 0.35363 & 0.84052 \end{bmatrix}\gamma$$

The problem is thus reduced from a β in 6-space with a 1-space constraint to an unconstrained γ in 5-space. In Exhibit 6 γ is estimated, then by transforming back to 6-space β is estimated, and finally we estimate $\mathbf{y}_2$. Exhibit 7 shows the completed loss rectangles.

One can see the effect of the constraint by comparing the completed loss rectangles of Exhibit 7 with those of Exhibit 4. The incurred losses of Exhibit 7 are a little higher than those of Exhibit 4, and the paid losses of Exhibit 7 are offsettingly lower. If one were to make a rate for future period AY4, according to the incurred betas the rate would be 66.894

+ 20.341 + 10.682 = 97 11/12; according to the paid betas it would be 51.439 + 27.159 + 19.318 = 97 11/12. The constraint of Model 2 has caused the paid and the incurred losses to work together enough to produce the same rate. But the ultimate paid losses for AY2 and AY3 still differ from the ultimate incurred losses.

What we need is a model which will make the ultimate paid losses equal to the ultimate incurred losses. Therefore, we must have a constraint for each accident year, AY_i:

$$AY_i \text{ Incremental: Incd @1} + \text{Incd @2} + \text{Incd @3} - \text{Paid @1} - \text{Paid @2} - \text{Paid @3} = 0$$

If $\mathbf{y}_{18\times1} = \begin{bmatrix} \mathbf{y}_{1\ 12\times1} \\ \mathbf{y}_{2\ 6\times1} \end{bmatrix}$, i.e., if $\mathbf{y}$ is ordered as in Exhibit 2, then the constraint would be $A\mathbf{y} = 0$,

where:

$$A_{3\times18} = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 & -1 & -1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & -1 & -1 & 0 & 1 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 1 & 1 & 0 & -1 & -1 \end{bmatrix}$$

Up until now, and in Appendix C, we treated constraints on β (i.e., $A\beta = b$). Conjoint-prediction of paid and incurred losses demands a new type of constraint, a constraint on $\mathbf{y}$.

Before returning to this particular constraint, consider what the general constraint $A\mathbf{y} = b$ means. Since $\mathbf{y} = X\beta + \mathbf{e}$, the constraint is also $AX\beta + A\mathbf{e} = b$. b is non-stochastic; therefore, $Var[A\mathbf{e}]$ must be zero. And since $E[A\mathbf{e}] = 0$, $AX\beta = b$. Therefore, the constraint $A\mathbf{y} = b$ is really two constraints: a constraint on β (viz., $AX\beta = b$) and a constraint on $\mathbf{e}$ (viz., $Var[A\mathbf{e}] = 0$, which implies that $A\mathbf{e} = E[A\mathbf{e}] = 0$). In the linear statistical model $Var[\mathbf{e}]$ was assumed to be Σ; and in general it is not true that $Var[A\mathbf{e}] = A\Sigma A' = 0$.

Therefore, if we want to constrain **y**, then we must modify Σ as $\Sigma^* = f(\Sigma, A)$ such that $A\Sigma^*A' = 0$. Obviously, if $\Sigma^* = 0$, then $A\Sigma^*A' = 0$. But this is a radical modification of Σ; we seek the least intrusive modification.

The answer is provided by the notion of quasi-observations, which is discussed in Appendix E. Think of $A\mathbf{e}$ as $\mathbf{e}_1$, and of $\mathbf{e}$ as $\mathbf{e}_2$. When we constrain that $A\mathbf{e} = 0$, it is as if we had observed $\mathbf{y}_1 = 0 + \mathbf{e}_1$ and found it to be 0. We want to estimate $\mathbf{y}_2 = 0 + \mathbf{e}_2$, so the model is:

$$\begin{aligned} 0 &= \mathbf{e}_1 \\ \mathbf{y}_2 &= \mathbf{e}_2 \end{aligned} \qquad \mathrm{Var}\begin{bmatrix} \mathbf{e}_1 \\ \mathbf{e}_2 \end{bmatrix} = \begin{bmatrix} \Sigma_{11} = A\Sigma A' & \Sigma_{12} = A\Sigma \\ \Sigma_{21} = \ \Sigma A' & \Sigma_{22} = \ \Sigma \end{bmatrix}$$

The model can be considered either as having a (0×1) β or as having a $(k\times1)$ β which is constrained to be 0. Either way, best linear unbiased estimation yields:

$$\begin{aligned} \hat{\mathbf{y}}_2 &= 0 \\ \mathrm{Var}[\mathbf{y}_2] &= \mathrm{Var}[\mathbf{y}_2 - 0] \\ &= \mathrm{Var}[\mathbf{y}_2 - \hat{\mathbf{y}}_2] \\ &= \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12} \\ &= \Sigma - \Sigma A'(A\Sigma A')^{-1}A\Sigma \end{aligned}$$

But $\mathrm{Var}[\mathbf{y}_2] = \mathrm{Var}[\mathbf{e}_2] = \mathrm{Var}[\mathbf{e}]$. If $A\Sigma A'$ happens to be singular, then use the MP inverse (Appendix B); so $\Sigma^* = f(\Sigma, A) = \Sigma - \Sigma A'(A\Sigma A')^{+}A\Sigma$. As a confirmation of the aptness of this modification, it can be proved that if A is $(m\times n)$ of rank m and Σ is $(n\times n)$ positive definite (Appendix A), then the rank of Σ^* will be $n - m$.[3] The modification is achieved with the least possible reduction of rank.

Therefore, Model 3 will be Model 2 with the additional constraint that $Ay = 0$, where A is the (3×18) matrix specified above. We keep the constraint of Model 2, viz., that [1 1 1 -1 -1 -1]β = [0], since this constrains the expected values of the losses, which is not the same as constraining the ultimate values. However, as can be seen in Exhibit 8, the additional constraint on β, viz., $AX\beta = 0$, happens to be a threefold repetition of the constraint of Model 2. Therefore, Model 3 differs from Model 2 only in that its variance matrix is $\Sigma^* = \sigma^2\Phi$, shown in Exhibit 8, whereas that of Model 2 is $\sigma^2 I_{18}$. It should be noticed that Φ is symmetric idempotent, so its rank is equal to its trace, which is $18^*(5/6) = 15$.

Model 3 is expressed in Exhibit 9, and in Exhibit 10 it is demonstrated that the (12×12) matrix Φ_{11} is non-singular, having a zero eigenvalue. As explained in Appendix C, the observed portion of the model can be transformed into a model with fewer observations, but with additional constraints (Exhibits 11 and 12). After transformation, observation 12 (Paid AY3@1) has no variance (and hence no covariance, as shown by the shaded region of Exhibit 12). It becomes a constraint, viz., [0.408 0.408 0.408 -0.408 -0.408 -0.408]β = [0], which happens not to add to the previous constraint [1 1 1 -1 -1 -1]β = [0].

Exhibit 13 is similar to Exhibit 5; however, the observed portion of the model has only eleven rows (since Paid AY3@1 became a constraint). Exhibit 14 is similar to Exhibit 6, but with a non-identity Φ matrix. Φ_{11}^*, Φ_{21}^*, Φ_{12}^* of Exhibit 14 are the Φ matrices of Exhibit 12 minus the shaded region. y_2 and its prediction error are derived in accordance with Appendix C, and the results are carried over to Exhibit 15. Finally we have a model whose ultimate paid losses and ultimate incurred losses are equal.

3. Efficiency Gains

Exhibit 16 compares some important variances among the three exemplary models. Each model has its own 6×6 prediction error variance matrix, $\mathrm{Var}[\mathbf{y}_2 - \hat{\mathbf{y}}_2]$ (Exhibits 3, 6, and 14). For now, let us call these matrices V_1, V_2, and V_3. Each row of the exhibit contains a (1×6) row vector A which determines a linear combination of the prediction errors, and the last three columns of the exhibit compare AV_iA'. So, for example, the row described as "Overall Balance" represents the variance of the difference of the total ultimate paid from the total ultimate incurred. Because the incurred design matrix is like the paid design matrix and σ^2 is the same for both paid and incurred losses, it is no surprise that for each V_i matrix the upper left (3×3) and lower right (3×3) submatrices are identical. Therefore, in Exhibit 16 the variances of the three "Incurred" rows are equal to those of the three "Paid" rows.

But from the exhibit it is evident that Model 2 has tighter prediction errors than does Model 1, and Model 3 is tighter still. Because the (3×3) off-diagonal covariance blocks of V_1 are zero, the "Balance" variances of Model 1 are the sums of the corresponding "Incurred" and "Paid" variances. In Model 2 there is some positive correlation between incurred and paid losses; hence, its "Balance" variances are less than those sums. And Model 3 was constructed with so much correlation between incurred and paid losses that its "Balance" variances must be zero.

The upper left (3×3) submatrices of the V_is fittingly represent the variance of the ultimate losses:

$$U_1 = [V_1]_{3\times3} = \begin{bmatrix} 198.6111 & 0 & 99.30556 \\ 0 & 148.9583 & 0 \\ 99.30556 & 0 & 198.6111 \end{bmatrix}$$

$$U_2 = [V_2]_{3\times3} = \begin{bmatrix} 147.444 & -11.6403 & 62.0819 \\ -11.6403 & 122.224 & -11.6403 \\ 62.0819 & -11.6403 & 147.444 \end{bmatrix}$$

$$U_3 = [V_3]_{3\times3} = \begin{bmatrix} 106.597 & 0.000 & 53.299 \\ 0.000 & 119.922 & -39.974 \\ 53.299 & -39.974 & 146.571 \end{bmatrix}$$

It can be shown that $U_1 - U_2$ is positive definite (PD), so $U_2 < U_1$ (cf. Appendix A). Therefore, Model 2 dominates Model 1. However, neither $U_1 - U_3$ nor $U_3 - U_1$ is PD; and similarly with U_2 and U_3. Exhibit 16 shows some important linear combinations of the predictions, and Model 3 dominates. But there are some (1×3) vectors A such that AU_1A' and/or $AU_2A' < AU_3A'$. Is there some measure of efficiency among the U_is, even though they rank differently depending on linear combination?

The determinant of a variance matrix, e.g., $|\Sigma|$, encapsulates in one scalar much of the information of the matrix (cf. Johnson [10: 104-108]). If **x** is an $(n\times1)$ random vector with mean μ and PD variance Σ, then $\{x: (x-\mu)'\Sigma^{-1}(x-\mu) < r^2\}$ is an n-ellipsoid which approximates a densest confidence region (cf. Appendix D). Σ^{-1} can be factored as W'W where W is non-singular. Define y as $(1/r)W(x-\mu)$, so $x = rW^{-1}y + \mu$. Then $\{x: (x-\mu)'\Sigma^{-1}(x-\mu) < r^2\}$ maps one-to-one onto $\{y: y'y < 1\}$, which region is a unit n-spheroid. Let dV_x

denote a differential volume in the x coordinate frame, and dV_y a differential volume in the y coordinate frame. Because of the one-to-one linear mapping between x and y, $dV_x = |rW^{-1}|dV_y = r^n|W^{-1}|dV_y = r^n|\Sigma|^{1/2}dV_y$.[4] If k_n denotes the n-volume enclosed by a unit n-spheroid:[5]

$$\begin{aligned}\int\limits_{(x-\mu)'\Sigma^{-1}(x-\mu)<r^2} dV_x &= \int\limits_{y'y<1} dV_x \\ &= \int\limits_{y'y<1} dV_y r^n|\Sigma|^{1/2} \\ &= k_n r^n |\Sigma|^{1/2}\end{aligned}$$

Thus we have the formula for the n-volume of $\{x: (x - \mu)'\Sigma^{-1}(x - \mu) < r^2\}$. If **x** is of high variance, then this volume will be great; if of low variance, it will be small. Therefore, one way to compare the efficiency of models is to compare the square roots of the determinants of their variance matrices. The determinants of the U_is are respectively 4,406,900, 2,162,922, and 1,362,668; the square roots are 2,099, 1,471, and 1,167. Efficiency could be defined as inversely proportional to the square roots. Hence, we could say that the efficiency of Model 2 is 2099/1471 = 1.43 times than of Model 1, or that Model 2 is 43 percent more efficient than Model 1. Similarly, Model 3 is 80 percent more efficient than Model 1, and 26 percent more efficient than Model 2.

Anderson [2: 259] calls $|\Sigma|$ the generalized variance of **x**. If Σ is diagonal, then $|\Sigma|$ is the product of the variances. Appendix A shows that $|\Sigma|$ is the product of the eigenvalues of Σ, which are the variances of a suitable orthogonal transformation of **x**. Therefore, according to Anderson's definition, the dimension of the generalized variance is variance raised to the n^{th} power. The author would like to modify the definition of generalized variance as $\sqrt[n]{|\Sigma|}$.

In other words, in this paper the generalized variance will be the geometric mean of the eigenvalues, and will have the dimension of variance. Accordingly, the generalized variances of the U_is are the cube roots of their determinants, or 163.950, 129.34, and 110.866. And, defining efficiency as inversely proportional to the generalized variance, we could say that the efficiency of Model 2 is 163.950/129.324 = 1.27 times that of Model 1, or that Model 2 is 27 percent more efficient than Model 1. Similarly, Model 3 is 48 percent more efficient than Model 1, and 17 percent more efficient than Model 2.

Though $|\Sigma|^{1/3}$ and $\sqrt[3]{|\Sigma|}$ yield different efficiency gains, at least they rank the models in the same order.

4. A Realistic Example of Conjoint Prediction

In an earlier paper [7] the author applied a linear statistical model to a medical provider that self-insured its workers compensation. In that paper the author dealt only with paid losses, which are reproduced in Exhibit 17. In this exhibit 'FY' means both 'Fund Year' and 'Fiscal Accident Year'. The fund's fiscal years commence on April 1. The losses are as of 31Mar1995 (FY 1994 as of 12 months). The exposures (employee payroll in units of a hundred dollars), adjusted to 1995 expected conditions,[6] are also shown, the exposure for FY 1995 being a forecast. As the earlier paper explained [7: 5], in self-insurance evaluations it is common to estimate not only the losses already incurred, but also the losses which will be incurred before the next evaluation. Actuaries commonly think of the former type of estimation as reserving, and of the latter type as ratemaking; but this distinction is immaterial to statistical modeling.

The portion of the paid loss rectangle above and to the left of the staircase shows observed payments, whether they be cumulative or incremental. In the exhibit the rectangle is completed in two ways, by the Chain Ladder Method and by the Additive Method. Both methods are well known to actuaries, and are described by Stanard [18: 130f.]. The development factors and the ratios of each method are weighted-averaged, i.e., rows are included in column sums only if both the columns being ratioed have observations. So, for example, the development factor @60, 1.091, equals (526,989 + 1,023,551 + 1,016,903) / (518,865 + 904,916 + 930,475). And the ratio @60, 0.514, equals (8,124 + 118,635 + 86,428) / (131,332.20 + 141,672.24 + 141,677.29). The age of the oldest FY is 84 months as of 31Mar95, so all movements after 84 months (movements to ultimate) are unobserved. For a reason which will be explained in Section 7 the author chose to call ultimate 108 months. Data from a rating bureau suggested that a FY at 84 months is 90 percent paid out; therefore, the development factor from 84 months to ultimate is 1.00/0.90 = 1.111. This also means that the payments after 84 months are expected to be ten percent of the total payments, and one ninth of the total payments before 84 months.

The Chain Ladder Method uses the development factors to project FY 1988-1994 losses to ultimate. Each FY's ultimate divided by its exposure yields its pure premium, and the weighted-average pure premium for 1988-1994 is 7.94. Since the exposures have been adjusted to 1995 conditions, the ultimate paid loss of FY 1995 may be estimated as 115,000

$\times$ 7.94 = 912,905. The Additive Method formed the ratios of incremental paid losses to exposures for ages 12 to 84. The ratio for age 108 equals one ninth of the sum of the other ratios. By these two methods ultimate paid losses are estimated, and the results are shown in the boxes of the exhibit.

Exhibit 18 is entirely analogous to Exhibit 17, but treats of incurred losses, or paid losses plus case reserves. The rating-bureau data suggested that at 84 months the case-incurred losses would be 95 percent of ultimate, which implies a development factor to ultimate of 1/0.95 = 1.053. Therefore, the incurred amounts after 84 months are expected to be one twentieth of the total incurred, and one nineteenth of the total incurred before 84 months. We now have two estimates of ultimate incurred losses, which makes for four estimates of ultimate losses. The incurred estimates are much lower than the paid; however, the actual amounts are not important. What is important for our purpose is that we in typical actuarial fashion have produced several estimates and are now faced with the problem of harmonizing them. Moreover, we have no idea as to how the much the ultimate losses will vary from any of our estimates, since such an idea requires statistical modeling.

We will work up to a conjoint model by degrees. As a first step, Exhibit 19 shows a simple paid loss model (Model 1). A model with 64 rows (28 observations and 36 predictions) is a bit large for a spreadsheet, and its solution will not be worked out, as were the exemplary models (Exhibits 3, 6, and 14). Appendix G contains the SAS® code used to generate the solutions of this and the following models. The solution of Model 1 is given in Exhibit 20. It is important to understand that since there is no observation at age 108, β_{108} could not be estimated without the constraint. Exhibits 21 and 22 are analogous to Exhibits 19 and 20, except that they pertain to incurred losses (Model 2).[7]

A first pass through the data indicates that the paid loss model, Model 1, is superior to the incurred loss model, Model 2. The estimate of σ^2 and the variances of Model 1 are much less than those of Model 2. It is for this reason that the author ignored the incurred losses in the earlier paper. However, it is a mistake to think that "not as good" means "no good;" as will be seen, the incurred losses are still useful.

Model 3 (Exhibit 23), which for the first time puts together paid and incurred losses, is like Exemplary Model 1 (Exhibit 2). That Model 3 has two constraints is immaterial, since the constraints refer to rows that are not observed. The difference of Model 3 from Exemplary Model 1 is that the variances of the incurred rows of Model 3 are not unity, but rather the ratio of the estimate of the σ^2 of the incurred model to that of the paid model (about 2.089). Because of the block-diagonal nature of the matrices X and Φ, Model 3 with its 128 rows will reproduce both Models 1 and 2, each with 64 rows. So as yet there is no commingling of paid and incurred losses.

Exhibit 24 shows the conditions which must be imposed on Model 3 in order for it to perform conjoint prediction. The matrix G differences incurred losses from paid losses for each FY. So the new Φ matrix, with the constraint that $GX\beta = 0$, guarantees that for each FY the ultimate paid and incurred losses will be equal.[8] The constraint $[\ 1_{1\times 8}\ \ -1_{1\times 8}]\beta = 0$ guarantees the equality of the paid and incurred pure premiums. GX is eight columns of exposure followed by eight columns of negative exposure; so $GX\beta = 0$ adds nothing the pure premium constraint. The results of Model are shown in Exhibit 25. It is satisfying to see that the ultimate paid losses and their variances are the same as the ultimate incurred.

Exhibit 26 gives an idea as to how much conjoint prediction (Model 4) improves upon the separate paid and incurred models (Models 1 and 2). The estimate of σ^2 of the incurred model is more than twice that of paid model; so it is surprising (at least to the author) how much the conjunction of the losses reduces the FY variances, about a one-third reduction from Model 1 variances for an efficiency gain of 50 percent.[9]

5. Comparing the Results of the Methods

Exhibit 27 summarizes the various methods of arriving at ultimate losses. Model 3 is not included because, as explained in the previous section, it combines but does not surpass Models 1 and 2. Except for the conjoint model, the paid losses project higher than do the incurred losses. Therefore, it is not surprising that the constraint of the conjoint model, that ultimate paid and incurred losses be equal by fund year, pushes the ultimate paid losses down and pulls the incurred losses up. The additive methods are not much different from Models 1 and 2, the difference being explained in footnote 7. Stanard [18] has noted that in simulations the chain ladder method with paid losses seems biased upward (cf. also [8: Section 2]). Unlike the development factors of paid losses, which are greater than one, the development factors of incurred losses may be less than one, as is the case with the development factor from 72 to 84 months in Exhibit 18. This can give the chain ladder method with incurred losses a downward bias, as here. Therefore, the ordering of the methods is reasonable:

Incd CL < Incd Additive ≈ Model 2 < Conjoint < Model 1 ≈ Paid Additive < Paid CL

The reduction of the incurred losses of FY 1988 between 72 and 84 months causes the development factor of Exhibit 18 to be less than one, as well as the ratio of the same exhibit and $\hat{\beta}_{84}$ of Model 2 (Exhibit 22) to be negative. Hence, the projected incremental incurred losses between 72 and 84 months for the subsequent fund years are negative. However, the incremental incurred projections of the conjoint method for this period are positive for all fund years except 1991 (Exhibit 25). As mentioned above, this is due to the upward pull of the conjoint method on the incurred losses. That FY 1991 is still negative is explained as follows: in Exhibit 27 the ultimate losses of Model 1 are on average about twenty percent higher than the corresponding losses of Model 2. But the FY 1991 ultimate loss of Model 1 is only 2.2 percent higher than that of Model 2. Therefore, the upward pull of the conjoint model on the incurred losses of FY 1991 is mild, and the 72:84 increment stays negative.

Exhibits 28 and 29 show paid and incurred loss patterns. Exhibits 17 and 18 actually display the chain-ladder (CL) and additive patterns. The other patterns can be derived from the FY 1995 ratios of cumulative sums to ultimate in Exhibits 20, 22, and 25. Every paid pattern is constrained by the rating-bureau data to be 90 percent at 84 months, as is every incurred pattern to be 95 percent. It is consistent with the ordering of the methods given above that the paid CL pattern is slower than the paid conjoint pattern, and that the reverse relation holds for the incurred patterns.

6. Quasi-Observations and Collateral Information

In Model 1 of the self-insured entity (Exhibits 19 and 20) bureau data was invoked that payments after 84 months are expected to be one ninth of the total payments before 84 months; thus $[\,1\ \ 1\ \ 1\ \ 1\ \ 1\ \ 1\ \ 1\ {-9}]\beta = [0]$. This is extraneous, or collateral, information. However, β_{108} is, so to speak, at the mercy of β_{12} through β_{84}. Even its variance and its covariances with the other elements of β are determined. There is another way to incorporate collateral information into a statistical model, viz., by quasi-observations. In effect, this is how the author treated collateral information in the earlier paper, although not explicitly. Since it is not obvious how the earlier treatment used quasi-observations and it makes for a good example, the problem will be redone here as Model 5.

Model 5 is like Model 1 (Exhibit 19); however, in the earlier paper [7: 9] the author introduced the constraint that the pure premium of the total payments before 84 months is 7.213, or $[\,1\ \ 1\ \ 1\ \ 1\ \ 1\ \ 1\ \ 1\ \ 0]\beta = [7.213]$. Unlike the constraint of Model 1, this constraint affects the observed rows the model. The bureau data was invoked that the remaining payments to ultimate should be one ninth of 7.213, or 0.801. Also, the author argued that the variance of the pure premium of the remaining payments is 0.2128, and that it does not covary with the pure premiums before age 84 [7: 12]. So, it is as if we had observed 0.801

as β_{108} plus an error term, where we knew the variance of the error term to be 0.2128. Exhibit 30 shows the quasi-observation appended to the real observations.

But the quasi-observation has a definite variance, 0.2128, whereas the variance of the true observations is σ^2, which without the quasi-observation was estimated as 6.2717×10^9 [7: 26].[10] Therefore, in the Φ matrix we make its variance relativity $0.2128/6.2717\times10^9$. Due to the block-diagonal relation between the real observations and the quasi-observation, the resulting estimate of σ^2 will not change;[11] it will still be 6.2717×10^9. The SAS® code for solving Model 5 also appears in Appendix G, and the results are shown in Exhibit 31. They are identical to the results of the earlier paper [7: 30].

Most of the variances of Model 5 are greater than those of Model 4. However, the seven and eight FY aggregate variances of Model 5 are less, and so too is the generalized variance. So by some measures Model 5 is superior to Model 4. However, Model 5 imposed an extra constraint, viz., that the pure premium of the total payments before 84 months is 7.213. Therefore, by all means, Method 4 compares favorably with Method 5.

The method of quasi-observations is well suited for the estimation of movements after the most aged observation (or, movements to ultimate), perhaps even better suited than the method of constraining the ultimate to be a function of the observed. But its disadvantage is that it requires more collateral information or assumptions.

7. Present Valuing the Future Loss Payments

Even in the unlikely event that incurred-loss data were deemed sufficient for the estimation of ultimate losses, it might still be necessary to estimate the future loss payments. Despite the slowness of accounting to recognize the present value of unpaid losses, more and more often actuaries are being asked for the present value of their loss predictions. If $\mathbf{y}_2$ represents the unobserved paid losses, of which $\hat{\mathbf{y}}_2$ is the best linear unbiased estimator (BLUE), then the present value of these paid losses will be $D\mathbf{y}_2$, where D is a diagonal matrix of discount factors. In Appendix C it was proved that if $\hat{\mathbf{y}}_2$ is the BLUE of $\mathbf{y}_2$, then the BLUE of $D\mathbf{y}_2$ is: $\widehat{(D\mathbf{y}_2)} = D\hat{\mathbf{y}}_2$. Moreover, as for the prediction error matrix:

$$\begin{aligned}
\mathrm{Var}[D\mathbf{y}_2 - \widehat{(D\mathbf{y}_2)}] &= \mathrm{Var}[D\mathbf{y}_2 - D\hat{\mathbf{y}}_2] \\
&= \mathrm{Var}[D(\mathbf{y}_2 - \hat{\mathbf{y}}_2)] \\
&= D\mathrm{Var}[\mathbf{y}_2 - \hat{\mathbf{y}}_2]D'
\end{aligned}$$

Therefore, the present valued estimates and prediction errors are readily obtainable from the nominal quantities.

Hence, one needs only to know the time of each payment and the discount factor appropriate to that time. The common opinion is that discount factors should depend somehow on the variability of the payments. But Appendix F argues that there is only one discount factor, viz., the so-called "risk free" discount factor. So loss estimates should be present valued as if they were certain. However, the sum of present value of the payments is disconnected from the value that they might have to someone, or from the price which at which they might be exchanged. Appendix F suggests that utility theory is the bridge from present value to price.

Exhibit 32 shows the present value of the loss estimates of Model 4. The present is the time of the evaluation, 31Mar95. As explained more fully in the earlier paper [7: 13f.], on that date yields on stripped US Treasury securities (effectively, risk-free zero-coupons) were obtained from the financial press. For simplicity it was assumed that a loss would be paid at the midpoint of its period. For example, FY 1995 @12 would be paid at FY 1995 @6. Since 31Mar95 is FY 1995 @0, this payment will occur six months into the future. The yield to a maturity of six months was found to be 6.03 percent; therefore, the discount factor is $(1.0603)^{-0.5} = 0.971$. The author decided that all payments after 84 months would be considered paid at 102 months, which effectively puts ultimate at 108 months. Therefore, every payment occurs either 0.5, 1.5, 2.5, 3.5, 4.5, 5.5, 6.5, 7.5, or 8.5 years from the present. The corresponding yields to maturity are 6.03, 6.36, 6.84, 6.99, 7.04, 7.14, 7.15, 7.21, and 7.21; therefore, the discount factors are 0.971, 0.912, 0.848, 0.789, 0.736, 0.684, .0638, 0.593, and 0.553. The SAS® code can be found at the end of Appendix G.

8. Refinements

Very little is perfect in this life; there is almost always room for improvement. The author disclaims that the models of this paper cannot be improved. Aside from the modified variance matrix $\Sigma^* = \Sigma - \Sigma A'(A\Sigma A')^{+}A\Sigma$, all variance matrices have been diagonal. Furthermore, most of these diagonal matrices have been of the form $\sigma^2\Phi = \sigma^2 I$. But quite likely, the variance of the observations varies by age. To account for this heteroskedasticity (viz., that $\Phi \neq I$, but is at least diagonal) one could have variance relativities, dependent on age, down the diagonal of Φ.

Another well known complication is autocorrelation, which means that Φ has non-zero elements off the diagonal. Autocorrelation is easily understandable in the context of conjoint prediction. Incurred losses involve the psychology of loss adjusters. It is possible that as paid losses are exceeding expectations, the loss adjusters are increasing case reserves. This would imply a positive correlation of incurred losses concurrent with paid losses. On the other hand, the adjusters might interpret this as a faster closure of claims, and decrease case reserves. This would imply a negative concurrent correlation. Also, it is possible that the loss adjusters might react to the paid losses of the previous period(s), which would give rise to correlation of incurred losses subsequent to paid losses. Or, the adjusters might be prescient, affecting the reserves before the payments, which would entail precedent correlation. Even more complicated, the correlation could be a local phenomenon, being, for example, concurrent at some accident or calendar periods and subsequent at others (perhaps even cyclical). The possibilities are limitless, and to investigate them is daunting.[12] The actuary must judge whether the possible gain in precision is worth the additions in effort, in the risk of error, and in the difficulty of explaining the model to others.

9. Conclusion

Blending predictions has always been both a problem and a challenge to actuaries. A statistical model, whether or not it be linear, can incorporate many types of data at once. Paid and incurred loss data are two obvious types. All that is needed is to deepen the theory of estimation to allow for constraints and singular variance matrices. Then one can estimate $\mathbf{y}_2$ and $\mathrm{Var}[\mathbf{y}_2 - \hat{\mathbf{y}}_2]$ in the very general model:

$$\mathbf{y} = \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix} = \begin{bmatrix} X_1\beta \\ X_2\beta \end{bmatrix} + \begin{bmatrix} \mathbf{e}_1 \\ \mathbf{e}_2 \end{bmatrix}, \text{ where } \mathrm{Var}\begin{bmatrix} \mathbf{e}_1 \\ \mathbf{e}_2 \end{bmatrix} = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix},$$

$$\begin{gathered} \text{subject to} \\ A\beta = b \\ Cy = d \end{gathered}$$

This deepening is performed in the appendices.

As for results, consider the example of the self-insured entity. It was fairly obvious that the loss reserves were inadequate. Many actuaries, like the author himself in the earlier paper, would write them off and deal only with the paid losses. However, even though the relative strength of the paid model to the incurred was as of a lion to a mouse, when the two were conjoined it was of appreciable benefit even to the lion. The moral here is that inferior data is still useful data, and that efficiency gains can be squeezed out of unlikely places. Moreover, when a statistical model predicts loss payments, it is easy to derive the first two moments of their present value, which is requisite to pricing.

This paper is long and at times difficult, but fairly self-contained. Even apart from conjoint prediction, the author hopes to impress upon the readers the enormous and untapped power of statistical modeling.

[1] Ancient wisdom on the subject of using different approaches is found in the biblical book of Ecclesiastes 4:9-12, especially v.12: "A threefold cord is not quickly broken." In a philosophical vein, according to Eric T. Bell (Men of Mathematics, New York, Simon and Schuster, 1965, p. 227) Carl Friedrich Gauss, "the prince of mathematicians," was not satisfied with a theorem until he had proved it in more than one way. Perhaps one reason for his doing this was to assure himself that he had not erred. But the major reason was that from several approaches he hoped to gain more understanding and leads to other theorems. In logic and mathematics one has assurance that different approaches will attain to the same truth; otherwise there is a contradiction in the assumptions. The actuary, however, has no such assurance when forecasting losses.

[2] The only thing estimated conjointly in Model 1 is σ^2, which is irrelevant to the purposes of this example. But for simplicity it was assumed that the paid and the incurred losses are homoskedastic, so that they have a common σ^2.

[3] Here is a proof that if A is $(m \times n)$ of rank m and Σ is $(n \times n)$ positive definite (Appendix A), then the rank of Σ^* will be $n - m$. At the end of Appendix A it is shown that under these conditions $A\Sigma A'$ is non-singular, so $\Sigma^* = \Sigma - \Sigma A'(A\Sigma A')^{-1}A\Sigma$ exists. Also in Appendix A it is shown that there exists an $(n \times n)$ matrix W such that $\Sigma = WW'$. The rank of W must be n. Then, $\Sigma^* = WW' - WW'A'(A\Sigma A')^{-1}AWW' = W(I_n - W'A'(A\Sigma A')^{-1}AW)W' = W(M)W'$. Because W is non-singular, the rank of Σ^* is equal to the rank of M. But M is a symmetric idempotent matrix (check it), and it was shown in Appendix B that the rank of such a matrix is equal to its trace. Therefore, using properties of the trace operator (Judge [11: 927]), $\rho(\Sigma^*) = \rho(M) = Tr(M) = Tr(I_n - W'A'(A\Sigma A')^{-1}AW) = Tr(I_n) - Tr(W'A'(A\Sigma A')^{-1}AW) = Tr(I_n) - Tr((A\Sigma A')^{-1}AWW'A') = Tr(I_n) - Tr((A\Sigma A')^{-1}A\Sigma A') = Tr(I_n) - Tr(I_m) = n - m$.

[4] If $y = Ax + b$, where A is non-singular, then $dV_y = |A|dV_x$. Schneider [16: 161-172] introduces the determinant as a volume function. This is a more intuitive and elegant way of developing the theory of determinants than the usual way of permutations. It makes for easy apprehension of the equation $dV_y = |A|dV_x$.

[5] Johnson [10: 107] states that $k_n = \dfrac{2\pi^{n/2}}{n\Gamma(n/2)}$, but does not mention that this is the formula for the volume of a unit n-spheroid. The recursive relation is $k_{n+1} = k_n \dfrac{n}{n+1}\dfrac{\Gamma(n/2)\Gamma(1/2)}{\Gamma((n+1)/2)}$, which uses the fact that $\Gamma(1/2) = \pi^{1/2}$. Once one knows k_n, by induction $k_{n+1} = \int_{-1}^{1} k_n(1-u^2)^{n/2}\,du$. This amounts to approximating a unit $(n+1)$-spheroid as many $(n+1)$-discs whose bases are n-spheroids, and will yield the recursive relation.

[6] In [7: 5-7] the author argued that it is the exposures (or the control variables, the columns of the design matrix) rather than the losses (or the observed quantities) that should be adjusted to account for changing conditions. The reader is urged to refer to the argument.

[7] The reader may be wondering why the estimate of β in Exhibit 20 is not the same as the "Ratio" row of Exhibit 17, even though it is close (similarly with Exhibits 18 and 22). The reason is that the observations of Exhibit 20 are assumed to be homoskedastic, so that each element of β is of the form $\frac{\sum x_i y_i}{\sum x_i x_i}$, whereas the ratios of Exhibit 17 are of the form $\frac{\sum y_i}{\sum x_i}$. If Exhibit 17 were modeled, its variance matrix would be heteroskedastic. Another common form is $\frac{\sum \frac{y_i}{x_i}}{\sum 1}$. All three forms imply different variance matrices. This has been frequently explained in the actuarial literature, e.g., Mack [12: 122], Murphy [13: 188, 232], and Peck [14: 104f.].

