

# Modeling Paid and Incurred Losses Together

Leigh J. Halliwell, FCAS, MAAA

---

## Abstract

The modeling skills of actuaries and academicians have developed to the point of their seeking joint models for paid and incurred losses, i.e., models in which paid and incurred losses will inform each other so that their confidence intervals will narrow and the two sets of ultimate losses will be equal. The key to such models is covariance; heteroskedastic models cannot serve the purpose. Properly accounting for covariance in the linear statistical model will provide an exact, sound, and elegant solution to the problem. Moreover, covariance is what distinguishes the same information from like information, and prevents the creation of information out of nothing.

**Key concepts:** linear statistical model, paid and incurred losses, seemingly unrelated regression (SUR), covariance, variance structure

---

## 1. INTRODUCTION

For setting the loss reserves of most casualty lines of insurance, actuaries must turn loss triangles into rectangles. Until the mid-1990s this was largely a deterministic exercise, which involved selecting development factors – perhaps with certain adjustments and sensitivity testing. Since then, business needs have demanded, and advances in theory and computing have allowed, probabilistic modeling of loss triangles. Deterministic methods are waning as actuaries are increasingly asked to estimate statistical properties of loss reserves, especially their probability distributions. However, attempts to answer this need are hampered by the duality (sometimes the multiplicity) of triangles for the same line of business. Triangles usually come in pairs, one of paid losses and another of (case-) incurred losses, which ultimately must reach equality. But when paid and incurred losses are modeled separately, any equality is accidental, and even then devoid of jointly statistical properties. A crisis of models, however artful, is not science, despite appeals to actuarial judgment.<sup>1</sup>

According to Gary Venter [2008, 348], “Formal modeling of paid and incurred simultaneously appears to have begun with Halliwell.” But the idea received scant attention until a paper by Quarg and Mack [2004 and 2008], which spurred the Venter paper, as well as a yet unpublished paper by Zhang and Clark [2009]. Unlike the Quarg and Mack approach, which “reduces the gap” between the projections, the Halliwell [1997] approach offered an exact

---

<sup>1</sup> Actuarial judgment functions *within* actuarial science, not to the transcendence or out-guessing of science. “Actuarial judgment is no antidote ..., as if actuaries possessed some expertise or intuition to herd or prod methods into correctness. ... Actuaries must not presume to judge what they cannot scientifically model.” [Halliwell, 2007, footnote 5]

solution. Moreover, Quarg and Mack sought the solution in the design matrix of their model, whereas the variance structure was the key for Halliwell. We believe that Halliwell was on the right track, but with some flaws. Our solution to the problem of modeling paid and incurred losses together improves on his variance approach, correcting its flaws. Furthermore, it is easier to understand, simpler to program, as well as theoretically streamlined and elegant.

In the next section we will introduce and comment upon Halliwell's version of the linear statistical model. In Section 3, while discussing basic properties of the model, we will introduce the distinction between statistical sameness and statistical similarity, which arises from differing variance structures and which ensures that information cannot be created *ex nihilo*. Section 4 will apply this theory to a simple example of a joint model of paid and incurred loss triangles, and Section 5 will show a more elaborate application to industry Workers' Compensation losses. Appendix A will deepen insights into the ideas of Section 3 by treating simultaneous equations as a subset of the linear statistical model. Finally, Appendix B will elaborate on permissible variance structures, viz., how with two random variables of known correlation a third may be correlated.

## 2. GENERAL FORMULATION OF THE LINEAR STATISTICAL MODEL

Many writers present versions of the linear statistical model; however, the version found in Halliwell [1997], despite its initial complexity, is most versatile and general. Moreover, most presentations stop at the estimation of the parameter  $\beta$ . But in his version this is just an intermediate step; the focus is on predictions based on  $\beta$  and their prediction-error variances. The basic form of the linear model is  $\mathbf{y}_{t \times 1} = X_{t \times k} \beta_{k \times 1} + \mathbf{e}_{t \times 1}$ . It is the error term  $\mathbf{e}$  that makes the model statistical (otherwise called probabilistic and stochastic); it is a random vector whose moments are:  $E[\mathbf{e}] = \mathbf{0}_{t \times 1}$ ,  $Var[\mathbf{e}] = \Sigma_{t \times t} = \sigma^2 \Phi_{t \times t}$ . For those unfamiliar with multivariate means and variances, especially with the quadratic form  $Var[\mathbf{Ae}] = \mathbf{A}Var[\mathbf{e}]\mathbf{A}'$  and with non-negative definite and positive definite matrices, the reader is advised to read Halliwell [1997; Appendix A], Healy [1986; Chapter 7], and Judge [1988, Appendix A].

The matrix  $X$  is known as the design, or regressor, matrix; its columns are called regressor variables and independent variables. The vector  $\mathbf{y}$  is called the response variable and the dependent variable, and  $\beta$  is the parameter of the model. But in this formulation, which emphasizes predictions rather than parameters, the  $t$  rows of the model are partitioned into  $t_1$  observations and  $t_2$  predictions. In matrix-partitioned form it is expressed:

$$\begin{bmatrix} \mathbf{y}_{1(t_1 \times 1)} \\ \mathbf{y}_{2(t_2 \times 1)} \end{bmatrix} = \begin{bmatrix} X_{1(t_1 \times k)} \\ X_{2(t_2 \times k)} \end{bmatrix} \boldsymbol{\beta}_{k \times 1} + \begin{bmatrix} \mathbf{e}_{1(t_1 \times 1)} \\ \mathbf{e}_{2(t_2 \times 1)} \end{bmatrix},$$

where:

$$E \begin{bmatrix} \mathbf{e}_{1(t_1 \times 1)} \\ \mathbf{e}_{2(t_2 \times 1)} \end{bmatrix} = \begin{bmatrix} \mathbf{0}_{t_1 \times 1} \\ \mathbf{0}_{t_2 \times 1} \end{bmatrix}, \quad \text{Var} \begin{bmatrix} \mathbf{e}_{1(t_1 \times 1)} \\ \mathbf{e}_{2(t_2 \times 1)} \end{bmatrix} = \begin{bmatrix} \Sigma_{11(t_1 \times t_1)} & \Sigma_{12(t_1 \times t_2)} \\ \Sigma_{21(t_2 \times t_1)} & \Sigma_{22(t_2 \times t_2)} \end{bmatrix} = \sigma^2 \begin{bmatrix} \Phi_{11(t_1 \times t_1)} & \Phi_{12(t_1 \times t_2)} \\ \Phi_{21(t_2 \times t_1)} & \Phi_{22(t_2 \times t_2)} \end{bmatrix}.$$

The noun ‘observations’ is imprecise. What distinguishes observations from predictions is that predictions are missing, or blank, elements of  $\mathbf{y}$  (which signals that predictions are desired), whereas “observations” are non-missing, real-valued elements. Usually, they happen to come from observation, such as with loss amounts. However, the key to a joint model of paid and incurred losses is what we will call “tautologous observations,” i.e., elements of  $\mathbf{y}$  that contain zeroes as the differences by exposure period of incurred ultimate losses from paid ultimate losses. We do not actually need to observe their ultimate equality to know that it will obtain; this is information that we know *a priori* and of which we should make use. The observed and predicted elements can appear in any order; the clearest presentation may not place all the observations in rows above those of the predictions; however, our software will reorder them. Of course, since the variance structure is symmetric, the columns of  $\text{Var}[\mathbf{e}]$  must likewise be reordered.