[8] Actuaries sometimes adjust paid and incurred n^{th}-to-ultimate development factors to force certain paid and incurred ultimate losses to be equal. But if the adjustment were to force a general equality, or an equality over several fund years, then there would not be equality in any particular fund year. And if the adjustment were to force an equality in one fund year, there would not be equality in the others. Conjoint prediction obviates the need for such an unsatisfactory adjustment.

[9] Because the paid model (Model 1) had more explanatory power than the incurred model (Model 2), it is not surprising that the incurred losses receive a tremendous efficiency gain from conjoint prediction. In explanatory power, Model 1 is to Model 2 as a lion is to a mouse. But there is a children's story of a lion and a mouse: the lion condescended to the mouse until a thorn lodged in his paw, which the mouse was able to remove. So here too, the incurred losses are of significant gain to the paid losses.

[10] Despite the similarity of Model 5 to Model 1, the estimate of σ^2 of Model 5 (6.2717×10^9) differs from that of Model 1 (6.5637×10^9) because in Model 5 the sum of β_{12} though β_{84} is constrained to be 7.213.

[11] As a proof, consider a model whose observations are block diagonal with no covariance between the blocks:

$$\mathbf{y} = X_{t\times k}\beta + \mathbf{e}$$

$$\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix} = \begin{bmatrix} X_{1\ t_1\times k_1} & \\ & X_{2\ t_2\times k_2} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} \mathbf{e}_1 \\ \mathbf{e}_2 \end{bmatrix}, \text{ where } \mathrm{Var}\begin{bmatrix} \mathbf{e}_1 \\ \mathbf{e}_2 \end{bmatrix} = \Sigma = \begin{bmatrix} \Sigma_{11} & \\ & \Sigma_{22} \end{bmatrix} = \sigma^2 \begin{bmatrix} \Phi_{11} & \\ & \Phi_{22} \end{bmatrix}$$

If X is of full column rank, then so too must be X_1 and X_2. Therefore:

$$\hat{\beta} = \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} = \begin{bmatrix} (X_1'\Phi_{11}^{-1}X_1)^{-1}X_1'\Phi_{11}^{-1}\mathbf{y}_1 \\ (X_2'\Phi_{22}^{-1}X_2)^{-1}X_2'\Phi_{22}^{-1}\mathbf{y}_2 \end{bmatrix}$$

$$\hat{\mathbf{e}} = \mathbf{y} - X\hat{\beta} = \begin{bmatrix} \mathbf{y}_1 - X_1\hat{\beta}_1 \\ \mathbf{y}_2 - X_2\hat{\beta}_2 \end{bmatrix} = \begin{bmatrix} \hat{\mathbf{e}}_1 \\ \hat{\mathbf{e}}_2 \end{bmatrix}$$

According to the formula of Appendix C:

$$\begin{aligned}
\hat{\sigma}^2 &= \hat{\mathbf{e}}'\Phi^{-1}\hat{\mathbf{e}}/(t-k) \\
&= (\hat{\mathbf{e}}_1'\Phi_{11}^{-1}\hat{\mathbf{e}}_1 + \hat{\mathbf{e}}_2'\Phi_{22}^{-1}\hat{\mathbf{e}}_2)/(t-k) \\
&= ((t_1-k_1)\frac{\hat{\mathbf{e}}_1'\Phi_{11}^{-1}\hat{\mathbf{e}}_1}{t_1-k_1} + (t_2-k_2)\frac{\hat{\mathbf{e}}_2'\Phi_{22}^{-1}\hat{\mathbf{e}}_2}{t_2-k_2})/(t-k) \\
&= ((t_1-k_1)\hat{\sigma}_1^2 + (t_2-k_2)\hat{\sigma}_2^2)/(t-k) \\
&= \frac{(t_1-k_1)\hat{\sigma}_1^2 + (t_2-k_2)\hat{\sigma}_2^2}{(t_1-k_1) \quad + (t_2-k_2)}
\end{aligned}$$

This shows that the estimate of σ^2 of the composite model is a weighted average of the estimates of the separate models, the weights being the degrees of freedom, $t_i - k_i$. If the second model is a quasi-observation, then $X_2 = I$, which is of full column rank, and $\hat{\mathbf{e}}_2 = 0$. Also, $t_2 = k_2$; hence, $t - k = (t_1 - k_1) + (t_2 - k_2) = t_1 - k_1$. Therefore:

$$\begin{aligned}
\hat{\sigma}^2 &= (\hat{\mathbf{e}}_1'\Phi_{11}^{-1}\hat{\mathbf{e}}_1 + \hat{\mathbf{e}}_2'\Phi_{22}^{-1}\hat{\mathbf{e}}_2)/(t-k) \\
&= (\hat{\mathbf{e}}_1'\Phi_{11}^{-1}\hat{\mathbf{e}}_1 + 0'\Phi_{22}^{-1}0)/(t-k) \\
&= \hat{\mathbf{e}}_1'\Phi_{11}^{-1}\hat{\mathbf{e}}_1/(t_1-k_1) \\
&= \hat{\sigma}_1^2
\end{aligned}$$

So the quasi-observation has no effect on the estimate of σ^2

[12] For discussions of incorporating heteroskedasticity and autocorrelation into the variance matrix see [8] and [11: Chapter 9]. One of the beauties of linear unbiased estimation is that even if one chooses a simplistic, or even an erroneous, variance structure, the estimates will still be unbiased. However, the estimates are no longer best.

References

1. Amemiya, Takeshi, Advanced Econometrics, Cambridge, MA, Harvard University Press, 1985.

2. Anderson, T. W., An Introduction to Multivariate Statistical Analysis (Second Edition), New York, John Wiley & Sons, 1984.

3. Bailey, Arthur, "A Generalized Theory of Credibility," *PCAS* XXXII, 1945, 13-20.

4. Bowers, N. L., Gerber, H. U., *et al.*, Actuarial Mathematics, 1986.

5. Eves, Howard, Elementary Matrix Theory, New York, Dover Publications, 1966.

6. Halliwell, Leigh J., "Portfolio Optimization and the Capital Asset Pricing Model: A Matrix Approach," Casualty Actuarial Society Forum Fall 1995, Casualty Actuarial Society, 1995, 355-368.

7. Halliwell, Leigh J., "Statistical and Financial Aspects of Self-Insurance Funding," Alternative Markets/Self Insurance, 1996 Discussion Paper Program, Casualty Actuarial Society, 1995, 1-46.

8. Halliwell, Leigh J., "Loss Prediction by Generalized Least Squares," *PCAS* LXXXIII, 1996.

9. Healy, M. J. R., Matrices for Statistics, Oxford, Clarendon Press, 1986.

10. Johnson, Richard A., and Wichern, Dean W., Applied Multivariate Statistical Analysis (Third Edition), Englewood Cliffs, NJ, Prentice Hall, 1992.

11. Judge, George G., Hill, R. C., *et al.*, Introduction to the Theory and Practice of Econometrics (Second Edition), New York, John Wiley & Sons, 1988.

12. Mack, Thomas, "Measuring the Variability of Chain Ladder Reserve Estimates," Casualty Actuarial Society Forum Spring 1994, Casualty Actuarial Society, 1994, 101-182.

13. Murphy, Daniel M., "Unbiased Loss Development Factors," Casualty Actuarial Society Forum Spring 1994, Casualty Actuarial Society, 1994, 183-245.

14. Peck, Edward F., "Discussion of Paper Published in Volume LXXII: A Simulation Test of Prediction Errors of Loss Reserve Estimation Techniques, by James N. Stanard," *PCAS* LXXXII, 1995, 104-120.

15. SAS/IML® Software: Usage and Reference, Version 6, First Edition, Cary, NC, SAS Institute Inc., 1990.

16. Schneider, Hans, and Barker, George Phillip, Matrices and Linear Algebra, New York, Dover Publications, 1973.

17. Searle, S. R., Linear Models, New York, John Wiley & Sons, 1971.

18. Stanard, James N., "A Simulation Test of Prediction Errors of Loss Reserve Estimation Techniques," *PCAS* LXXII, 1985, 124-148.

19. Venter, Gary G., "Credibility," Foundations of Casualty Actuarial Science, Casualty Actuarial Society, 1990, 375-483.

Exhibit 1

Exemplary Data

Cumulative Losses

		@1	@2	@3
AY1	Incd	75	90	100
	Paid	50	80	100
AY2	Incd	75	100	
	Paid	60	85	
AY3	Incd	50		
	Paid	45		

Incremental Losses

		@1	@2	@3
AY1	Incd	75	15	10
	Paid	50	30	20
AY2	Incd	75	25	
	Paid	60	25	
AY3	Incd	50		
	Paid	45		

Exhibit 2

Exemplary Model 1: $\mathbf{y} = X\beta + \mathbf{e}$, where $\mathrm{Var}[\mathbf{e}] = \sigma^2 I_{18}$

$$
\begin{array}{ll}
 & \\
\text{Incd} & \text{AY1@1} \\
 & \text{AY1@2} \\
 & \text{AY1@3} \\
 & \text{AY2@1} \\
 & \text{AY2@2} \\
 & \text{AY3@1} \\
\text{Paid} & \text{AY1@1} \\
 & \text{AY1@2} \\
 & \text{AY1@3} \\
 & \text{AY2@1} \\
 & \text{AY2@2} \\
 & \text{AY3@1}
\end{array}
\overset{\mathbf{y}_1}{\begin{bmatrix} 75 \\ 15 \\ 10 \\ 75 \\ 25 \\ 50 \\ 50 \\ 30 \\ 20 \\ 60 \\ 25 \\ 45 \end{bmatrix}}
=
\overset{X_1}{\begin{bmatrix}
1 & & & & & \\
 & 1 & & & & \\
 & & 1 & & & \\
1 & & & & & \\
 & 1 & & & & \\
1 & & & & & \\
 & & & 1 & & \\
 & & & & 1 & \\
 & & & & & 1 \\
 & & & 1 & & \\
 & & & & 1 & \\
 & & & 1 & &
\end{bmatrix}}
{}^{*}\beta + \mathbf{e}_1
$$

$$
\begin{array}{ll}
 & \\
\text{Incd} & \text{AY2@3} \\
 & \text{AY3@2} \\
 & \text{AY3@3} \\
\text{Paid} & \text{AY2@3} \\
 & \text{AY3@2} \\
 & \text{AY3@3}
\end{array}
\overset{\mathbf{y}_2}{\begin{bmatrix} \\ \\ \\ \\ \\ \\ \end{bmatrix}}
=
\overset{X_2}{\begin{bmatrix}
 & & 1 & & & \\
 & 1 & & & & \\
 & & 1 & & & \\
 & & & & & 1 \\
 & & & & 1 & \\
 & & & & & 1
\end{bmatrix}}
{}^{*}\beta + \mathbf{e}_2
$$

Exhibit 3

Exemplary Model 1 (Cont'd): Best Linear Unbiased Estimation

$X_1'X_1$

3	0	0	0	0	0
0	2	0	0	0	0
0	0	1	0	0	0
0	0	0	3	0	0
0	0	0	0	2	0
0	0	0	0	0	1

$X_1'y_1$

200
40
10
155
55
20

$(X_1'X_1)^{-1}$

0.333333	0	0	0	0	0
0	0.5	0	0	0	0
0	0	1	0	0	0
0	0	0	0.333333	0	0
0	0	0	0	0.5	0
0	0	0	0	0	1

$\beta = (X_1'X_1)^{-1}X_1'y_1$

66.66667
20
10
51.66667
27.5
20

t_1	12
k	6
df = t_1-k	6
$\sigma^2 = e_1'e_1/df$	99.30556

$e_1 = y_1 - X_1\beta$

8.333333
-5
0
8.333333
5
-16.6667
-1.66667
2.5
0
8.333333
-2.5
-6.66667

$Var[\beta] = \sigma^2(X_1'X_1)^{-1}$

33.10185	0	0	0	0	0
0	49.65278	0	0	0	0
0	0	99.30556	0	0	0
0	0	0	33.10185	0	0
0	0	0	0	49.65278	0
0	0	0	0	0	99.30556

$Q = X_2 - \Sigma_{21}\Sigma_{11}^{-1}X_1$

0	0	1	0	0	0
0	1	0	0	0	0
0	0	1	0	0	0
0	0	0	0	0	1
0	0	0	0	1	0
0	0	0	0	0	1

$Var[y_2 - y_2] = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12} + QVar[\beta]Q'$

198.6111	0	99.30556	0	0	0
0	148.9583	0	0	0	0
99.30556	0	198.6111	0	0	0
0	0	0	198.6111	0	99.30556
0	0	0	0	148.9583	0
0	0	0	99.30556	0	198.6111

$y_2 = X_2\beta + \Sigma_{21}\Sigma_{11}^{-1}e_1$

10
20
10
20
27.5
20

Exhibit 4

Exemplary Model 1 (Cont'd): Results

Cumulative Losses

		@1	@2	@3
AY1	Incd	75	90	100
	Paid	50	80	100
AY2	Incd	75	100	110
	Paid	60	85	105
AY3	Incd	50	70	80
	Paid	45	72.5	92.5

Incremental Losses

		@1	@2	@3
AY1	Incd	75	15	10
	Paid	50	30	20
AY2	Incd	75	25	10
	Paid	60	25	20
AY3	Incd	50	20	10
	Paid	45	27.5	20

Exhibit 5

Exemplary Model 2: $\mathbf{y} = X\beta + \mathbf{e}$, where $Var[\mathbf{e}] = \sigma^2 I_{18}$,
subject to the constraint that $[\,1\ \ 1\ \ 1\ -1\ -1\ -1\,]\beta = [0]$

Reduced Model: $\mathbf{y} = XV\gamma + \mathbf{e} = X^*\gamma + \mathbf{e}$, where $\beta = V\gamma$ and $Var[\mathbf{e}] = \sigma^2 I_{18}$,

V

0.88808	0.00000	0.00000	0.19954	0.06949
-0.26202	0.80783	-0.11865	0.30139	0.08472
-0.26202	-0.30116	0.75893	0.30139	0.08472
0.26202	0.50667	0.64028	-0.30139	-0.08472
0.05958	0.00000	0.00000	0.75008	-0.51688
0.04244	0.00000	0.00000	0.35363	0.84052

$\mathbf{y}_1 = X_1V\gamma + \mathbf{e}_1$, where $X_1V = X_1^*$:

		$\mathbf{y}_1$		$X_1V = X_1^*$					
Incd	AY1@1	75		0.8881	0.0000	0.0000	0.1995	0.0695	
	AY1@2	15		-0.2620	0.8078	-0.1187	0.3014	0.0847	
	AY1@3	10		-0.2620	-0.3012	0.7589	0.3014	0.0847	
	AY2@1	75		0.8881	0.0000	0.0000	0.1995	0.0695	
	AY2@2	25		-0.2620	0.8078	-0.1187	0.3014	0.0847	
	AY3@1	50	=	0.8881	0.0000	0.0000	0.1995	0.0695	$*\gamma + \mathbf{e}_1$
Paid	AY1@1	50		0.2620	0.5067	0.6403	-0.3014	-0.0847	
	AY1@2	30		0.0596	0.0000	0.0000	0.7501	-0.5169	
	AY1@3	20		0.0424	0.0000	0.0000	0.3536	0.8405	
	AY2@1	60		0.2620	0.5067	0.6403	-0.3014	-0.0847	
	AY2@2	25		0.0596	0.0000	0.0000	0.7501	-0.5169	
	AY3@1	45		0.2620	0.5067	0.6403	-0.3014	-0.0847	

		$\mathbf{y}_2$		$X_2V = X_2^*$					
Incd	AY2@3			-0.2620	-0.3012	0.7589	0.3014	0.0847	
	AY3@2			-0.2620	0.8078	-0.1187	0.3014	0.0847	
	AY3@3		=	-0.2620	-0.3012	0.7589	0.3014	0.0847	$*\gamma + \mathbf{e}_2$
Paid	AY2@3			0.0424	0.0000	0.0000	0.3536	0.8405	
	AY3@2			0.0596	0.0000	0.0000	0.7501	-0.5169	
	AY3@3			0.0424	0.0000	0.0000	0.3536	0.8405	

Exhibit 6

Exemplary Model 2 (Cont'd): Best Linear Unbiased Estimation

$X_1^{*\prime}X_1^*$

2.78689	0.05385	0.36662	0.1622	0.02604
0.05385	2.16602	0.55297	-0.06194	-0.01741
0.36662	0.55297	1.83398	-0.42171	-0.11854
0.1622	-0.06194	-0.42171	1.91476	-0.28337
0.02604	-0.01741	-0.11854	-0.28337	1.29835

$X_1^{*\prime}y_1$

209.254
107.835
102.086
56.5892
-6.61458

$(X_1^{*\prime}X_1^*)^{-1}$

0.37438	0.0127	-0.09337	-0.05602	-0.02809
0.0127	0.50137	-0.15958	-0.02191	-0.01288
-0.09337	-0.15958	0.6552	0.16106	0.0947
-0.05602	-0.02191	0.16106	0.58289	0.14275
-0.02809	-0.01288	0.0947	0.14275	0.8104

$\gamma = (X_1^{*\prime}X_1^*)^{-1}X_1^{*\prime}y_1$

67.1949
39.2778
38.6286
34.3976
5.11848

t_1	12
k	5
$df = t_1 - k$	7
$\sigma^2 = e_1'e_1/df$	85.3626

$e_1 = y_1 - X_1^*\gamma$

8.10606
-5.34091
-0.68182
8.10606
4.65909
-16.8939
-1.43939
2.84091
0.68182
8.56061
-2.15909
-6.43939

$Var[\gamma] = \sigma^2(X_1^{*\prime}X_1^*)^{-1}$

31.9583	1.08418	-7.97003	-4.78226	-2.39782
1.08418	42.7983	-13.622	-1.87025	-1.0997
-7.97003	-13.622	55.9291	13.7486	8.08411
-4.78226	-1.87025	13.7486	49.7572	12.1856
-2.39782	-1.0997	8.08411	12.1856	69.1779

$Var[\beta] = VVar[\gamma]V'$

25.8674	-3.88012	-7.76023	2.58674	3.88012	7.76023
-3.88012	36.8611	-11.6403	3.88012	5.82017	11.6403
-7.76023	-11.6403	62.0819	7.76023	11.6403	23.2807
2.58674	3.88012	7.76023	25.8674	-3.88012	-7.76023
3.88012	5.82017	11.6403	-3.88012	36.8611	-11.6403
7.76023	11.6403	23.2807	-7.76023	-11.6403	62.0819

$\beta = V\gamma$

66.8939
20.3409
10.6818
51.4394
27.1591
19.3182

$Q = X_2 - \Sigma_{21}\Sigma_{11}^{-1}X_1$

0	0	1	0	0	0
0	1	0	0	0	0
0	0	1	0	0	0
0	0	0	0	0	1
0	0	0	0	1	0
0	0	0	0	0	1

$Var[y_2 - y_2] = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12} + QVar[\beta]Q'$

147.444	-11.6403	62.0819	23.2807	11.6403	23.2807
-11.6403	122.224	-11.6403	11.6403	5.82017	11.6403
62.0819	-11.6403	147.444	23.2807	11.6403	23.2807
23.2807	11.6403	23.2807	147.444	-11.6403	62.0819
11.6403	5.82017	11.6403	-11.6403	122.224	-11.6403
23.2807	11.6403	23.2807	62.0819	-11.6403	147.444

$y_2 = X_2\beta + \Sigma_{21}\Sigma_{11}^{-1}e_1$

10.6818
20.3409
10.6818
19.3182
27.1591
19.3182

Exhibit 7

Exemplary Model 2 (Cont'd): Results

		Cumulative Losses		
		@1	@2	@3
AY1	Incd	75	90	100
	Paid	50	80	100
AY2	Incd	75	100	110.6818
	Paid	60	85	104.3182
AY3	Incd	50	70.34091	81.02273
	Paid	45	72.15909	91.47727

		Incremental Losses		
		@1	@2	@3
AY1	Incd	75	15	10
	Paid	50	30	20
AY2	Incd	75	25	10.68182
	Paid	60	25	19.31818
AY3	Incd	50	20.34091	10.68182
	Paid	45	27.15909	19.31818

Exhibit 8

Exemplary Model 3: $\mathbf{y} = X\beta + \mathbf{e}$, where $Var[\mathbf{e}] = \Sigma = \sigma^2\Phi$,
subject to the constraints that $[\,1\ \ 1\ \ 1\ -1\ -1\ -1]\beta = [0]$ and $A\mathbf{y} = 0$

A

1	1	1	0	0	0	-1	-1	-1	0	0	0	0	0	0	0	0	0
0	0	0	1	1	0	0	0	0	-1	-1	0	1	0	0	-1	0	0
0	0	0	0	0	1	0	0	0	0	0	-1	0	1	1	0	-1	-1

AX

1	1	1	-1	-1	-1
1	1	1	-1	-1	-1
1	1	1	-1	-1	-1

Partitions of Φ:

Φ_{11} 12×12	Φ_{12} 12×6
Φ_{21} 6×12	Φ_{22} 6×6

$\Phi = I_{18} - A'(AA')^{-1}A$

0.833	-0.167	-0.167				0.167	0.167	0.167									
-0.167	0.833	-0.167				0.167	0.167	0.167									
-0.167	-0.167	0.833				0.167	0.167	0.167									
			0.833	-0.167					0.167	0.167		-0.167			0.167		
			-0.167	0.833					0.167	0.167		-0.167			0.167		
					0.833						0.167		-0.167	-0.167		0.167	0.167
0.167	0.167	0.167				0.833	-0.167	-0.167									
0.167	0.167	0.167				-0.167	0.833	-0.167									
0.167	0.167	0.167				-0.167	-0.167	0.833									
			0.167	0.167					0.833	-0.167		0.167			-0.167		
			0.167	0.167					-0.167	0.833		0.167			-0.167		
					0.167						0.833		0.167	0.167		-0.167	-0.167
			-0.167	-0.167					0.167	0.167		0.833			0.167		
					-0.167						0.167		0.833	-0.167		0.167	0.167
					-0.167						0.167		-0.167	0.833		0.167	0.167
			0.167	0.167					-0.167	-0.167		0.167			0.833		
					0.167						-0.167		0.167	0.167		0.833	-0.167
					0.167						-0.167		0.167	0.167		-0.167	0.833

Exhibit 9

Exemplary Model 3 (Cont'd): $\mathbf{y} = X\beta + \mathbf{e}$, where $\mathrm{Var}[\mathbf{e}] = \Sigma = \sigma^2\Phi$,
subject to the constraint that $[\,1\ \ 1\ \ 1\ {-1}\ {-1}\ {-1}\,]\beta = [0]$

$$
\begin{array}{ll}
 & \\
\text{Incd} & \text{AY1@1} \\
 & \text{AY1@2} \\
 & \text{AY1@3} \\
 & \text{AY2@1} \\
 & \text{AY2@2} \\
 & \text{AY3@1} \\
\text{Paid} & \text{AY1@1} \\
 & \text{AY1@2} \\
 & \text{AY1@3} \\
 & \text{AY2@1} \\
 & \text{AY2@2} \\
 & \text{AY3@1}
\end{array}
\overset{\mathbf{y}_1}{\begin{bmatrix} 75 \\ 15 \\ 10 \\ 75 \\ 25 \\ 50 \\ 50 \\ 30 \\ 20 \\ 60 \\ 25 \\ 45 \end{bmatrix}}
=
\overset{X_1}{\begin{bmatrix}
1 & & & & & \\
 & 1 & & & & \\
 & & 1 & & & \\
1 & & & & & \\
 & 1 & & & & \\
1 & & & & & \\
 & & & 1 & & \\
 & & & & 1 & \\
 & & & & & 1 \\
 & & & 1 & & \\
 & & & & 1 & \\
 & & & 1 & &
\end{bmatrix}} * \beta + \mathbf{e}_1
$$

$$
\begin{array}{ll}
 & \\
\text{Incd} & \text{AY2@3} \\
 & \text{AY3@2} \\
 & \text{AY3@3} \\
\text{Paid} & \text{AY2@3} \\
 & \text{AY3@2} \\
 & \text{AY3@3}
\end{array}
\overset{\mathbf{y}_2}{\begin{bmatrix} \ \\ \ \\ \ \\ \ \\ \ \\ \ \end{bmatrix}}
=
\overset{X_2}{\begin{bmatrix}
 & & 1 & & & \\
 & 1 & & & & \\
 & & 1 & & & \\
 & & & & & 1 \\
 & & & & 1 & \\
 & & & & & 1
\end{bmatrix}} * \beta + \mathbf{e}_2
$$

Exhibit 10

Exemplary Model 3 (Cont'd): $\mathbf{y} = X\beta + \mathbf{e}$, where $\mathrm{Var}[\mathbf{e}] = \Sigma = \sigma^2\Phi$,
subject to the constraint that $[\,1\ \ 1\ \ 1\ -1\ -1\ -1\,]\beta = [0]$

Diagonalization of $\Phi_{11} = W\Lambda W'$

W is orthogonal ($WW' = W'W = I_{12}$), and Λ is diagonal.

W

0.125	0.110	0.107	0.857			0.208	0.103	0.082			0.408
0.125	0.277	0.107	-0.393			0.208	0.103	0.722			0.408
-0.009	0.506	0.107	-0.272			0.208	0.103	-0.662			0.408
-0.649	0.269		0.095	0.289		-0.308	0.237	0.116		0.500	
0.711	-0.036		-0.045	0.289		-0.308	0.237	-0.079		0.500	
					0.707				0.707		
0.052	0.194	-0.519	0.042			0.491	0.529	0.031			-0.408
0.188	0.700	0.007	0.150			-0.093	-0.515	0.111			-0.408
		0.834				0.226	0.294				-0.408
				0.866						-0.500	
0.063	0.233		0.050	-0.289		-0.616	0.474	0.037		-0.500	
					0.707				-0.707		

Λ

1											
	1										
		1									
			1								
				1							
					1						
						1					
							1				
								1			
									0.6667		
										0.3333	
											0

Exhibit 11

Exemplary Model 3 (Cont'd): **y**= Xβ+**e**, where Var[**e**] = $\Sigma = \sigma^2\Phi$,
subject to the constraint that [1 1 1 -1 -1 -1]β = [0]

Transformation of Observations

		$y_1^* = W'y_1$		$X_1^* = W'X_1$						
Incd	AY1@1	-9.941		-0.524	0.836	-0.009	0.052	0.251		
	AY1@2	73.274		0.379	0.242	0.506	0.194	0.933		
	AY1@3	1.685		0.107	0.107	0.107	-0.519	0.007	0.834	
	AY2@1	69.480		0.952	-0.438	-0.272	0.042	0.200		
	AY2@2	73.612		0.289	0.289		0.866	-0.289		
	AY3@1	67.175	=	0.707			0.707			$*\beta + W'e_1$
Paid	AY1@1	0.842		-0.100	-0.100	0.208	0.491	-0.710	0.226	
	AY1@2	62.702		0.340	0.340	0.103	0.529	-0.041	0.294	
	AY1@3	22.794		0.197	0.643	-0.662	0.031	0.148		
	AY2@1	3.536		0.707			-0.707			
	AY2@2	7.500		0.500	0.500		-0.500	-0.500		
	AY3@1	0.000		0.408	0.408	0.408	-0.408	-0.408	-0.408	

		y_2		X_2						
Incd	AY2@3					1				
	AY3@2				1					
	AY3@3		=			1				$*\beta + e_2$
Paid	AY2@3								1	
	AY3@2							1		
	AY3@3								1	

Exhibit 12

Exemplary Model 3 (Cont'd): $y = X\beta + e$, where $Var[e] = \Sigma = \sigma^2\Phi$,
subject to the constraint that $[\,1\ \ 1\ \ 1\ -1\ -1\ -1\,]\beta = [0]$

Transformation of the Variance Matrix

$\Phi_{11}{}^{*} = W'\Phi_{11}W = W'(W\Lambda W')W = \Lambda$

1											
	1										
		1									
			1								
				1							
					1						
						1					
							1				
								1			
									0.667		
										0.333	
											0

$\Phi_{12}{}^{*} = W'\Phi_{12}$

	-0.236	-0.236		0.236	0.236
-0.333			0.333		

$\Phi_{21}{}^{*} = \Phi_{12}W$

										-0.333	
									-0.236		
									-0.236		
										0.333	
									0.236		
									0.236		

$\Phi_{22}{}^{*} = \Phi_{22}$

0.833			0.167		
	0.833	-0.167		0.167	0.167
	-0.167	0.833		0.167	0.167
0.167			0.833		
	0.167	0.167		0.833	-0.167
	0.167	0.167		-0.167	0.833

Exhibit 13

Exemplary Model 3 (Cont'd): $y^* = X^*\beta + e^*$, where $Var[e^*] = \Sigma = \sigma^2\Phi^*$, subject to the constraint that $[1\ 1\ 1\ -1\ -1\ -1]\beta = [0]$

Reduced Model: $y^* = X^*V\gamma + e^* = X^\#\gamma + e^*$, where $\beta = V\gamma$

V

0.88808	0.00000	0.00000	0.19954	0.06949
-0.26202	0.80783	-0.11865	0.30139	0.08472
-0.26202	-0.30116	0.75893	0.30139	0.08472
0.26202	0.50667	0.64028	-0.30139	-0.08472
0.05958	0.00000	0.00000	0.75008	-0.51688
0.04244	0.00000	0.00000	0.35363	0.84052

$y_1^* = X_1^*V\gamma + e_1^*$, where $X_1^*V = X_1^\#$

		y_1^*		$X_1^\#$					
Incd	AY1@1	-9.941		-0.6535	0.7045	-0.0728	0.3169	-0.1003	
	AY1@2	73.274		0.2472	0.1410	0.4795	0.9422	-0.4089	
	AY1@3	1.685		-0.0610	-0.2084	-0.2634	0.5426	0.7675	
	AY2@1	69.480		1.0540	-0.2512	-0.1274	0.1137	-0.1011	
	AY2@2	73.612		0.3904	0.6720	0.5202	-0.3329	0.1204	
	AY3@1	67.175	=	0.8132	0.3583	0.4527	-0.0720	-0.0108	$*\gamma + e_1^*$
Paid	AY1@1	0.842		-0.0212	0.1052	0.4841	-0.5879	0.5171	
	AY1@2	62.702		0.3344	0.5116	0.3764	0.1148	0.2840	
	AY1@3	22.794		0.1967	0.7345	-0.5591	0.1351	-0.0668	
	AY2@1	3.536		0.4427	-0.3583	-0.4527	0.3542	0.1090	
	AY2@2	7.500		0.1522	0.1506	-0.3795	0.0261	0.3779	

$y_2 = X_2V\gamma + e_2$, where $X_2V = X_2^\#$

		y_2		$X_2^\#$					
Incd	AY2@3			-0.2620	-0.3012	0.7589	0.3014	0.0847	
	AY3@2			-0.2620	0.8078	-0.1187	0.3014	0.0847	
	AY3@3		=	-0.2620	-0.3012	0.7589	0.3014	0.0847	$*\gamma + e_2$
Paid	AY2@3			0.0424	0.0000	0.0000	0.3536	0.8405	
	AY3@2			0.0596	0.0000	0.0000	0.7501	-0.5169	
	AY3@3			0.0424	0.0000	0.0000	0.3536	0.8405	

Exhibit 14

Exemplary Model 3 (Cont'd): Best Linear Unbiased Estimation

$X_1^{\#\prime}\Phi_{11}^{*-1}X_1^{\#}$

2.931	0.020	0.151	0.249	0.165
0.020	2.276	0.520	-0.118	0.077
0.151	0.520	2.224	-0.522	-0.430
0.249	-0.118	-0.522	1.979	-0.244
0.165	0.077	-0.430	-0.244	1.590

$X_1^{\#\prime}\Phi_{11}^{*-1}y_1^{*}$

212.320
109.461
95.594
57.607
-0.753

$(X_1^{\#\prime}\Phi_{11}^{*-1}X_1^{\#})^{-1}$

0.353	0.008	-0.053	-0.065	-0.061
0.008	0.469	-0.125	-0.013	-0.059
-0.053	-0.125	0.559	0.170	0.189
-0.065	-0.013	0.170	0.575	0.142
-0.061	-0.059	0.189	0.142	0.711

$\gamma = (X_1^{\#\prime}\Phi_{11}^{*-1}X_1^{\#})^{-1}X_1^{\#\prime}\Phi_{11}^{*-1}y_1$

66.968
40.333
38.149
33.912
6.146

t_1	11
k	5
$df = t_1 - k$	6
$\sigma^2 = e_1^{*\prime}\Phi_{11}^{*-1}e_1^{*}/df$	106.597

$e_1^{*} = y_1^{*} - X_1^{\#}\gamma$

-1.939
3.301
1.113
10.651
11.066
-16.499
-3.688
-0.327
-2.847
-7.071
2.500

$Var[\gamma] = \sigma^2(X_1^{\#\prime}\Phi_{11}^{*-1}X_1^{\#})^{-1}$

37.611	0.813	-5.641	-6.971	-6.545
0.813	50.015	-13.295	-1.417	-6.316
-5.641	-13.295	59.544	18.101	20.115
-6.971	-1.417	18.101	61.297	15.108
-6.545	-6.316	20.115	15.108	75.793

$Var[\beta] = VVar[\gamma]V'$

29.610	-5.922	-5.922	5.922	5.922	5.922
-5.922	42.935	-10.364	5.922	10.364	10.364
-5.922	-10.364	69.584	5.922	10.364	37.013
5.922	5.922	5.922	29.610	-5.922	-5.922
5.922	10.364	10.364	-5.922	42.935	-10.364
5.922	10.364	37.013	-5.922	-10.364	69.584

$\beta = V\gamma$

66.667
21.250
10.000
51.667
26.250
20.000

$Q = X_2^{\#} - \Phi_{21}^{*}\Phi_{11}^{*-1}X_1^{\#}$

-0.110	-0.151	0.379	0.328	0.463
-0.106	0.681	-0.279	0.427	0.123
-0.106	-0.428	0.599	0.427	0.123
-0.110	-0.151	0.379	0.328	0.463
-0.097	0.127	0.160	0.625	-0.555
-0.114	0.127	0.160	0.228	0.802

$Var[y_2 - y_2] = \sigma^2(\Phi_{22} - \Phi_{21}^{*}\Phi_{11}^{*-1}\Phi_{12}^{*}) + QVar[\gamma]Q'$