The known, or specified, elements of the linear model are the entire<sup>2</sup> design matrix  $X$ , the entire variance structure, whether in absolute form  $\Sigma$  or in relative form  $\Phi$ , and the observed elements of  $\mathbf{y}$ . The modeler desires an estimate of  $\mathbf{y}_2$ , viz.,  $\hat{\mathbf{y}}_2$ . However, that estimate will turn out to be in error by the vector  $\mathbf{y}_2 - \hat{\mathbf{y}}_2$ , which we will call the prediction error. The formulæ for an estimator of  $\mathbf{y}_2$  and the variance of its prediction error, viz.,  $\hat{\mathbf{y}}_2$  and  $\text{Var}[\mathbf{y}_2 - \hat{\mathbf{y}}_2]$ , are:

$$\hat{\mathbf{y}}_2 = X_2 \hat{\boldsymbol{\beta}} + \Phi_{21} \Phi_{11}^{-1} (\mathbf{y}_1 - X_1 \hat{\boldsymbol{\beta}})$$

$$\text{Var}[\mathbf{y}_2 - \hat{\mathbf{y}}_2] = (X_2 - \Phi_{21} \Phi_{11}^{-1} X_1) \text{Var}[\hat{\boldsymbol{\beta}}] (X_2 - \Phi_{21} \Phi_{11}^{-1} X_1)' + \sigma^2 (\Phi_{22} - \Phi_{21} \Phi_{11}^{-1} \Phi_{12}),$$

where  $\hat{\boldsymbol{\beta}} = (X_1' \Phi_{11}^{-1} X_1)^{-1} X_1' \Phi_{11}^{-1} \mathbf{y}_1$  and  $\text{Var}[\hat{\boldsymbol{\beta}}] = \sigma^2 (X_1' \Phi_{11}^{-1} X_1)^{-1}$ . Or one may use the absolute-variance form, in which ‘ $\sigma^2$ ’ is omitted and ‘ $\Sigma$ ’ replaces ‘ $\Phi$ ’. This estimator algebraically

---

<sup>2</sup> This conflicts with many linear models of loss triangles, in particular with the Quarg/Mack [2004 and 2008] model, whose predictions two or more periods into the future depend on predictions of the previous periods. The feedback loop ‘predictions → regressors → predictions’, which Judge [1988; Chapter 13] calls “Stochastic Regressors,” is undesirable for theoretical, numerical-analytic, and aesthetic reasons, some of which Halliwell [2007; Section 5] discusses.

reduces to a linear function of  $\mathbf{y}_1$ , and Halliwell [1997; Appendix C] gives a version of the Gauss-Markov theorem in proof that  $\hat{\mathbf{y}}_2$  is the best linear unbiased estimator (BLUE) of  $\mathbf{y}_2$ .<sup>3</sup>

A few minor conditions need to be made explicit. First, the variance structure ( $\Sigma$  or  $\Phi$ ) must be non-negative definite. Otherwise some random variable consisting of a linear combination of the elements of  $\mathbf{e}$  would have a negative variance. At the very least this implies that that none of the diagonal elements of the variance structure is negative. Second,  $\Sigma_{11}$  or  $\Phi_{11}$  must be positive definite. Being a block-diagonal part of the variance structure, it must be non-negative definite. However, a positive definite  $\Sigma_{11}$  or  $\Phi_{11}$  has no variance degeneracy, which guarantees the existence of its inverse. Third,  $X_1$  must be of full column rank, i.e.,  $\text{rank}(X_1) = k$ . The second and third conditions together guarantee that  $(X_1' \Phi_{11}^{-1} X_1)^{-1}$  exists.

Usually the variance structure is known only to within a scale factor; most models posit relative, not absolute, variances. In that case one must estimate  $\sigma^2$  as  $\hat{\sigma}^2 = (\mathbf{y}_1 - X_1 \hat{\beta})' \Phi_{11}^{-1} (\mathbf{y}_1 - X_1 \hat{\beta}) / (t_1 - k)$ , a matrix-weighted “sum of squared residuals” divided by the degrees of freedom. Our software will display a  $3 \times 3$  matrix titled “SSCP,” which stands for “(matrix-weighted) sums of squares and cross products.” Define the  $t_1 \times 3$  partitioned matrix  $V = \begin{bmatrix} \mathbf{y}_1 & | & X_1 \hat{\beta} & | & \mathbf{y}_1 - X_1 \hat{\beta} \end{bmatrix}$ . Then  $\text{SSCP} = V' \Phi_{11}^{-1} V$ , and  $\hat{\sigma}^2 = \text{SSCP}_{33} / (t_1 - k)$ . Theorems of the linear model state that  $\text{SSCP}_{11} = \text{SSCP}_{22} + \text{SSCP}_{33}$  and  $\text{SSCP}_{32} = \text{SSCP}_{23} = 0$ .  $\text{SSCP}_{22} / \text{SSCP}_{11}$  is the portion of the observations that the model “explains.”<sup>4</sup>

### 3. BASIC PROPERTIES OF THE LINEAR STATISTICAL MODEL

Before introducing the joint model of paid and incurred losses we must discuss some basic properties of the model and its predictions. First consider the model:

$$\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix} = \begin{bmatrix} X_1 \\ I_{k \times k} \end{bmatrix} \beta + \begin{bmatrix} \mathbf{e}_1 \\ 0 \end{bmatrix}, \quad \text{Var} \begin{bmatrix} \mathbf{e}_1 \\ 0 \end{bmatrix} = \begin{bmatrix} \Sigma_{11} & | & 0 \\ 0 & | & 0 \end{bmatrix}$$

One here wants just to predict the parameter ( $\mathbf{y}_2 = \beta$ ), so this is a check of the parameter of

---

<sup>3</sup> This estimator is unbiased in that  $E[\hat{\mathbf{y}}_2] = \mathbf{y}_2$ . There are infinitely many linear-in- $\mathbf{y}_1$ , unbiased estimators (LUE) of  $\mathbf{y}_2$ . But according to the Gauss-Markov theorem, none of them is as good as  $\hat{\mathbf{y}}_2$ ;  $\hat{\mathbf{y}}_2$  is the best (B). This means that its prediction-error variance is less than theirs, in the sense that the difference of its prediction-error variance from theirs is positive definite. To the philosophically inclined it is amazing for an estimator to exist that positive-definitely dominates in the linear unbiased universe. It is no less amazing that this BLUE estimator is identical to the maximum-likelihood estimator under the assumption that the error terms are multivariate-normally distributed – an assumption not necessary to the linear statistical model. Such feelings of amazement incline most mathematicians toward a Platonic belief that mathematics is discovered, rather than toward a formalist belief that it is invented. Then again, in this abstruse realm what might be the difference between discovery and invention?

<sup>4</sup> Though we will call this a rho-square statistic, it differs from the commonly defined statistic that was devised for regression models containing an intercept and that strips away the explanatory power of the intercept. Our definition is appropriate to our intercept-free models.

the model. Applying the formulæ, we confirm:

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

In fact, Halliwell [1997; 331] began with a “wordier” version of the solution and derived in this manner the streamlined version of the solution that employs  $\hat{\boldsymbol{\beta}}$ .

The second model is:

$$\begin{bmatrix} \frac{A\mathbf{y}_1}{\mathbf{y}_2} \end{bmatrix} = \begin{bmatrix} \frac{AX_1}{X_2} \end{bmatrix} \boldsymbol{\beta} + \begin{bmatrix} \frac{A\mathbf{e}_1}{\mathbf{e}_2} \end{bmatrix}, \quad \text{Var} \begin{bmatrix} \frac{A\mathbf{e}_1}{\mathbf{e}_2} \end{bmatrix} = \begin{bmatrix} \frac{A\Sigma_{11}A'}{\Sigma_{21}A'} \mid \frac{A\Sigma_{12}}{\Sigma_{22}} \end{bmatrix}$$

The observed part of this model has been transformed by matrix A. The transformation affects even the error term,  $\mathbf{e}_1 \rightarrow A\mathbf{e}_1$ , and one purpose of this exercise is to sensitize the reader to the variance structure. For in general,  $\text{Cov}[A\mathbf{e}_1, B\mathbf{e}_2] = A\text{Cov}[\mathbf{e}_1, \mathbf{e}_2]B'$ . If A is a nonsingular, or invertible, matrix:

$$\begin{aligned}\hat{\boldsymbol{\beta}} &= \left( (AX_1)' (A\Sigma_{11}A')^{-1} (AX_1) \right)^{-1} (AX_1)' (A\Sigma_{11}A')^{-1} (A\mathbf{y}_1) \\ &= (X_1' A' A^{-1} \Sigma_{11}^{-1} A^{-1} AX_1)^{-1} X_1' A' A^{-1} \Sigma_{11}^{-1} A^{-1} A \mathbf{y}_1 \\ &= (X_1' \Sigma_{11}^{-1} X_1)^{-1} X_1' \Sigma_{11}^{-1} \mathbf{y}_1 \\ &= \text{Var}[\hat{\boldsymbol{\beta}}] X_1' \Sigma_{11}^{-1} \mathbf{y}_1\end{aligned}$$

At this point we’ve demonstrated that the parameter estimate is unaffected. A similar cancellation of A with its inverse continues into the rest of the prediction, as the reader can verify. Hence, a one-to-one transformation of the observations has no effect on the predictions. A corollary to this is that incremental or cumulative models of loss triangles will yield the same predictions, as long as the variance structure is correctly handled.