106.597	0.000	53.299	106.597	0.000	53.299
0.000	119.922	-39.974	0.000	39.974	39.974
53.299	-39.974	146.571	53.299	39.974	66.623
106.597	0.000	53.299	106.597	0.000	53.299
0.000	39.974	39.974	0.000	119.922	-39.974
53.299	39.974	66.623	53.299	-39.974	146.571

$y_2 = X_2^{\#}\gamma + \Phi_{21}^{*}\Phi_{11}^{*-1}e_1^{*}$

7.5
23.75
12.5
22.5
23.75
17.5

Exhibit 15

Exemplary Model 3 (Cont'd): Results

Cumulative Losses

		@1	@2	@3
AY1	Incd	75	90	100
	Paid	50	80	100
AY2	Incd	75	100	107.5
	Paid	60	85	107.5
AY3	Incd	50	73.75	86.25
	Paid	45	68.75	86.25

Incremental Losses

		@1	@2	@3
AY1	Incd	75	15	10
	Paid	50	30	20
AY2	Incd	75	25	7.5
	Paid	60	25	22.5
AY3	Incd	50	23.75	12.5
	Paid	45	23.75	17.5

Exhibit 16

Exemplary Models: Comparison of Variances

	A_{1x6}						$AVar[\mathbf{y}_2\text{-}\mathbf{y}_2]A'$		
	Incurred			Paid					
Description	AY2@3	AY3@2	AY3@3	AY2@3	AY3@2	AY3@3	Model 1	Model 2	Model 3
Overall Balance	1	1	1	-1	-1	-1	1489.583	698.421	0.000
AY2 Balance	1	0	0	-1	0	0	397.222	248.327	0.000
AY3 Balance	0	1	1	0	-1	-1	695.139	388.012	0.000
Overall Incurred	1	1	1	0	0	0	744.792	494.715	399.740
AY2 Incurred	1	0	0	0	0	0	198.611	147.444	106.597
AY3 Incurred	0	1	1	0	0	0	347.569	246.387	186.545
Overall Paid	0	0	0	1	1	1	744.792	494.715	399.740
AY2 Paid	0	0	0	1	0	0	198.611	147.444	106.597
AY3 Paid	0	0	0	0	1	1	347.569	246.387	186.545

Exhibit 17

Self-Insured Entity: Paid Workers Compensation Losses

Chain Ladder Method: Cumulative Losses

FY	Exposure	@12	@24	@36	@48	@60	@72	@84	@108 (Ult)	Pure Prem
1988	131,332.20	266,354	432,926	465,255	518,865	526,989	543,913	583,022	647,802	4.93
1989	141,672.24	246,981	606,361	835,377	904,916	1,023,551	1,123,843	1,204,651	1,338,501	9.45
1990	141,677.29	203,178	578,946	855,563	930,475	1,016,903	1,093,778	1,172,424	1,302,693	9.19
1991	142,577.99	395,630	656,273	823,982	1,094,674	1,193,801	1,284,049	1,376,376	1,529,306	10.73
1992	143,285.58	207,698	382,313	544,953	630,669	687,778	739,772	792,964	881,071	6.15
1993	138,261.75	167,681	447,859	594,230	687,696	749,970	806,665	864,667	960,741	6.95
1994	121,857.69	215,740	450,281	597,444	691,416	754,026	811,028	869,344	965,937	7.93
1995	115,000.00								912,905	7.94
Development Factor:			2.087	1.327	1.157	1.091	1.076	1.072	1.111	
Pattern:		22.3%	46.6%	61.9%	71.6%	78.1%	84.0%	90.0%	100.0%	

	Paid	Unpaid	Ultimate
FY 1988-1994:	5,026,994	2,599,058	7,626,052
FY 1995:		912,905	912,905
Total	5,026,994	3,511,964	8,538,958

Additive Method: Incremental Losses

FY	Exposure	@12	@24	@36	@48	@60	@72	@84	@108	Ultimate
1988	131,332.20	266,354	166,572	32,329	53,610	8,124	16,924	39,109	98,912	681,934
1989	141,672.24	246,981	359,380	229,016	69,539	118,635	100,292	42,188	106,699	1,272,730
1990	141,677.29	203,178	375,768	276,617	74,912	86,428	60,830	42,190	106,703	1,226,626
1991	142,577.99	395,630	260,643	167,709	270,692	73,299	61,217	42,458	107,381	1,379,029
1992	143,285.58	207,698	174,615	162,640	120,528	73,663	61,520	42,669	107,914	951,247
1993	138,261.75	167,681	280,178	171,372	116,302	71,080	59,363	41,173	104,131	1,011,280
1994	121,857.69	215,740	205,132	151,040	102,504	62,647	52,320	36,288	91,776	917,446
1995	115,000.00	203,895	193,588	142,540	96,735	59,121	49,376	34,245	86,611	866,112
Ratio:		1.773	1.683	1.239	0.841	0.514	0.429	0.298	0.753	
Cumulative:		1.773	3.456	4.696	5.537	6.051	6.480	6.778	7.531	
Pattern:		23.5%	45.9%	62.4%	73.5%	80.3%	86.0%	90.0%	100.0%	

	Paid	Unpaid	Ultimate
FY 1988-1994:	5,026,994	2,413,298	7,440,292
FY 1995:		866,112	866,112
Total	5,026,994	3,279,411	8,306,405

Exhibit 18

Self-Insured Entity: Incurred Workers Compensation Losses

Chain Ladder Method: Cumulative Losses

FY	Exposure	@12	@24	@36	@48	@60	@72	@84	@108 (Ult)	Pure Prem
1988	131,332.20	422,076	506,045	525,616	581,603	569,929	633,054	583,022	613,707	4.67
1989	141,672.24	457,750	855,007	1,006,522	1,068,468	1,204,197	1,147,426	1,056,742	1,112,360	7.85
1990	141,677.29	345,084	727,598	1,069,667	1,081,962	1,024,044	1,027,712	946,489	996,304	7.03
1991	142,577.99	591,842	880,633	1,032,814	1,340,032	1,372,471	1,377,387	1,268,528	1,335,293	9.37
1992	143,285.58	379,033	536,841	611,605	685,215	701,803	704,316	648,652	682,792	4.77
1993	138,261.75	308,803	448,780	543,512	608,927	623,667	625,901	576,434	606,773	4.39
1994	121,857.69	215,851	340,842	412,790	462,472	473,667	475,364	437,794	460,836	3.78
1995	115,000.00								695,276	6.05
Development Factor:			1.579	1.211	1.120	1.024	1.004	0.921	1.053	
Pattern:		46.8%	74.0%	89.6%	100.4%	102.8%	103.2%	95.0%	100.0%	

	Incurred	IBNR	Ultimate
FY 1988-1994:	5,370,760	437,305	5,808,065
FY 1995:		695,276	695,276
Total	5,370,760	1,132,581	6,503,341

Additive Method: Incremental Losses

FY	Exposure	@12	@24	@36	@48	@60	@72	@84	@108	Ultimate
1988	131,332.20	422,076	83,969	19,571	55,987	-11,674	63,125	-50,032	41,368	624,390
1989	141,672.24	457,750	397,257	151,515	61,946	135,729	-56,771	-53,971	44,625	1,138,080
1990	141,677.29	345,084	382,514	342,069	12,295	-57,918	3,297	-53,973	44,627	1,017,995
1991	142,577.99	591,842	288,791	152,181	307,218	22,740	3,318	-54,316	44,911	1,356,684
1992	143,285.58	379,033	157,808	74,764	112,478	22,852	3,335	-54,586	45,133	740,818
1993	138,261.75	308,803	139,977	146,068	108,535	22,051	3,218	-52,672	43,551	719,531
1994	121,857.69	215,851	183,969	128,738	95,658	19,435	2,836	-46,423	38,384	638,448
1995	115,000.00	325,660	173,616	121,493	90,274	18,341	2,677	-43,810	36,224	724,475
Ratio:		2.832	1.510	1.056	0.785	0.159	0.023	-0.381	0.315	
Cumulative:		2.832	4.342	5.398	6.183	6.342	6.366	5.985	6.300	
Pattern:		45.0%	68.9%	85.7%	98.1%	100.7%	101.0%	95.0%	100.0%	

	Incurred	IBNR	Ultimate
FY 1988-1994:	5,370,760	865,187	6,235,947
FY 1995:		724,475	724,475
Total	5,370,760	1,589,662	6,960,422

Exhibit 19

Self-Insured Entity Model 1: Paid Losses

$y_{paid} = X\beta + e$, where $Var[e] = \sigma^2 I_{64}$, subject to the constraint that $[1\ 1\ 1\ 1\ 1\ 1\ 1\ {-9}]\beta = [0]$

FY	Age	y_{paid}	X							
1988	12	266,354	131,332.20	0	0	0	0	0	0	0
1988	24	166,572	0	131,332.20	0	0	0	0	0	0
1988	36	32,329	0	0	131,332.20	0	0	0	0	0
1988	48	53,610	0	0	0	131,332.20	0	0	0	0
1988	60	8,124	0	0	0	0	131,332.20	0	0	0
1988	72	16,924	0	0	0	0	0	131,332.20	0	0
1988	84	39,109	0	0	0	0	0	0	131,332.20	0
1988	108	.	0	0	0	0	0	0	0	131,332.20
1989	12	246,981	141,672.24	0	0	0	0	0	0	0
1989	24	359,380	0	141,672.24	0	0	0	0	0	0
1989	36	229,016	0	0	141,672.24	0	0	0	0	0
1989	48	69,539	0	0	0	141,672.24	0	0	0	0
1989	60	118,635	0	0	0	0	141,672.24	0	0	0
1989	72	100,292	0	0	0	0	0	141,672.24	0	0
1989	84	.	0	0	0	0	0	0	141,672.24	0
1989	108	.	0	0	0	0	0	0	0	141,672.24
1990	12	203,178	141,677.29	0	0	0	0	0	0	0
1990	24	375,768	0	141,677.29	0	0	0	0	0	0
1990	36	276,617	0	0	141,677.29	0	0	0	0	0
1990	48	74,912	0	0	0	141,677.29	0	0	0	0
1990	60	86,428	0	0	0	0	141,677.29	0	0	0
1990	72	.	0	0	0	0	0	141,677.29	0	0
1990	84	.	0	0	0	0	0	0	141,677.29	0
1990	108	.	0	0	0	0	0	0	0	141,677.29
1991	12	395,630	142,577.99	0	0	0	0	0	0	0
1991	24	260,643	0	142,577.99	0	0	0	0	0	0
1991	36	167,709	0	0	142,577.99	0	0	0	0	0
1991	48	270,692	0	0	0	142,577.99	0	0	0	0
1991	60	.	0	0	0	0	142,577.99	0	0	0
1991	72	.	0	0	0	0	0	142,577.99	0	0
1991	84	.	0	0	0	0	0	0	142,577.99	0
1991	108	.	0	0	0	0	0	0	0	142,577.99
1992	12	207,698	143,285.58	0	0	0	0	0	0	0
1992	24	174,615	0	143,285.58	0	0	0	0	0	0
1992	36	162,640	0	0	143,285.58	0	0	0	0	0
1992	48	.	0	0	0	143,285.58	0	0	0	0
1992	60	.	0	0	0	0	143,285.58	0	0	0
1992	72	.	0	0	0	0	0	143,285.58	0	0
1992	84	.	0	0	0	0	0	0	143,285.58	0
1992	108	.	0	0	0	0	0	0	0	143,285.58
1993	12	167,681	138,261.75	0	0	0	0	0	0	0
1993	24	280,178	0	138,261.75	0	0	0	0	0	0
1993	36	.	0	0	138,261.75	0	0	0	0	0
1993	48	.	0	0	0	138,261.75	0	0	0	0
1993	60	.	0	0	0	0	138,261.75	0	0	0
1993	72	.	0	0	0	0	0	138,261.75	0	0
1993	84	.	0	0	0	0	0	0	138,261.75	0
1993	108	.	0	0	0	0	0	0	0	138,261.75
1994	12	215,740	121,857.69	0	0	0	0	0	0	0
1994	24	.	0	121,857.69	0	0	0	0	0	0
1994	36	.	0	0	121,857.69	0	0	0	0	0
1994	48	.	0	0	0	121,857.69	0	0	0	0
1994	60	.	0	0	0	0	121,857.69	0	0	0
1994	72	.	0	0	0	0	0	121,857.69	0	0
1994	84	.	0	0	0	0	0	0	121,857.69	0
1994	108	.	0	0	0	0	0	0	0	121,857.69
1995	12	.	115,000.00	0	0	0	0	0	0	0
1995	24	.	0	115,000.00	0	0	0	0	0	0
1995	36	.	0	0	115,000.00	0	0	0	0	0
1995	48	.	0	0	0	115,000.00	0	0	0	0
1995	60	.	0	0	0	0	115,000.00	0	0	0
1995	72	.	0	0	0	0	0	115,000.00	0	0
1995	84	.	0	0	0	0	0	0	115,000.00	0
1995	108	.	0	0	0	0	0	0	0	115,000.00

Exhibit 20

Self-Insured Entity Model 1 (Cont'd): Results

	β
@12	1.773
@24	1.934
@36	1.253
@48	0.850
@60	0.525
@72	0.440
@84	0.298
@108	0.786
Total	7.859

Var[β]

0.0496	0	0	0	0	0	0	0.0055
0	0.0559	0	0	0	0	0	0.0062
0	0	0.0668	0	0	0	0	0.0074
0	0	0	0.0845	0	0	0	0.0094
0	0	0	0	0.1144	0	0	0.0127
0	0	0	0	0	0.1759	0	0.0195
0	0	0	0	0	0	0.3805	0.0423
0.0055	0.0062	0.0074	0.0094	0.0127	0.0195	0.0423	0.0115

σ^2

6.5637E+09

Incremental Paid Losses

FY	@12	@24	@36	@48	@60	@72	@84	@108	Ultimate	Variance
1988	266,354	166,572	32,329	53,610	8,124	16,924	39,109	103,209	686,231	6.761E+09
1989	246,981	359,380	229,016	69,539	118,635	100,292	42,188	111,335	1,277,366	2.269E+10
1990	203,178	375,768	276,617	74,912	86,428	62,379	42,190	111,339	1,232,810	3.357E+10
1991	395,630	260,643	167,709	270,692	74,825	62,776	42,458	112,047	1,386,779	4.315E+10
1992	207,698	174,615	162,640	121,864	75,196	63,087	42,669	112,603	960,371	5.200E+10
1993	167,681	280,178	173,226	117,591	72,560	60,875	41,173	108,655	1,021,938	5.881E+10
1994	215,740	235,631	152,674	103,639	63,951	53,653	36,288	95,763	957,338	6.205E+10
1995	203,877	222,371	144,082	97,807	60,352	50,633	34,245	90,374	903,741	6.765E+10

	Paid	Unpaid	Ultimate	Variance	Std. Dev.
FY 1988-1994:	5,026,994	2,495,840	7,522,834	7.114E+11	843,448
FY 1995:		903,741	903,741	6.765E+10	260,105
Total	5,026,994	3,399,580	8,426,574	9.468E+11	973,022

Generalized Variance 7.718E+09

Exhibit 21

Self-Insured Entity Model 2: Incurred Losses

$y_{incd} = X\beta + e$, where $Var[e] = \sigma^2 I_{64}$, subject to the constraint that $[\,1\ \ 1\ \ 1\ \ 1\ \ 1\ \ 1\ \ 1\ \ -19\,]\beta = [0]$

FY	Age	y_{incd}	X							
1988	12	422,076	131,332.20	0	0	0	0	0	0	0
1988	24	83,969	0	131,332.20	0	0	0	0	0	0
1988	36	19,571	0	0	131,332.20	0	0	0	0	0
1988	48	55,987	0	0	0	131,332.20	0	0	0	0
1988	60	-11,674	0	0	0	0	131,332.20	0	0	0
1988	72	63,125	0	0	0	0	0	131,332.20	0	0
1988	84	-50,032	0	0	0	0	0	0	131,332.20	0
1988	108	.	0	0	0	0	0	0	0	131,332.20
1989	12	457,750	141,672.24	0	0	0	0	0	0	0
1989	24	397,257	0	141,672.24	0	0	0	0	0	0
1989	36	151,515	0	0	141,672.24	0	0	0	0	0
1989	48	61,946	0	0	0	141,672.24	0	0	0	0
1989	60	135,729	0	0	0	0	141,672.24	0	0	0
1989	72	-56,771	0	0	0	0	0	141,672.24	0	0
1989	84	.	0	0	0	0	0	0	141,672.24	0
1989	108	.	0	0	0	0	0	0	0	141,672.24
1990	12	345,084	141,677.29	0	0	0	0	0	0	0
1990	24	382,514	0	141,677.29	0	0	0	0	0	0
1990	36	342,069	0	0	141,677.29	0	0	0	0	0
1990	48	12,295	0	0	0	141,677.29	0	0	0	0
1990	60	-57,918	0	0	0	0	141,677.29	0	0	0
1990	72	.	0	0	0	0	0	141,677.29	0	0
1990	84	.	0	0	0	0	0	0	141,677.29	0
1990	108	.	0	0	0	0	0	0	0	141,677.29
1991	12	591,842	142,577.99	0	0	0	0	0	0	0
1991	24	288,791	0	142,577.99	0	0	0	0	0	0
1991	36	152,181	0	0	142,577.99	0	0	0	0	0
1991	48	307,218	0	0	0	142,577.99	0	0	0	0
1991	60	.	0	0	0	0	142,577.99	0	0	0
1991	72	.	0	0	0	0	0	142,577.99	0	0
1991	84	.	0	0	0	0	0	0	142,577.99	0
1991	108	.	0	0	0	0	0	0	0	142,577.99
1992	12	379,033	143,285.58	0	0	0	0	0	0	0
1992	24	157,808	0	143,285.58	0	0	0	0	0	0
1992	36	74,764	0	0	143,285.58	0	0	0	0	0
1992	48	.	0	0	0	143,285.58	0	0	0	0
1992	60	.	0	0	0	0	143,285.58	0	0	0
1992	72	.	0	0	0	0	0	143,285.58	0	0
1992	84	.	0	0	0	0	0	0	143,285.58	0
1992	108	.	0	0	0	0	0	0	0	143,285.58
1993	12	308,803	138,261.75	0	0	0	0	0	0	0
1993	24	139,977	0	138,261.75	0	0	0	0	0	0
1993	36	.	0	0	138,261.75	0	0	0	0	0
1993	48	.	0	0	0	138,261.75	0	0	0	0
1993	60	.	0	0	0	0	138,261.75	0	0	0
1993	72	.	0	0	0	0	0	138,261.75	0	0
1993	84	.	0	0	0	0	0	0	138,261.75	0
1993	108	.	0	0	0	0	0	0	0	138,261.75
1994	12	215,851	121,857.69	0	0	0	0	0	0	0
1994	24	.	0	121,857.69	0	0	0	0	0	0
1994	36	.	0	0	121,857.69	0	0	0	0	0
1994	48	.	0	0	0	121,857.69	0	0	0	0
1994	60	.	0	0	0	0	121,857.69	0	0	0
1994	72	.	0	0	0	0	0	121,857.69	0	0
1994	84	.	0	0	0	0	0	0	121,857.69	0
1994	108	.	0	0	0	0	0	0	0	121,857.69
1995	12	.	115,000.00	0	0	0	0	0	0	0
1995	24	.	0	115,000.00	0	0	0	0	0	0
1995	36	.	0	0	115,000.00	0	0	0	0	0
1995	48	.	0	0	0	115,000.00	0	0	0	0
1995	60	.	0	0	0	0	115,000.00	0	0	0
1995	72	.	0	0	0	0	0	115,000.00	0	0
1995	84	.	0	0	0	0	0	0	115,000.00	0
1995	108	.	0	0	0	0	0	0	0	115,000.00

Exhibit 22

Self-Insured Entity Model 2 (Cont'd): Results

	β
@12	2.850
@24	1.744
@36	1.068
@48	0.794
@60	0.165
@72	0.007
@84	-0.381
@108	0.329
Total	6.575

Var[β]

0.1037	0	0	0	0	0	0	0.0055
0	0.1168	0	0	0	0	0	0.0061
0	0	0.1395	0	0	0	0	0.0073
0	0	0	0.1764	0	0	0	0.0093
0	0	0	0	0.2389	0	0	0.0126
0	0	0	0	0	0.3674	0	0.0193
0	0	0	0	0	0	0.7949	0.0418
0.0055	0.0061	0.0073	0.0093	0.0126	0.0193	0.0418	0.0054

σ^2

1.3710E+10

Incremental Incurred Losses

FY	@12	@24	@36	@48	@60	@72	@84	@108	Ultimate	Variance
1988	422,076	83,969	19,571	55,987	-11,674	63,125	-50,032	43,176	626,198	1.380E+10
1989	457,750	397,257	151,515	61,946	135,729	-56,771	-53,971	46,575	1,140,030	4.516E+10
1990	345,084	382,514	342,069	12,295	-57,918	939	-53,973	46,577	1,017,587	6.702E+10
1991	591,842	288,791	152,181	307,218	23,576	945	-54,316	46,873	1,357,111	8.643E+10
1992	379,033	157,808	74,764	113,702	23,693	950	-54,586	47,106	742,470	1.045E+11
1993	308,803	139,977	147,632	109,715	22,863	917	-52,672	45,454	722,689	1.186E+11
1994	215,851	212,472	130,116	96,698	20,150	808	-46,423	40,061	669,733	1.262E+11
1995	327,792	200,515	122,794	91,256	19,016	763	-43,810	37,807	756,132	1.381E+11

	Incurred	IBNR	Ultimate	Variance	Std. Dev.
FY 1988-1994:	5,370,760	905,058	6,275,818	1.338E+12	1,156,753
FY 1995:		756,132	756,132	1.381E+11	371,584
Total	5,370,760	1,661,190	7,031,950	1.782E+12	1,334,794

Generalized Variance 1.609E+10

Exhibit 23

Self-Insured Entity Model 3: Unrelated Paid and Incurred Losses

y= Xβ+**e**, where Var[**e**]=$\sigma^2\Phi$, subject to the constraint that

$$\begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & -9 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & -19 \end{bmatrix}\beta = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\mathbf{y} = \begin{bmatrix} \mathbf{y}_{paid} \\ \mathbf{y}_{incd} \end{bmatrix} \qquad X = \begin{bmatrix} X_{64x8} & \\ & X_{64x8} \end{bmatrix} \qquad \Phi = \begin{bmatrix} I_{64} & \\ & \sigma^2\text{ratio}\, I_{64} \end{bmatrix}$$

Note 1: X $_{64x8}$ is the same as X in Exhibits 19 and 21.

Note 2: σ^2ratio = σ^2_{incd} (Exhibit 22)/σ^2_{paid} (Exhibit 20) = 1.3710E+10/6.5637E+09 = 2.089

Note 3: σ^2 will equal σ^2_{paid} , and Model 3 will reproduce the results of Exhibits 20 and 22.

Exhibit 24

Self-Insured Entity Model 4: Paid and Incurred Losses have same Ultimate

Similar to Model 3 except that new $\Phi = \Phi\text{-}\Phi G'(G\Phi G')^{-1}G\Phi$, where

$$G_{8\times 128} = \begin{bmatrix} 1_{1\times 8} & & & & & & & & -1_{1\times 8} & & & & & & & \\ & 1_{1\times 8} & & & & & & & & -1_{1\times 8} & & & & & & \\ & & 1_{1\times 8} & & & & & & & & -1_{1\times 8} & & & & & \\ & & & 1_{1\times 8} & & & & & & & & -1_{1\times 8} & & & & \\ & & & & 1_{1\times 8} & & & & & & & & -1_{1\times 8} & & & \\ & & & & & 1_{1\times 8} & & & & & & & & -1_{1\times 8} & & \\ & & & & & & 1_{1\times 8} & & & & & & & & -1_{1\times 8} & \\ & & & & & & & 1_{1\times 8} & & & & & & & & -1_{1\times 8} \end{bmatrix},$$

and with additional constraints:

$$G\,X\,\beta = 0$$

$$\begin{bmatrix} 1_{1\times 8} & -1_{1\times 8} \end{bmatrix}\beta = 0$$

Exhibit 25

Self-Insured Entity Model 4 (Cont'd). Results

Incremental Paid Losses

FY	@12	@24	@36	@48	@60	@72	@84	@108	Ultimate	Variance
1988	266,354	166,572	32,329	53,610	8,124	16,924	39,109	81,406	664,428	4.557E+09
1989	246,981	359,380	229,016	69,539	118,635	100,292	5,628	99,174	1,228,645	1.518E+10
1990	203,178	375,768	276,617	74,912	86,428	42,446	2,593	96,142	1,158,085	2.248E+10
1991	395,630	260,643	167,709	270,692	80,042	60,977	20,871	115,015	1,371,579	2.893E+10
1992	207,698	174,615	162,640	112,706	67,410	48,250	7,945	102,556	883,820	3.489E+10
1993	167,681	280,178	162,551	105,069	61,361	42,873	3,981	95,275	918,969	3.952E+10
1994	215,740	227,745	143,256	92,594	54,072	37,777	3,500	83,962	858,646	4.181E+10
1995	203,084	220,742	141,007	93,197	56,842	41,465	9,116	85,051	850,505	4.565E+10

	Paid	Unpaid	Ultimate	Variance	Std. Dev.
FY 1988-1994:	5,026,994	2,057,178	7,084,172	4.667E+11	683,148
FY 1995:		850,505	850,505	4.565E+10	213,656
Total	5,026,994	2,907,683	7,934,677	6.212E+11	788,147

Generalized Variance 7.112E+09

Incremental Incurred Losses

FY	@12	@24	@36	@48	@60	@72	@84	@108	Ultimate	Variance
1988	422,076	83,969	19,571	55,987	-11,674	63,125	-50,032	81,406	664,428	4.557E+09
1989	457,750	397,257	151,515	61,946	135,729	-56,771	17,128	64,091	1,228,645	1.518E+10
1990	345,084	382,514	342,069	12,295	-57,918	40,140	23,468	70,433	1,158,085	2.248E+10
1991	591,842	288,791	152,181	307,218	11,085	2,252	-14,527	32,737	1,371,579	2.893E+10
1992	379,033	157,808	74,764	131,648	38,356	29,479	12,617	60,115	883,820	3.489E+10
1993	308,803	139,977	169,028	134,730	44,709	36,143	19,872	65,705	918,969	3.952E+10
1994	215,851	228,277	148,993	118,764	39,424	31,874	17,534	57,929	858,646	4.181E+10
1995	328,889	203,287	128,465	99,937	25,062	17,937	4,404	42,525	850,505	4.565E+10

	Incurred	IBNR	Ultimate	Variance	Std. Dev.
FY 1988-1994:	5,370,760	1,713,412	7,084,172	4.667E+11	683,148
FY 1995:		850,505	850,505	4.565E+10	213,656
Total	5,370,760	2,563,917	7,934,677	6.212E+11	788,147

Generalized Variance 1.262E+10

Exhibit 26

Self-Insured Entity Models: Comparison of Variances

Paid Losses				
	Variance		Efficiency Gain	
FY	Model 1	Model 4	Model 4 : Model 1	
1988	6.761E+09	4.557E+09	1.484	48.4%
1989	2.269E+10	1.518E+10	1.495	49.5%
1990	3.357E+10	2.248E+10	1.493	49.3%
1991	4.315E+10	2.893E+10	1.492	49.2%
1992	5.200E+10	3.489E+10	1.490	49.0%
1993	5.881E+10	3.952E+10	1.488	48.8%
1994	6.205E+10	4.181E+10	1.484	48.4%
1995	6.765E+10	4.565E+10	1.482	48.2%
1988-1994	7.114E+11	4.667E+11	1.524	52.4%
1995	6.765E+10	4.565E+10	1.482	48.2%
Total	9.468E+11	6.212E+11	1.524	52.4%
Gen. Var.	7.718E+09	7.112E+09	1.085	8.5%

Incurred Losses				
	Variance		Efficiency Gain	
FY	Model 2	Model 4	Model 4 : Model 2	
1988	1.380E+10	4.557E+09	3.029	202.9%
1989	4.516E+10	1.518E+10	2.975	197.5%
1990	6.702E+10	2.248E+10	2.982	198.2%
1991	8.643E+10	2.893E+10	2.988	198.8%
1992	1.045E+11	3.489E+10	2.994	199.4%
1993	1.186E+11	3.952E+10	3.002	200.2%
1994	1.262E+11	4.181E+10	3.017	201.7%
1995	1.381E+11	4.565E+10	3.025	202.5%
1988-1994	1.338E+12	4.667E+11	2.867	186.7%
1995	1.381E+11	4.565E+10	3.025	202.5%
Total	1.782E+12	6.212E+11	2.868	186.8%
Gen. Var.	1.609E+10	1.262E+10	1.276	27.6%

Exhibit 27

Self-Insured Entity Models: Comparison of Methods

Paid Losses								
	Chain Ladder Exhibit 17		Additive Exhibit 17		Model 1 Exhibit 20		Model 4 (Conjoint) Exhibit 25	
FY	Ultimate	Relation	Ultimate	Relation	Ultimate	Relation	Ultimate	Relation
1988	647,802	97%	681,934	103%	686,231	103%	664,428	100%
1989	1,338,501	109%	1,272,730	104%	1,277,366	104%	1,228,645	100%
1990	1,302,693	112%	1,226,626	106%	1,232,810	106%	1,158,085	100%
1991	1,529,306	111%	1,379,029	101%	1,386,779	101%	1,371,579	100%
1992	881,071	100%	951,247	108%	960,371	109%	883,820	100%
1993	960,741	105%	1,011,280	110%	1,021,938	111%	918,969	100%
1994	965,937	112%	917,446	107%	957,338	111%	858,646	100%
1995	912,905	107%	866,112	102%	903,741	106%	850,505	100%
1988-1994	7,626,052	108%	7,440,292	105%	7,522,834	106%	7,084,172	100%
1995	912,905	107%	866,112	102%	903,741	106%	850,505	100%
Total	8,538,958	108%	8,306,405	105%	8,426,574	106%	7,934,677	100%

Incurred Losses								
	Chain Ladder Exhibit 18		Additive Exhibit 18		Model 2 Exhibit 22		Model 4 (Conjoint) Exhibit 25	
FY	Ultimate	Relation	Ultimate	Relation	Ultimate	Relation	Ultimate	Relation
1988	613,707	92%	624,390	94%	626,198	94%	664,428	100%
1989	1,112,360	91%	1,138,080	93%	1,140,030	93%	1,228,645	100%
1990	996,304	86%	1,017,995	88%	1,017,587	88%	1,158,085	100%
1991	1,335,293	97%	1,356,684	99%	1,357,111	99%	1,371,579	100%
1992	682,792	77%	740,818	84%	742,470	84%	883,820	100%
1993	606,773	66%	719,531	78%	722,689	79%	918,969	100%
1994	460,836	54%	638,448	74%	669,733	78%	858,646	100%
1995	695,276	82%	724,475	85%	756,132	89%	850,505	100%
1988-1994	5,808,065	82%	6,235,947	88%	6,275,818	89%	7,084,172	100%
1995	695,276	82%	724,475	85%	756,132	89%	850,505	100%
Total	6,503,341	82%	6,960,422	88%	7,031,950	89%	7,934,677	100%

Exhibit 28

Self-Insured Entity Models: Comparison of Paid-Loss Patterns

Paid Losses				
FY	CL	Additive	Model 1	Conjoint
@12	22.3%	23.5%	22.6%	23.9%
@24	46.6%	45.9%	47.2%	49.8%
@36	61.9%	62.4%	63.1%	66.4%
@48	71.6%	73.5%	73.9%	77.4%
@60	78.1%	80.3%	80.6%	84.1%
@72	84.0%	86.0%	86.2%	88.9%
@84	90.0%	90.0%	90.0%	90.0%
@Ult	100.0%	100.0%	100.0%	100.0%

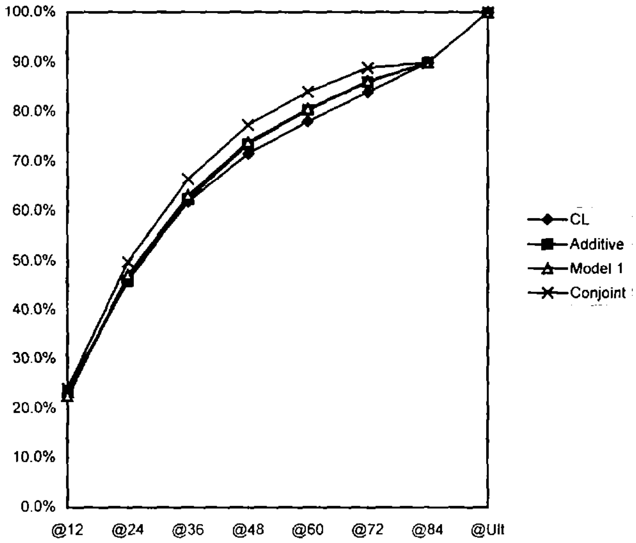

Exhibit 29

Self-Insured Entity Models: Comparison of Incurred-Loss Patterns

Incurred Losses				
FY	CL	Additive	Model 2	Conjoint
@12	46.8%	45.0%	43.4%	38.7%
@24	74.0%	68.9%	69.9%	62.6%
@36	89.6%	85.7%	86.1%	77.7%
@48	100.4%	98.1%	98.2%	89.4%
@60	102.8%	100.7%	100.7%	92.4%
@72	103.2%	101.0%	100.8%	94.5%
@84	95.0%	95.0%	95.0%	95.0%
@Ult	100.0%	100.0%	100.0%	100.0%

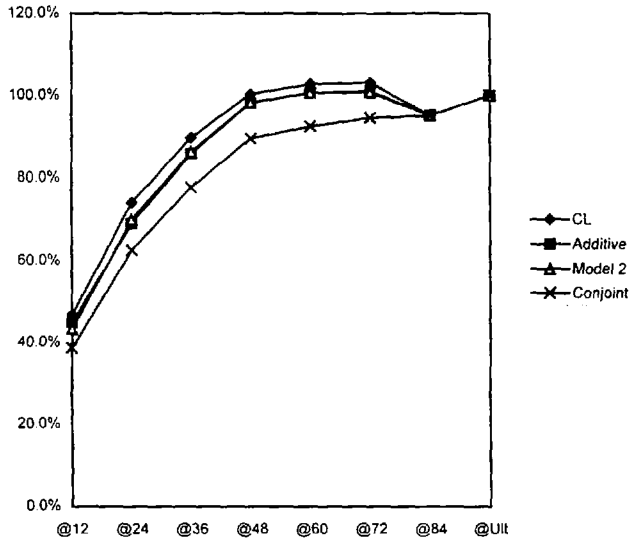

Exhibit 30

Self-Insured Entity Model 5: Paid Losses

$y_{paid} = X\beta + e$, where $Var[e] = \sigma^2\Phi$, with quasi-observation of β_{108}, subject to the constraint that $[1\ 1\ 1\ 1\ 1\ 1\ 1\ 0]\beta = [7.213]$,

$$\text{where } \Phi = \begin{bmatrix} I_{64} & \\ & \sigma^2 \text{ratio } I_1 \end{bmatrix}$$

FY	Age	y_{paid}	X							
1988	12	266,354	131,332.20	0	0	0	0	0	0	0
1988	24	166,572	0	131,332.20	0	0	0	0	0	0
1988	36	32,329	0	0	131,332.20	0	0	0	0	0
1988	48	53,610	0	0	0	131,332.20	0	0	0	0
1988	60	8,124	0	0	0	0	131,332.20	0	0	0
1988	72	16,924	0	0	0	0	0	131,332.20	0	0
1988	84	39,109	0	0	0	0	0	0	131,332.20	0
1988	108	.	0	0	0	0	0	0	0	131,332.20
1989	12	246,981	141,672.24	0	0	0	0	0	0	0
1989	24	359,380	0	141,672.24	0	0	0	0	0	0
1989	36	229,016	0	0	141,672.24	0	0	0	0	0
1989	48	69,539	0	0	0	141,672.24	0	0	0	0
1989	60	118,635	0	0	0	0	141,672.24	0	0	0
1989	72	100,292	0	0	0	0	0	141,672.24	0	0
1989	84	.	0	0	0	0	0	0	141,672.24	0
1989	108	.	0	0	0	0	0	0	0	141,672.24
1990	12	203,178	141,677.29	0	0	0	0	0	0	0
1990	24	375,768	0	141,677.29	0	0	0	0	0	0
1990	36	276,617	0	0	141,677.29	0	0	0	0	0
1990	48	74,912	0	0	0	141,677.29	0	0	0	0
1990	60	86,428	0	0	0	0	141,677.29	0	0	0
1990	72	.	0	0	0	0	0	141,677.29	0	0
1990	84	.	0	0	0	0	0	0	141,677.29	0
1990	108	.	0	0	0	0	0	0	0	141,677.29
1991	12	395,630	142,577.99	0	0	0	0	0	0	0
1991	24	260,643	0	142,577.99	0	0	0	0	0	0
1991	36	167,709	0	0	142,577.99	0	0	0	0	0
1991	48	270,692	0	0	0	142,577.99	0	0	0	0
1991	60	.	0	0	0	0	142,577.99	0	0	0
1991	72	.	0	0	0	0	0	142,577.99	0	0
1991	84	.	0	0	0	0	0	0	142,577.99	0
1991	108	.	0	0	0	0	0	0	0	142,577.99
1992	12	207,698	143,285.58	0	0	0	0	0	0	0
1992	24	174,615	0	143,285.58	0	0	0	0	0	0
1992	36	162,640	0	0	143,285.58	0	0	0	0	0
1992	48	.	0	0	0	143,285.58	0	0	0	0
1992	60	.	0	0	0	0	143,285.58	0	0	0
1992	72	.	0	0	0	0	0	143,285.58	0	0
1992	84	.	0	0	0	0	0	0	143,285.58	0
1992	108	.	0	0	0	0	0	0	0	143,285.58
1993	12	167,681	138,261.75	0	0	0	0	0	0	0
1993	24	280,178	0	138,261.75	0	0	0	0	0	0
1993	36	.	0	0	138,261.75	0	0	0	0	0
1993	48	.	0	0	0	138,261.75	0	0	0	0
1993	60	.	0	0	0	0	138,261.75	0	0	0
1993	72	.	0	0	0	0	0	138,261.75	0	0
1993	84	.	0	0	0	0	0	0	138,261.75	0
1993	108	.	0	0	0	0	0	0	0	138,261.75
1994	12	215,740	121,857.69	0	0	0	0	0	0	0
1994	24	.	0	121,857.69	0	0	0	0	0	0
1994	36	.	0	0	121,857.69	0	0	0	0	0
1994	48	.	0	0	0	121,857.69	0	0	0	0
1994	60	.	0	0	0	0	121,857.69	0	0	0
1994	72	.	0	0	0	0	0	121,857.69	0	0
1994	84	.	0	0	0	0	0	0	121,857.69	0
1994	108	.	0	0	0	0	0	0	0	121,857.69
1995	12	.	115,000.00	0	0	0	0	0	0	0
1995	24	.	0	115,000.00	0	0	0	0	0	0
1995	36	.	0	0	115,000.00	0	0	0	0	0
1995	48	.	0	0	0	115,000.00	0	0	0	0
1995	60	.	0	0	0	0	115,000.00	0	0	0
1995	72	.	0	0	0	0	0	115,000.00	0	0
1995	84	.	0	0	0	0	0	0	115,000.00	0
1995	108	.	0	0	0	0	0	0	0	115,000.00
Quasi-Observation		0.801	0	0	0	0	0	0	0	1

Exhibit 31

Self-Insured Entity Model 5 (Cont'd): Results

	β
@12	1.780
@24	1.942
@36	1.263
@48	0.863
@60	0.542
@72	0.467
@84	0.355
@108	0.801
Total	8.014

Var[β]

0.0449	-0.0029	-0.0034	-0.0043	-0.0058	-0.0090	-0.0195	0
-0.0029	0.0502	-0.0038	-0.0049	-0.0066	-0.0101	-0.0219	0
-0.0034	-0.0038	0.0592	-0.0058	-0.0079	-0.0121	-0.0262	0
-0.0043	-0.0049	-0.0058	0.0733	-0.0099	-0.0153	-0.0331	0
-0.0058	-0.0066	-0.0079	-0.0099	0.0958	-0.0207	-0.0448	0
-0.0090	-0.0101	-0.0121	-0.0153	-0.0207	0.1362	-0.0689	0
-0.0195	-0.0219	-0.0262	-0.0331	-0.0448	-0.0689	0.2144	0
0	0	0	0	0	0	0	0.2128

σ^2

6.2717E+09

Incremental Paid Losses

FY	@12	@24	@36	@48	@60	@72	@84	@108	Ultimate	Variance
1988	266,354	166,572	32,329	53,610	8,124	16,924	39,109	105,254	688,276	9.941E+09
1989	246,981	359,380	229,016	69,539	118,635	100,292	50,335	113,541	1,287,719	2.112E+10
1990	203,178	375,768	276,617	74,912	86,428	66,144	50,337	113,545	1,246,929	2.736E+10
1991	395,630	260,643	167,709	270,692	77,289	66,565	50,657	114,267	1,403,452	3.302E+10
1992	207,698	174,615	162,640	123,692	77,672	66,895	50,908	114,834	978,955	3.848E+10
1993	167,681	280,178	174,622	119,355	74,949	64,550	49,123	110,808	1,041,266	4.341E+10
1994	215,740	236,661	153,904	105,194	66,057	56,891	43,295	97,661	975,403	4.773E+10
1995	204,739	223,342	145,243	99,274	62,339	53,690	40,858	92,165	921,651	5.299E+10

	Paid	Unpaid	Ultimate	Variance	Std. Dev.
FY 1988-1994:	5,026,994	2,595,006	7,622,000	4.325E+11	657,623
FY 1995:		921,651	921,651	5.299E+10	230,189
Total	5,026,994	3,516,658	8,543,652	5.325E+11	729,701
Generalized Variance				7.503E+09	

Exhibit 32

Self-Insured Entity Model 4: Results at Present Value

Incremental Unpaid Losses at Present Value

FY	@12	@24	@36	@48	@60	@72	@84	@108	Unpaid	Variance
1988								74,215	74,215	3.787E+09
1989							5,466	84,055	89,520	1.317E+10
1990						41,222	2,364	75,895	119,481	1.837E+10
1991					77,733	55,590	17,689	84,683	235,695	2.203E+10
1992				109,454	61,455	40,895	6,272	70,182	288,258	2.474E+10
1993			157,861	95,787	52,007	33,844	2,931	60,818	403,247	2.622E+10
1994		221,175	130,600	78,478	42,684	27,815	2,395	49,810	552,957	2.615E+10
1995	197,225	201,242	119,511	73,570	41,852	28,376	5,819	47,063	714,657	2.697E+10

	Unpaid	Variance	Std. Dev.
FY 1988-1994:	1,763,374	3.256E+11	570,628
FY 1995:	714,657	2.697E+10	164,224
Total	2,478,031	4.136E+11	643,147

Generalized Variance 4.714E+09

Appendix A

Basic Multivariate Statistical Concepts

A random matrix is a matrix whose elements are random scalars. Suppose $\mathbf{Y}$ to be an $(m \times n)$ random matrix, whose ij^{th} element, $\{\mathbf{Y}\}_{ij}$, is the random scalar y_{ij}. Then the expectation of $\mathbf{Y}$, denoted $\mathrm{E}[\mathbf{Y}]$, is an $(m \times n)$ matrix whose ij^{th} element, $\{\mathrm{E}[\mathbf{Y}]\}_{ij}$, is $\mathrm{E}[y_{ij}]$. In other words, the expectation of a matrix is the matrix of the expectations.

Now consider $\mathrm{A}\mathbf{Y}$, where A is a $(p \times m)$ non-stochastic matrix. Then $\mathrm{A}\mathbf{Y}$ is a $(p \times n)$ random matrix, whose ij^{th} element is:

$$\begin{aligned}\{\mathrm{A}\mathbf{Y}\}_{ij} &= \sum_{k=1}^{m} \{\mathrm{A}\}_{ik} \{\mathbf{Y}\}_{kj} \\ &= \sum_{k=1}^{m} a_{ik} y_{kj}\end{aligned}$$

Since the expectation of $\mathrm{A}\mathbf{Y}$ is the matrix of the expectations, the ij^{th} element of $\mathrm{E}[\mathrm{A}\mathbf{Y}]$ is:

$$\begin{aligned}\{\mathrm{E}[\mathrm{A}\mathbf{Y}]\}_{ij} &= \mathrm{E}\left[\{\mathrm{A}\mathbf{Y}\}_{ij}\right] \\ &= \mathrm{E}\left[\sum_{k=1}^{m} a_{ik} y_{kj}\right] \\ &= \sum_{k=1}^{m} \mathrm{E}[a_{ik} y_{kj}] \\ &= \sum_{k=1}^{m} a_{ik} \mathrm{E}[y_{kj}] \\ &= \sum_{k=1}^{m} \{\mathrm{A}\}_{ik} \{\mathrm{E}[\mathbf{Y}]\}_{kj} \\ &= \{\mathrm{A}\mathrm{E}[\mathbf{Y}]\}_{ij}\end{aligned}$$

Therefore, if A is non-stochastic, then $\mathrm{E}[\mathrm{A}\mathbf{Y}] = \mathrm{A}\mathrm{E}[\mathbf{Y}]$. By similar reasoning, if B is non-stochastic $\mathrm{E}[\mathbf{Y}\mathrm{B}] = \mathrm{E}[\mathbf{Y}]\mathrm{B}$. Also, $\mathrm{E}[\mathrm{A}\mathbf{Y}\mathrm{B}] = \mathrm{E}[\mathrm{A}(\mathbf{Y}\mathrm{B})] = \mathrm{A}\mathrm{E}[\mathbf{Y}\mathrm{B}] = \mathrm{A}\mathrm{E}[\mathbf{Y}]\mathrm{B}$.

Let $\mathbf{x}_1$ be an $(m\times 1)$ random column vector whose expectation exists, viz., $E[\mathbf{x}_1] = \mu_1$. Similarly, let $\mathbf{x}_2$ be an $(n\times 1)$ random vector whose expectation is μ_2. The covariance of $\mathbf{x}_1$ with $\mathbf{x}_2$ is defined as $\mathrm{Cov}[\mathbf{x}_1, \mathbf{x}_2] = E[(\mathbf{x}_1-\mu_1)(\mathbf{x}_2-\mu_2)']$, an $(m\times n)$ matrix. But $\mathrm{Cov}[\mathbf{x}_1, \mathbf{x}_2] = E[(\mathbf{x}_1-\mu_1)(\mathbf{x}_2-\mu_2)'] = E[((\mathbf{x}_1-\mu_1)(\mathbf{x}_2-\mu_2)')']' = E[(\mathbf{x}_2-\mu_2)(\mathbf{x}_1-\mu_1)']' = \mathrm{Cov}[\mathbf{x}_2, \mathbf{x}_1]'$. Also,

$$\begin{aligned}
\mathrm{Cov}[A\mathbf{x}_1, B\mathbf{x}_2] &= E[(A\mathbf{x}_1 - A\mu_1)(B\mathbf{x}_1 - B\mu_2)'] \\
&= E[A(\mathbf{x}_1 - \mu_1)(B(\mathbf{x}_1 - \mu_2))'] \\
&= E[A(\mathbf{x}_1 - \mu_1)(\mathbf{x}_1 - \mu_2)'B'] \\
&= AE[(\mathbf{x}_1 - \mu_1)(\mathbf{x}_1 - \mu_2)']B' \\
&= A\mathrm{Cov}[\mathbf{x}_1, \mathbf{x}_2]B'
\end{aligned}$$

The variance of $\mathbf{x}$ is defined as $\mathrm{Var}[\mathbf{x}] = \mathrm{Cov}[\mathbf{x}, \mathbf{x}] = E[(\mathbf{x}-\mu)(\mathbf{x}-\mu)']$. Therefore, $\mathrm{Var}[\mathbf{x}] = \mathrm{Cov}[\mathbf{x}, \mathbf{x}] = \mathrm{Cov}[\mathbf{x}, \mathbf{x}]' = \mathrm{Var}[\mathbf{x}]'$. In other words, a variance matrix must be symmetric. Also, $\mathrm{Var}[A\mathbf{x}] = \mathrm{Cov}[A\mathbf{x}, A\mathbf{x}] = A\mathrm{Cov}[\mathbf{x}, \mathbf{x}]A' = A\mathrm{Var}[\mathbf{x}]A'$.

Throughout this paper we assume that the elements of all matrices belong to the real numbers, i.e., that no complex numbers are allowed. If we now restrict A to being a $(1\times n)$ non-stochastic row vector, then $A\mathbf{x}$ will be a real-valued (1×1) linear combination of the elements of $\mathbf{x}$. Its variance, $A\mathrm{Var}[\mathbf{x}]A'$, is a (1×1) matrix whose element must be greater than or equal to zero (since the variance of a real-valued random scalar cannot be negative). This shows that the variance of a real-valued random vector, $\Sigma = \mathrm{Var}[\mathbf{x}]$, must have the property that for every conformable row vector A, $A\Sigma A' \geq [0]$. A symmetric matrix Σ with this property is said to be *non-negative definite*. Of course, if A is a zero vector, then $A\Sigma A' = [0]$; so [0] can always be obtained. However, if for all non-zero A, $A\Sigma A' > [0]$, then Σ is said to be *positive definite*.

Consider the following symmetric (2×2) matrices:

$$\Sigma_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$\Sigma_2 = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$$

$$\Sigma_3 = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}$$

Letting A be $[a_1 \;\; a_2]$, we have $A\Sigma_1 A' = [a_1^2 + a_2^2]$. This is greater than or equal to zero, and equal to zero if and only if a_1 and a_2 are zero. Therefore Σ_1 is positive definite. And $A\Sigma_2 A' = [a_1^2 + 2a_1a_2 + a_2^2] = [(a_1 + a_2)^2]$. This is greater than or equal to zero, but is equal to zero whenever $a_2 = -a_1$. Therefore, Σ_2 is not positive definite, but only non-negative definite. Finally, $A\Sigma_3 A' = [a_1^2 + 4a_1a_2 + a_2^2]$. This is negative for many values of a_1 and a_2, for example, for $a_1 = 1$ and $a_2 = -1$. Therefore, Σ_3 is not even non-negative definite. Σ_3 cannot be a variance matrix for some $\mathbf{x}$; otherwise $[1 \; -1]\mathbf{x}$ would be a real-valued random variable with a negative variance. Σ_2 can be a variance matrix for some $\mathbf{x}$; however, $[1 \; -1]\mathbf{x}$ has a variance of zero. If the variance of a random vector is positive definite, then no non-zero linear combination of the random vector is without some variance; whereas if it is only non-negative definite, then some non-zero linear combination degenerates to a constant.

Consider two $(n\times 1)$ random vectors, $\mathbf{x}_1$ and $\mathbf{x}_2$, with respective $(n\times n)$ variances Σ_1 and Σ_2. How can we decide which variance is smaller when $n > 1$? Order relations are defined only for real-valued scalars and (1×1) matrices. So consider $A\mathbf{x}_1$ and $A\mathbf{x}_2$, where A is $(1\times n)$. These are (1×1) random vectors, whose (1×1) variances $A\Sigma_1 A'$ and $A\Sigma_2 A'$ are susceptible to ordering. We could say that the n-dimensional random variables $\mathbf{x}_1$ and $\mathbf{x}_2$ have been collapsed into 1-space by the linear combination A. If the variance of $\mathbf{x}_1$ is less than or equal to that of $\mathbf{x}_2$ after any such collapse into 1-space, then we could define $\mathbf{x}_1$'s variance to be the smaller. In other words, $\Sigma_1 \le \Sigma_2$. means that for any $(1\times n)$ vector A, $A\Sigma_1A' \le A\Sigma_2A'$. But $A\Sigma_1A' \le A\Sigma_2A'$ if and only if $[0] \le A\Sigma_2A' - A\Sigma_1A' = A(\Sigma_2 - \Sigma_1)A'$. This means that $\Sigma_2 - \Sigma_1$ is non-negative definite. Similarly, $\Sigma_1 < \Sigma_2$ means that $\Sigma_2 - \Sigma_1$ is positive definite. Every two real-valued scalars are able to be ordered, i.e., the one is either less than, equal to, or greater than the other. This is called the law of trichotomy. For symmetric matrices in general, the law of trichotomy does not hold: $A\Sigma_1A'$ may be less than $A\Sigma_2A'$ for some A, and greater for others. However, whenever we have two unbiased estimators and the difference between their variance matrices is non-zero and non-negative definite, then we have reason for preferring one estimator to the other. This is the basis for determining what is "Best" in Best Linear Unbiased Estimation (BLUE).

The many theorems about non-negative definite (NND) and positive definite (PD) matrices can be proved easily by matrix diagonalization. It is a theorem of matrix algebra that if Σ is an $(n\times n)$ symmetric real-valued matrix, then Σ can be factored as $W\Lambda W'$, where W is orthogonal and Λ is diagonal. W and Λ are real-valued. An orthogonal matrix is one whose inverse is its transpose, so $WW' = W'W = I_n$. To demonstrate this factorization (cf. Healy [9: 56-59] and Judge [11: Appendix A, 951-957]) involves eigenvalues, and the diagonal elements of Λ (the non-diagonal elements are zeroes) are called the eigenvalues of Σ. Diagonalization helps in this way: whether or not a symmetric matrix Σ is NND (or PD) depends on the behavior of $A\Sigma A'$. But, by diagonalization, $A\Sigma A' = A(W\Lambda W')A' =$

$(AW)\Lambda(W'A') = (AW)\Lambda(AW)'$. But AW, a $(1\times n)$ row vector, is an orthogonal one-to-one transformation of A, an arbitrary $(1\times n)$ row vector. And $AW = 0$ if and only if $A = 0$. Therefore, Σ is NND (or PD) if and only if Λ is NND (or PD). Therefore, if one knows the eigenvalues of a symmetric matrix, one will know as to its definiteness.

The definiteness of a diagonal matrix is easily determined:

$$
\begin{aligned}
A\Lambda A' &= \begin{bmatrix} a_1 & \dots & a_n \end{bmatrix} \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix} \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} \\
&= \left[\sum_{i=1}^{n} \lambda_i a_i^2 \right]
\end{aligned}
$$

Barring $A = 0$, the element will always be positive if and only if all the λ_is (the eigenvalues) are positive. Therefore, any symmetric real-valued Σ is positive definite if and only if all its eigenvalues are positive. It is non-negative definite if and only if all its eigenvalues are non-negative. Also, since Σ can be factored as $W\Lambda W'$, $|\Sigma| = |W\Lambda W'| = |W||\Lambda||W'| = |\Lambda||W'||W| = |\Lambda||W'W| = |\Lambda||I_n| = |\Lambda|$. But the determinant of Λ, a diagonal matrix, is the product of the diagonal elements. Therefore, the determinant of Σ is equal to the product of its eigenvalues; so that if Σ is NND (or PD), then $|\Sigma| \geq 0$ (or > 0).

If Σ is NND, then all its eigenvalues, being non-negative, have non-negative square roots. Let $\Lambda^{1/2}$ be the diagonal matrix containing these square roots. Then $\Sigma = W\Lambda W' = W\Lambda^{1/2}\Lambda^{1/2}W'$ $= W\Lambda^{1/2}\Lambda^{1/2}{}'W' = W\Lambda^{1/2}(W\Lambda^{1/2})'$. This shows that every NND matrix can be factored as the product of some real-valued matrix and its transpose. The converse is also true, viz., that if a matrix can be factored as the product of some real-valued matrix and its transpose, then it is NND. For suppose that $\Sigma = QQ'$, for some $(n \times p)$ Q. Then $A\Sigma A' = AQQ'A' = (AQ)(AQ)' \geq [0]$, since AQ is a real-valued $(1 \times p)$ row vector, the sum of the squares of whose elements equals the element of $(AQ)(AQ)'$. Therefore, a matrix is NND if and only if it can be factored as the product some real-valued matrix and its transpose.

If $\Sigma = W\Lambda W'$ is PD, then all the diagonal elements of Λ are positive. Then Λ^{-1} exists, its diagonal being the reciprocal of the diagonal of Λ. All the elements of Λ^{-1} are positive, so Λ^{-1} is positive definite. However, since W is orthogonal:

$$\begin{aligned}\Sigma(W\Lambda^{-1}W') &= (W\Lambda W')(W\Lambda^{-1}W') \\ &= W\Lambda(W'W)\Lambda^{-1}W' \\ &= W\Lambda(I_n)\Lambda^{-1}W' \\ &= W\Lambda\Lambda^{-1}W' \\ &= WW' \\ &= I_n\end{aligned}$$

Similarly, $(W\Lambda^{-1}W')\Sigma = I_n$. Therefore, Σ has an inverse, which inverse, having positive eigenvalues, is also positive definite. So Σ is PD if and only if Σ^{-1} exists and is PD. Also, if a NND matrix has an inverse, then it is also PD. For if the NND matrix has an inverse, then its determinant is non-zero. But its determinant is the product of its eigenvalues. Therefore, no eigenvalue is zero. Since the eigenvalues of the NND matrix are non-negative, and none is zero, then they are all positive, which makes the matrix also PD.

Finally, consider the expression $U\Sigma U'$, where U is $(m \times n)$ and Σ is an $(n \times n)$ NND matrix. As shown above, there is an $(n \times n)$ matrix Q, such that $\Sigma = QQ'$. Then $U\Sigma U' = UQQ'U' = (UQ)(UQ)'$, so we know that $U\Sigma U'$ is an $(m \times m)$ NND matrix.

Now, in addition suppose that Σ is PD and that the rank of U, or $\rho(U)$, is m. It is a theorem of matrix algebra (cf. Eves [5: 84f.]) that if U is real-valued, then $\rho(UU') = \rho(U) = m$. Then UU' is an $(m \times m)$ matrix of rank m, which means that $(UU')^{-1}$ exists. $U\Sigma U'$ is known to be NND; it will also be PD if and only if, for any non-zero $(1 \times m)$ A, $AU\Sigma U'A' = (AU)\Sigma(AU)' > [0]$. Since Σ is PD, $(AU)\Sigma(AU)' > [0]$ if and only if AU is non-zero. Now, if $A = 0$, then $AU = 0$. Conversely, if $AU = 0$, then $A = AUU'(UU')^{-1} = 0U'(UU')^{-1} = 0$. Therefore, due to the additional suppositions, AU is non-zero if and only if A is non-zero. Therefore, if U is of full row rank and Σ is PD, then $U\Sigma U'$ is PD.

Appendix B

The Moore-Penrose Inverse and the Solution of Linear Equations

A matrix G is an inverse of an $(m \times n)$ matrix A if and only if AG and GA are identity matrices. For AG and GA to exist, G must be $(n \times m)$. Therefore, $AG = I_m$, and $GA = I_n$. It is a theorem of matrix algebra (cf. Eves [5: 82]) that the rank of a product of matrices is less than or equal to the rank of each of the matrices, e.g., $\rho(AG) \leq \min(\rho(A), \rho(G))$. Also, both $\rho(A)$ and $\rho(G)$ are less than or equal to $\min(m, n)$. Therefore, $\rho(AG) \leq \min(m, n)$. But $m = \rho(I_m) = \rho(AG) \leq \min(\rho(A), \rho(G)) \leq \min(m, n) \leq m$. Hence, both A and G must be of rank m. Similarly, $n = \rho(I_n) = \rho(GA) \leq \min(\rho(G), \rho(A)) \leq \min(m, n) \leq n$, and both A and G must be of rank n. Since the rank of a matrix is unique, $m = n$. Therefore, a matrix has an inverse if and only if it is square and of full rank.

But the concept of an inverse has been generalized to any $(m \times n)$ matrix A, and the *generalized inverse* of a matrix is very useful. This appendix will treat one type of generalized inverse, the Moore-Penrose inverse, and its application to the solution of linear equations. All this will be relevant to the linear statistical model.

Searle [17: 1] defines a generalized inverse thus: G is a generalized inverse of A if and only if $AGA = A$. If A is $(m \times n)$, then G must be $(n \times m)$. G may not be unique; in fact, if $A = 0$, then any $(n \times m)$ matrix will be its generalized inverse. This definition does not demand that $AG = I_m$, a condition too restrictive. However, it does demand that AG behave like I_m when premultiplying A, since $(AG)A = A = (I_m)A$; so in a sense AG is like I_m.

The traditional matrix inverse has the property that the inverse of the inverse of A is A, or $(A^{-1})^{-1} = A$. The Moore-Penrose (MP) inverse goes beyond the generalized inverse by demanding that if G is an MP inverse of A, then A is an MP-inverse of G. This is equivalent to demanding that GAG = G. Again, it is not demanded that $GA = I_n$, only that it behave like I_n when premultiplying G.

The MP inverse will make G as close as possible to the traditional inverse. Two more demands are made: although AG and GA may not be identity matrices, they should at least have the symmetry of identity matrices. Thus we have the definition: G is an MP inverse of A if and only if:

1. AGA = A
2. GAG = G
3. AG is symmetric, or AG = (AG)'
4. GA is symmetric, or GA = (GA)'

The first task is to establish that every matrix has an MP inverse, i.e., the proof of existence. To do this we start with an $(m\times n)$ matrix A whose rank is r. We need to prove that A can be factored as BC, where B is $(m\times r)$, C is $(r\times n)$, and $\rho(B) = \rho(C) = r$. A fundamental theorem of matrix algebra (cf. Eves [5: 74]) states that there exist non-singular matrices P and Q such that:

$$A = P\begin{bmatrix} I_r & \\ & \end{bmatrix}Q$$

The right side of the equation is called a canonical form of A, and r (the rank) is the same for all such forms of A. The matrix in between P and Q is zero, except that the first r elements of its diagonal are ones. Then A can be factored as:

$$
\begin{aligned}
A &= P_{(m\times m)}\begin{bmatrix} I_r & \\ & \end{bmatrix}_{(m\times n)} Q_{(n\times n)} \\
&= P_{(m\times m)}\begin{bmatrix} I_r \\ \end{bmatrix}_{(m\times r)} \begin{bmatrix} I_r & \end{bmatrix}_{(r\times n)} Q_{(n\times n)} \\
&= \left(P\begin{bmatrix} I_r \\ \end{bmatrix}\right)_{(m\times r)} \left(\begin{bmatrix} I_r & \end{bmatrix}Q\right)_{(r\times n)} \\
&= \left(P\begin{bmatrix} I_r \\ \end{bmatrix}I_r\right)_{(m\times r)} \left(I_r\begin{bmatrix} I_r & \end{bmatrix}Q\right)_{(r\times n)} \\
&= B_{(m\times r)}C_{(r\times n)}
\end{aligned}
$$

Since P, Q, and I_r are non-singular, B and C are names of canonical forms; so they are both of rank r. As explained in the last paragraph of Appendix A, $(B'B)^{-1}$ and $(CC')^{-1}$ exist. Then $G = C'(CC')^{-1}(B'B)^{-1}B'$ is an MP inverse of A, since it satisfies the four conditions:

$$
\begin{aligned}
AGA &= BC(C'(CC')^{-1}(B'B)^{-1}B')BC \\
&= B(CC')(CC')^{-1}(B'B)^{-1}(B'B)C \\
&= BC \\
&= A
\end{aligned}
$$

$$
\begin{aligned}
GAG &= (C'(CC')^{-1}(B'B)^{-1}B')BC(C'(CC')^{-1}(B'B)^{-1}B') \\
&= C'(CC')^{-1}(B'B)^{-1}(B'B)(CC')(CC')^{-1}(B'B)^{-1}B' \\
&= C'(CC')^{-1}(B'B)^{-1}B' \\
&= G
\end{aligned}
$$

$$
\begin{aligned}
AG &= BC(C'(CC')^{-1}(B'B)^{-1}B') \\
&= B(CC')(CC')^{-1}(B'B)^{-1}B' \\
&= B(B'B)^{-1}B' \\
&= (B(B'B)^{-1}B')' \\
&= (AG)'
\end{aligned}
$$

$$
\begin{aligned}
GA &= (C'(CC')^{-1}(B'B)^{-1}B')BC \\
&= C'(CC')^{-1}(B'B)^{-1}(B'B)C \\
&= C'(CC')^{-1}C \\
&= (C'(CC')^{-1}C)' \\
&= (GA)'
\end{aligned}
$$

Therefore, every matrix has an MP inverse.

The second task is to establish the uniqueness of the MP inverse. Suppose that G_1 and G_2 are MP inverses of A. Then, by a series of small steps involving the MP conditions and matrix transpositions:

$$
\begin{aligned}
G_1 &= G_1AG_1 \\
&= G_1(AG_1)' \\
&= G_1G_1'A' \\
&= G_1G_1'(AG_2A)' \\
&= G_1G_1'A'G_2'A' \\
&= G_1(AG_1)'G_2'A' \\
&= G_1AG_1G_2'A' \\
&= G_1G_2'A' \\
&= G_1(AG_2)' \\
&= G_1AG_2 \\
&= G_1AG_1AG_2 \\
&= (G_1A)'G_1AG_2 \\
&= A'G_1'G_1AG_2 \\
&= (AG_2A)'G_1'G_1AG_2 \\
&= A'G_2'A'G_1'G_1AG_2 \\
&= (G_2A)'(G_1A)'G_1AG_2 \\
&= (G_2A)(G_1A)G_1AG_2 \\
&= G_2AG_1(AG_1A)G_2 \\
&= G_2AG_1AG_2 \\
&= G_2AG_2 \\
&= G_2
\end{aligned}
$$

Thus, the MP inverse is unique. Because of its existence and uniqueness we can denote the MP inverse of A as A^{+} [11: Appendix A, 939].

A few theorems will now be proved:

<u>Theorem</u>: $(A^{+})^{+} = A$

Proof: A satisfies the four conditions for being an MP inverse of A^{+}:

1. $A^{+}AA^{+} = A^{+}$
2. $AA'A = A$
3. $A^{+}A$ is symmetric
4. AA^{+} is symmetric

Theorem: If A^{-1} exists, then $A^{+} = A^{-1}$.

Proof: Again, by satisfying the four MP conditions, since A^{-1} exists:

1. $AA^{-1}A = A$
2. $A^{-1}AA^{-1} = A^{-1}$
3. $AA^{-1} = I$, which is symmetric
4. $A^{-1}A = I$, which is symmetric

Theorem: $(A')^{+} = (A^{+})'$

Proof: Again, by satisfying the four MP conditions:

1. $A'(A^{+})'A' = (AA^{+}A)' = A'$
2. $(A^{+})'A'(A^{+})' = (A^{+}AA^{+})' = (A^{+})'$
3. $A'(A^{+})' = (A^{+}A)' = A^{+}A$, which is symmetric
4. $(A^{+})'A' = (AA^{+})' = AA^{+}$, which is symmetric

Theorem: $A' = A'AA^{+} = A^{+}AA'$

Proof: $A' = (AA^{+}A)' = A'(AA^{+})' = A'AA^{+}$

$A' = (AA^{+}A)' = (A^{+}A)'A' = A^{+}AA'$

Theorem: $(A'A)^{+} = A^{+}(A')^{+}$

Proof: By satisfying the four MP conditions, with earlier theorems:

1. $A'A(A^{+}(A')^{+})A'A = (A'AA^{+})((A')^{+}A'A) = (A'AA^{+})((A^{+})'A'A)$
   $= (A'AA^{+})(A'AA^{+})' = (A')(A')' = A'A$
2. $(A^{+}(A')^{+})A'A(A^{+}(A')^{+}) = A^{+}(A')^{+}(A'AA^{+})(A')^{+} = A^{+}(A')^{+}A'(A')^{+}$
   $= A^{+}(A')^{+}$
3. $A'A(A^{+}(A')^{+}) = (A'AA^{+})(A')^{+} = A'(A')^{+}$, which is symmetric
4. $(A^{+}(A')^{+})A'A = A^{+}((A')^{+}A'A) = A^{+}((A^{+})'A'A) = A^{+}(A'AA^{+})'$
   $= A^{+}(A')' = A^{+}A$, which is symmetric

Theorem: $A^{+} = (A'A)^{+}A' = A'(AA')^{+}$

Proof: $A^{+} = A^{+}(AA^{+}) = A^{+}(AA^{+})' = A^{+}(A^{+})'A' = A^{+}(A')^{+}A' = (A'A)^{+}A'$

$A^{+} = (A^{+}A)A^{+} = (A^{+}A)'A^{+} = A'(A^{+})'A^{+} = A'(A')^{+}A^{+} = A'(AA')^{+}$

Theorem: $A(A'A)^{+}A'A = A$

Proof: $A(A'A)^{+}A'A = A((A'A)^{+}A')A = A(A^{+})A = A$

Theorem: $\rho(A^{+}) = \rho(AA^{+}) = \rho(A^{+}A) = \rho(A)$

Proof: Because the rank of a product is less than or equal to the rank of any of the factors, $\rho(A) = \rho(AA^{+}A) \le \rho(AA^{+}) \le \rho(A)$. Therefore, $\rho(AA^{+}) = \rho(A)$. So, $\rho(A) = \rho(AA^{+}) \le \rho(A^{+})$. Similarly, $\rho(A^{+}) = \rho(A^{+}AA^{+}) \le \rho(A^{+}A) \le \rho(A)$. Hence, $\rho(A) = \rho(A^{+}A) = \rho(A^{+})$.