Now consider a transformation of the predictions:

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

In this case:

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

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

So the BLUE of a linear combination of predictions is the linear combination of the BLUE of the predictions. Note that here  $B$ , unlike the previous  $A$ , does not have to be invertible. The most common linear combinations of loss-triangle cells are exposure-period subtotals, i.e., unpaid, IBNR, or even ultimate losses. When we care only for these subtotals, as in the Workers' Compensation model of Section 5, this theorem allows us to bypass the extra time and space of cell-by-cell prediction and to predict them directly.

A special pair of models that will illustrate the effect of covariance is the following:

$$\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix} = \begin{bmatrix} X_1 \\ X_1 \end{bmatrix} \beta + \begin{bmatrix} \mathbf{e}_1 \\ \mathbf{e}_2 \end{bmatrix}, \quad \text{Var} \begin{bmatrix} \mathbf{e}_1 \\ \mathbf{e}_2 \end{bmatrix} = \begin{bmatrix} \Sigma_{11} & 0 \\ 0 & \Sigma_{11} \end{bmatrix}$$

Here the prediction is *like* the observation, the solution being:

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

The prediction is *like*, because the variance of  $\mathbf{e}_2$  is like that of  $\mathbf{e}_1$ , ‘like’ in the sense of identically distributed, but nonetheless uncorrelated. But changing the variance structure to

$$\text{Var} \begin{bmatrix} \mathbf{e}_1 \\ \mathbf{e}_2 \end{bmatrix} = \begin{bmatrix} \Sigma_{11} & \Sigma_{11} \\ \Sigma_{11} & \Sigma_{11} \end{bmatrix} \text{ alters the solution drastically:}$$

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

$$\begin{aligned} \text{Var}[\mathbf{y}_2 - \hat{\mathbf{y}}_2] &= (X_1 - \Sigma_{11} \Sigma_{11}^{-1} X_1) \text{Var}[\hat{\boldsymbol{\beta}}] (X_1 - \Sigma_{11} \Sigma_{11}^{-1} X_1)' + (\Sigma_{11} - \Sigma_{11} \Sigma_{11}^{-1} \Sigma_{11}) \\ &= (X_1 - X_1) \text{Var}[\hat{\boldsymbol{\beta}}] (X_1 - X_1)' + (\Sigma_{11} - \Sigma_{11}) \\ &= 0 \end{aligned}$$

This version predicts not something *like*  $\mathbf{y}_1$ , but rather something *identical* to  $\mathbf{y}_1$ . Covariance is the key to distinguishing between likeness and sameness.<sup>5</sup>

And finally, we consider pooling two models. For this purpose and for here only, the subscripts ‘1’ and ‘2’ identify the models (e.g., paid and incurred), rather than observations and predictions (here dismissed with a dot ‘•’). We might be tempted to solve the “super” model:

$$\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \bullet \end{bmatrix} = \begin{bmatrix} X_1 & 0 \\ 0 & X_2 \\ \bullet & \bullet \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{bmatrix} + \begin{bmatrix} \mathbf{e}_1 \\ \mathbf{e}_2 \\ \bullet \end{bmatrix}, \quad \text{Var} \begin{bmatrix} \mathbf{e}_1 \\ \mathbf{e}_2 \\ \bullet \end{bmatrix} = \begin{bmatrix} \Sigma_{11} & 0 & \bullet \\ 0 & \Sigma_{22} & \bullet \\ \bullet & \bullet & \bullet \end{bmatrix}$$

However, the solution for the parameter is:

---

<sup>5</sup> Appendix A shows covariance to be the solution to the “paradox” of why writing data twice does not increase information. It would increase, if it were *like*, or similar, information; just as repeated sampling of independent, identically-distributed random variables increases information. But repetitions of the *same* information are perfectly correlated with the original, and provide nothing new, not even when the repetition is disguised by a linear

transformation. In matrix algebra,  $\text{Var} \begin{bmatrix} \mathbf{e} \\ \mathcal{A}\mathbf{e} \end{bmatrix} = \begin{bmatrix} \Sigma & \Sigma\mathcal{A}' \\ \mathcal{A}\Sigma & \mathcal{A}\Sigma\mathcal{A}' \end{bmatrix} = \begin{bmatrix} I \\ \mathcal{A} \end{bmatrix} \Sigma \begin{bmatrix} I & \mathcal{A}' \end{bmatrix}$ , and the rank of this larger matrix is still equal to the rank of  $\Sigma$ . If the off-block-diagonal elements were zero the rank would increase, depending on  $\mathcal{A}$ , to as much as twice the rank of  $\Sigma$ .

$$\begin{aligned}
 \text{Var} \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} &= \left( \begin{bmatrix} X_1 & 0 \\ 0 & X_2 \end{bmatrix}' \begin{bmatrix} \Sigma_{11} & 0 \\ 0 & \Sigma_{22} \end{bmatrix}^{-1} \begin{bmatrix} X_1 & 0 \\ 0 & X_2 \end{bmatrix} \right)^{-1} \\
 &= \left( \begin{bmatrix} X_1' & 0 \\ 0 & X_2' \end{bmatrix} \begin{bmatrix} \Sigma_{11}^{-1} & 0 \\ 0 & \Sigma_{22}^{-1} \end{bmatrix} \begin{bmatrix} X_1 & 0 \\ 0 & X_2 \end{bmatrix} \right)^{-1} \\
 &= \left( \begin{bmatrix} X_1' \Sigma_{11}^{-1} X_1 & 0 \\ 0 & X_2' \Sigma_{22}^{-1} X_2 \end{bmatrix} \right)^{-1} \\
 &= \begin{bmatrix} (X_1' \Sigma_{11}^{-1} X_1)^{-1} & 0 \\ 0 & (X_2' \Sigma_{22}^{-1} X_2)^{-1} \end{bmatrix} \\
 &= \begin{bmatrix} \text{Var}[\hat{\beta}_1] & 0 \\ 0 & \text{Var}[\hat{\beta}_2] \end{bmatrix} \\
 \\ \\
 \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} &= \text{Var} \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} \left( \begin{bmatrix} X_1 & 0 \\ 0 & X_2 \end{bmatrix}' \begin{bmatrix} \Sigma_{11} & 0 \\ 0 & \Sigma_{22} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix} \right) \\
 &= \begin{bmatrix} \text{Var}[\hat{\beta}_1] & 0 \\ 0 & \text{Var}[\hat{\beta}_2] \end{bmatrix} \begin{bmatrix} X_1' \Sigma_{11}^{-1} \mathbf{y}_1 \\ X_2' \Sigma_{22}^{-1} \mathbf{y}_2 \end{bmatrix} \\
 &= \begin{bmatrix} \text{Var}[\hat{\beta}_1] X_1' \Sigma_{11}^{-1} \mathbf{y}_1 \\ \text{Var}[\hat{\beta}_2] X_2' \Sigma_{22}^{-1} \mathbf{y}_2 \end{bmatrix} \\
 &= \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix}
 \end{aligned}$$

Hence, simply juxtaposing two (or more) models provides no additional information.<sup>6</sup> There is no covariance between the two; they are “frictionless” and “like ships passing in the night.” But if the off-block-diagonal variance were not zero, the combined model would not reduce to the separate submodels. Judge [1988; Section 11.2] calls covariance-linked models “seemingly unrelated regression” (SUR) models, for they seem to be unrelated if one considers only the model design and ignores the variance structure. The Zhang/Clark [2009] model is an SUR model, tying paid and incurred losses together with covariance; but it does not guarantee the equality of ultimate paid and incurred. The Halliwell model [1997] also qualifies as SUR. The model that we will present in the next section is not an SUR one; rather, tautologous equations

---

<sup>6</sup> However, one caveat: The combined model in relative-variance format would invite the modeler to estimate an overall variance scale  $\hat{\sigma}^2$ , which would be an average of the scale estimates of the submodels weighted according to their respective degrees of freedom.



will be the glue between paid and incurred, and the only tricky part will be the effect of these additional equations on the variance structure.