Recall that A is a real-valued ($m \times n$) matrix. If $\rho(A) = m$ (i.e., A is of full row rank), then $(AA')^{-1}$ exists; so $A^+ = A'(AA')^+ = A'(AA')^{-1}$. Also, $AA^+ = I_m$. Similarly, if $\rho(A) = n$ (i.e., A is of full column rank), then $(A'A)^{-1}$ exists; so $A^+ = (A'A)^+A' = (A'A)^{-1}A'$, and $A^+A = I_n$. After the analogy of the traditional inverse, one might suppose that $(BC)^+ = C^+B^+$. However, the analogue for the MP inverse has an explicit condition: if B ($m \times r$) and C ($r \times n$) are both of rank r, then $(BC)^+ = C^+B^+$. This is proved by noting that BC is the factorization discussed in the existence of the MP inverse. Therefore, $(BC)^+ = C'(CC')^{-1}(B'B)^{-1}B'$, which because of the full ranks of B and C is equal to C^+B^+.

Practically speaking, how does one calculate the MP inverse, rather than just blindly accept the results of a software package (e.g., the SAS/IML® function for A^+ is GINV(A))? In the case of a diagonal matrix, A^+ is easily determined: first, invert the diagonal elements, except that any zeroes on the diagonal are left as zeroes; then transpose. This is the essence of a generalized inverse, viz., that the inverse of a non-zero scalar is its reciprocal and the inverse of zero is zero. If A is not diagonal, things are not so easy. But due to the theorem that $A^+ = (A'A)^+A' = A'(AA')^+$, we can calculate A^+ if we know $(A'A)^+$ or $(AA')^+$. This reduces the problem to calculating the MP inverse of a symmetric real-valued matrix. But, as was mentioned in Appendix A, any such matrix can be diagonalized, or factored as $W\Lambda W'$, where W is orthogonal and Λ is diagonal. One can show that $W\Lambda^+W'$ satisfies the four MP conditions. So even here the problem reduces to the simple case of a diagonal matrix (however, the practicalities of the eigenvalue problem are formidable).

Consider now the matrices AA^+ and A^+A. One is $(m \times m)$ and the other is $(n \times n)$, but according to a theorem above they are of the same rank as $\rho(A)$. The third and fourth MP conditions state that both matrices are symmetric. But note that $(AA^+)(AA^+) = (AA^+A)A^+ = AA^+$. Similarly, $(A^+A)(A^+A) = A^+(AA^+A) = A^+A$. By definition, if a matrix M has the property that MM = M, then it is said to be *idempotent*. Therefore, AA^+ and A^+A are *symmetric idempotent* matrices.

Symmetric idempotent matrices have three useful properties. First, if M is symmetric idempotent, then $M = MM = MM'$, which implies that M is non-negative definite (NND, cf. Appendix A). Second, since M is symmetric, it can be factored as $W\Lambda W'$, where W is orthogonal and Λ is diagonal. Then $W\Lambda W' = M = MM = (W\Lambda W')(W\Lambda W') = W\Lambda(W'W)\Lambda W' = W\Lambda\Lambda W'$. Therefore, $W\Lambda W' = W\Lambda\Lambda W'$; hence, $\Lambda = W'(W\Lambda W')W = W'(W\Lambda\Lambda W')W = \Lambda\Lambda$. This can happen if and only if every diagonal element of Λ is either zero or one. Therefore, every eigenvalue of a symmetric idempotent matrix is zero or one. And third, the rank of Λ is equal to be number of non-zero diagonal elements. Since these elements must be either zero or one, the rank of Λ is equal to sum of its diagonal elements. The trace of a matrix is a scalar equal to the sum of the diagonal elements; therefore, $\rho(\Lambda) = Tr(\Lambda)$. However, the trace has the property that $Tr(AB) = Tr(BA)$ (cf. Judge [11: Appendix A, 927]). Therefore, since the ranks of M and Λ are equal, $\rho(M) = \rho(\Lambda) = Tr(\Lambda) = Tr(\Lambda I) = Tr(\Lambda W'W) = Tr(W\Lambda W') = Tr(M)$, i.e., the rank of a symmetric idempotent matrix is equal to its trace.

Moreover, if $(n \times n)$ M is symmetric idempotent, then so too is $I_n - M$; because $I_n - M$ is symmetric, and $(I_n - M)(I_n - M) = I_n - M - M + MM = I_n - M - M + M = I_n - M$. Every symmetric idempotent matrix is non-negative definite; therefore $0_{(n \times n)} \leq M$ and $0_{(n \times n)} \leq I_n - M$, which last inequality implies that $M \leq I_n$. Therefore, $0_{(n \times n)} \leq M \leq I_n$. All this shows that a symmetric idempotent matrix is an orthogonal transformation of a matrix which is like an identity matrix except possibly incomplete. A symmetric idempotent matrix is the closest thing to an identity matrix; so it augurs well for the MP inverse that AA^+ and A^+A are symmetric idempotent.

The MP inverse helps in the solution of linear equations. Let $Ax = b$ be a linear equation to be solved for x, where A is $(m \times n)$, x is $(n \times p)$, and b is $(m \times p)$. It is usual for x to be $(n \times 1)$, or a row vector, but at this point we can be general. An important theorem is that $Ax = b$ has a solution for x if and only if $AA^+b = b$. Obviously, if $AA^+b = b$, then a solution exists, viz., $x = A^+b$. Conversely, if there is an x such that $Ax = b$, then $AA^+b = AA^+(Ax) = Ax = b$. This is to say that whether $Ax = b$ is consistent or inconsistent depends on whether or not $AA^+b = b$.

The function SOLUTION(A, b) of Appendix G returns $x = A^+b$ if $Ax = AA^+b = b$. However, due to limitations of computing precision, AA^+b may not exactly equal b. It can be shown that $AA^+b = b$ if and only if $(AA^+b)'(AA^+b) = b'AA^+b = b'b$. And because AA^+ and $I_m - AA^+$ are symmetric idempotent, $(AA^+b)'(AA^+b) = b'b$ if and only if their diagonals are equal. The function performs a relative check, viz., that the quotient of each pair of diagonal elements is tolerably close to one.

If $Ax = b$ is consistent (i.e., having at least one solution), then there is something special about the solution A^+b, viz., for every solution x, $(A^+b)'(A^+b) \le x'x$, with equality obtaining if and only if $x = A^+b$. The proof of this makes use of theorems from both Appendix A and this appendix, as well as of the fact that $Ax = b$:

$$
\begin{aligned}
0 &\le (x - A^+b)'(x - A^+b) \qquad \text{(with equality if and only if } x = A^+b)\\
&= (x - A'(AA')^+b)'(x - A'(AA')^+b)\\
&= (x' - b'(AA')^+A)(x - A'(AA')^+b)\\
&= x'x - x'A'(AA')^+b - b'(AA')^+Ax + b'(AA')^+AA'(AA')^+b\\
&= x'x - (Ax)'(AA')^+b - b'(AA')^+Ax + b'(AA')^+AA'(AA')^+b\\
&= x'x - b'(AA')^+b - b'(AA')^+b + b'(AA')^+AA'(AA')^+b\\
&= x'x - b'(AA')^+b - b'(AA')^+b + b'(AA')^+b\\
&= x'x - b'(AA')^+b\\
&= x'x - b'(A')^+A^+b\\
&= x'x - b'(A^+)'A^+b\\
&= x'x - (A^+b)'(A^+b)
\end{aligned}
$$

If for the moment we restrict x to an $(n\times 1)$ row vector, then x'x represents the square of the distance of x from the origin of n-space. Therefore, $x = A^+b$ is the solution closest to the origin. It can now be seen that the previous theorem, that $Ax = b$ has a solution for x if and only if $AA^+b = b$, is tantamount to stating that if $Ax = b$ has a solution, then it has a solution closest to the origin.

Another way of understanding how the MP inverse works is to consider the mapping from n-space (x) to m-space (y): $x \rightarrow (Ax = y)$. An inverse should undo the mapping: $(x = A^+y) \leftarrow y$. However, the mapping from x to y may be many-to-one, in which case returning to x from y is impossible because of the multiplicity of candidates for x. In this situation the

MP inverse does the next best thing; it maps back to the candidate closest to the origin of *n*-space.

The MP inverse helps in the solution of the least-squares problem, viz., to find the value of β which minimizes $(y - X\beta)'(y - X\beta)$. Searle [17: 80] and Judge [11: 190-192] show that the minimizing value solves the equation $X'X\beta = X'y$ (frequently called the normal equation). If X is of full column rank, then $(X'X)^{-1}$ will exist and the solution will be unique: $\beta = (X'X)^{-1}X'y$. However, the normal equation has a solution, irrespective of the rank of X, because $X'X(X'X)^{+}X'y = X'y$:

$$\begin{aligned} X'X(X'X)^{+}X'y &= X'XX^{+}(X')^{+}X'y \\ &= (X'XX^{+})(X')^{+}X'y \\ &= X'(X')^{+}X'y \\ &= X'y \end{aligned}$$

So the general solution of the minimization problem is $\beta = (X'X)^{+}X'y = X^{+}y$. When X is not of full column rank (in other words, X is multicollinear), then this solution is not unique; however, of all the solutions it is the one closest to the origin.

The MP inverse plays a role in the multivariate Cauchy-Schwartz inequality. Let Σ be NND, and partitioned into quadrants. Then, by the rules of partitioned matrix multiplication (Eves [5: 37-40]):

$$
\begin{aligned}
0 &\le A\Sigma A' \\
&\le \begin{bmatrix} I_{11} & -\Sigma_{12}\Sigma_{22}^{+} \end{bmatrix} \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \begin{bmatrix} I_{11} \\ -\Sigma_{22}^{+}\Sigma_{21} \end{bmatrix} \\
&\le \begin{bmatrix} \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{+}\Sigma_{21} & \Sigma_{12} - \Sigma_{12}\Sigma_{22}^{+}\Sigma_{22} \end{bmatrix} \begin{bmatrix} I_{11} \\ -\Sigma_{22}^{+}\Sigma_{21} \end{bmatrix} \\
&\le \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{+}\Sigma_{21} - \Sigma_{12}\Sigma_{22}^{+}\Sigma_{21} + \Sigma_{12}\Sigma_{22}^{+}\Sigma_{22}\Sigma_{22}^{+}\Sigma_{21} \\
&\le \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{+}\Sigma_{21} - \Sigma_{12}\Sigma_{22}^{+}\Sigma_{21} + \Sigma_{12}\Sigma_{22}^{+}\Sigma_{21} \\
&\le \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{+}\Sigma_{21}
\end{aligned}
$$

If Σ is PD, then all the instances of '$\le$' in the previous proof may be replaced with '$<$'.

Finally, we will use the MP inverse to build the solution set $\{x\text{: } Ax = b\}$. As before, A is $(m \times n)$, x is $(n \times p)$, and b is $(m \times p)$. Unless $AA^{+}b = b$ there is no solution; so we will assume that $A^{+}b$ is one solution. First, $Ax = b$ if and only if $A^{+}Ax = A^{+}b$. For if $Ax = b$, then $A^{+}Ax = A^{+}b$. Conversely, if $A^{+}Ax = A^{+}b$, then $Ax = A(A^{+}Ax) = A(A^{+}b) = b$. The form $A^{+}Ax = A^{+}b$ is convenient because the $(n \times n)$ matrix $A^{+}A$ is symmetric idempotent. Also we know the rank of A, which we will call r, because $r = \rho(A) = \rho(A^{+}A) = \mathrm{Tr}(A^{+}A)$. And let $s = n - r$.

Next, $A^{+}A$ can be factored as $W\Lambda W'$, where W is orthogonal and Λ is diagonal. In SAS/IML® the subroutine EIGEN(Λ, W, GINV(A)*A) returns W and Λ. As shown above, the diagonal elements of Λ must be zeroes and ones; and W and Λ can be arranged such that the ones occupy the first r places of the diagonal. (The subroutine EIGEN orders the eigenvalues in descending order.) What we have done to this point is expressed as follows:

$$
\begin{aligned}
Ax &= b \\
A^{+}Ax &= A^{+}b \\
W\Lambda W'x &= A^{+}b \\
W\begin{bmatrix} I_{r\times r} & 0_{r\times s} \\ 0_{s\times r} & 0_{s\times s} \end{bmatrix} W'x &= A^{+}b
\end{aligned}
$$

Now multiply both sides by W', and let $y = W'x$ and $c = W'A^{+}b$. Since W is orthogonal, this multiplication is a transformation which looses no information; i.e., it is reversible. The solution set of y corresponds one-to-one with the solution set of x. Partition the first r rows of y as y_1, the remaining s rows as y_2; and do similarly with c. Employing partitioned matrix multiplication at the end, we have:

$$
\begin{aligned}
W'W\begin{bmatrix} I_{r\times r} & 0_{r\times s} \\ 0_{s\times r} & 0_{s\times s} \end{bmatrix} W'x &= W'A^{+}b \\
\begin{bmatrix} I_{r\times r} & 0_{r\times s} \\ 0_{s\times r} & 0_{s\times s} \end{bmatrix} W'x &= W'A^{+}b \\
\begin{bmatrix} I_{r\times r} & 0_{r\times s} \\ 0_{s\times r} & 0_{s\times s} \end{bmatrix} y &= c \\
\begin{bmatrix} I_{r\times r} & 0_{r\times s} \\ 0_{s\times r} & 0_{s\times s} \end{bmatrix} \begin{bmatrix} y_{1\; r\times p} \\ y_{2\; s\times p} \end{bmatrix} &= \begin{bmatrix} c_{1\; r\times p} \\ c_{2\; s\times p} \end{bmatrix} \\
\begin{bmatrix} I_{r\times r}y_1 + 0_{r\times s}y_2 \\ 0_{s\times r}y_1 + 0_{s\times s}y_2 \end{bmatrix} &= \begin{bmatrix} c_{1\; r\times p} \\ c_{2\; s\times p} \end{bmatrix} \\
\begin{bmatrix} y_{1\; r\times p} \\ 0_{\; s\times p} \end{bmatrix} &= \begin{bmatrix} c_{1\; r\times p} \\ c_{2\; s\times p} \end{bmatrix}
\end{aligned}
$$

Therefore, the solution set of y is any $(n\times p)$ vector whose first r rows equal c_1; the last s rows of y are irrelevant. That c_2, the last s rows of $c = W'A^{+}b$, must be zero is a consequence of the fact that A^+b is a solution for x; if c_2 were not zero, then $Ax = b$ would be inconsistent.

The last step is to transform back to $x = I_n x = W(W'x) = Wy$, also partitioning the first r columns of W as W_1 and the last s columns as W_2:

$$
\begin{aligned}
x_{n\times p} &= W_{n\times n} y_{n\times p} \\
&= W\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \\
&= W\begin{bmatrix} c_1 \\ y_2 \end{bmatrix} \\
&= W\left(\begin{bmatrix} c_1 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ y_2 \end{bmatrix}\right) \\
&= W\left(c + \begin{bmatrix} 0 \\ y_2 \end{bmatrix}\right) \\
&= Wc + W\begin{bmatrix} 0 \\ y_2 \end{bmatrix} \\
&= WW'A^+b + W\begin{bmatrix} 0 \\ y_2 \end{bmatrix} \\
&= A^+b + W\begin{bmatrix} 0 \\ y_2 \end{bmatrix} \\
&= A^+b + \begin{bmatrix} W_{1\,n\times r} & W_{2\,n\times s} \end{bmatrix}\begin{bmatrix} 0_{\,r\times p} \\ y_{2\,s\times p} \end{bmatrix} \\
&= A^+b + W_2 y_2
\end{aligned}
$$

It should be noted that $W_1'W_1 = I_r$, $W_1'W_2 = 0$, and $W_2'W_2 = I_s$ due to the orthogonality of W:

$$
\begin{aligned}
I_n &= \begin{bmatrix} I_{r\times r} & 0_{r\times s} \\ 0_{s\times r} & I_{s\times s} \end{bmatrix} \\
&= W'W \\
&= \begin{bmatrix} W_1'{}_{\,r\times n} \\ W_2'{}_{\,s\times n} \end{bmatrix} \begin{bmatrix} W_1{}_{\,n\times r} & W_2{}_{\,n\times s} \end{bmatrix} \\
&= \begin{bmatrix} W_1'W_1{}_{\,r\times r} & W_1'W_2{}_{\,r\times s} \\ W_2'W_1{}_{\,s\times r} & W_2'W_2{}_{\,s\times s} \end{bmatrix}
\end{aligned}
$$

Therefore, $\rho(W_2) = s$; and $y_2 = W_2'(W_2y_2) = W_2'(x - A^+b)$.

To summarize, to generate all the solutions of $Ax = b$, calculate A^+ and determine whether A^+b is a solution. If it is not, then the equation is inconsistent. But if A^+b is a solution, then calculate A^+A and $r = \mathrm{Tr}(A^+A)$. Next, diagonalize A^+A as $W\Lambda W'$, and note the columns of Λ which contain the zero eigenvalues. Select the same columns of W, and call this submatrix V ($V'V = I$). Then x is a solution if and only if there exists a y such that $x = A^+b + Vy$. If Λ has no zero eigenvalue, then A^+b is the only solution. If x is $(n\times 1)$, then $\{x: Ax = b\}$ will be an $n - r$ dimensional subspace of n-space.

Appendix C

The Linear Statistical Model and Best Linear Unbiased Estimation

A *statistical model* is an explanation of a random vector (y) as the sum of a known function (f) of an unknown vector (β) and an error vector (**e**):

$$\mathbf{y}_{t\times 1} = f(\beta_{k\times 1}) + \mathbf{e}_{t\times 1}$$

Although β is not known, it is not random; an estimator of β is random, but β itself is not. What injects randomness into **y** and an estimator of β is the error term **e**. The purpose of $f(\beta)$ is to specify the expectation of **y**; hence, E[**e**] = 0. The variance of **e** is known at least to within a proportionality constant; i.e., Var[**e**] = $\Sigma = \sigma^2\Phi$, where at least Φ is known, and possibly σ^2 also. When we say 'known' in this context, we mean 'taken for granted' rather than 'known for certain'. In this appendix no assumption is made as to the probability distribution of **e**.

A *linear statistical model* is one whose function is linear:

$$f(\beta_{k\times 1}) = X\beta$$
$$\therefore \quad \mathbf{y}_{t\times 1} = X_{t\times k}\beta_{k\times 1} + \mathbf{e}$$

X is often called the design matrix, and its columns are sometimes called control variables. The elements of β are variously called parameters and effects. There is much talk nowadays about non-linear processes; so one might be inclined to disparage a linear model. However, in practice, the functions of non-linear models are differentiable; and Taylor's theorem shows that such functions are by locale approximately linear. In multivariate terms, in the neighborhood of β_0 (Judge [11: 508-511]):

$$\mathbf{y} \approx f(\beta_0) + \left.\frac{\partial f}{\partial \beta'}\right|_{\beta=\beta_0} (\beta - \beta_0) + \mathbf{e}$$

Therefore, we consider linearity to be not much of a drawback, and to be the phase analogous to walking, which comes in between crawling and running.

Most presentations of the linear statistical model dwell on how to estimate β. But here a wider approach will be taken. Suppose that the t rows of the $\mathbf{y}$ are of two types, those which have been observed and those which have not. The observed portion of $\mathbf{y}$ we will call $\mathbf{y}_1$ and say that it is $(t_1 \times 1)$; the unobserved will be $\mathbf{y}_2$ and $(t_2 \times 1)$. Of course, $t_1 + t_2 = t$. We can also arrange the rows of the model such that the observed portion comes first. Similarly partition X and $\mathbf{e}$, so that the model looks like:

$$\begin{bmatrix} \mathbf{y}_1 & {}_{t_1 \times 1} \\ \mathbf{y}_2 & {}_{t_2 \times 1} \end{bmatrix} = \begin{bmatrix} X_1 & {}_{t_1 \times k} \\ X_2 & {}_{t_2 \times k} \end{bmatrix} \beta_{k \times 1} + \begin{bmatrix} \mathbf{e}_1 & {}_{t_1 \times 1} \\ \mathbf{e}_2 & {}_{t_2 \times 1} \end{bmatrix} = \begin{bmatrix} X_1\beta \\ X_2\beta \end{bmatrix} + \begin{bmatrix} \mathbf{e}_1 \\ \mathbf{e}_2 \end{bmatrix}, \text{ where } \mathrm{Var}\begin{bmatrix} \mathbf{e}_1 \\ \mathbf{e}_2 \end{bmatrix} = \begin{bmatrix} \Sigma_{11} \; {}_{t_1 \times t_1} & \Sigma_{12} \; {}_{t_1 \times t_2} \\ \Sigma_{21} \; {}_{t_2 \times t_1} & \Sigma_{22} \; {}_{t_2 \times t_2} \end{bmatrix}$$

Being unobserved, $\mathbf{y}_2$ contains missing values. The error term $\mathbf{e}$, like β, is not knowable; however, its expectation is zero and its variance is known to be Σ (or proportional thereto). Both $\mathbf{y}_1$ and X are known, the first by observation and the second by design. It will be our task to formulate an estimator of $\mathbf{y}_2$ based on $\mathbf{y}_1$, X, and Σ. In particular, we want the estimator to be linear in $\mathbf{y}_1$, to be unbiased, and to be in some way optimal; i.e., we want the best linear unbiased estimator (BLUE) of $\mathbf{y}_2$.

To be linear the estimator must be of the form $\tilde{\mathbf{y}}_2 = A\mathbf{y}_1 + b$. However, in the case that $\mathbf{y}_1$ were observed to be zero, it would be most natural to assume that $\beta = 0$ and $\Sigma = 0$, which would make zero the most reasonable estimator for $\mathbf{y}_2$. Therefore, we can consider b to be zero, and the estimator to be of the form $\tilde{\mathbf{y}}_2 = A\mathbf{y}_1$, where A is $(t_2 \times t_1)$ and some function of X_1, X_2, Σ_{11}, Σ_{12}, Σ_{21} ($= \Sigma_{12}'$, cf. Appendix A), and Σ_{22}.

For the estimator to be unbiased the following must hold, regardless of the value of β:

$$\begin{aligned}
E[\tilde{\mathbf{y}}_2] &= E[\mathbf{y}_2] \\
\text{But } E[\tilde{\mathbf{y}}_2] &= E[A\mathbf{y}_1] \\
&= AE[\mathbf{y}_1] \\
&= AE[X_1\beta + \mathbf{e}_1] \\
&= A(X_1\beta + E[\mathbf{e}_1]) \\
&= AX_1\beta \\
\text{And } E[\mathbf{y}_2] &= E[X_2\beta + \mathbf{e}_2] \\
&= X_2\beta + E[\mathbf{e}_2] \\
&= X_2\beta
\end{aligned}$$

Therefore, for all β, $AX_1\beta = X_2\beta$. β is a $(k \times 1)$ row vector, but we could join columnwise l such vectors to form a $(k \times l)$ matrix B, and it would still be true that for all B, $AX_1B = X_2B$. From the case that $B = I_k$, we conclude that $AX_1 = X_2$. The converse is obvious; therefore, $A\mathbf{y}_l$ is an unbiased estimator of $\mathbf{y}_2$ if and only if $AX_l = X_2$.

We do not know how $\mathbf{y}_2$, when it will have been observed, will differ from its estimator. If we had information about this, we could use it to improve upon the estimator. But we can say that of competing linear unbiased estimators the best estimator is the one the variance of whose prediction error is smallest:

$$\begin{aligned}
\mathrm{Var}[\mathbf{y}_2 - \hat{\mathbf{y}}_2] &\leq \mathrm{Var}[\mathbf{y}_2 - \tilde{\mathbf{y}}_2], \text{ or} \\
0 &\leq \mathrm{Var}[\mathbf{y}_2 - \tilde{\mathbf{y}}_2] - \mathrm{Var}[\mathbf{y}_2 - \hat{\mathbf{y}}_2]
\end{aligned}$$

As explained in Appendix A, this means that the right-hand side of the second inequality is non-negative definite (NND). The estimator with the caret is at least as good as the one with the tilde; and if the expression is non-zero, it is better.

We now show that when $A_0 = \Sigma_{21}\Sigma_{11}^{-1} + (X_2 - \Sigma_{21}\Sigma_{11}^{-1}X_1)(X_1'\Sigma_{11}^{-1}X_1)^{-1}X_1'\Sigma_{11}^{-1}$, $\hat{\mathbf{y}}_2 = A_0\mathbf{y}_1$ is the best of all the linear unbiased estimators of $\mathbf{y}_2$. The only requirement for A_0 to exist is that both Σ_{11} and $X_1'\Sigma_{11}^{-1}X_1$ be nonsingular. This makes both Σ_{11} and Σ_{11}^{-1} positive definite; so, according to the last theorem of Appendix A, $X_1'\Sigma_{11}^{-1}X_1$ is nonsingular if and only if X_1 is of full column rank. In practice this is usually met; and later it will be shown how to handle a non-singular Σ_{11} and an X_1 of less than full column rank. A_0 depends only on the partitions of X and Σ, things which are known. (Even if only $\Phi = \Sigma/\sigma^2$ is known, A_0 is invariant with respect to σ^2.) Moreover, A_0 makes for an unbiased estimator, since:

$$\begin{aligned}
A_0X_1 &= (\Sigma_{21}\Sigma_{11}^{-1} + (X_2 - \Sigma_{21}\Sigma_{11}^{-1}X_1)(X_1'\Sigma_{11}^{-1}X_1)^{-1}X_1'\Sigma_{11}^{-1})X_1 \\
&= \Sigma_{21}\Sigma_{11}^{-1}X_1 + (X_2 - \Sigma_{21}\Sigma_{11}^{-1}X_1)(X_1'\Sigma_{11}^{-1}X_1)^{-1}(X_1'\Sigma_{11}^{-1}X_1) \\
&= \Sigma_{21}\Sigma_{11}^{-1}X_1 + (X_2 - \Sigma_{21}\Sigma_{11}^{-1}X_1) \\
&= X_2
\end{aligned}$$

The proof that the estimator with A_0 is best begins as follows:

$$\begin{aligned}
\mathrm{Var}[\mathbf{y}_2 - \tilde{\mathbf{y}}_2] &= \mathrm{Var}[\mathbf{y}_2 - \hat{\mathbf{y}}_2 + \hat{\mathbf{y}}_2 - \tilde{\mathbf{y}}_2] \\
&= \mathrm{Cov}[(\mathbf{y}_2 - \hat{\mathbf{y}}_2) + (\hat{\mathbf{y}}_2 - \tilde{\mathbf{y}}_2), (\mathbf{y}_2 - \hat{\mathbf{y}}_2) + (\hat{\mathbf{y}}_2 - \tilde{\mathbf{y}}_2)] \\
&= \mathrm{Cov}[(\mathbf{y}_2 - \hat{\mathbf{y}}_2), (\mathbf{y}_2 - \hat{\mathbf{y}}_2)] + \mathrm{Cov}[(\mathbf{y}_2 - \hat{\mathbf{y}}_2), (\hat{\mathbf{y}}_2 - \tilde{\mathbf{y}}_2)] + \\
&\quad \mathrm{Cov}[(\hat{\mathbf{y}}_2 - \tilde{\mathbf{y}}_2), (\mathbf{y}_2 - \hat{\mathbf{y}}_2)] + \mathrm{Cov}[(\hat{\mathbf{y}}_2 - \tilde{\mathbf{y}}_2), (\hat{\mathbf{y}}_2 - \tilde{\mathbf{y}}_2)] \\
&= \mathrm{Var}[\mathbf{y}_2 - \hat{\mathbf{y}}_2] + \mathrm{Cov}[\mathbf{y}_2 - \hat{\mathbf{y}}_2, \hat{\mathbf{y}}_2 - \tilde{\mathbf{y}}_2] + \mathrm{Cov}[\hat{\mathbf{y}}_2 - \tilde{\mathbf{y}}_2, \mathbf{y}_2 - \hat{\mathbf{y}}_2] + \mathrm{Var}[\hat{\mathbf{y}}_2 - \tilde{\mathbf{y}}_2] \\
&= \mathrm{Var}[\mathbf{y}_2 - \hat{\mathbf{y}}_2] + \mathrm{Cov}[\mathbf{y}_2 - \hat{\mathbf{y}}_2, \hat{\mathbf{y}}_2 - \tilde{\mathbf{y}}_2] + \mathrm{Cov}[\mathbf{y}_2 - \hat{\mathbf{y}}_2, \hat{\mathbf{y}}_2 - \tilde{\mathbf{y}}_2]' + \mathrm{Var}[\hat{\mathbf{y}}_2 - \tilde{\mathbf{y}}_2]
\end{aligned}$$

But:

$$
\begin{aligned}
\operatorname{Var}[\hat{\mathbf{y}}_2 - \tilde{\mathbf{y}}_2] &= \operatorname{Var}[\mathrm{A}_0\mathbf{y}_1 - \mathrm{A}\mathbf{y}_1] \\
&= \operatorname{Var}[(\mathrm{A}_0 - \mathrm{A})\mathbf{y}_1] \\
&= (\mathrm{A}_0 - \mathrm{A})\operatorname{Var}[\mathbf{y}_1](\mathrm{A}_0 - \mathrm{A})' \\
&= (\mathrm{A}_0 - \mathrm{A})\Sigma_{11}(\mathrm{A}_0 - \mathrm{A})' \\
&\geq 0
\end{aligned}
$$

Σ_{11} is an NND matrix. But requisite to the existence of A_0, it is also non-singular. According to Appendix A, a non-singular NND matrix must be positive definite (PD). Therefore, $(\mathrm{A}_0 - \mathrm{A})\Sigma_{11}(\mathrm{A}_0 - \mathrm{A})' = 0$ if and only if $\mathrm{A} = \mathrm{A}_0$. As for the covariance term:

$$
\begin{aligned}
\operatorname{Cov}[\mathbf{y}_2 - \hat{\mathbf{y}}_2, \hat{\mathbf{y}}_2 - \tilde{\mathbf{y}}_2] &= \operatorname{Cov}[\mathbf{y}_2 - \mathrm{A}_0\mathbf{y}_1, \mathrm{A}_0\mathbf{y}_1 - \mathrm{A}\mathbf{y}_1] \\
&= \operatorname{Cov}[\mathbf{y}_2 - \mathrm{A}_0\mathbf{y}_1, (\mathrm{A}_0 - \mathrm{A})\mathbf{y}_1] \\
&= \operatorname{Cov}[\mathbf{y}_2 - \mathrm{A}_0\mathbf{y}_1, \mathbf{y}_1](\mathrm{A}_0 - \mathrm{A})' \\
&= (\operatorname{Cov}[\mathbf{y}_2, \mathbf{y}_1] - \operatorname{Cov}[\mathrm{A}_0\mathbf{y}_1, \mathbf{y}_1])(\mathrm{A}_0 - \mathrm{A})' \\
&= (\operatorname{Cov}[\mathbf{y}_2, \mathbf{y}_1] - \mathrm{A}_0\operatorname{Cov}[\mathbf{y}_1, \mathbf{y}_1])(\mathrm{A}_0 - \mathrm{A})' \\
&= (\Sigma_{21} - \mathrm{A}_0\Sigma_{11})(\mathrm{A}_0 - \mathrm{A})' \\
&= (\Sigma_{21} - (\Sigma_{21}\Sigma_{11}^{-1} + (\mathrm{X}_2 - \Sigma_{21}\Sigma_{11}^{-1}\mathrm{X}_1)(\mathrm{X}_1'\Sigma_{11}^{-1}\mathrm{X}_1)^{-1}\mathrm{X}_1'\Sigma_{11}^{-1})\Sigma_{11})(\mathrm{A}_0 - \mathrm{A})' \\
&= (\Sigma_{21} - (\Sigma_{21} + (\mathrm{X}_2 - \Sigma_{21}\Sigma_{11}^{-1}\mathrm{X}_1)(\mathrm{X}_1'\Sigma_{11}^{-1}\mathrm{X}_1)^{-1}\mathrm{X}_1'))(\mathrm{A}_0 - \mathrm{A})' \\
&= -(\mathrm{X}_2 - \Sigma_{21}\Sigma_{11}^{-1}\mathrm{X}_1)(\mathrm{X}_1'\Sigma_{11}^{-1}\mathrm{X}_1)^{-1}\mathrm{X}_1'(\mathrm{A}_0 - \mathrm{A})' \\
&= -(\mathrm{X}_2 - \Sigma_{21}\Sigma_{11}^{-1}\mathrm{X}_1)(\mathrm{X}_1'\Sigma_{11}^{-1}\mathrm{X}_1)^{-1}((\mathrm{A}_0 - \mathrm{A})\mathrm{X}_1)' \\
&= -(\mathrm{X}_2 - \Sigma_{21}\Sigma_{11}^{-1}\mathrm{X}_1)(\mathrm{X}_1'\Sigma_{11}^{-1}\mathrm{X}_1)^{-1}(\mathrm{A}_0\mathrm{X}_1 - \mathrm{A}\mathrm{X}_1)' \\
&= -(\mathrm{X}_2 - \Sigma_{21}\Sigma_{11}^{-1}\mathrm{X}_1)(\mathrm{X}_1'\Sigma_{11}^{-1}\mathrm{X}_1)^{-1}(\mathrm{X}_2 - \mathrm{X}_2)' \\
&= 0
\end{aligned}
$$

Therefore:

$$
\begin{aligned}
\operatorname{Var}[\mathbf{y}_2 - \tilde{\mathbf{y}}_2] &= \operatorname{Var}[\mathbf{y}_2 - \hat{\mathbf{y}}_2] + \operatorname{Cov}[\mathbf{y}_2 - \hat{\mathbf{y}}_2, \hat{\mathbf{y}}_2 - \tilde{\mathbf{y}}_2] + \operatorname{Cov}[\mathbf{y}_2 - \hat{\mathbf{y}}_2, \hat{\mathbf{y}}_2 - \tilde{\mathbf{y}}_2]' + \operatorname{Var}[\hat{\mathbf{y}}_2 - \tilde{\mathbf{y}}_2] \\
&= \operatorname{Var}[\mathbf{y}_2 - \hat{\mathbf{y}}_2] + 0 + 0' + (\mathrm{A}_0 - \mathrm{A})\Sigma_{11}^{-1}(\mathrm{A}_0 - \mathrm{A})' \\
&= \operatorname{Var}[\mathbf{y}_2 - \hat{\mathbf{y}}_2] + (\mathrm{A}_0 - \mathrm{A})\Sigma_{11}^{-1}(\mathrm{A}_0 - \mathrm{A})' \\
&\geq \operatorname{Var}[\mathbf{y}_2 - \hat{\mathbf{y}}_2] \quad \text{(with equality if and only if } \mathrm{A} = \mathrm{A}_0\text{)}
\end{aligned}
$$

Thus, $\hat{\mathbf{y}}_2 = (\Sigma_{21}\Sigma_{11}^{-1} + (\mathrm{X}_2 - \Sigma_{21}\Sigma_{11}^{-1}\mathrm{X}_1)(\mathrm{X}_1'\Sigma_{11}^{-1}\mathrm{X}_1)^{-1}\mathrm{X}_1'\Sigma_{11}^{-1})y_1$ is the best linear unbiased estimator of $\mathbf{y}_2$.