#### 4. TAUTOLOGY AND A SIMPLE JOINT PAID-INCURRED MODEL

The following example comes from Halliwell [1997; Exhibit 1]. It consists of incremental paid and incurred losses for three accident years at three years of development, development being complete at the third year:

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

The AY exposures are equal, and the cells are homoskedastic. The first accident year is mature, and both paid and incurred losses accumulate to 100. The paid and incurred models have the same design matrix (viz., additive, cf. Halliwell [2007; 228]), and the parameter elements are pure premiums by loss type and by age or development period. All cells have the same variance relativity and zero covariance. In symbols, the model of the  $hij^{th}$  cell is  $y_{hij} = 1 \cdot \beta_{hj} + e_{hij}$ ,  $Var[e_{hij}] = \sigma^2 \cdot 1$ , where  $h \in \{1 = \text{Paid}, 2 = \text{Incurred}\}$ ,  $i \in \{1 = \text{AY1}, 2 = \text{AY2}, 3 = \text{AY3}\}$ , and  $j \in \{1 = \text{Age1}, 2 = \text{Age2}, 3 = \text{Age3}\}$ . If the accident years were not of equal exposure, the model would be  $y_{hij} = \xi_i \beta_{hj} + e_{hij}$ ,  $Var[e_{hij}] = \sigma^2 \xi_i$ , where  $\xi_i$  is the exposure of the  $i^{th}$  accident year. In this simple, juxtaposed model one would estimate paid development at ages 2 and 3 as 27.5 and 20, and incurred development likewise as 20 and 10. The paid ultimate losses of AY2 and AY3 would be 95 and 92.5, as compared with incurred ultimate losses of 110 and 80.

Exhibit 1 contains the joint model in the standard form:

$$\begin{bmatrix} \mathbf{y}_{1(14 \times 1)} \\ \mathbf{y}_{2(6 \times 1)} \end{bmatrix} = \begin{bmatrix} \mathbf{X}_{1(14 \times 6)} \\ \mathbf{X}_{2(6 \times 6)} \end{bmatrix} \boldsymbol{\beta}_{6 \times 1} + \begin{bmatrix} \mathbf{e}_{1(14 \times 1)} \\ \mathbf{e}_{2(6 \times 1)} \end{bmatrix}, \quad Var \begin{bmatrix} \mathbf{e}_1 \\ \mathbf{e}_2 \end{bmatrix} = \sigma^2 \begin{bmatrix} \Phi_{11(14 \times 14)} & | & \Phi_{12(14 \times 6)} \\ \hline \Phi_{21(6 \times 14)} & | & \Phi_{22(6 \times 6)} \end{bmatrix}$$

The reader will recognize the  $\mathbf{y}$ ,  $\mathbf{X}$ , and  $\Phi$  matrices within the exhibit, and the dotted lines partition the matrices into observations and predictions. Zeroes are present, but not shown. Except for rows 13 and 14, marked ‘Diff’, and corresponding columns 13 and 14 of  $\Phi$ , the model would consist of two, unrelated paid and incurred homoskedastic submodels. The ‘1’s in the design matrix represent the unitary exposure slotted into the  $hj^{th}$  column of  $\mathbf{X}$  so as to interact

with the  $h_j^{th}$  element of the model parameter  $\beta$  (implicit in the exhibit, but not shown).

But it is the rows marked ‘Diff’ that join, or laminate, the paid and the incurred losses. Although we have not observed six of the eighteen cells, we do know that total paid must equal total incurred by AY. For AY2 the difference is:

$$\begin{aligned} 0 &= \mathbf{y}_{121} + \mathbf{y}_{122} + \mathbf{y}_{123} - \mathbf{y}_{221} - \mathbf{y}_{222} - \mathbf{y}_{223} \\ &= \beta_{11} + \beta_{12} + \beta_{13} - \beta_{21} - \beta_{22} - \beta_{23} + \{e_{121} + e_{122} + e_{123} - e_{221} - e_{222} - e_{223}\} \\ &= [1 \ 1 \ 1 \ -1 \ -1 \ -1]\beta + \{e_{Diff2}\} \end{aligned}$$

So this tautology, or difference equation, is equivalent to a zero observation, the design  $[1 \ 1 \ 1 \ -1 \ -1 \ -1]$ , and error term that involves six other error terms. Because of heteroskedasticity, the variance of this error term is 6. However, because these are *same* error terms, not *like* ones,  $Cov[e_{Diff2}, \pm e_{hij}] = \pm Var[e_{hij}]$ . So the reader should now understand row 13 of the model, and similarly, row 14, the tautology for AY3. We could have added a tautology for AY1; however, its variance structure involves no predictions. It would add no new observation, and the resulting  $15 \times 15$  matrix  $\Phi_{11}$  would still be of rank 14 (see footnote 5) and thus non-invertible.

In Exhibit 2 we solve for the estimates of  $\beta$  and  $\sigma$ , and derive the SSCP matrix, as explained in Section 2. In the spreadsheet we simplified the notation by dropping subscripts (which are all ‘1’, pertaining to observations) and carets. Both paid and incurred total pure premiums equal  $97.91\bar{6}$ , which is the average of the stand-alone total pure premiums of paid  $99.1\bar{6}$  and incurred  $96.\bar{6}$ .

The predictions are estimated in Exhibit 3 according to the formulæ:

$$\hat{\mathbf{y}}_2 = X_2 \hat{\beta} + \Phi_{21} \Phi_{11}^{-1} (\mathbf{y}_1 - X_1 \hat{\beta})$$

$$Var[\mathbf{y}_2 - \hat{\mathbf{y}}_2] = (X_2 - \Phi_{21} \Phi_{11}^{-1} X_1) Var[\hat{\beta}] (X_2 - \Phi_{21} \Phi_{11}^{-1} X_1)' + \sigma^2 (\Phi_{22} - \Phi_{21} \Phi_{11}^{-1} \Phi_{12})$$

To help the reader who wishes to reproduce the results, in the bottom half of the exhibit are intermediate calculations. The prediction and the variance of the prediction error are:

Type	AY	Age	$\hat{\mathbf{y}}_2$	$Var[\mathbf{y}_2 - \hat{\mathbf{y}}_2]$					
Paid	2	3	22.50	79.95	0	39.97	79.95	0	39.97
Paid	3	2	23.75	0	89.94	-29.98	0	29.98	29.98
Paid	3	3	17.50	39.97	-29.98	109.93	39.97	29.98	49.97
Incd	2	3	7.50	79.95	0	39.97	79.95	0	39.97
Incd	3	2	23.75	0	29.98	29.98	0	89.94	-29.98
Incd	3	3	12.50	39.97	29.98	49.97	39.97	-29.98	109.93

In the topmost figure of the exhibit AY subtotals are formed and combined with the paid and incurred to date. For example, the calculation of the prediction-error standard deviation of the

AY3 IBNR is  $11.83 = \sqrt{89.94 - 29.98 - 29.98 + 109.93}$ . Although it may seem a wonder that the means and variances of the ultimate losses are identical whether one builds them from the paid to date or from the incurred, the model was constructed for this purpose. The identity serves only to confirm that the model was solved without mistake.

This technique is more general than the tautology of  $0 = Paid - Incd$ . In addition to the observations  $\mathbf{y}_1$  of the model  $\mathbf{y} = X\beta + \mathbf{e}$ , one may know by means other than observation that  $\mathbf{z}_{t_3 \times 1} = Q_{t_3 \times t} \mathbf{y}_{t \times 1}$ . Hence:

$$\mathbf{z} = Q(X\beta + \mathbf{e}) = QX\beta + Q\mathbf{e}$$

$$Var[\mathbf{z}] = Var[Q\mathbf{e}] = Q\Sigma Q'$$

$$Cov[\mathbf{z}, \mathbf{e}_1] = Cov[Q\mathbf{e}, \mathbf{e}_1] = QCov[\mathbf{e}, \mathbf{e}_1] = QCov\left[\begin{bmatrix} \mathbf{e}_1 \\ \mathbf{e}_2 \end{bmatrix}, \mathbf{e}_1\right] = Q\begin{bmatrix} Cov[\mathbf{e}_1, \mathbf{e}_1] \\ Cov[\mathbf{e}_2, \mathbf{e}_1] \end{bmatrix} = Q\begin{bmatrix} \Sigma_{11} \\ \Sigma_{21} \end{bmatrix}$$

$$Cov[\mathbf{z}, \mathbf{e}_2] = Cov[Q\mathbf{e}, \mathbf{e}_2] = \dots = Q\begin{bmatrix} \Sigma_{12} \\ \Sigma_{22} \end{bmatrix}$$

Augmenting the model for the  $t_3$  new observations, we arrive at the form:

$$\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{z} \\ \mathbf{y}_2 \end{bmatrix} = \begin{bmatrix} X_1 \\ QX \\ X_2 \end{bmatrix} \beta + \begin{bmatrix} \mathbf{e}_1 \\ Q\mathbf{e} \\ \mathbf{e}_2 \end{bmatrix}, \quad Var\begin{bmatrix} \mathbf{e}_1 \\ Q\mathbf{e} \\ \mathbf{e}_2 \end{bmatrix} = \begin{bmatrix} \Sigma_{11} & [\Sigma_{11} \ \Sigma_{12}]Q' & \\ Q\begin{bmatrix} \Sigma_{11} \\ \Sigma_{21} \end{bmatrix} & Q\Sigma Q' & Q\begin{bmatrix} \Sigma_{12} \\ \Sigma_{22} \end{bmatrix} \\ \Sigma_{21} & [\Sigma_{21} \ \Sigma_{22}]Q' & \Sigma_{22} \end{bmatrix}$$

From the juxtaposed model the simple joint model will arise according to this form, if  $\mathbf{z}$  is zero and:

$$Q = \begin{bmatrix} 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & -1 & -1 & 0 & 1 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 1 & 1 & 0 & -1 & -1 \end{bmatrix}$$

It is a powerful extension of the linear statistical model; nevertheless, the same two questions inform every statistical model: “What is the equation for each row?” and “How does each row covary with itself and the other rows?”<sup>7</sup> Next we will apply the joint model to ten accident years

<sup>7</sup> Halliwell’s solution [1997; Exhibit 14] differs from ours only in the prediction error variance, and there only because of a disagreement over the estimate of  $\sigma^2$ : his 106.597 versus our 79.948. And this is due to a difference of degrees of freedom, his six versus our eight ( $106.597/79.948 = 8/6$ ). Halliwell [1997; 247-249] both constrained the variance

of industry Workers' Compensation losses.

## 5. TAUTOLOGY AND A JOINT WORKERS' COMPENSATION MODEL

Exhibit 4 contains net paid and case-incurred (incurred less bulk and IBNR) triangles for U.S. Worker's Compensation, along with net earned premium and ultimate loss. Ignoring the prior-to-1998 line, we have ten accident years at ten evaluations. But we will project beyond the tenth evaluation (@120 months) to ultimate, our models assuming that at 120 months paid losses are 85.5% of ultimate and incurred are 95.0%. So our model will work with two 10×11 rectangles, each with fifty-five observations. However, to save space we will not predict each future cell, but only the unpaid and IBNR totals by AY. Thus, each of the paid and incurred submodels will have fifty-five observations and ten predictions.

It is not necessary to do so, but we will use the same design matrix for both submodels, an additive design with pure premiums by age (cf. Halliwell [2007; Section 7 and Exhibits 7A and 7B]). Because we have no exposure data, net earned premium will have to suffice. However, due to the underwriting cycle, we must adjust it to a constant loss ratio. Without rate-change information, we needed to develop the triangles to ultimate with standard deterministic methods. It's not desirable, perhaps it even smacks of cheating; but frequently it's a necessary evil, and its circularity does not seem to be vicious. The adjusted, or on-level, premium summary appears in Exhibit 5. The "Selected" losses are the simple average of the booked ultimates and four development methods. The overall loss ratio is 71%, and "Adj Prem" is simply the selected losses divided by this loss ratio, which conserves total premium. Since premium is the exposure base, our pure-premium betas are actually loss ratios.

Recognizing that the volatility, or unit-variance, of the incremental losses varies by age, we must also derive variances, or at least variance relativities, for a heteroskedastic model. We use the additive method on adjusted premium in Exhibits 6. First we derive pure premiums by age in Exhibit 6.1. The pure premium from 120 months to ultimate is calculated as:

$$\beta_{\text{ult}} = \left( \frac{1}{m} - 1 \right) (\beta_{12} + \dots + \beta_{120}),$$

where  $m$ , the maturity at 120 months, is 0.855 for paid and 0.950 for incurred. We assume that variance is proportional to exposure, and the "Selected" rows of Exhibit 6.2 derive the paid and incurred unit variances with some judgmental smoothing and extrapolation. The selected unit variances are then multiplied in Exhibit 6.3 by the adjusted premiums. We will treat these

---

structure and imposed constraints on  $\beta$ . This seems to have double-counted some observations and reduced the degrees of freedom. Our approach is more easily understood, stays closer to the empirical data, and does not require eigen-decomposition of variance matrices. Moreover, it will not disturb the variance relativities of non-homoskedastic models.

as absolute variances, and the “Unpaid” and “IBNR” columns contain the sums of the variances of AY predictions.

Exhibits 7.1 and 7.2 present the separate paid and incurred models. As mentioned, they share the same design matrix. The column marked ‘ $\Sigma$ ’ contains the homoskedastic variances; the column header ‘ $\Sigma$ ’, rather than ‘ $\Phi$ ’, signals our modeling software to take these as absolute variances. Each model has sixty-six rows, fifty-five observations, ten predictions of AY totals, and one row marked ‘Constraint’. Since there is no observation beyond 120 months, the constraint (note its zero variance) allows for the estimation of  $\beta_{ult}$ . In keeping with the two maturities, each constraint is  $0 = (1 - m)(\beta_{12} + \dots + \beta_{120}) - m\beta_{ult}$ .

Part of the joint paid-incurred model appears in Exhibit 8.1. A new column ‘Type’ identifies the paid and incurred submodels, which one can recognize in the block diagonal form. To these one hundred thirty-two rows were added the tautologous observations, ‘Ult =’ for each accident year. The negative exposure in the incurred half of these ten rows indicates that the zero ‘ $y$ ’ values are the difference of incurred from paid (paid minus incurred). The variance ‘ $\Sigma$ ’ of each new row is the sum of all the paid and incurred variances of its AY. However, the variance structure of the model at this point is the ‘ $\Sigma$ ’ column distributed down the main diagonal of a 142×142 matrix. Variance matrices, though large (sometimes exceeding the limit of 256 columns of a Excel spreadsheet) are often sparse. Our software allows us to specify only the columns of non-zero covariance, and to treat the rest of the covariance as zero. In Exhibit 8.2 we have specified how the last ten rows, the tautologous observations covary with all 142 rows. Each column (e.g., the 1998 ‘Covariance’ column) records the ‘ $\Sigma$ ’ values of its AY, except that ‘Type’ = ‘Incd’ rows must be negated, since incurred rows are subtracted in the tautology. Let  $C$  be the 142×10 covariance matrix of this exhibit. The software knows to insert  $C$  into the last ten columns and to insert  $C'$  into the last ten rows of the 142×142 variance structure. This completes the joining of the two types of losses. The two modeling questions are answered: “What is the expected value of each row?” and “How does each row vary with itself and covary with the others?”

Exhibit 9.1 presents some diagnostics. The model had  $122 = 2(55 + 1) + 10$  observations; however, two of these were constraints. After transformation (see footnote 9), a model with 122 observations in 22 parameters becomes one with 120 observations and 20 parameters. Readers wishing to solve the model in Excel can do so by reformulating ‘@Ult’ predictions in terms of paid and incurred  $\beta_{12}, \dots, \beta_{120}$  (if Excel will invert a 120×120 matrix). Two paid observations, AY 1998@12 and AY 2001@60, are more than two standard units away from expected. This is apparent even from the paid residuals of Exhibit 6.2. However, the diagnostics are not unusual, and we will focus on the estimates of  $\beta$  and  $y_2$ , as found in Exhibit 9.2. As for ‘Betahat’, the first

eleven elements are paid, and the second eleven incurred. They both sum to 0.696. This pure-premium equality would not obtain, if the exposures were adjusted for inflation and changes in claim processing; nonetheless, the ultimate AY equality would still be preserved. The reader may verify that  $\hat{\beta}_{12} + \dots + \hat{\beta}_{120}$  equals  $0.855 \times 0.696$  for paid and  $0.950 \times 0.696$  for incurred, as required by the constraints.

The unpaid and IBNR predictions are transferred from Exhibit 9.2 to summary Exhibit 10. The ‘Joint Paid-Incurred’ box shows agreement of ultimate loss and the standard deviation of its prediction error throughout the ten accident years and in total. The total ‘Std Dev’ of  $\pm 1,196,054$  equals the square root of the sum of the diagonal  $11 \times 11$  blocks of the ‘VarPrdErr’ matrix of Exhibit 9.2. We also ran separate paid and incurred models from Exhibits 7, and summarized them in the right side of Exhibit 10. Not surprisingly, the joint model mediates between the separate models:  $218,374,758 \in (217,725,562, 219,198,836)$ . This holds true by accident year except for AY 2000. But the joint modeling produces second-moment estimates that dominate those of the submodels, both in total and by accident year.