The variance of the prediction error is determined as follows:

$$\begin{aligned}
\mathrm{Var}[\mathbf{y}_2 - \hat{\mathbf{y}}_2] &= \mathrm{Var}[\mathbf{y}_2 - A_0\mathbf{y}_1] \\
&= \mathrm{Cov}[\mathbf{y}_2 - A_0\mathbf{y}_1, \mathbf{y}_2 - A_0\mathbf{y}_1] \\
&= \mathrm{Cov}[\mathbf{y}_2, \mathbf{y}_2] - \mathrm{Cov}[\mathbf{y}_2, A_0\mathbf{y}_1] - \mathrm{Cov}[A_0\mathbf{y}_1, \mathbf{y}_2] + \mathrm{Cov}[A_0\mathbf{y}_1, A_0\mathbf{y}_1] \\
&= \mathrm{Cov}[\mathbf{y}_2, \mathbf{y}_2] - \mathrm{Cov}[\mathbf{y}_2, \mathbf{y}_1]A_0' - A_0\mathrm{Cov}[\mathbf{y}_1, \mathbf{y}_2] + A_0\mathrm{Cov}[\mathbf{y}_1, \mathbf{y}_1]A_0' \\
&= \Sigma_{22} - \Sigma_{21}A_0' - A_0\Sigma_{12} + A_0\Sigma_{11}A_0' \\
&= \Sigma_{22} - A_0\Sigma_{12} + (A_0\Sigma_{11} - \Sigma_{21})A_0'
\end{aligned}$$

But:

$$\begin{aligned}
(A_0\Sigma_{11} - \Sigma_{21})A_0' &= ((\Sigma_{21}\Sigma_{11}^{-1} + (X_2 - \Sigma_{21}\Sigma_{11}^{-1}X_1)(X_1'\Sigma_{11}^{-1}X_1)^{-1}X_1'\Sigma_{11}^{-1})\Sigma_{11} - \Sigma_{21})A_0' \\
&= (\Sigma_{21} + (X_2 - \Sigma_{21}\Sigma_{11}^{-1}X_1)(X_1'\Sigma_{11}^{-1}X_1)^{-1}X_1' - \Sigma_{21})A_0' \\
&= (X_2 - \Sigma_{21}\Sigma_{11}^{-1}X_1)(X_1'\Sigma_{11}^{-1}X_1)^{-1}X_1'A_0' \\
&= (X_2 - \Sigma_{21}\Sigma_{11}^{-1}X_1)(X_1'\Sigma_{11}^{-1}X_1)^{-1}(A_0X_1)' \\
&= (X_2 - \Sigma_{21}\Sigma_{11}^{-1}X_1)(X_1'\Sigma_{11}^{-1}X_1)^{-1}X_2'
\end{aligned}$$

Also:

$$\begin{aligned}
-A_0\Sigma_{12} &= -(\Sigma_{21}\Sigma_{11}^{-1} + (X_2 - \Sigma_{21}\Sigma_{11}^{-1}X_1)(X_1'\Sigma_{11}^{-1}X_1)^{-1}X_1'\Sigma_{11}^{-1})\Sigma_{12} \\
&= -\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12} - (X_2 - \Sigma_{21}\Sigma_{11}^{-1}X_1)(X_1'\Sigma_{11}^{-1}X_1)^{-1}X_1'\Sigma_{11}^{-1}\Sigma_{12}
\end{aligned}$$

Therefore,

$$\begin{aligned}
\mathrm{Var}[\mathbf{y}_2 - \hat{\mathbf{y}}_2] &= \Sigma_{22} - A_0\Sigma_{12} + (A_0\Sigma_{11} - \Sigma_{21})A_0' \\
&= \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12} - (X_2 - \Sigma_{21}\Sigma_{11}^{-1}X_1)(X_1'\Sigma_{11}^{-1}X_1)^{-1}X_1'\Sigma_{11}^{-1}\Sigma_{12} + \\
&\quad (X_2 - \Sigma_{21}\Sigma_{11}^{-1}X_1)(X_1'\Sigma_{11}^{-1}X_1)^{-1}X_2' \\
&= \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12} + (X_2 - \Sigma_{21}\Sigma_{11}^{-1}X_1)(X_1'\Sigma_{11}^{-1}X_1)^{-1}(X_2' - X_1'\Sigma_{11}^{-1}\Sigma_{12}) \\
&= \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12} + (X_2 - \Sigma_{21}\Sigma_{11}^{-1}X_1)(X_1'\Sigma_{11}^{-1}X_1)^{-1}(X_2 - \Sigma_{21}\Sigma_{11}^{-1}X_1)'
\end{aligned}$$

These formulas for the best linear unbiased estimator of $\mathbf{y}_2$ and the variance of its prediction error are complicated; however, an insight will reduce them to more familiar terms. As a special case of $\mathbf{y}_2 = X_2\beta + \mathbf{e}_2$ consider $X_2 = I_k$, and $\mathrm{Var}[\mathbf{e}_2] = \Sigma_{22} = 0$. If $\Sigma_{22} = 0$, then for the whole $(t \times t)$ Σ variance matrix to remain NND the covariances Σ_{12} and Σ_{21} must also be zero. Then $\mathbf{y}_2 = \beta$, and the estimator of $\mathbf{y}_2$ will be the estimator of β. The formulas reduce:

$$\begin{aligned}
\hat{\beta} &= \hat{\mathbf{y}}_2 \\
&= (\Sigma_{21}\Sigma_{11}^{-1} + (X_2 - \Sigma_{21}\Sigma_{11}^{-1}X_1)(X_1'\Sigma_{11}^{-1}X_1)^{-1}X_1'\Sigma_{11}^{-1})\mathbf{y}_1 \\
&= (0\Sigma_{11}^{-1} + (I_k - 0\Sigma_{11}^{-1}X_1)(X_1'\Sigma_{11}^{-1}X_1)^{-1}X_1'\Sigma_{11}^{-1})\mathbf{y}_1 \\
&= (X_1'\Sigma_{11}^{-1}X_1)^{-1}X_1'\Sigma_{11}^{-1}\mathbf{y}_1
\end{aligned}$$

$$\begin{aligned}
\mathrm{Var}[\hat{\beta}] &= \mathrm{Var}[\beta - \hat{\beta}] \\
&= \mathrm{Var}[\mathbf{y}_2 - \hat{\mathbf{y}}_2] \\
&= \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12} + (X_2 - \Sigma_{21}\Sigma_{11}^{-1}X_1)(X_1'\Sigma_{11}^{-1}X_1)^{-1}(X_2 - \Sigma_{21}\Sigma_{11}^{-1}X_1)' \\
&= 0 - 0\Sigma_{11}^{-1}0 + (I_k - 0\Sigma_{11}^{-1}X_1)(X_1'\Sigma_{11}^{-1}X_1)^{-1}(I_k - 0\Sigma_{11}^{-1}X_1)' \\
&= (X_1'\Sigma_{11}^{-1}X_1)^{-1}
\end{aligned}$$

These formulas regarding the estimator of β are then inserted into the general formulas:

$$\begin{aligned}
\hat{\mathbf{y}}_2 &= (\Sigma_{21}\Sigma_{11}^{-1} + (X_2 - \Sigma_{21}\Sigma_{11}^{-1}X_1)(X_1'\Sigma_{11}^{-1}X_1)^{-1}X_1'\Sigma_{11}^{-1})\mathbf{y}_1 \\
&= \Sigma_{21}\Sigma_{11}^{-1}\mathbf{y}_1 + (X_2 - \Sigma_{21}\Sigma_{11}^{-1}X_1)(X_1'\Sigma_{11}^{-1}X_1)^{-1}X_1'\Sigma_{11}^{-1}\mathbf{y}_1 \\
&= \Sigma_{21}\Sigma_{11}^{-1}\mathbf{y}_1 + (X_2 - \Sigma_{21}\Sigma_{11}^{-1}X_1)\hat{\beta} \\
&= \Sigma_{21}\Sigma_{11}^{-1}\mathbf{y}_1 + X_2\hat{\beta} - \Sigma_{21}\Sigma_{11}^{-1}X_1\hat{\beta} \\
&= X_2\hat{\beta} + \Sigma_{21}\Sigma_{11}^{-1}(\mathbf{y}_1 - X_1\hat{\beta})
\end{aligned}$$

$$\begin{aligned}
\mathrm{Var}[\mathbf{y}_2 - \hat{\mathbf{y}}_2] &= \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12} + (X_2 - \Sigma_{21}\Sigma_{11}^{-1}X_1)(X_1'\Sigma_{11}^{-1}X_1)^{-1}(X_2 - \Sigma_{21}\Sigma_{11}^{-1}X_1)' \\
&= (\Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}) + (X_2 - \Sigma_{21}\Sigma_{11}^{-1}X_1)\mathrm{Var}(\hat{\beta})(X_2 - \Sigma_{21}\Sigma_{11}^{-1}X_1)'
\end{aligned}$$

These are the more familiar forms that are found in other works, e.g., in Halliwell [8: Appendices A and C], forms that are derived from the least squares principle. Thus, best linear unbiased estimation gives the same results as does least squares.

It is instructive to see how the formulas work on the following simple model:

$$\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix} = \begin{bmatrix} X_1\beta \\ X_1\beta \end{bmatrix} + \begin{bmatrix} \mathbf{e}_1 \\ \mathbf{e}_2 \end{bmatrix}, \text{ where } \mathrm{Var}\begin{bmatrix} \mathbf{e}_1 \\ \mathbf{e}_2 \end{bmatrix} = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{11} \end{bmatrix}$$

The estimated rows are of the same form as the observed rows; even the variance of the estimated rows, Σ_{22}, is equal to the variance of the observed rows, Σ_{11}. As a first case of this model let $\Sigma_{21} = \Sigma_{12}'$ be zero. Therefore, the observed rows and the estimated rows are like two random samples from the same distribution. Accordingly,

$$\begin{aligned}
\hat{\mathbf{y}}_2 &= X_2\hat{\beta} + \Sigma_{21}\Sigma_{11}^{-1}(\mathbf{y}_1 - X_1\hat{\beta}) \\
&= X_1\hat{\beta} + 0\Sigma_{11}^{-1}(\mathbf{y}_1 - X_1\hat{\beta}) \\
&= X_1\hat{\beta} \\
\mathrm{Var}[\mathbf{y}_2 - \hat{\mathbf{y}}_2] &= (\Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}) + (X_2 - \Sigma_{21}\Sigma_{11}^{-1}X_1)\mathrm{Var}(\hat{\beta})(X_2 - \Sigma_{21}\Sigma_{11}^{-1}X_1)' \\
&= (\Sigma_{11} - 0\Sigma_{11}^{-1}0) + (X_1 - 0\Sigma_{11}^{-1}X_1)\mathrm{Var}(\hat{\beta})(X_1 - 0\Sigma_{11}^{-1}X_1)' \\
&= \Sigma_{11} + X_1\mathrm{Var}(\hat{\beta})X_1'
\end{aligned}$$

As a second case, let $\Sigma_{21} = \Sigma_{12}' = \Sigma_{11}$. The simplest interpretation of this case is that the observed rows and the estimated rows are the same sample. Then:

$$\begin{aligned}
\hat{\mathbf{y}}_2 &= X_2\hat{\beta} + \Sigma_{21}\Sigma_{11}^{-1}(\mathbf{y}_1 - X_1\hat{\beta}) \\
&= X_1\hat{\beta} + \Sigma_{11}\Sigma_{11}^{-1}(\mathbf{y}_1 - X_1\hat{\beta}) \\
&= X_1\hat{\beta} + (\mathbf{y}_1 - X_1\hat{\beta}) \\
&= \mathbf{y}_1 \\
\mathrm{Var}[\mathbf{y}_2 - \hat{\mathbf{y}}_2] &= (\Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}) + (X_2 - \Sigma_{21}\Sigma_{11}^{-1}X_1)\mathrm{Var}(\hat{\beta})(X_2 - \Sigma_{21}\Sigma_{11}^{-1}X_1)' \\
&= (\Sigma_{11} - \Sigma_{11}\Sigma_{11}^{-1}\Sigma_{11}) + (X_1 - \Sigma_{11}\Sigma_{11}^{-1}X_1)\mathrm{Var}(\hat{\beta})(X_1 - \Sigma_{11}\Sigma_{11}^{-1}X_1)' \\
&= (\Sigma_{11} - \Sigma_{11}) + (X_1 - X_1)\mathrm{Var}(\hat{\beta})(X_1 - X_1)' \\
&= 0
\end{aligned}$$

This makes sense: If the model dictated that the same thing would happen twice, then the best estimate would be what had been observed; and there could be no different outcome the second time.

Usually Σ is known only to within a proportionality constant, i.e., Φ of the equation Var[**e**] $= \Sigma = \sigma^2\Phi$ is known and σ^2 must be estimated. σ^2 can be estimated as follows. First, let:

$$\begin{aligned}
\hat{\mathbf{e}}_1 &= \mathbf{y}_1 - X_1\hat{\beta} \\
&= \mathbf{y}_1 - X_1(X_1'\Phi_{11}^{-1}X_1)^{-1}X_1'\Phi_{11}^{-1}\mathbf{y}_1 \\
&= (I_{n_1} - X_1(X_1'\Phi_{11}^{-1}X_1)^{-1}X_1'\Phi_{11}^{-1})\mathbf{y}_1 \\
\text{Also, } E[\hat{\mathbf{e}}_1] &= E[(I_{n_1} - X_1(X_1'\Phi_{11}^{-1}X_1)^{-1}X_1'\Phi_{11}^{-1})\mathbf{y}_1] \\
&= (I_{n_1} - X_1(X_1'\Phi_{11}^{-1}X_1)^{-1}X_1'\Phi_{11}^{-1})E[\mathbf{y}_1] \\
&= (I_{n_1} - X_1(X_1'\Phi_{11}^{-1}X_1)^{-1}X_1'\Phi_{11}^{-1})X_1\beta \\
&= (X_1 - X_1(X_1'\Phi_{11}^{-1}X_1)^{-1}X_1'\Phi_{11}^{-1}X_1)\beta \\
&= (X_1 - X_1)\beta \\
&= 0
\end{aligned}$$

Then, using the trace operator of Appendix B and the fact that Var[$\mathbf{y}_1$] = Var[$X_1\beta + \mathbf{e}_1$] = Var[$\mathbf{e}_1$] $= \Sigma_{11} = \sigma^2\Phi_{11}$, we have:

$$
\begin{aligned}
E[\hat{\mathbf{e}}_1'\Phi_{11}^{-1}\hat{\mathbf{e}}_1] &= E[Tr(\hat{\mathbf{e}}_1'\Phi_{11}^{-1}\hat{\mathbf{e}}_1)] \\
&= E[Tr(\Phi_{11}^{-1}\hat{\mathbf{e}}_1\hat{\mathbf{e}}_1')] \\
&= Tr(E[\Phi_{11}^{-1}\hat{\mathbf{e}}_1\hat{\mathbf{e}}_1']) \\
&= Tr(\Phi_{11}^{-1}E[\hat{\mathbf{e}}_1\hat{\mathbf{e}}_1']) \\
&= Tr(\Phi_{11}^{-1}Var[\hat{\mathbf{e}}_1]) \\
&= Tr(\Phi_{11}^{-1}Var[(I_{t_1} - X_1(X_1'\Phi_{11}^{-1}X_1)^{-1}X_1'\Phi_{11}^{-1})\mathbf{y}_1]) \\
&= Tr(\Phi_{11}^{-1}(I_{t_1} - X_1(X_1'\Phi_{11}^{-1}X_1)^{-1}X_1'\Phi_{11}^{-1})Var[\mathbf{y}_1](I_{t_1} - X_1(X_1'\Phi_{11}^{-1}X_1)^{-1}X_1'\Phi_{11}^{-1})') \\
&= Tr(\Phi_{11}^{-1}(I_{t_1} - X_1(X_1'\Phi_{11}^{-1}X_1)^{-1}X_1'\Phi_{11}^{-1})\sigma^2\Phi_{11}(I_{t_1} - X_1(X_1'\Phi_{11}^{-1}X_1)^{-1}X_1'\Phi_{11}^{-1})') \\
&= \sigma^2 Tr(\Phi_{11}^{-1}(I_{t_1} - X_1(X_1'\Phi_{11}^{-1}X_1)^{-1}X_1'\Phi_{11}^{-1})\Phi_{11}(I_{t_1} - X_1(X_1'\Phi_{11}^{-1}X_1)^{-1}X_1'\Phi_{11}^{-1})') \\
&= \sigma^2 Tr(\Phi_{11}^{-1}(\Phi_{11} - X_1(X_1'\Phi_{11}^{-1}X_1)^{-1}X_1')(I_{t_1} - X_1(X_1'\Phi_{11}^{-1}X_1)^{-1}X_1'\Phi_{11}^{-1})') \\
&= \sigma^2 Tr((I_{t_1} - \Phi_{11}^{-1}X_1(X_1'\Phi_{11}^{-1}X_1)^{-1}X_1')(I_{t_1} - X_1(X_1'\Phi_{11}^{-1}X_1)^{-1}X_1'\Phi_{11}^{-1})') \\
&= \sigma^2 Tr((I_{t_1} - \Phi_{11}^{-1}X_1(X_1'\Phi_{11}^{-1}X_1)^{-1}X_1')(I_{t_1} - \Phi_{11}^{-1}X_1(X_1'\Phi_{11}^{-1}X_1)^{-1}X_1')) \\
&= \sigma^2 Tr(I_{t_1} - \Phi_{11}^{-1}X_1(X_1'\Phi_{11}^{-1}X_1)^{-1}X_1' - \Phi_{11}^{-1}X_1(X_1'\Phi_{11}^{-1}X_1)^{-1}X_1' + \Phi_{11}^{-1}X_1(X_1'\Phi_{11}^{-1}X_1)^{-1}X_1') \\
&= \sigma^2 Tr(I_{t_1} - \Phi_{11}^{-1}X_1(X_1'\Phi_{11}^{-1}X_1)^{-1}X_1') \\
&= \sigma^2 Tr(I_{t_1}) - \sigma^2 Tr(\Phi_{11}^{-1}X_1(X_1'\Phi_{11}^{-1}X_1)^{-1}X_1') \\
&= \sigma^2 Tr(I_{t_1}) - \sigma^2 Tr((X_1'\Phi_{11}^{-1}X_1)^{-1}X_1'\Phi_{11}^{-1}X_1) \\
&= \sigma^2 Tr(I_{t_1}) - \sigma^2 Tr(I_k) \\
&= \sigma^2(t_1 - k)
\end{aligned}
$$

Therefore, one can estimate σ^2 as $\hat{\sigma}^2 = \dfrac{\hat{\mathbf{e}}_1'\Phi_{11}^{-1}\hat{\mathbf{e}}_1}{t_1 - k}$, which estimator is unbiased because

$$E[\hat{\sigma}^2] = \frac{E[\hat{\mathbf{e}}_1'\Phi_{11}^{-1}\hat{\mathbf{e}}_1]}{t_1 - k} = \frac{\sigma^2(t_1 - k)}{t_1 - k} = \sigma^2 .$$

This approach to the linear statistical model and estimation should be compared with Gary Venter's profound chapter on credibility [19]. Venter concentrates on what he, borrowing from Arthur Bailey in the 1940s, calls "greatest accuracy credibility." This subject, he says, has two subdivisions, least squares credibility and Bayesian analysis [19: 383-387]. As an illustration of least squares credibility, he set forth N risks, each one being observed over n years. X_{iu} represents the pure premium of the i^{th} risk in year u [19: 416-418], and an estimator is sought for X_{g0}, the pure premium of the g^{th} risk in a future year. The desired estimator will be linear in the X_{iu}s, using "the weights (a's) that minimize" [19: 418]:

$$E[X_{g0} - (a_0 + \sum_{i,u} a_{iu} X_{iu})]^2$$

Venter assumes a variance structure which Searle calls a one-way random effects model [17: 473], and minimizes the mean squared error by differentiation with respect to the a's. He calls the result a credibility formula:

> Thus the best linear estimate of X_{g0} turns out to be a credibility formula. This formula can alternatively be derived as the least squares linear estimate having $a_0 = 0$ but constrained to be unbiased [19: 423].

Thus Venter is essentially doing best linear unbiased estimation on a linear model. The author hopes that actuaries will begin to see the subject of credibility from the perspective of statistical modeling.

A desirable property of best linear unbiased estimation is that it is a linear operation, which will now be proved. Consider the linear model:

$$\begin{matrix} \mathbf{y}_1 = X_1\beta + \mathbf{e}_1 \\ \mathbf{y}_2 = X_2\beta + \mathbf{e}_2 \\ \mathbf{y}_3 = X_3\beta + \mathbf{e}_3 \end{matrix} \qquad \Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} & \Sigma_{13} \\ \Sigma_{21} & \Sigma_{22} & \Sigma_{23} \\ \Sigma_{31} & \Sigma_{32} & \Sigma_{33} \end{bmatrix}$$

The second and third groups are unobserved, and we wish to estimate the linear combination $\mathbf{y}_4 = D_2\mathbf{y}_2 + D_3\mathbf{y}_3 = (D_2X_2 + D_3X_3)\beta + D_2\mathbf{e}_2 + D_3\mathbf{e}_3 = (D_2X_2 + D_3X_3)\beta + \mathbf{e}_4$. The variance matrix is:

$$\Sigma^{*} = \begin{bmatrix} \Sigma_{11} & \Sigma_{14} \\ \Sigma_{41} & \Sigma_{44} \end{bmatrix} = \begin{bmatrix} \Sigma_{11} & \Sigma_{12}D_2' + \Sigma_{13}D_3' \\ D_2\Sigma_{21} + D_3\Sigma_{31} & \Sigma_{44} \end{bmatrix}$$

Then:

$$\begin{aligned}
\widehat{(D_2\mathbf{y}_2 + D_3\mathbf{y}_3)} &= (\Sigma_{41}\Sigma_{11}^{-1} + ((D_2X_2 + D_3X_3) - \Sigma_{41}\Sigma_{11}^{-1}X_1)(X_1'\Sigma_{11}^{-1}X_1)^{-1}X_1'\Sigma_{11}^{-1})\mathbf{y}_1 \\
&= ((D_2\Sigma_{21} + D_3\Sigma_{31})\Sigma_{11}^{-1} + ((D_2X_2 + D_3X_3) - (D_2\Sigma_{21} + D_3\Sigma_{31})\Sigma_{11}^{-1}X_1)(X_1'\Sigma_{11}^{-1}X_1)^{-1}X_1'\Sigma_{11}^{-1})\mathbf{y}_1 \\
&= ((D_2\Sigma_{21})\Sigma_{11}^{-1} + ((D_2X_2) - (D_2\Sigma_{21})\Sigma_{11}^{-1}X_1)(X_1'\Sigma_{11}^{-1}X_1)^{-1}X_1'\Sigma_{11}^{-1})\mathbf{y}_1 + \\
&\quad ((D_3\Sigma_{31})\Sigma_{11}^{-1} + ((D_3X_3) - (D_3\Sigma_{31})\Sigma_{11}^{-1}X_1)(X_1'\Sigma_{11}^{-1}X_1)^{-1}X_1'\Sigma_{11}^{-1})\mathbf{y}_1 \\
&= D_2(\Sigma_{21}\Sigma_{11}^{-1} + (X_2 - \Sigma_{21}\Sigma_{11}^{-1}X_1)(X_1'\Sigma_{11}^{-1}X_1)^{-1}X_1'\Sigma_{11}^{-1})\mathbf{y}_1 + \\
&\quad D_3(\Sigma_{31}\Sigma_{11}^{-1} + (X_3 - \Sigma_{31}\Sigma_{11}^{-1}X_1)(X_1'\Sigma_{11}^{-1}X_1)^{-1}X_1'\Sigma_{11}^{-1})\mathbf{y}_1 \\
&= D_2\hat{\mathbf{y}}_2 + D_3\hat{\mathbf{y}}_3
\end{aligned}$$

Now consider the linear statistical model with a constraint on β:

$$\begin{aligned}
\mathbf{y}_1 &= X_1\beta + \mathbf{e}_1 \\
b_{m\times 1} &= A_{m\times k}\beta_{k\times 1} \\
\mathbf{y}_2 &= X_2\beta + \mathbf{e}_2
\end{aligned}$$

where $\rho(A) = r$. Let $s = k - r$. Assume that the constraint is consistent. Then, according to Appendix B, there exists a ($k\times s$) matrix V such that $V'V = I_S$, and every solution will be of the form $\beta = A^{+}b + V\gamma$, for some ($s\times 1$) γ. Therefore, we can transform the model into an unconstrained model in γ:

$$\begin{aligned}
\mathbf{y}_1 &= X_1(A^{+}b + V\gamma) + \mathbf{e}_1 \\
\mathbf{y}_2 &= X_2(A^{+}b + V\gamma) + \mathbf{e}_2
\end{aligned}$$

$$\begin{aligned}
(\mathbf{y}_1 - X_1A^{+}b) &= (X_1V)\gamma + \mathbf{e}_1 \\
(\mathbf{y}_2 - X_2A^{+}b) &= (X_2V)\gamma + \mathbf{e}_2
\end{aligned}$$

It was required that $X_1'\Sigma_{11}^{-1}X_1$ be nonsingular. Since it is NND, then it must be also PD. And since $V'V = I_S$, $\rho(V) = s$. From the last theorem of Appendix A we can infer that $V'(X_1'\Sigma_{11}^{-1}X_1)V = (X_1V)'\Sigma_{11}^{-1}(X_1V)$ is PD, which implies that it is non-singular. Hence, the transformed model satisfies the requirement.

Hence, relying on previous theorems of the appendix, we have:

$$
\begin{aligned}
\hat{y}_2 &= (\hat{y}_2 - X_2A^+b) + X_2A^+b \\
&= (\hat{y}_2 - \widehat{(X_2A^+b)}) + X_2A^+b \\
&= (\widehat{y_2 - X_2A^+b}) + X_2A^+b \\
&= (X_2V)\hat{\gamma} + \Sigma_{21}\Sigma_{11}^{-1}((y_1 - X_1A^+b) - (X_1V)\hat{\gamma}) + X_2A^+b \\
&= X_2(A^+b + V\hat{\gamma}) + \Sigma_{21}\Sigma_{11}^{-1}(y_1 - X_1(A^+b + V\hat{\gamma})) \\
&= X_2\widehat{(A^+b + V\gamma)} + \Sigma_{21}\Sigma_{11}^{-1}(y_1 - X_1\widehat{(A^+b + V\gamma)}) \\
&= X_2\hat{\beta} + \Sigma_{21}\Sigma_{11}^{-1}(y_1 - X_1\hat{\beta})
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{Var}[y_2 - \hat{y}_2] &= (X_2V - \Sigma_{21}\Sigma_{11}^{-1}X_1V)\mathrm{Var}(\hat{\gamma})(X_2V - \Sigma_{21}\Sigma_{11}^{-1}X_1V)' + (\Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}) \\
&= (X_2 - \Sigma_{21}\Sigma_{11}^{-1}X_1)V\mathrm{Var}(\hat{\gamma})V'(X_2 - \Sigma_{21}\Sigma_{11}^{-1}X_1)' + (\Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}) \\
&= (X_2 - \Sigma_{21}\Sigma_{11}^{-1}X_1)\mathrm{Var}(V\hat{\gamma})(X_2 - \Sigma_{21}\Sigma_{11}^{-1}X_1)' + (\Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}) \\
&= (X_2 - \Sigma_{21}\Sigma_{11}^{-1}X_1)\mathrm{Var}(A^+b + V\hat{\gamma})(X_2 - \Sigma_{21}\Sigma_{11}^{-1}X_1)' + (\Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}) \\
&= (X_2 - \Sigma_{21}\Sigma_{11}^{-1}X_1)\mathrm{Var}(\hat{\beta})(X_2 - \Sigma_{21}\Sigma_{11}^{-1}X_1)' + (\Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12})
\end{aligned}
$$

Therefore, even though the transformation puts the model into γ-space, the results can be transformed back into β-space.

Before proceeding to relaxing the requirement that Σ_{11} be non-singular, consider the transformation of the observations by a $(t_1 \times t_1)$ non-singular matrix D:

$$\begin{aligned} D\mathbf{y}_1 &= DX_1\beta + D\mathbf{e}_1 \\ \mathbf{y}_2 &= X_2\beta + \mathbf{e}_2 \end{aligned} \qquad \Sigma^* = \begin{bmatrix} D\Sigma_{11}D' & D\Sigma_{12} \\ \Sigma_{21}D' & \Sigma_{22} \end{bmatrix}$$

Then:

$$\begin{aligned} \hat{\mathbf{y}}_2 &= (\Sigma_{21}D'(D\Sigma_{11}D')^{-1} + (X_2 - \Sigma_{21}D'(D\Sigma_{11}D')^{-1}DX_1)(X_1'D'(D\Sigma_{11}D')^{-1}DX_1)^{-1}X_1'D'(D\Sigma_{11}D')^{-1})D\mathbf{y}_1 \\ &= (\Sigma_{21}(D\Sigma_{11})^{-1} + (X_2 - \Sigma_{21}(D\Sigma_{11})^{-1}DX_1)(X_1'(D\Sigma_{11})^{-1}DX_1)^{-1}X_1'(D\Sigma_{11})^{-1})D\mathbf{y}_1 \\ &= (\Sigma_{21}\Sigma_{11}^{-1}D^{-1} + (X_2 - \Sigma_{21}\Sigma_{11}^{-1}X_1)(X_1'\Sigma_{11}^{-1}X_1)^{-1}X_1'\Sigma_{11}^{-1}D^{-1})D\mathbf{y}_1 \\ &= (\Sigma_{21}\Sigma_{11}^{-1} + (X_2 - \Sigma_{21}\Sigma_{11}^{-1}X_1)(X_1'\Sigma_{11}^{-1}X_1)^{-1}X_1'\Sigma_{11}^{-1})\mathbf{y}_1 \\ \mathrm{Var}[\mathbf{y}_2 - \hat{\mathbf{y}}_2] &= \Sigma_{22} - \Sigma_{21}D'(D\Sigma_{11}D')^{-1}D\Sigma_{12} + (X_2 - \Sigma_{21}D'(D\Sigma_{11}D')^{-1}DX_1)(X_1'D'(D\Sigma_{11}D')^{-1}X_1)^{-1}(X_2 - \Sigma_{21}D'(D\Sigma_{11}D')^{-1}DX_1)' \\ &= (X_2 - \Sigma_{21}\Sigma_{11}^{-1}X_1)\mathrm{Var}(\hat{\beta})(X_2 - \Sigma_{21}\Sigma_{11}^{-1}X_1)' + (\Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}) \end{aligned}$$

Therefore, a non-singular transformation of the observations has no effect on the estimator of $\mathbf{y}_2$.

Now suppose that $(t_1 \times t_1)$ Σ_{11} is singular; in fact, let $\rho(\Sigma_{11}) = r < t_1$, and let $s = t_1 - r$. Then (cf. Appendix A) Σ_{11} can be diagonalized as $W\Lambda W'$, where W is orthogonal. Moreover, Λ can be arranged such that the eigenvalues are in descending order, which makes the first r diagonal elements of Λ positive and the remaining s zero. According to the previous theorem, the observations can be transformed by W' without affecting the estimator:

$$\begin{aligned} W'\mathbf{y}_1 &= W'X_1\beta + W'\mathbf{e}_1 \\ \mathbf{y}_2 &= X_2\beta + \mathbf{e}_2 \end{aligned} \qquad \Sigma^* = \begin{bmatrix} W'\Sigma_{11}W & W'\Sigma_{12} \\ \Sigma_{21}W & \Sigma_{22} \end{bmatrix}$$

$$\text{where } W'\Sigma_{11}W = W'(W\Lambda W')W = \Lambda = \begin{bmatrix} \Lambda_r & 0_{r\times s} \\ 0_{s\times r} & 0_{s\times s} \end{bmatrix}$$

Let W_1' be the first r rows of W', and W_2' be the last s rows. Equivalently, let W_1 be the first r columns of W, and W_2 be the last s columns. As mentioned in Appendix B, because

W is orthogonal, $W_1'W_1 = I_r$, $W_1'W_2 = 0_{r \times s}$, and $W_2'W_2 = I_s$. Then the model can be written as:

$$\begin{aligned} W_1'y_1 &= W_1'X_1\beta + W_1'e_1 \\ W_2'y_1 &= W_2'X_1\beta + W_2'e_1 \\ y_2 &= X_2\beta + e_2 \end{aligned} \qquad \Sigma^* = \begin{bmatrix} \Lambda_r & 0_{r \times s} & W_1'\Sigma_{12\ r \times t_2} \\ 0_{s \times r} & 0_{s \times s} & W_2'\Sigma_{12\ s \times t_2} \\ \Sigma_{21}W_{1\ t_2 \times r} & \Sigma_{21}W_{2\ t_2 \times s} & \Sigma_{22\ t_2 \times t_2} \end{bmatrix}$$

But whenever a NND matrix has a zero diagonal element, the entire row and column intersecting that element must be zero. Therefore, $W_2'\Sigma_{12}$ and its transpose must be zero. Thus the transformed model becomes a model with a constraint on β:

$$\begin{aligned} W_1'y_1 &= W_1'X_1\beta + W_1'e_1 \\ W_2'y_1 &= W_2'X_1\beta \\ y_2 &= X_2\beta + e_2 \end{aligned} \qquad \Sigma^* = \begin{bmatrix} \Lambda_r & W_1'\Sigma_{12\ r \times t_2} \\ & \\ \Sigma_{21}W_{1\ t_2 \times r} & \Sigma_{22\ t_2 \times t_2} \end{bmatrix}$$

Since X_1 is still assumed to be of full column rank, the requirement that both Λ_r and $(W_1'X_1)'\Lambda_r^{-1}(W_1'X_1)$ be non-singular is satisfied; therefore, as long as the constraint is consistent, there is a best linear unbiased estimator of y_2. For a slightly different treatment of this subject see Amemiya [1: 185].