## 6. CONCLUSION

In a footnote of the introduction we quoted, “Actuaries must not presume to judge what they cannot scientifically model.” For a time science may endure competing theories, but eventually one will prevail. Likewise, the *ad hoc* blending of models, especially those arising from paid and incurred data sources, is a stopgap. For at some point it puts knowledge at the mercy of intuition at best, and of whim at worst. Hence actuaries have begun to search for joint models. Here we have shown that the linear statistical model is versatile enough to satisfy the search. The key, as always, is to ask first what all the equations are and second how they covary with each other. We may dub these the “first and second moment” questions. Actuaries have made rapid progress on the first-moment question; we hope that this paper will spur progress on the second.<sup>8</sup>

---

<sup>8</sup> Since 2000 the topic of generalized linear models (GLM) has received much attention from actuaries and academicians. One is easily lulled into thinking that generalizing is only in one direction. Consider a linear statistical model (LSM)  $\mathbf{y} = \mathbf{Xb} + \mathbf{e}$ ,  $\text{Var}[\mathbf{e}] = \text{diag}(\sigma)$ , i.e., a heteroskedastic model. GLM generalizes it with a link function and with a distributional form of  $\mathbf{e}$ , for example as:  $\mathbf{y} = g^{-1}(\mathbf{Xb}) + \mathbf{e} - E[\mathbf{e}]$ ,  $\mathbf{e} \sim \text{distribution}(\Theta)$  [Anderson, 2004; 13-14]. The LSM is a GLM whose link is the identity function and whose distribution is multivariate normal. Now the link function can be accommodated with a *non-linear* statistical model (Halliwell [1997; 325-326] and Judge [1988; Chapter 12]). Hence, many consider the advantage of GLM over LSM to reside in non-normal error terms. However, the density of  $\mathbf{e}$  is invariably assumed to be the product of the densities of the elements of  $\mathbf{e}$ , which implies zero covariance. GLM does not generalize the variance structure beyond heteroskedasticity. Our linear model generalizes the LSM in a different direction from the one in which GLM generalizes it. We do not wish to gainsay GLM; certainly, no one has a panacea. However, to him whose only tool is a hammer everything looks like a nail. Enthusiasm over GLM may distract actuaries from asking the second-moment question. If covariance is key to the joint paid-incurred model, GLM will not provide an acceptable solution.

## **Acknowledgement**

The author thanks David R. Clark, FCAS, MAAA, for his helpful editing and critique of this paper.

## **REFERENCES**

- [1.] Anderson, Duncan, Feldblum, Sholom, et al., "A Practitioner's Guide to Generalized Linear Models," 2004 Discussion Paper Program: Applying and Evaluating Generalized Linear Models, Casualty Actuarial Society, 2004, 1-116, [www.casact.org/pubs/dpp/dpp04/04dpp1.pdf](http://www.casact.org/pubs/dpp/dpp04/04dpp1.pdf).
- [2.] Halliwell, Leigh J., "Conjoint Prediction of Paid and Incurred Losses," 1997 Loss Reserving Discussion Papers, Casualty Actuarial Society, 1997, 241-379, [www.casact.org/pubs/forum/97sforum/97sf1241.pdf](http://www.casact.org/pubs/forum/97sforum/97sf1241.pdf).
- [3.] "Chain-Ladder Bias: Its Reason and Meaning," *Variance*, 1:2, 2007, 214-247, [www.variancejournal.org/issues/01-02/214.pdf](http://www.variancejournal.org/issues/01-02/214.pdf).
- [4.] Healy, M. J. R., *Matrices for Statistics*, Oxford, Clarendon Press, 1986.
- [5.] Judge, George G., Hill, R. C., et al., *Introduction to the Theory and Practice of Econometrics* (Second Edition), New York, John Wiley & Sons, 1988.
- [6.] Quarg, Gerhard, and Mack, Thomas, "Munich Chain Ladder: A Reserving Method that Reduces the Gap between IBNR Projections Based on Paid Losses and IBNR Projections Based on Incurred Losses," *Variance*, 2:2, 2008, 266-299, [www.variancejournal.org/issues/02-02/266.pdf](http://www.variancejournal.org/issues/02-02/266.pdf). Reprinted from *Blätter der Deutschen Gesellschaft für Versicherungs- und Finanzmathematik*, 26:4, 2004, 597-630.
- [7.] Venter, Gary G., "Distribution and Value of Reserves Using Paid and Incurred Triangles," *Casualty Actuarial Society E-Forum*, Fall 2008, 348-375, [www.casact.org/pubs/forum/08fforum/15Venter.pdf](http://www.casact.org/pubs/forum/08fforum/15Venter.pdf).
- [8.] Zhang, Yanwei, and Clark, David R., "A Bivariate Approach to Setting Reserves," unpublished as of March 2009.

Exhibit 1

Joint Paid-Incurred Model

Type	AY	Age	y	Paid			Incd			$\Phi$								
				1	2	3	1	2	3		X							
Paid	1	1	50	1						1								
Paid	1	2	30		1						1							
Paid	1	3	20			1						1						
Paid	2	1	60	1									1					
Paid	2	2	25		1									1				
Paid	3	1	45	1											1			
Incd	1	1	75				1											
Incd	1	2	15					1										
Incd	1	3	10						1									
Incd	2	1	75				1			1						-1		
Incd	2	2	25					1			1						-1	
Incd	3	1	50						1			1						-1
Diff	2	Ult	0	1	1	1	-1	-1	-1									
Diff	3	Ult	0	1	1	1	-1	-1	-1									
Paid	2	3				1												
Paid	3	2			1													
Paid	3	3				1												
Incd	2	3																
Incd	3	2						1										
Incd	3	3							1									



Exhibit 2

Solution of the Joint Paid-Incurred Model

$X'\Phi^{-1}y$	$X'\Phi^{-1}X$					
155	3	0	0	0	0	0
56.25	0	2.25	0.25	0	-0.25	-0.25
28.75	0	0.25	1.75	0	-0.25	-0.75
200	0	0	0	3	0	0
38.75	0	-0.25	-0.25	0	2.25	0.25
1.25	0	-0.25	-0.75	0	0.25	1.75

	$\beta (X'\Phi^{-1}X)^{-1}$						
Paid1	51.666667	0.3333333	0	0	0	0	0
Paid2	26.25	0	0.4583333	-0.041667	0	0.0416667	0.0416667
Paid3	20	0	-0.041667	0.7083333	0	0.0416667	0.2916667
Incd1	66.666667	0	0	0	0.3333333	0	0
Incd2	21.25	0	0.0416667	0.0416667	0	0.4583333	-0.041667
Incd3	10	0	0.0416667	0.2916667	0	-0.041667	0.7083333

Var[ $\beta$ ]						
26.649306	0	0	0	0	0	0
0	36.642795	-3.331163	0	3.3311632	3.3311632	0
0	-3.331163	56.629774	0	3.3311632	23.318142	0
0	0	0	26.649306	0	0	0
0	3.3311632	3.3311632	0	36.642795	-3.331163	0
0	3.3311632	23.318142	0	-3.331163	56.629774	0

Type	AY	Age	y	X $\beta$	e
Paid	1	1	50	51.666667	-1.666667
Paid	1	2	30	26.25	3.75
Paid	1	3	20	20	0
Paid	2	1	60	51.666667	8.3333333
Paid	2	2	25	26.25	-1.25
Paid	3	1	45	51.666667	-6.666667
Incd	1	1	75	66.666667	8.3333333
Incd	1	2	15	21.25	-6.25
Incd	1	3	10	10	0
Incd	2	1	75	66.666667	8.3333333
Incd	2	2	25	21.25	3.75
Incd	3	1	50	66.666667	-16.666667
Diff	2	Ult	0	0	0
Diff	3	Ult	0	0	0

SSCP		
24868.75	24229.167	639.58333
24229.167	24229.167	0
639.58333	0	639.58333

100.0%	97.4%	2.6%
--------	-------	------

$t$	14
$k$	6
$df$	8
$\sigma^2$	79.947917

Exhibit 3

Prediction of the Joint Paid-Incurred Model

AY	Paid	Incd	Unpaid	IBNR	Ultimate
1	100	100	0 ± 0	0 ± 0	100 ± 0
2	85	100	22.50 ± 8.94	7.50 ± 8.94	107.50 ± 8.94
3	45	50	41.25 ± 11.83	36.25 ± 11.83	86.25 ± 11.83
Total	230	250	63.75 ± 17.31	43.75 ± 17.31	293.75 ± 17.31

Type	AY	Age	$\hat{y}_2$	Std [PE]	Var [Prediction Error]							
Paid	2	3	22.50	± 8.94	79.95	0	39.97	79.95	0	39.97		
Paid	3	2	23.75	± 9.48	0	89.94	-29.98	0	29.98	29.98		
Paid	3	3	17.50	± 10.48	39.97	-29.98	109.93	39.97	29.98	49.97		
Incd	2	3	7.50	± 8.94	79.95	0	39.97	79.95	0	39.97		
Incd	3	2	23.75	± 9.48	0	29.98	29.98	0	89.94	-29.98		
Incd	3	3	12.50	± 10.48	39.97	29.98	49.97	39.97	-29.98	109.93		

$$X_2 - \Phi_{21} \Phi_{11}^{-1} X_1$$

0	0	0.5	0	0	0.5
0	0.75	-0.25	0	0.25	0.25
0	-0.25	0.75	0	0.25	0.25
0	0	0.5	0	0	0.5
0	0.25	0.25	0	0.75	-0.25
0	0.25	0.25	0	-0.25	0.75

$$\Phi_{22} - \Phi_{21} \Phi_{11}^{-1} \Phi_{12}$$

0.5	0	0	0.5	0	0
0	0.75	-0.25	0	0.25	0.25
0	-0.25	0.75	0	0.25	0.25
0.5	0	0	0.5	0	0
0	0.25	0.25	0	0.75	-0.25
0	0.25	0.25	0	-0.25	0.75

$$\Phi_{21} \Phi_{11}^{-1}$$

0	0	0	-0.5	-0.5	0	0	0	0	0.5	0.5	0	0.5	0
0	0	0	0	0	-0.25	0	0	0	0	0	0.25	0	0.25
0	0	0	0	0	-0.25	0	0	0	0	0	0.25	0	0.25
0	0	0	0.5	0.5	0	0	0	0	-0.5	-0.5	0	-0.5	0
0	0	0	0	0	0.25	0	0	0	0	0	-0.25	0	-0.25
0	0	0	0	0	0.25	0	0	0	0	0	-0.25	0	-0.25

Modeling Paid and Incurred Losses Together

Exhibit 4

Industry Workers' Compensation Net Losses (000)

AY	EarnPrem	Cumulative Paid									
		@12	@24	@36	@48	@60	@72	@84	@96	@108	@120
1998	23,278,084	4,651,588	9,585,142	12,606,256	14,094,760	15,268,042	16,074,584	16,668,642	17,106,835	17,422,941	17,738,999
1999	21,555,421	4,211,880	9,632,480	12,750,495	14,618,989	15,637,068	16,221,974	16,753,957	17,200,779	17,557,257	
2000	23,495,444	4,553,584	10,366,172	13,709,157	15,579,342	16,724,292	17,365,134	17,961,104	18,432,885		
2001	25,864,065	4,556,995	10,343,323	13,761,573	15,619,782	16,358,074	16,800,979	17,289,118			
2002	29,134,414	4,262,115	9,525,796	12,527,871	14,177,862	15,284,598	15,899,281				
2003	32,391,860	4,274,440	9,451,725	12,390,213	14,138,206	15,283,538					
2004	36,533,278	4,624,395	9,798,635	12,473,626	14,134,508						
2005	39,208,849	4,865,363	9,946,876	12,789,801							
2006	42,065,555	5,130,174	10,724,002								
2007	40,220,014	5,211,936									

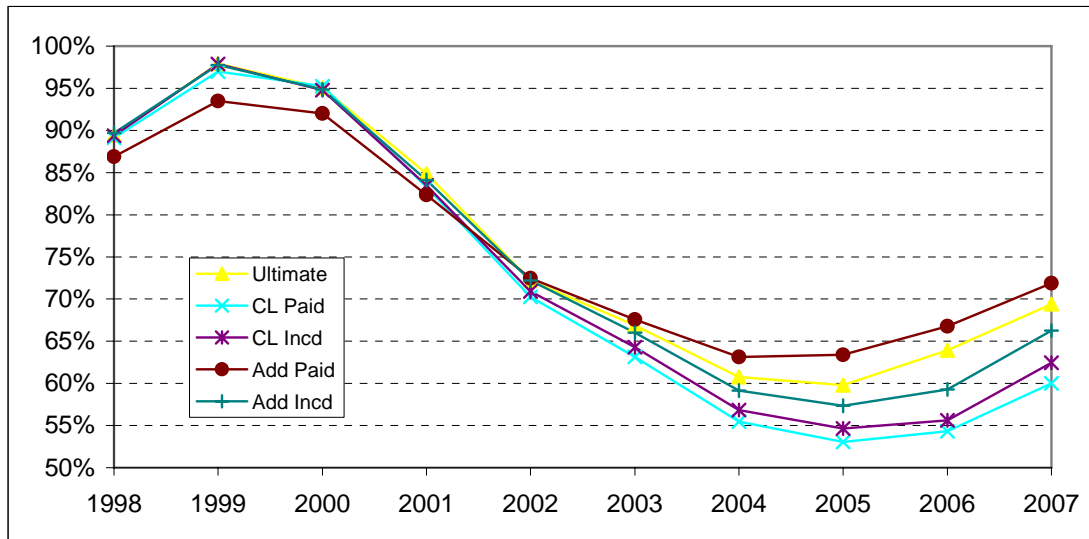
AY	Ultimate	Cumulative Case-Incurred									
		@12	@24	@36	@48	@60	@72	@84	@96	@108	@120
1998	20,815,720	10,440,449	14,526,669	16,215,164	17,259,403	18,111,150	18,727,822	19,147,843	19,469,090	19,540,774	19,765,070
1999	21,107,246	10,104,076	14,366,317	16,374,957	17,641,331	18,407,246	18,857,403	19,336,715	19,548,265	19,812,936	
2000	22,339,113	10,614,330	15,701,665	17,701,687	18,844,583	19,668,677	20,094,561	20,402,031	20,739,446		
2001	21,958,321	11,104,926	15,846,924	17,963,819	18,932,871	19,179,055	19,484,216	19,811,447			
2002	21,039,160	10,379,583	15,108,660	16,994,756	17,685,953	18,233,994	18,589,679				
2003	21,658,869	10,932,703	15,324,420	16,898,562	17,695,675	18,314,350					
2004	22,204,956	11,239,343	15,320,398	16,843,250	17,632,172						
2005	23,445,324	11,978,411	15,632,319	17,221,257							
2006	26,885,991	12,468,437	16,822,179								
2007	27,906,944	12,931,177									

Exhibit 5

Comparison of Ultimates and On-Level Premium

AY	EarnPrem	Ultimate	CL Paid	CL Incd	Add Paid	Add Incd	Selected
1998	23,278,084	20,815,720	20,747,367	20,805,337	20,226,368	20,881,833	20,695,325
1999	21,555,421	21,107,246	20,907,312	21,095,112	20,153,220	21,074,491	20,867,476
2000	23,495,444	22,339,113	22,380,335	22,271,937	21,614,963	22,286,578	22,178,585
2001	25,864,065	21,958,321	21,545,915	21,589,768	21,305,566	21,762,406	21,632,395
2002	29,134,414	21,039,160	20,472,774	20,661,071	21,107,195	21,042,136	20,864,467
2003	32,391,860	21,658,869	20,446,996	20,823,394	21,885,290	21,389,016	21,240,713
2004	36,533,278	22,204,956	20,265,744	20,762,892	23,064,618	21,607,790	21,581,200
2005	39,208,849	23,445,324	20,806,146	21,420,815	24,850,676	22,488,817	22,602,355
2006	42,065,555	26,885,991	22,848,638	23,395,295	28,090,567	24,939,425	25,231,983
2007	40,220,014	27,906,944	24,129,598	25,119,943	28,910,427	26,647,699	26,542,922
Total	313,746,984	229,361,644	214,550,827	217,945,563	231,208,890	224,120,191	223,437,423

AY	Adj Prem	Ultimate	CL Paid	CL Incd	Add Paid	Add Incd	Selected
1998	29,060,019	89%	89%	89%	87%	90%	89%
1999	29,301,751	98%	97%	98%	93%	98%	97%
2000	31,142,788	95%	95%	95%	92%	95%	94%
2001	30,375,837	85%	83%	83%	82%	84%	84%
2002	29,297,526	72%	70%	71%	72%	72%	72%
2003	29,825,844	67%	63%	64%	68%	66%	66%
2004	30,303,950	61%	55%	57%	63%	59%	59%
2005	31,737,838	60%	53%	55%	63%	57%	58%
2006	35,430,317	64%	54%	56%	67%	59%	60%
2007	37,271,114	69%	60%	62%	72%	66%	66%
Total	313,746,984	73%	68%	69%	74%	71%	71%