According to the last theorem of Appendix A, If Σ_{11} is non-singular and X_1 is of full column rank, then $X_1'\Sigma_{11}^{-1}X_1$ is PD, and hence non-singular. But if Σ_{11} is non-singular and X_1 is not of full column rank (or, X_1 is multicollinear) then $X_1'\Sigma_{11}^{-1}X_1$ will be singular. It is natural to try $(X_1'\Sigma_{11}^{-1}X_1)^+$ in the definition of A_0:

$$\hat{y}_2 = A_0 y_1 = (\Sigma_{21}\Sigma_{11}^{-1} + (X_2 - \Sigma_{21}\Sigma_{11}^{-1}X_1)(X_1'\Sigma_{11}^{-1}X_1)^+ X_1'\Sigma_{11}^{-1})y_1$$

The proof that this estimator is best will still hold when the traditional inverse is replaced with the MP inverse. However, the unbaisedness of the estimator is imperiled. This will be demonstrated in the following, which uses MP theorems of Appendix B and the theorem that the PD matrix Σ_{11}^{-1} can be factored as W'W. First:

$$\begin{aligned}
\Sigma_{11}^{-1}X_1(X_1'\Sigma_{11}^{-1}X_1)^{+}X_1'\Sigma_{11}^{-1}X_1 &= W'WX_1(X_1'W'WX_1)^{+}X_1'W'WX_1 \\
&= W'(WX_1)((WX_1)'(WX_1))^{+}(WX_1)'(WX_1) \\
&= W'(WX_1) \\
&= \Sigma_{11}^{-1}X_1
\end{aligned}$$

Then:

$$\begin{aligned}
A_0X_1 &= (\Sigma_{21}\Sigma_{11}^{-1} + (X_2 - \Sigma_{21}\Sigma_{11}^{-1}X_1)(X_1'\Sigma_{11}^{-1}X_1)^{+}X_1'\Sigma_{11}^{-1})X_1 \\
&= \Sigma_{21}\Sigma_{11}^{-1}X_1 + X_2(X_1'\Sigma_{11}^{-1}X_1)^{+}X_1'\Sigma_{11}^{-1}X_1 - \Sigma_{21}\Sigma_{11}^{-1}X_1(X_1'\Sigma_{11}^{-1}X_1)^{+}X_1'\Sigma_{11}^{-1}X_1 \\
&= \Sigma_{21}\Sigma_{11}^{-1}X_1 + X_2(X_1'\Sigma_{11}^{-1}X_1)^{+}X_1'\Sigma_{11}^{-1}X_1 - \Sigma_{21}\Sigma_{11}^{-1}X_1 \\
&= X_2(X_1'\Sigma_{11}^{-1}X_1)^{+}X_1'\Sigma_{11}^{-1}X_1
\end{aligned}$$

But the estimator is unbiased if and only if $A_0X_1 = X_2$. When X_1 is multicollinear, the equality is spoiled by the addition of the symmetric idempotent $(X_1'\Sigma_{11}^{-1}X_1)^{+}X_1'\Sigma_{11}^{-1}X_1$. As mentioned in Appendix B, such a matrix is the closest thing to an identity matrix. But all we can say is that the estimator is close to being unbiased. However, if there is a matrix D such that $X_2 = DX_1$, then the estimator is guaranteed to be unbiased:

$$\begin{aligned}
A_0X_1 &= X_2(X_1'\Sigma_{11}^{-1}X_1)^{+}X_1'\Sigma_{11}^{-1}X_1 \\
&= DX_1(X_1'\Sigma_{11}^{-1}X_1)^{+}X_1'\Sigma_{11}^{-1}X_1 \\
&= D\Sigma_{11}\Sigma_{11}^{-1}X_1(X_1'\Sigma_{11}^{-1}X_1)^{+}X_1'\Sigma_{11}^{-1}X_1 \\
&= D\Sigma_{11}\Sigma_{11}^{-1}X_1 \\
&= DX_1 \\
&= X_2
\end{aligned}$$

In other words, if the design matrix of the estimations has the same linear dependencies as the design matrix of the observations, then multicollinearity is inconsequential.

Appendix D

Confidence Ellipsoids and Chebyshev's Inequality

Appendix C touched on Gary Venter's discussion of greatest accuracy credibility, but was not the place to dwell on the subject. The term "greatest accuracy credibility" originated with Arthur Bailey [3: 20]. Venter subdivides this into two parts, least squares (LS) credibility and Bayesian analysis [19: 383f.]. Though best linear unbiased estimation (BLUE) provides a more general approach to the linear statistical model than does LS, Appendix C noted that the two approaches produce the same results. Therefore, Venter's evaluation of least squares credibility applies in part to BLUE; and the author wishes to defend BLUE against some of his criticisms of LS credibility.

In comparing LS credibility with Bayesian analysis Venter says:

> An apparent advantage of credibility over Bayesian analysis is that distributional assumptions are not needed for credibility. That is, credibility gives the best linear least squares answer for any distribution, whereas a Bayesian analysis will be different for different distributions. There are two problems with this conclusion, however. First, . . . when the Bayesian estimates are not linear, as in the case of most highly skewed distributions, credibility errors can be substantially greater than the Bayes' errors. Second, when Bayes' estimates are linear functions of the data, postulating normal or gamma distributions will give the same answer as credibility, because the Bayesian predictive means are linear in the observations for these distributions. Thus credibility analysis gives the same answer as assuming normal (or gamma) distributions and doing a Bayesian analysis, and it gives a useful answer only in those cases where normal or gamma distributions would be reasonable.
>
> . . . An advantage of Bayesian analysis is that it gives a distribution around the estimate, so that the degree of likely deviation from the estimate can be quantified. [19: 386f.]

Obviously, if the distribution underlying the error vector $\mathbf{e}$ of a statistical model is highly skewed, then employing knowledge of the distribution will improve the estimation. (Since $E[\mathbf{e}] = 0$, conventional distributions must be centered around their means.) But actuaries usually are content to work with such light-tailed distributions as the gamma; perhaps only in reinsurance (Venter's specialty) is there need to use heavy-tailed distributions, the "grand-daddy" of them all being the transformed beta distribution [19: 480].

The great advantage of BLUE, which cannot be said of Venter's LS formulation, is that it is multivariate. A $(t_2 \times 1)$ random vector, $\mathbf{y}_2$, is estimated *as a unit*; there are not t_2 separate estimations. This is the basis for the "conjoint prediction" of the body of this paper. And BLUE minimizes the variance of the prediction error *as a unit*, relying on the concept of a non-negative definite matrix (Appendix A). This does justice to Arthur Bailey's criterion that estimates should be optimal *overall* and not necessarily *piece-by-piece* [3: 13].

Also, when loss triangles are involved, whether in ratemaking or in reserving, invariably the sum of many cells in the triangle is more important than the individual cells themselves. So even if the error of each cell is highly skewed, the sum of many such cells may not be significantly skewed. Moreover, Venter himself says:

> A useful alternative when working with highly skewed distributions . . . is to transform the data, e.g., by taking logs, before doing the analysis. . . . The purpose of this . . . is to get distributions which are closer to the normal or gamma, so that the credibility estimator is more nearly optimal [19: 387].

This is good advice, and statisticians and econometricians frequently transform their data for the reason specified by Venter. The only drawback when this is applied to loss triangles is that if one desires the estimate of a sum of cells, e.g., $X_1 + X_2$, and one has log-transformed the data, it is not true that $X_1 + X_2 = \exp(\ln(X_1) + \ln(X_2))$.

But what Venter says about LS credibility, that it does not give a distribution around the estimate, is not true; at least it is not true when translated into BLUE, since BLUE yields the variance of the prediction error, $\mathrm{Var}[\mathbf{y}_2 - \hat{\mathbf{y}}_2]$.

Multivariate variance is extremely powerful. Consider an $(n\times 1)$ random vector $\mathbf{x}$, whose $(n\times 1)$ mean is μ and $(n\times n)$ variance is Σ. Stating that $\mathbf{x}$ is within a certain distance from μ to a certain confidence has little relation to how close each element x_i is to the corresponding μ_i. In fact, the notion of multivariate distance needs clarification. For an $(n\times 1)$ vector x, x'x represents the square of the distance of x from the origin of n-space, and $(x - \mu)'(x - \mu) = r^2$ represents a sphere in n-space with center μ and radius r. It is logical to use this measure of distance with a random vector of mean μ and variance I_n.

Now suppose that $\mathbf{x}$ has mean μ and positive definite variance Σ. Then according to Appendix A Σ is non-singular and there exists a nonsingular W such that $\Sigma = WW'$. And if $\mathbf{y} = W^{-1}(\mathbf{x} - \mu)$, which happens if and only if $\mathbf{x} = W\mathbf{y} + \mu$, then $E[\mathbf{y}] = 0$, and $\mathrm{Var}[\mathbf{y}] = \mathrm{Var}[W^{-1}(\mathbf{x} - \mu)] = W^{-1}\mathrm{Var}[(\mathbf{x} - \mu)](W^{-1})' = W^{-1}\Sigma(W^{-1})' = W^{-1}WW'(W^{-1})' = (W^{-1}W)(W^{-1}W)' = I_n$. Since $\mathbf{y}'\mathbf{y}$ is the appropriate measure for $\mathbf{y}$, then the appropriate measure for $\mathbf{x}$ is:

$$\begin{aligned}
\mathbf{y}'\mathbf{y} &= (W^{-1}(\mathbf{x}-\mu))'(W^{-1}(\mathbf{x}-\mu)) \\
&= (\mathbf{x}-\mu)'(W^{-1})'W^{-1}(\mathbf{x}-\mu) \\
&= (\mathbf{x}-\mu)'(W')^{-1}W^{-1}(\mathbf{x}-\mu) \\
&= (\mathbf{x}-\mu)'(WW')^{-1}(\mathbf{x}-\mu) \\
&= (\mathbf{x}-\mu)'\Sigma^{-1}(\mathbf{x}-\mu)
\end{aligned}$$

The last expression defines an ellipsoid with center μ and axes of lengths proportional to the square roots of the eigenvalues of Σ (cf. Johnson [10: 131]). Σ also determines how the axes of the ellipsoid are oriented with respect to the axes of n-space.

In the case of the multivariate normal distribution (cf. Johnson [10: 126-133] and Judge [11: 970-973]) the densest confidence regions are ellipsoids. But regardless of the distribution Chebyshev's inequality provides an ellipsoid guaranteed to enclose a certain confidence level. The proof depends on theorems from Appendices A and B, using the notation of the previous paragraph:

$$
\begin{aligned}
n &= \mathrm{Tr}(\mathrm{I}_n) \\
&= \mathrm{Tr}(\mathrm{Var}[\mathrm{W}^{-1}(\mathbf{x}-\mu)]) \\
&= \mathrm{Tr}(\mathrm{E}[(\mathrm{W}^{-1}(\mathbf{x}-\mu))(\mathrm{W}^{-1}(\mathbf{x}-\mu))']) \\
&= \mathrm{E}[\mathrm{Tr}((\mathrm{W}^{-1}(\mathbf{x}-\mu))(\mathrm{W}^{-1}(\mathbf{x}-\mu))')] \\
&= \mathrm{E}[\mathrm{Tr}((\mathrm{W}^{-1}(\mathbf{x}-\mu))'(\mathrm{W}^{-1}(\mathbf{x}-\mu)))] \\
&= \mathrm{E}[\mathrm{Tr}((\mathbf{x}-\mu)'(\mathrm{W}^{-1})'\mathrm{W}^{-1}(\mathbf{x}-\mu))] \\
&= \mathrm{E}[\mathrm{Tr}((\mathbf{x}-\mu)'\Sigma^{-1}(\mathbf{x}-\mu))] \\
&= \mathrm{E}[(\mathbf{x}-\mu)'\Sigma^{-1}(\mathbf{x}-\mu)] \\
&= \int (\mathrm{x}-\mu)'\Sigma^{-1}(\mathrm{x}-\mu)\,\mathrm{dF}(\mathrm{x}) \\
&= \int_{(\mathrm{x}-\mu)'\Sigma^{-1}(\mathrm{x}-\mu)<r^2} (\mathrm{x}-\mu)'\Sigma^{-1}(\mathrm{x}-\mu)\,\mathrm{dF}(\mathrm{x}) + \\
&\qquad \int_{(\mathrm{x}-\mu)'\Sigma^{-1}(\mathrm{x}-\mu)\geq r^2} (\mathrm{x}-\mu)'\Sigma^{-1}(\mathrm{x}-\mu)\,\mathrm{dF}(\mathrm{x}) \\
&\geq \int_{(\mathrm{x}-\mu)'\Sigma^{-1}(\mathrm{x}-\mu)\geq r^2} (\mathrm{x}-\mu)'\Sigma^{-1}(\mathrm{x}-\mu)\,\mathrm{dF}(\mathrm{x}) \\
&\geq \int_{(\mathrm{x}-\mu)'\Sigma^{-1}(\mathrm{x}-\mu)\geq r^2} r^2 \quad \mathrm{dF}(\mathrm{x}) = r^2\mathrm{Prob}[(\mathbf{x}-\mu)'\Sigma^{-1}(\mathbf{x}-\mu)\geq r^2]
\end{aligned}
$$

$$
\therefore\ \mathrm{Prob}[(\mathbf{x}-\mu)'\Sigma^{-1}(\mathbf{x}-\mu) < r^2] \geq 1 - \frac{n}{r^2}
$$

Hence, if one wishes to enclose at least $p = 1 - n/r^2$ of probability, then one should choose r as sqrt(n/(1 - p)).

Appendix E

The Allocation Problem

An interesting and important application of the linear statistical model involves what might be called quasi-observations. The form of this model is:

$$\begin{matrix} A\mathbf{y}_2 = AX_2\beta + A\mathbf{e}_2 \\ \mathbf{y}_2 = X_2\beta + \mathbf{e}_2 \end{matrix} \qquad \mathrm{Var}[\mathbf{e}] = \Sigma = \begin{bmatrix} A\Sigma_{22}A' & A\Sigma_{22} \\ \Sigma_{22}A' & \Sigma_{22} \end{bmatrix}$$

It can be seen that the observed rows of the model are a matrix times the unobserved rows. Therefore, the subscript will be dropped:

$$\begin{matrix} A\mathbf{y} = AX\beta + A\mathbf{e} \\ \mathbf{y} = X\beta + \mathbf{e} \end{matrix} \qquad \begin{bmatrix} A\Sigma A' & A\Sigma \\ \Sigma A' & \Sigma \end{bmatrix}$$

Ay might be observed, or it might be demanded from theory or from some other source. Also, frequently the model places constraints on β, even to a single value. Whether β is known or unknown, the best linear unbiased estimator (cf. Appendix C) of $\mathbf{y}$ is:

$$\begin{aligned} \hat{\mathbf{y}} &= X\hat{\beta} + \Sigma_{21}\Sigma_{11}^{-1}(A\mathbf{y} - AX\hat{\beta}) \\ &= X\hat{\beta} + (\Sigma A')(A\Sigma A')^{-1}(A\mathbf{y} - AX\hat{\beta}) \end{aligned}$$

The estimator premultiplied by A will have the value of the quasi-observation:

$$\begin{aligned} A\hat{\mathbf{y}} &= AX\hat{\beta} + A(\Sigma A')(A\Sigma A')^{-1}(A\mathbf{y} - AX\hat{\beta}) \\ &= AX\hat{\beta} + (A\Sigma A')(A\Sigma A')^{-1}(A\mathbf{y} - AX\hat{\beta}) \\ &= AX\hat{\beta} + (A\mathbf{y} - AX\hat{\beta}) \\ &= A\mathbf{y} \end{aligned}$$

The allocation problem is the simplest form of this model. Letting J_t denote a $(t\times 1)$ vector of ones, we state the problem as:

$$\begin{matrix} J_t'\mathbf{y} = J_t'\mu + J_t'\mathbf{e} \\ \mathbf{y} = I_t\mu + \mathbf{e} \end{matrix} \quad \Sigma = \begin{bmatrix} J_t'\Sigma J_t & J_t'\Sigma \\ \Sigma J_t & \Sigma \end{bmatrix},$$

where μ is known. We do not know $\mathbf{y}$, but we do know the quasi-observation $J_t'\mathbf{y}$, which is the sum of the elements of $\mathbf{y}$. The estimator of $\mathbf{y}$ is:

$$\begin{aligned} \hat{\mathbf{y}} &= I_t\mu + (\Sigma J_t)(J_t'\Sigma J_t)^{-1}(J_t'\mathbf{y} - J_t'\mu) \\ &= \mu + (\Sigma J_t)(J_t'\Sigma J_t)^{-1}(J_t'\mathbf{y} - J_t'\mu) \end{aligned}$$

This estimator answers the question: Knowing that $\mathbf{y}$ has mean μ and variance Σ, how should our expectation of $\mathbf{y}$ be modified when we know the sum of the elements of $\mathbf{y}$?

Consider the allocation of surplus. An insurance company has t liabilities, represented as a $(t\times 1)$ vector $\mathbf{y}$, whose mean μ and variance Σ are known. Also, the company has a surplus of s, considered as a (1×1) vector. We wish to allocate the surplus to the t liabilities. Think of what would be expected to happen to the liabilities if the total liability were exactly to consume the surplus:

$$\begin{aligned} \hat{\mathbf{y}} &= \mu + (\Sigma J_t)(J_t'\Sigma J_t)^{-1}(J_t'\mathbf{y} - J_t'\mu) \\ &= \mu + (\Sigma J_t)(J_t'\Sigma J_t)^{-1}((J_t'\mu + s) - J_t'\mu) \\ &= \mu + (\Sigma J_t)(J_t'\Sigma J_t)^{-1} s \end{aligned}$$

The second term of the last line shows how the losses consume the surplus, or how the surplus is allocated. ΣJ_t represents the sums across the rows of Σ; $J_t'\Sigma J_t$ represents the sum of all the elements of Σ. Therefore, surplus is allocated to the i^{th} liability according to the ratio of the sum of the i^{th} row of Σ to the sum of all the rows of Σ. As a check, if a liability had zero variance, then (since Σ is non-negative definite) all its covariances would be zero, and zero percent of the surplus would be allocated to it. Moreover, if Σ were diagonal, then surplus would be allocated according to the variances.

Another example of the allocation problem concerns price. As before, **y** is a $(t\times1)$ vector whose mean μ and variance Σ are known, and **y** represents the present value of the t objects. Because buyers are not normally risk-neutral, prices usually differ from μ. Suppose that we know the sum of the prices of the t objects, M, which might also be called the price of the whole market. Then the price P ought to be allocated as:

$$P = \mu + (\Sigma J_t)(J_t'\Sigma J_t)^{-1}(M - J_t'\mu)$$

Because of the relation between the estimator and the quasi-observation, $J_t'P = M$. Finally, if one defines $\mathbf{y}_M$ as $J_t'\mathbf{y}$, or as the present value of the whole market, then the equation can be written as:

$$\begin{aligned} P &= \mu + (\mathrm{Cov}[\mathbf{y},\mathbf{y}_M])(\mathrm{Cov}[\mathbf{y}_M,\mathbf{y}_M])^{-1}(M - J_t'\mu) \\ &= \mu + \mathrm{Cov}[\mathbf{y},\mathbf{y}_M]\mathrm{Var}[\mathbf{y}_M,\mathbf{y}_M]^{-1}(M - J_t'\mu) \end{aligned}$$

This looks like the Capital Asset Pricing Model [6: 360], except that the CAPM refers to rates of return, whereas the equation above refers to values and prices.

Appendix F

The Value of a Stochastic Cash Flow

For every time t, there is an amount $a(t)$ such that an investor is indifferent to receiving one dollar now (time 0) or receiving $a(t)$ dollars at time t. Because more money is better that less money, the investor would prefer receiving one dollar now to receiving less than $a(t)$ dollars at time t; and he would prefer receiving more than $a(t)$ dollars at time t to receiving one dollar now. Of course, $a(0) = 1$. Let $v(t) = a(t)^{-1}$. This represents the investor's indifference to receiving $v(t)$ dollars now or receiving one dollar at time t. These indifference functions presuppose that the reception of the dollars is certain, or risk-free.

For every cash flow there is an amount such that the investor indifferent to receiving that amount now or receiving the cash flow. This is true whether or not the cash flow is stochastic. We will call this amount the *certainty equivalent value* (CEV) of the cash flow, a fancy word for price.

Cash flows can be extremely complicated; but we will call the smallest unit of a cash flow an atom of cash flow, or an atomic cash flow. An atom of cash flow is the receipt of x units of money at time t. Let $\mathbf{u}_{(2\times 1)} = \begin{bmatrix} x \\ t \end{bmatrix}$, with mean $\mu = \begin{bmatrix} \mu_x \\ \mu_t \end{bmatrix}$. The atom is certain, or non-stochastic, if and only if $\mathrm{Var}[\mathbf{u}] = 0$; otherwise, the atom is uncertain, or stochastic. If the atom is non-stochastic, the only possible value is the mean. The investor will be indifferent to receiving μ_X at time μ_t or receiving $\mu_X \cdot v(\mu_t)$ now. Therefore, the CEV of a non-stochastic atomic cash flow is $\mu_X \cdot v(\mu_t)$.

When the atom is stochastic it makes sense to consider $x \cdot v(t)$ as a random variable. Letting f be the probability distribution function of **u**, we have:

$$\mathrm{E}[x \cdot v(t)] = \int x \cdot v(t) f(\mathrm{u}) d\mathrm{u}$$
$$\mathrm{Var}[x \cdot v(t)] = \int (x \cdot v(t) - \mathrm{E}[x \cdot v(t)])^2 f(\mathrm{u}) d\mathrm{u}$$

We will call $\mathrm{E}[x \cdot v(t)]$ the *actuarial present value* (APV) of the cash flow. In the case of an non-stochastic cash flow APV = $\mu_X \cdot v(\mu_t)$ = CEV.

Consider a simple atom of cash flow: a coin is tossed, and if it lands heads, one dollar will be received now; if tails, zero dollars will be received. This is a stochastic cash flow, because x varies, even though t is always zero. $x \cdot v(t)$ is a Bernoulli random variable with probability of one-half; therefore, APV = $\mathrm{E}[x \cdot v(t)] = ½$, and $\mathrm{Var}[x \cdot v(t)] = ¼$.

However, it is a fact of life that an investor may not be indifferent to receiving this cash flow or receiving the APV of half a dollar now. When the coin lands heads the investor gains half a dollar more than he would with certainty option; and when the coin lands tails he loses the half a dollar that he would have with the certainty option. If the investor desires the half a dollar gain more than he dreads the half a dollar loss, then he will prefer receiving the stochastic cash flow. Such an investor is *risk-inclined.* But if the investor dreads the loss more than he desires the gain, he will prefer receiving the APV of half a dollar now, and will be *risk-averse*. Only a *risk-neutral* investor is indifferent to the two cash flows.

Therefore, we can say that the CEV of a stochastic cash flow is less than the APV to a risk-averse investor, is equal to the APV to a risk-neutral investor, and is greater than the APV to a risk-inclined investor. Another fact of life is that most investment decisions are risk-averse.

Because CEV = APV for non-stochastic cash flows, financial theorists have tried to modify the APV formula to serve for stochastic cash flows. Think of the case of the risk averse investor, to whom CEV = APV* < APV. How should the formula $\mathrm{APV} = \int x \cdot v(t) f(\mathrm{u}) d\mathrm{u}$ be modified (denoted by the asterisk)? The theorists fastened on to the factor $v(t)$, so $\mathrm{APV}^* = \int x \cdot v^*(t) f(\mathrm{u}) d\mathrm{u}$. If you want APV* < APV, then make $v^*(t) < v(t)$. Since $v(t)$ is a discount function, the theorists decided to deepen the discount function to compensate for risk. Ignoring the term structure of interests rates, we may consider $v(t)$ to be $(1 + i)^{-t}$ for some i. Then the solution consisted in choosing some $r > i$, and letting $v^*(t)$ to be $(1 + r)^{-t}$. This seemed to be in keeping with the thinking of investors, who demanded a higher return (r) on a stochastic investment than that (i) on a non-stochastic investment.

Therefore, according to modern financial theory, the CEV of an atomic cash flow results from the correct choice of r in the equation:

$$\mathrm{PV} = \int x \cdot (1+r)^{-t} f(\mathrm{u}) d\mathrm{u}$$

'PV' stands for 'present value'; it is the accepted language, and for now we shall reserve *actuarial* present value for the case in which the discount function within the integral is $v(t)$. Therefore, it is accepted theory that the CEV is obtained by present valuing the cash flow at the correct *risk-adjusted rate of return*. It is important to understand that this theory means more than the statement that there exists an r such that $\mathrm{CEV} = \int x \cdot (1+r)^{-t} f(\mathrm{u}) d\mathrm{u}$. This would merely be an application of the intermediate value theorem. Rather, the theory states that the PV equation is the cause of the CEV's being what it is, that the correct way to arrive at the CEV is the PV equation, and that it should be easier to determine the correct r than to determine the CEV. The theory is canonized by the Actuarial Standards Board in its "Actuarial Standard of Practice No. 19: Actuarial Appraisals."

But the theory will not stand up to three very simple problems. The first involves the term structure of interest rates. The true $v(t)$ is not equal to $(1 + i)^{-t}$; rather there is a yield curve $i(t)$ such that $v(t) = (1 + i(t))^{-t}$. It stands to reason that there should be a risk-adjusted yield curve as well, $r(t)$, so that $v^*(t) = (1 + r(t))^{-t}$. In the case of a flat yield curve, there is some constant γ such that $v^*(t)/v(t) = \gamma^{-t}$. Therefore, the financial theorists at the very least chose wrongly on the form of the modified discount function; instead of $(1 + r)^{-t}$, it should be $v(t)\gamma^{-t}$.

The first problem demonstrates a flaw in the accepted theory which many would not consider to be grave. In fact, the demonstration also specified how to fix the flaw. But the second and third problems will be devastating to the theory, and will also point toward the true solution of the CEV of a stochastic cash flow. In the second problem the accepted theory does not discount when it should. This is like a type I error, wherein a test does not accept the null hypothesis when it should. In the third problem the accepted theory discounts when it should not. This is like a type II error, wherein a test accepts the null hypothesis when it should not.

The second problem uses the example of the coin toss. If the coin lands heads, one dollar will be received now; if tails, then nothing is received. The APV of this stochastic cash flow is half a dollar. But the true theory should be able to explain how to some investor the CEV is, for example, forty cents. Using the PV equation, we have:

$$
\begin{aligned}
\text{CEV} &= \int x \cdot (1+r)^{-t} f(\text{u}) du \\
&= \sum_i x_i \cdot (1+r)^{-t_i} p_i \\
&= 1 \cdot (1+r)^{-0}(1/2) + 0 \cdot (1+r)^{-0}(1/2) \\
&= 1 \cdot (1+r)^{-0}(1/2) \\
&= 1 \cdot 1(1/2) \\
&= 1/2
\end{aligned}
$$

This problem shows that the accepted theory cannot perform the requisite discounting of ten cents on present cash receipts. It is to no avail to argue that the money will be received at some $t > 0$. For then let the time of receipt be ε, and take the limit as $\varepsilon \to 0^+$. The theory just won't budge from the APV for cash flows in the present, or in the immediate future.

For the third problem imagine the following atomic cash flow: a time t is picked at random, as the result of which $x = a(t)$ dollars will be received at time t. This atom is stochastic as respects both amount and time. But because $a(t)$ represents the investor's indifference, he is indifferent to receiving $a(t)$ dollars at time t or $a(t) \cdot v(t) = 1$ dollar now. No matter what t is picked, the investor will receive a cash flow which, were it on its own as non-stochastic, would have a CEV of one dollar. Therefore, the CEV of this cash flow has to be one dollar. If one used the PV equation with a risk-adjusted rate of return, one would obtain the wrong answer of a CEV of less than one dollar.

The key to valuing a stochastic cash flow is evident from these two problems. The CEV of the second problem is less than the APV to risk-averse investor because $\mathrm{Var}[x \cdot v(t)] > 0$. In the third problem $\mathrm{Var}[x \cdot v(t)] = 0$, and the CEV is not less than the APV. The risk to which the investor is averse is not the variability of x and t, but the variability of $x \cdot v(t)$. The following table summarizes the relation of CEV to APV:

	$\mathrm{Var}[x \cdot v(t)] = 0$	$\mathrm{Var}[x \cdot v(t)] > 0$
Risk-averse Investor	CEV = APV	CEV < APV
Risk-neutral Investor	CEV = APV	CEV = APV
Risk-inclined Investor	CEV = APV	CEV > APV

Let us redefine the present value of an atomic cash flow as $y = x \cdot v(t)$. This means that from now on the identity is severed between certainty equivalent value and present value. It also means that there is only one present value, the one which uses $v(t)$, rather than a present value for every risk-adjusted rate of return.

Keeping in mind this definition of present value, we now pose the question: How do we determine the CEV of a cash flow whose variance is positive? Here is the author's suggested solution: Do it from the top down. Take a universe of cash flows and atomize it into n atoms. Let y_i be the present value of the i^{th} atomic cash flow, and let the $(n\times 1)$ random vector $\mathbf{y}$ have the y_is as its elements. Let the mean and variance of $\mathbf{y}$ be μ and Σ. Then the present value of the universe of cash flows (or the market) is the sum of the elements of $\mathbf{y}$, or, to use the notation of Appendix E, $\mathbf{y}_M = J_n'\mathbf{y}$. The mean and the variance of $\mathbf{y}_M$ are $J_n'\mu$ and $J_n'\Sigma J_n$.

The investor, looking at things from the top level, asks what should be the CEV of $\mathbf{y}_M$. The author suggests that this question is answered in the realm of utility theory (cf. Bowers [4: 2-15]). If the investor's utility function is χ, then the CEV of $\mathbf{y}_M$ is the solution of the equation $\chi(CEV_M) = E[\chi(\mathbf{y}_M)]$, or $CEV_M = \chi^{-1}(E[\chi(\mathbf{y}_M)])$. Once CEV_M is determined, it is allocated to the atoms according to the manner in which price was allocated in Appendix E:

$$\begin{aligned}
CEV[\mathbf{y}] &= E[\mathbf{y}] + Cov[\mathbf{y},\mathbf{y}_M]Cov[\mathbf{y}_M,\mathbf{y}_M]^{-1}(CEV_M - E[\mathbf{y}_M]) \\
&= \mu + \Sigma J_n(J_n'\Sigma J_n)^{-1}(CEV_M - J_n'\mu)
\end{aligned}$$

As a check of the correctness of this theory, the CEV of a linear combination of atomic cash flows (which would be a complicated cash flow) equals the linear combination of the CEVs of the atoms:

$$\begin{aligned}
CEV[A\mathbf{y}] &= E[A\mathbf{y}] + Cov[A\mathbf{y},\mathbf{y}_M]Cov[\mathbf{y}_M,\mathbf{y}_M]^{-1}(CEV_M - E[\mathbf{y}_M]) \\
&= AE[\mathbf{y}] + ACov[\mathbf{y},\mathbf{y}_M]Cov[\mathbf{y}_M,\mathbf{y}_M]^{-1}(CEV_M - E[\mathbf{y}_M]) \\
&= A(E[\mathbf{y}] + Cov[\mathbf{y},\mathbf{y}_M]Cov[\mathbf{y}_M,\mathbf{y}_M]^{-1}(CEV_M - E[\mathbf{y}_M])) \\
&= A(CEV[\mathbf{y}])
\end{aligned}$$

Therefore, the value (CEV) of a cash flow equals its expected present value plus a beta times the difference of the expected present value of the universe from the CEV of the universe, where beta is the covariance of the cash flow with the universe divided by the variance of the universe. If the covariance is zero, a special case of which is if the cash flow is non-stochastic, then the CEV is equal to the expected present value.

Present valuing is a prerequisite to certainty equivalent valuing; it was a mistake for the financial theorists to try to make it explain certainty equivalence, which the author suggests to be founded on utility theory. By the same token, it is a mistake for actuaries to set risk margins on loss reserves by present valuing at supposed risk-adjusted rates of return.

Appendix G

A SAS® Subroutine for the Linear Statistical Model

The examples of this paper were performed with the following SAS® subroutine. The author believes that a programming language, particularly one designed for statistical work, is necessary for all but the simplest linear models. Some spreadsheets do not have even simple matrix functions, much less generalized inverse and eigenvalue routines; and spreadsheet recalculation with matrices of rank greater than twenty can be very time-consuming. Subroutine LINMOD should be understandable to someone who is acquainted with the interactive matrix language of SAS® (SAS/IML® [15]) and who has studied Appendix C of this paper. In particular, the notation of the subroutine closely follows the notation of Appendix C. Sometimes, especially at the end of the subroutine, the code is cluttered with 'if' statements in order to handle degenerate situations involving $(0\times n)$ and $(m\times 0)$ matrices. These situations fit well into matrix theory, but most software does not accommodate them. Also, the function FUZZ(A) will round each element of A which is within 10^{-12} of an integer to that integer. This is a helpful function in matrix programming, where rounding errors can be significant.

```
proc iml;

   start solution (A, b);
   **********************************************************;
   ** Returns x=ginv(A)*b if it approximately solves A*x=b **;
   **********************************************************;

   G       =ginv(A)                        ;
   Gb      =G*b                            ;
   if any(nmiss(Gb)) then return (.)       ;
```

```
dbtb     =b[##,]                                  ;
if any(dbtb=0) then zero=loc(dbtb=0)              ;
               else zero=     .                   ;
if any(dbtb>0) then pos =loc(dbtb>0)              ;
               else pos =     .                   ;
AGb      =A*Gb                                    ;
dbtAGb   =AGb[##,]                                ;
if n(pos ) then do                                ;
   reldiff=dbtAGb[pos]/dbtb[pos]-1                ;
   if any(fuzz(reldiff)) then return (.)          ;
end                                               ;
if n(zero) then Gb[,zero]=0                       ;
return (Gb)                                       ;

finish solution;


start genvar (S);
***************************************************;
***** Returns the generalized variance of S  *****;
***************************************************;

scale=exp((log(vecdiag(S)))[:]);
return (scale*det(S/scale)**(1/nrow(S)));

finish genvar;


start linmod (analysis,y2est,y,X,Phi,A,b);
***************************************************;
** Linear model: y=X*beta+e where Var[e]=s2*Phi **;
**                subject to A*beta=b            **;
** Inputs: y,X,Phi,A,b. Outputs: analysis,y2est **;
***************************************************;

analysis=  .  ;
y2est   =  .  ;

t=nrow(y);
k=ncol(X);

/********** Error trapping **********/

if (type(y)^='N')|(type(X)^='N')|(type(Phi)^='N')|
   (type(A)^='N')|(type(b)^='N') then do;
   analysis='Faulty Model: Non-Numeric Matrix';
```

```
   return;
end;

if (nrow(X)^=t)|(nrow(Phi)^=t) then do;
   analysis='Faulty Model: Non-Conformable Matrices in Model';
   return;
end;

if (ncol(y)^=1)|(ncol(b)^=1) then do;
   analysis='Faulty Model: y or b is not a column vector';
   return;
end;

if any(nmiss(X))|any(nmiss(Phi)) then do;
   analysis='Faulty Model: Misssing Values in X or Phi';
   return;
end;

if nrow(phi)^=ncol(phi) then do;
   analysis='Faulty Model: Non-Symmetric Phi';
   return;
end;

if any(fuzz(phi`-phi)) then do;
   analysis='Faulty Model: Non-Symmetric Phi';
   return;
end;

if any(fuzz(eigval(phi))<0) then do;
   analysis='Faulty Model: Phi is not Non-Negative Definite';
   return;
end;

if all(^A) & all(^b) then do;
   A=shape(0,1,k);                /* dummy constraint */
   b=shape(0,1,1);
end;

if (nrow(A)^=nrow(b))|(ncol(A)^=k) then do;
   analysis='Faulty Model: Non-Conformable Matrices in Constraint';
   return;
end;

if any(nmiss(A))|any(nmiss(b)) then do;
   analysis='Faulty Model: Missing Values in Constraint';
   return;
```

```
end;

if any(nmiss(solution(A,b))) then do;
   analysis='Faulty Model: Inconsistent Constraint';
   return;
end;

/******** End Error trapping ********/

row1=n    (y);                     /* the   observed rows */
row2=nmiss(y);                     /* the unobserved rows */

if any(row2) then do;
   row2 =loc(row2     );
   y2   =  y[row2,    ];
   X2   =  X[row2,    ];
   Phi22=Phi[row2,row2];
   t2   =    nrow   (y2);
end;
else t2=0;

if any(row1) then do;
   row1 =loc(row1);
   y1   =  y[row1,    ];
   X1   =  X[row1,    ];
   Phi11=Phi[row1,row1];
   if any(row2) then do;
      Phi12=Phi[row1,row2];
      Phi21=Phi[row2,row1];
   end;
   t1   =    nrow   (y1);

   call eigen(lambda, W, Phi11);  /* Phi11=W*Diag(lambda)*W` */

   lambda=fuzz(lambda);
   if any(lambda<0) then do;
      analysis='Faulty Model: Phi11 not Non-Negative Definite';
      return;
   end;
   if any(lambda=0) then zero    =loc(lambda=0); else zero    =.;
   if any(lambda>0) then positive=loc(lambda>0); else positive=.;

   y1   =W`*y1;
   X1   =W`*X1;
   Phi11=Diag(lambda);            /* equal to W`*Phi11*W */
```

```
   if t2 then do;
      Phi12=W`*Phi12  ;
      Phi21=   Phi21*W;
   end;

   if any(zero) then do;          /* Phi11 not of full rank */
      A =A//(X1[zero,]);          /*  augmented constraint  */
      b =b//(y1[zero,]);
      if any(positive) then do;
         y1   =   y1[positive,        ];
         X1   =   X1[positive,        ];
         Phi11=Phi11[positive,positive];
         if t2 then do;
            Phi12=Phi12[positive,        ];
            Phi21=Phi21[        ,positive];
         end;
         t1   =nrow(y1);
      end;
      else t1=0;                  /* Phi11 was zero */
   end;
end;
else t1=0;

/*******************************************************
   The model now is:
   y1=X1*beta+e1       Phi11 Phi12
   y2=X2*beta+e2       Phi21 Phi22
   subject to A*beta=b.
   Phi11 is diagonal and positive definite.
   t1 and t2 can be zero.
*******************************************************/

G    =ginv(A);
GA   =G*A    ;
Gb   =G*b    ;
if any(nmiss(solution(A,b))) then do;
   analysis='Faulty Model: Inconsistent Augmented Constraint';
   return;
end;
call eigen(lambda, W, GA);
lambda=fuzz(lambda);
if any(lambda>0) then do;         /* beta has r constraints */
   positive =loc(lambda>0);
   U        =W[,positive]`;
   r        =nrow(U)      ;
   c        =U*Gb         ;       /* Reduced constraint:  U*beta=c */
```

```
   U        =fuzz(U)      ;
   c        =fuzz(c)      ;
end;
else r=0;
if any(lambda=0) then do;          /* beta has s degrees of freedom */
   zero     =loc(lambda=0);
   V        =W[,zero    ] ;
   s        =ncol(V)      ;
end;
else s=0;                          /* beta is fully constrained */

/*****************************************************************
   beta is a solution of A*beta=b if and only if there
   exists a gamma such that beta=Gb+V*gamma, where V
   is (kxs) and of full column rank

   Transformed model:
   p1=y1-X1*Gb=X1*V*gamma+e1=Q1*gamma+e1   Phi11 Phi12
   p2=y2-X2*Gb=X2*V*gamma+e2=Q2*gamma+e2   Phi21 Phi22

   X1 is (t1xk)  Q1 is (t1xs)  Phi11 is (t1xt1)   Phi12 is (t1xt2)
   X2 is (t2xk)  Q2 is (t2xs)  Phi21 is (t2xt1)   Phi22 is (t2xt2)

   Phi11 is (t1xt1) diagonal and positive definite.
   t1, t2, and s can be zero.

   p2 is estimable if there exists a D such that Q2=D*Q1, or
   Q1`*D`=Q2`.  If t2=0 or s=0, a solution exists.  If t2>0
   and s>0 and t1=0, a solution exists if and only if Q2=0.
*****************************************************************/

if  t1        then InvPhi11=diag(1/vecdiag(Phi11));    * (t1xt1);
if  t1        then p1      =y1-X1*Gb;                 * (t1x 1);
if  t1    & s then Q1      =   X1*V ;                 * (t1x s);
if      t2    then p2      =y2-X2*Gb;                 * (t2x 1);
if      t2& s then Q2      =   X2*V ;                 * (t2x s);

if t2& s then do;
   if (^t1& any(Q2))|(t1& any(nmiss(solution(Q1`,Q2`)))) then do;
      analysis='Faulty Model: y2 not estimable';
      return;
   end;
end;

if  t1    & s then QtIPhi  =Q1`*InvPhi11     ;        * (s xt1);
if  t1    & s then QtIPhiQ =QtIPhi*Q1        ;        * (s x s);
```

```
if ^t1    & s then QtIPhiQ =shape(0, s, s)  ;          * (s x s);
if  t1    & s then QtIPhip =QtIPhi*p1       ;          * (s x 1);
if ^t1    & s then QtIPhip =shape(0, s, 1)  ;          * (s x 1);
if          s then GQtIPhiQ=ginv(QtIPhiQ)   ;          * (s x s);
if          s then gamma   =GQtIPhiQ*QtIPhip;          * (s x 1);

if  t1    & s then p1hat   =Q1*gamma        ;          * (t1x 1);
if  t1    &^s then p1hat   =shape(0,t1, 1)  ;          * (t1x 1);
if  t1        then e1hat   =p1-p1hat        ;          * (t1x 1);
                   df      =t1-s            ;
if  t1& df    then s2=(e1hat`*InvPhi11*e1hat)/df;
              else s2=            .            ;
if          s then Vargamma=s2#GQtIPhiQ     ;          * (s x s);
if          s then beta    =Gb+V*gamma      ;          * (k x 1);
              else beta    =Gb              ;          * (k x 1);
if          s then Varbeta =V*Vargamma*V`   ;          * (k x k);
              else Varbeta =shape(0, k, k)  ;          * (k x k);
if  t1        then rsqrd   =(p1hat`*InvPhi11*p1hat)
                           /(p1    *InvPhi11*p1    );
              else rsqrd   =          .             ;
if  t1        then rbarsqrd=1-(t1/df)*(1-rsqrd)     ;
              else rbarsqrd=          .             ;
if  t1& t2    then Phi2111 =Phi21*InvPhi11     ;       * (t2xt1);
if  t1& t2    then y2hat   =X2*beta
                           +Phi2111*(y1-X1*beta);      * (t2x 1);
if ^t1& t2    then y2hat   =X2*beta            ;       * (t2x 1);
if  t1& t2    then X2X1    =X2-Phi2111*X1      ;       * (t2x k);
if ^t1& t2    then X2X1    =X2                 ;       * (t2x k);
if      t2    then Vary2hat=s2#(Phi22-Phi2111*Phi12)
                           +X2X1*Varbeta*X2X1`;        * (t2xt2);

analysis=df||s2||rsqrd||rbarsqrd||r;
analysis=shape(analysis,k,ncol(analysis));
analysis=beta||Varbeta||analysis;
constrnt=shape(0,k,k+1);
if r then constrnt[(1:r),]=U||c;
analysis=analysis||constrnt;
if t2 then do;
  y2est=row2`||y2hat||Vary2hat;
end;
else y2est=.;
return;

finish linmod;
```

The following statements will solve the exemplary models of Section 2:

```
y={75 15 10,
   75 25  .,
   50  .  .,
   50 30 20,
   60 25  .,
   45  .  .};
y=shape(y,18,1);

X={1 0 0 0 0 0,
   0 1 0 0 0 0,
   0 0 1 0 0 0,
   1 0 0 0 0 0,
   0 1 0 0 0 0,
   0 0 1 0 0 0,
   1 0 0 0 0 0,
   0 1 0 0 0 0,
   0 0 1 0 0 0,
   0 0 0 1 0 0,
   0 0 0 0 1 0,
   0 0 0 0 0 1,
   0 0 0 1 0 0,
   0 0 0 0 1 0,
   0 0 0 0 0 1,
   0 0 0 1 0 0,
   0 0 0 0 1 0,
   0 0 0 0 0 1};

Phi=i(18);

print /;
print
"Variation 1: Unrelated Paid and Incurred Losses";
A=.;
b=.;
call linmod(analysis,y2est,y,X,Phi,A,b);
print analysis, y2est /;

print
"Variation 2: Expected Paid and Incurred Losses have same Ultimate";
A={ 1  1  1 -1 -1 -1};
b={0};
call linmod(analysis,y2est,y,X,Phi,A,b);
```

```
print analysis, y2est /;

print
"Variation 3: Paid and Incurred Losses have same Ultimate";
G={ 1  1  1  0  0  0  0  0  0 -1 -1 -1  0  0  0  0  0  0,
    0  0  0  1  1  1  0  0  0  0  0  0 -1 -1 -1  0  0  0,
    0  0  0  0  0  0  1  1  1  0  0  0  0  0  0 -1 -1 -1};
A=G*X;
b={0, 0, 0};
Phi=Phi-Phi*G`*inv(G*Phi*G`)*G*Phi;
call linmod(analysis,y2est,y,X,Phi,A,b);
print analysis, y2est;
```

The following statements apply the subroutine to the self-insured entity of Sections 4-7:

```
exposure={1988 131332.20,
          1989 141672.24,
          1990 141677.29,
          1991 142577.99,
          1992 143285.58,
          1993 138261.75,
          1994 121857.69,
          1995 115000.00};

age ={   .      12      24      36      48      60      72      84     108};

paid={1988 266354 166572  32329  53610   8124  16924  39109      .,
      1989 246981 359380 229016  69539 118635 100292      .      .,
      1990 203178 375768 276617  74912  86428      .      .      .,
      1991 395630 260643 167709 270692      .      .      .      .,
      1992 207698 174615 162640      .      .      .      .      .,
      1993 167681 280178      .      .      .      .      .      .,
      1994 215740      .      .      .      .      .      .      .,
      1995      .      .      .      .      .      .      .      .};

incd={1988 422076  83969  19571  55987 -11674  63125 -50032      .,
      1989 457750 397257 151515  61946 135729 -56771      .      .,
      1990 345084 382514 342069  12295 -57918      .      .      .,
      1991 591842 288791 152181 307218      .      .      .      .,
      1992 379033 157808  74764      .      .      .      .      .,
      1993 308803 139977      .      .      .      .      .      .,
      1994 215851      .      .      .      .      .      .      .,
      1995      .      .      .      .      .      .      .      .};

fyid   =shape(repeat(paid[,1],1,8),64,1);
```

```
ageid  =shape(repeat(age[2:9],8,1),64,1);
X      =exposure[,2]@i(8);
Phi    =i(64);

print /;
print "Model 1: Paid Losses";
ypaid  =shape(paid[,2:9],64,1);
A1     ={ 1  1  1  1  1  1  1 -9};
b1     =0;
call linmod(analyze1,y2est1,ypaid,X,Phi,A1,b1);
print analyze1 /;

print "Model 2: Incurred Losses";
yincd  =shape(incd[,2:9],64,1);
A2     ={  1   1   1   1   1   1   1 -19};
b2     =0;
call linmod(analyze2,y2est2,yincd,X,Phi,A2,b2);
print analyze2 /;

print "Model 3: Unrelated Paid and Incurred Losses";
s2ratio=analyze2[1,11]/analyze1[1,11];
y      =ypaid//yincd;
X      =block(X,X);
Phi    =block(i(64), s2ratio#i(64));
A3     =block(A1,A2);
b3     =b1//b2;
call linmod(analyze3,y2est3,y,X,Phi,A3,b3);
print analyze3 /;

print "Model 4: Paid and Incurred Losses have same Ultimate";
G      =i(8)@{1 1 1 1 1 1 1 1};
G      =G||(-G);
Phi    =Phi-Phi*G`*inv(G*Phi*G`)*G*Phi;
A4     =A3//{[8]  1 [8] -1}//      (G*X)      ;
b4     =b3//       {0}       //{0,0,0,0,0,0,0,0};
call linmod(analyze4,y2est4,y,X,Phi,A4,b4);
print analyze4 /;

print "Model 5: Paid Losses with Ultimate Quasi-Observation";
A5     ={1 1 1 1 1 1 1 0};
b5     =7.2129233260;
yquasi =ypaid              //(b5/9)           ;
X      =(exposure[,2]@i(8))//{0 0 0 0 0 0 0 1};
s2ratio=0.212750769/6271655116;
Phi    =block(  i(64)       ,      s2ratio     );
call linmod(analyze5,y2est5,yquasi,X,Phi,A5,b5);
```

```
print analyze5 /;

/***Predictions***/
unpd1   =y2est1[ 1:36,2];
ibnr2   =y2est2[ 1:36,2];
unpd3   =y2est3[ 1:36,2];
ibnr3   =y2est3[37:72,2];
unpd4   =y2est4[ 1:36,2];
ibnr4   =y2est4[37:72,2];
unpd5   =y2est5[ 1:36,2];

/***Variances of Prediction Errors***/
varunpd1=y2est1[ 1:36, 3:38];
varibnr2=y2est2[ 1:36, 3:38];
varunpd3=y2est3[ 1:36, 3:38];
varibnr3=y2est3[37:72,39:74];
varunpd4=y2est4[ 1:36, 3:38];
varibnr4=y2est4[37:72,39:74];
varunpd5=y2est5[ 1:36, 3:38];

fysums=shape(0,10,36);
fysums[ 1, 1:1 ]=1;          /* FY 1988      */
fysums[ 2, 2:3 ]=1;          /* FY 1989      */
fysums[ 3, 4:6 ]=1;          /* FY 1990      */
fysums[ 4, 7:10]=1;          /* FY 1991      */
fysums[ 5,11:15]=1;          /* FY 1992      */
fysums[ 6,16:21]=1;          /* FY 1993      */
fysums[ 7,22:28]=1;          /* FY 1994      */
fysums[ 8,29:36]=1;          /* FY 1995      */
fysums[ 9, 1:28]=1;          /* FY 1988-1994 */
fysums[10, 1:36]=1;          /* FY 1988-1995 */

/***Variances by Fund Year***/
fyvunpd1=vecdiag(fysums*varunpd1*fysums`);
fyvibnr2=vecdiag(fysums*varibnr2*fysums`);
fyvunpd3=vecdiag(fysums*varunpd3*fysums`);
fyvibnr3=vecdiag(fysums*varibnr3*fysums`);
fyvunpd4=vecdiag(fysums*varunpd4*fysums`);
fyvibnr4=vecdiag(fysums*varibnr4*fysums`);
fyvunpd5=vecdiag(fysums*varunpd5*fysums`);

rows={'1988', '1989', '1990', '1991', '1992', '1993',
      '1994', '1995', '1988-1994', '1988-1995'};
cols={'@12' '@24' '@36' '@48' '@60' '@72' '@84' '@108'};

triunpd1=shape(.,10,8);
```

```
triunpd1[y2est1[,1]]=unpd1;
genvar=genvar(varunpd1);
print "Conjoint Prediction of Paid and Incurred Losses: Exhibit 20";
print triunpd1 [format=comma9.0 colname=cols rowname=rows]
      fyvunpd1 [format=e10.3];
print genvar [format=e10.3] /;

triibnr2=shape(.,10,8);
triibnr2[y2est2[,1]]=ibnr2;
genvar=genvar(varibnr2);
print "Conjoint Prediction of Paid and Incurred Losses: Exhibit 22";
print triibnr2 [format=comma9.0 colname=cols rowname=rows]
      fyvibnr2 [format=e10.3];
print genvar [format=e10.3] /;

triunpd4=shape(.,10,8);
triunpd4[(y2est4[,1])[1:36]]=unpd4;
genvar=genvar(varunpd4);
print "Conjoint Prediction of Paid and Incurred Losses: Exhibit 25";
print triunpd4 [format=comma9.0 colname=cols rowname=rows]
      fyvunpd4 [format=e10.3];
print genvar [format=e10.3] /;

triibnr4=shape(.,10,8);
triibnr4[(y2est4[,1])[1:36]]=ibnr4;
genvar=genvar(varibnr4);
print "Conjoint Prediction of Paid and Incurred Losses: Exhibit 25";
print triibnr4 [format=comma9.0 colname=cols rowname=rows]
      fyvibnr4 [format=e10.3];
print genvar [format=e10.3] /;

triunpd5=shape(.,10,8);
triunpd5[y2est5[,1]]=unpd5;
genvar=genvar(varunpd5);
print "Conjoint Prediction of Paid and Incurred Losses: Exhibit 31";
print triunpd5 [format=comma9.0 colname=cols rowname=rows]
      fyvunpd5 [format=e10.3];
print genvar [format=e10.3] /;


/***How to present value the predictions***/

maturity={ 0.5  1.5  2.5  3.5  4.5  5.5  6.5  7.5  8.5};
yield   ={6.03 6.36 6.84 6.99 7.04 7.14 7.15 7.21 7.21};
dscfac  =(1+yield/100)##-maturity;
```

```
time0   =(1995+    0/12)       ;
time    =(fyid+ageid/12)-time0;
time    =time[loc(time>0)]    ;
pv      =diag(dscfac[time])   ;

/*** An example of discounting***/;
dscunpd4=pv*   unpd4    ;
dvarupd4=pv*varunpd4*pv`;
fydvupd4=vecdiag(fysums*dvarupd4*fysums`);
tridupd4=shape(.,10,8);
tridupd4[(y2est4[,1])[1:36]]=dscunpd4;
genvar=genvar(dvarupd4);
print "Conjoint Prediction of Paid and Incurred Losses: Exhibit 32";
print tridupd4 [format=comma9.0 colname=cols rowname=rows]
      fydvupd4 [format=e10.3];
print genvar [format=e10.3] /;
```

Appendix H

Conjoint Prediction and the Minimum Bias Method

One of the reviewers of this paper, Al Weller, FCAS, commented, "The paper presents a special case of simultaneous estimation and should therefore include the Bailey-Simon paper among its references." Both the Bailey-Simon paper [1] (references, notes, and exhibits follow) and the later Bailey [2] paper showed how to determine class differentials simultaneously over two or more dimensions so that the resulting rates have "minimum bias." Recently, Robert Brown [3] elaborated on these papers, and Gary Venter [5] contributed much to the subject in his discussion of Brown's paper. Although the reviewer did not ask for a comparison of conjoint prediction with minimum bias ratemaking, the author believes that a brief comparison will be valuable.

The comparison will use the simple example (Exhibit 1) of a (3×2) matrix **Y** of pure premiums. The row dimension (i) could represent type of driver, and the column dimension (j) could represent territory. For simplicity it is assumed that the variances of all six cells are equal. This would normally mean that the exposures underlying the cells are equal. The pure premiums were simulated by the additive formula $y_{ij} = i + j + e_{ij}$, where the e_{ij}s are independent standard normal variables.

The most obvious approach, called "Two-Way ANOVA" (Analysis of Variance) in the exhibit, is to calculate the column means $y_{.j}$ (2.3484 4.4835) and the row means $y_{i.}$ (1.2621

4.3618 4.6239). The grand mean $y_{..}$ is 3.4160. A prediction of y_{ij} is then $y_{i.} + y_{.j} - y_{..}$. In a rating manual the pure premiums would most naturally be expressed as the grand mean plus row and column surcharges: $y_{ij} = y_{..} + (y_{i.} - y_{..}) + (y_{.j} - y_{..})$. The average surcharge is zero.

In this additive example the row and column averages are calculated separately from each other. Bailey's idea was to calculate them simultaneously. Having assumed starting values, one cycles through the dimensions solving a balance equation for the values of one dimension in terms of the values of the others. These cycles are continued until convergence is achieved (hopefully). This is illustrated in the second part of the exhibit. Convergence is achieved after three iterations. The grand mean will count in the dimension of the first iteration only, in this case the row dimension $y_{i.}$. Since the prediction of Bailey's method is the same as the ANOVA prediction, what is gained by Bailey's method? In this simple example nothing; but the method can be valuable to less simple problems.

Also, we can consider this as a least-squares problem. Representing the row dimension as α and the column dimension as β, we can model y_{ij} as $\alpha_i + \beta_j + e_{ij}$. Assuming the e_{ij}s to be non-covarying and of equal variance, we can find the αs and βs which minimize the function:

$$f = \sum_i \sum_j \left(y_{ij} - \alpha_i - \beta_j \right)^2$$

The function is minimized for αs and βs which set the first derivatives to zero:

$$\frac{\partial f}{\partial \alpha_i} = \sum_j 2\left(y_{ij} - \alpha_i - \beta_j\right)(-1)$$

$$\therefore 0 = \sum_j \left(y_{ij} - \alpha_i - \beta_j\right)$$

$$\sum_j \alpha_i = \sum_j \left(y_{ij} - \beta_j\right)$$

$$\alpha_i = \frac{\sum_j \left(y_{ij} - \beta_j\right)}{\sum_j 1}$$

and similarly for β_j. But the equations for the αs and βs are the same as Bailey's balance equations. Therefore, Bailey's minimum bias method, at least for this example, is really an iterative method of solving a least-squares problem. It can be shown that in general his method is equivalent to a weighted least-squares problem which may be non-linear in the αs and βs. But it was mentioned in Appendix C that the best linear unbiased estimator of β in the linear model $\mathbf{y} = X\beta + \mathbf{e}$, where $\mathrm{Var}[\mathbf{e}] = \Sigma$, minimizes the least-squares equation $f(\beta) = (y - X\beta)'\Sigma^{-1}(y - X\beta)$. Therefore, Bailey's linear models can be expressed as linear statistical models.[1]

Exhibit 2 shows the simple additive example as a statistical model. If the model were fully expressed, it would appear as:

$$\begin{bmatrix} 0.9122 \\ 1.6120 \\ 3.4770 \\ 5.2467 \\ 2.6560 \\ 6.5918 \end{bmatrix} = \mathbf{y} = X\beta + \mathbf{e} = \begin{bmatrix} 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \beta_1 \\ \beta_2 \end{bmatrix} + \mathbf{e}$$

In this form the model is obviously a two-way ANOVA problem.[2] However, the solution of the parameters is not unique, since one could obtain another solution simply by subtracting any amount δ from each α and adding it to each β. Therefore, in Exhibit 2 α_1 was arbitrarily set to zero, which eliminates the first column from the (6×5) design matrix X and allows for X'X to be non-singular.[3] The exhibit then estimates the parameters of the model, and arrives at the same predictions as did Exhibit 1.

The minimum bias method, therefore, makes an *n*-dimensional ANOVA model out of *n* 1-dimensional models. The *n*-dimensionality shows up in the design matrix. But "minimum bias" is a misnomer. The estimates of an ANOVA model are best linear unbiased estimates (BLUE). It is not that they are of ***minimum*** bias; rather, they are *un*biased. In certain situations, an *n*-dimensional model has more explanatory power than do *n* 1-dimensional models; but this is not to say that the *n*-dimensional model is less biased. Nor is it momentous to say that the dimensions of the *n*-dimensional model are estimated simultaneously. It is true that whereas they had been treated as *n* separate β vectors in *n* models, now they are assumed into one great β vector in the composite model. But to say that the dimensions are estimated simultaneously is of no greater moment than to say of any vector operation that the elements are calculated simultaneously. What really matters is that lesser models are combined into a greater model, and that by way of the design matrix.

The conjoint prediction of this paper combines lesser models, but in a different way. One can place matrix brackets around linear statistical models and treat them as one grand model, and yet the models can be unrelated. Consider the following example:

$$\underbrace{\begin{bmatrix} \mathbf{y}_{1\,(t_1\times 1)} \\ \mathbf{y}_{2\,(t_2\times 1)} \end{bmatrix} = \begin{bmatrix} X_{1\,(t_1\times k_1)} & \\ & X_{2\,(t_2\times k_2)} \end{bmatrix} \begin{bmatrix} \beta_{1\,(k_1\times 1)} \\ \beta_{2\,(k_2\times 1)} \end{bmatrix} + \begin{bmatrix} \mathbf{e}_{1\,(t_1\times 1)} \\ \mathbf{e}_{2\,(t_2\times 1)} \end{bmatrix}}$$

$$\mathbf{y}_{(t\times 1)} = X_{(t\times k)}\beta_{(k\times 1)} + \mathbf{e}_{(t\times 1)}$$

This may look like an imposing grand model of $t = t_1 + t_2$ observations and $k = k_1 + k_2$ parameters.[4] But if $\mathbf{e}_1$ does not covary with $\mathbf{e}_2$, i.e., if $\operatorname{Var}\begin{bmatrix} \mathbf{e}_1 \\ \mathbf{e}_2 \end{bmatrix} = \begin{bmatrix} \Sigma_{11} & 0 \\ 0 & \Sigma_{22} \end{bmatrix} = \Sigma$, then, according to Appendix C:

$$\begin{aligned}
\hat{\beta} &= (X'\Sigma X)^{-1} X'\Sigma \mathbf{y} \\
&= \left(\begin{bmatrix} X_1' & \\ & X_2' \end{bmatrix} \begin{bmatrix} \Sigma_{11} & \\ & \Sigma_{22} \end{bmatrix}^{-1} \begin{bmatrix} X_1 & \\ & X_2 \end{bmatrix} \right)^{-1} \begin{bmatrix} X_1' & \\ & X_2' \end{bmatrix} \begin{bmatrix} \Sigma_{11} & \\ & \Sigma_{22} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix} \\
&= \left(\begin{bmatrix} X_1' & \\ & X_2' \end{bmatrix} \begin{bmatrix} \Sigma_{11}^{-1} & \\ & \Sigma_{22}^{-1} \end{bmatrix} \begin{bmatrix} X_1 & \\ & X_2 \end{bmatrix} \right)^{-1} \begin{bmatrix} X_1' & \\ & X_2' \end{bmatrix} \begin{bmatrix} \Sigma_{11}^{-1} & \\ & \Sigma_{22}^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix} \\
&= \left(\begin{bmatrix} X_1'\Sigma_{11}^{-1}X_1 & \\ & X_2'\Sigma_{22}^{-1}X_2 \end{bmatrix} \right)^{-1} \begin{bmatrix} X_1'\Sigma_{11}^{-1}\mathbf{y}_1 \\ X_2'\Sigma_{22}^{-1}\mathbf{y}_2 \end{bmatrix} \\
&= \begin{bmatrix} \left(X_1'\Sigma_{11}^{-1}X_1\right)^{-1} & \\ & \left(X_2'\Sigma_{22}^{-1}X_2\right)^{-1} \end{bmatrix} \begin{bmatrix} X_1'\Sigma_{11}^{-1}\mathbf{y}_1 \\ X_2'\Sigma_{22}^{-1}\mathbf{y}_2 \end{bmatrix} \\
&= \begin{bmatrix} \left(X_1'\Sigma_{11}^{-1}X_1\right)^{-1} X_1'\Sigma_{11}^{-1}\mathbf{y}_1 \\ \left(X_2'\Sigma_{22}^{-1}X_2\right)^{-1} X_2'\Sigma_{22}^{-1}\mathbf{y}_2 \end{bmatrix} \\
&= \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix}
\end{aligned}$$

This means that the grand model produces the same estimates as do the lesser models, but at greater expense due to its larger (and sparse) matrices.

But conjoint prediction of paid and incurred losses requires that the ultimate paid and incurred losses of each exposure period be equal. This causes the paid observations to covary with the incurred, i.e., $\Sigma_{12} = \Sigma_{12}' \neq 0$. So conjoint prediction combines models not by way of the design matrix (which is block diagonal), but rather by way of the variance of the error term. Because one normally assumes this variance to be diagonal, or block diagonal, statisticians refer to models like conjoint prediction as *seemingly* unrelated models (cf. Judge [4: Chapter11]). The combination of models in conjoint prediction is more subtle than that of class-differential models.

Notes to Appendix H

[1] Also, Bailey's non-linear models can be expressed as non-linear statistical models. The beginning of Appendix C mentions how to generalize from linear to non-linear models.

[2] Venter [5:340] notes that at least one author has seen the connection between Bailey's method and ANOVA. His reference is to Chamberlain, C., "Relativity Pricing through Analysis of Variance," 1980 Discussion Paper Program, Casualty Actuarial Society.

[3] Venter [5:338] mentions that fully expressed ANOVA models in two or more dimensions are overspecified. As discussed in Appendix C, this means that the design matrix X is not of full column rank. The simplest solution is to set some parameters to arbitrary values. Another solution, which is more complicated, is to introduce an intercept to the model by adding a column of 1s to the design matrix, and then to constrain the sums of the parameters by dimension to be zero.

[4] The notation here differs from that of Appendix C. Here the subscripts 1 and 2 refer to models 1 and 2, and the grand model considers only observations. In Appendix C the subscript 1 refers to observations and the subscript 2 refers to predictions.

References of Appendix H

1. Bailey, Robert A., and Simon, LeRoy J., "Two Studies in Automobile Insurance," *PCAS* XLVII, 1960.

2. Bailey, Robert A., "Insurance Rates with Minimum Bias," *PCAS* L, 1963.

3. Brown, Robert L., "Minimum Bias with Generalized Linear Models," *PCAS* LXXV, 1988, 187-207.

4. Judge, George G., Hill, R. C., *et al.*, Introduction to the Theory and Practice of Econometrics (Second Edition), New York, John Wiley & Sons, 1988.

5. Venter, Gary G., "Discussion of Paper Published in Volume LXXV: Minimum Bias with Generalized Linear Models, by Robert L. Brown," *PCAS* LXXVII, 1990, 337-349.

Appendix H

Exhibit 1

Two-Way ANOVA

$i \setminus j$	1	2	$\mathbf{y}_{i.}$
1	0.9122	1.6120	1.2621
2	3.4770	5.2467	4.3618
3	2.6560	6.5918	4.6239
$\mathbf{y}_{.j}$	2.3484	4.4835	3.4160

Prediction	1	2
1	0.1945	2.3297
2	3.2943	5.4294
3	3.5563	5.6915

Bailey's Iterative Solution

First Iteration

$i \setminus j$	1	2	$\mathbf{y}_{i.}$
1	0.9122	1.6120	1.2621
2	3.4770	5.2467	4.3618
3	2.6560	6.5918	4.6239
$\mathbf{y}_{.j}$	0.0000	0.0000	

Second Iteration

$i \setminus j$	1	2	$\mathbf{y}_{i.}$
1	0.9122	1.6120	1.2621
2	3.4770	5.2467	4.3618
3	2.6560	6.5918	4.6239
$\mathbf{y}_{.j}$	-1.0676	1.0676	

Third Iteration

$i \setminus j$	1	2	$\mathbf{y}_{i.}$
1	0.9122	1.6120	1.2621
2	3.4770	5.2467	4.3618
3	2.6560	6.5918	4.6239
$\mathbf{y}_{.j}$	-1.0676	1.0676	

Prediction	1	2
1	0.1945	2.3297
2	3.2943	5.4294
3	3.5563	5.6915

Appendix H

Exhibit 2

Equivalent Linear Model

y

0.9122
1.6120
3.4770
5.2467
2.6560
6.5918

X

0	0	1	0
0	0	0	1
1	0	1	0
1	0	0	1
0	1	1	0
0	1	0	1

$e = y - X\beta$

0.7176
-0.7176
0.1827
-0.1827
-0.9004
0.9004

X'y

8.7237
9.2478
7.0451
13.4506

X'X

2	0	1	1
0	2	1	1
1	1	3	0
1	1	0	3

t	6
k	4
df	2
$\sigma^2 = e'e/df$	1.3591

$\beta = (X'X)^{-1}X'y$

3.0998
3.3618
0.1945
2.3297

$(X'X)^{-1}$

1.0000	0.5000	-0.5000	-0.5000
0.5000	1.0000	-0.5000	-0.5000
-0.5000	-0.5000	0.6667	0.3333
-0.5000	-0.5000	0.3333	0.6667

$Var[\beta] = \sigma^2(X'X)^{-1}$

1.3591	0.6795	-0.6795	-0.6795
0.6795	1.3591	-0.6795	-0.6795
-0.6795	-0.6795	0.9060	0.4530
-0.6795	-0.6795	0.4530	0.9060

Prediction	1	2	
1	0.1945	2.3297	0.0000
2	3.2943	5.4294	3.0998
3	3.5563	5.6915	3.3618
	0.1945	2.3297	