Modeling Paid and Incurred Losses Together

Exhibit 6.1

Additive Projections with On-Level Premium

AY	Adj Prem	Incremental Paid											Ultimate	LR
		@12	@24	@36	@48	@60	@72	@84	@96	@108	@120	@Ult		
1998	29,060,019	4,651,588	4,933,554	3,021,114	1,488,504	1,173,282	806,542	594,058	438,193	316,106	316,058	2,924,110	20,663,109	71%
1999	29,301,751	4,211,880	5,420,600	3,118,015	1,868,494	1,018,079	584,906	531,983	446,822	356,478	318,687	2,948,434	20,824,378	71%
2000	31,142,788	4,553,584	5,812,588	3,342,985	1,870,185	1,144,950	640,842	595,970	471,781	358,902	338,710	3,133,685	22,264,182	71%
2001	30,375,837	4,556,995	5,786,328	3,418,250	1,858,209	738,292	442,905	488,139	460,466	350,063	330,369	3,056,512	21,486,528	71%
2002	29,297,526	4,262,115	5,263,681	3,002,075	1,649,991	1,106,736	614,683	540,138	444,120	337,636	318,641	2,948,009	20,487,825	70%
2003	29,825,844	4,274,440	5,177,285	2,938,488	1,747,993	1,145,332	617,774	549,878	452,129	343,725	324,387	3,001,170	20,572,601	69%
2004	30,303,950	4,624,395	5,174,240	2,674,991	1,660,882	1,071,056	627,677	558,692	459,376	349,235	329,587	3,049,279	20,579,410	68%
2005	31,737,838	4,865,363	5,081,513	2,842,925	1,841,463	1,121,735	657,376	585,128	481,113	365,759	345,182	3,193,561	21,381,119	67%
2006	35,430,317	5,130,174	5,593,828	3,580,408	2,055,705	1,252,242	733,858	653,204	537,087	408,313	385,342	3,565,110	23,895,269	67%
2007	37,271,114	5,211,936	6,503,618	3,766,430	2,162,510	1,317,302	771,986	687,141	564,991	429,527	405,362	3,750,337	25,571,140	69%
Pure Prem		0.148	0.174	0.101	0.058	0.035	0.021	0.018	0.015	0.012	0.011	0.101	217,725,562	69%

AY	Adj Prem	Incremental Case-Incurred											Ultimate	LR
		@12	@24	@36	@48	@60	@72	@84	@96	@108	@120	@Ult		
1998	29,060,019	10,440,449	4,086,220	1,688,495	1,044,239	851,747	616,672	420,021	321,247	71,684	224,296	1,015,137	20,805,337	72%
1999	29,301,751	10,104,076	4,262,241	2,008,640	1,266,374	765,915	450,157	479,312	211,550	264,671	226,162	1,023,581	21,093,787	72%
2000	31,142,788	10,614,330	5,087,335	2,000,022	1,142,896	824,094	425,884	307,470	337,415	179,484	240,372	1,087,893	22,272,950	72%
2001	30,375,837	11,104,926	4,741,998	2,116,895	969,052	246,184	305,161	327,231	295,330	175,064	234,452	1,061,102	21,596,099	71%
2002	29,297,526	10,379,583	4,729,077	1,886,096	691,197	548,041	355,685	374,902	284,846	168,850	226,129	1,023,434	20,678,323	71%
2003	29,825,844	10,932,703	4,391,717	1,574,142	797,113	618,675	430,571	381,663	289,983	171,895	230,207	1,041,889	20,861,756	70%
2004	30,303,950	11,239,343	4,081,055	1,522,852	788,922	652,563	437,473	387,781	294,631	174,650	233,897	1,058,590	20,855,966	69%
2005	31,737,838	11,978,411	3,653,908	1,588,938	1,015,906	683,441	458,173	406,129	308,573	182,914	244,964	1,108,680	21,601,428	68%
2006	35,430,317	12,468,437	4,353,742	2,114,552	1,134,100	762,954	511,478	453,379	344,473	204,195	273,464	1,237,667	23,811,342	67%
2007	37,271,114	12,931,177	5,309,716	2,224,415	1,193,022	802,594	538,052	476,935	362,370	214,804	287,672	1,301,970	25,621,849	69%
Pure Prem		0.358	0.142	0.060	0.032	0.022	0.014	0.013	0.010	0.006	0.008	0.035	219,198,836	70%

Modeling Paid and Incurred Losses Together

Exhibit 6.2

Variances from Additive Method

AY	1/AdjPrem	Paid Residuals										
		@12	@24	@36	@48	@60	@72	@84	@96	@108	@120	@Ult
1998	3.44E-08	359,234	-137,270	84,456	-197,589	146,191	204,630	58,299	-2,327	-18,793	0	
1999	3.41E-08	-116,179	307,595	156,928	168,375	-17,556	-22,012	-8,233	2,638	18,793		
2000	3.21E-08	-46,408	378,332	195,853	63,247	44,246	-4,209	21,812	-311			
2001	3.29E-08	70,286	485,901	348,622	95,770	-335,305	-186,261	-71,879				
2002	3.41E-08	-65,320	151,413	41,415	-49,883	71,250	7,852					
2003	3.35E-08	-131,031	-27,171	-75,561	17,466	91,174						
2004	3.30E-08	148,305	-113,643	-387,373	-97,386							
2005	3.15E-08	177,478	-456,577	-364,340								
2006	2.82E-08	-103,114	-588,580									
2007	2.68E-08	-293,250										
Zero check		0	0	0	0	0	0	0	0	0	0	
WSSR		10,180	33,825	15,703	3,150	4,962	2,602	305	0	24	0	
df		9	8	7	6	5	4	3	2	1	0	
Unit Var		1,131	4,228	2,243	525	992	651	102	0	24		
Selected		1,131	4,228	2,243	992	992	651	102	102	55	27	1,928

AY	1/AdjPrem	Incurred Residuals										
		@12	@24	@36	@48	@60	@72	@84	@96	@108	@120	@Ult
1998	3.44E-08	48,816	-53,727	-45,865	114,048	225,970	197,156	48,158	38,710	-95,797	0	
1999	3.41E-08	-373,998	87,856	259,853	328,446	134,933	27,152	104,356	-73,338	95,797		
2000	3.21E-08	-522,084	650,673	141,358	146,037	153,467	-23,699	-91,045	34,628			
2001	3.29E-08	242,767	414,597	304,004	-3,257	-407,927	-133,350	-61,469				
2002	3.41E-08	-96,980	555,294	137,561	-246,596	-82,850	-67,259					
2003	3.35E-08	267,218	142,669	-205,924	-157,591	-23,593						
2004	3.30E-08	402,891	-236,105	-285,748	-181,086							
2005	3.15E-08	629,211	-867,527	-305,240								
2006	2.82E-08	-201,163	-693,730									
2007	2.68E-08	-396,678										
Zero check		0	0	0	0	0	0	0	0	0	0	
WSSR		41,458	69,959	13,759	8,805	8,866	2,121	842	274	629	0	
df		9	8	7	6	5	4	3	2	1	0	
Unit Var		4,606	8,745	1,966	1,467	1,773	530	281	137	629		
Selected		4,606	8,745	1,966	1,773	1,773	530	291	291	291	200	1,072

Modeling Paid and Incurred Losses Together

Exhibit 6.3

Absolute Variances from Additive Method

AY	Adj Prem	Paid Variance											Unpaid
		12	24	36	48	60	72	84	96	108	120	@Ult	
1998	29,060,019	3.29E+10	1.23E+11	6.52E+10	2.88E+10	2.88E+10	1.89E+10	2.95E+09	2.95E+09	1.59E+09	7.97E+08	5.60E+10	5.60E+10
1999	29,301,751	3.31E+10	1.24E+11	6.57E+10	2.91E+10	2.91E+10	1.91E+10	2.98E+09	2.98E+09	1.61E+09	8.04E+08	5.65E+10	5.73E+10
2000	31,142,788	3.52E+10	1.32E+11	6.99E+10	3.09E+10	3.09E+10	2.03E+10	3.16E+09	3.16E+09	1.71E+09	8.55E+08	6.01E+10	6.26E+10
2001	30,375,837	3.44E+10	1.28E+11	6.81E+10	3.01E+10	3.01E+10	1.98E+10	3.08E+09	3.08E+09	1.67E+09	8.33E+08	5.86E+10	6.42E+10
2002	29,297,526	3.31E+10	1.24E+11	6.57E+10	2.91E+10	2.91E+10	1.91E+10	2.98E+09	2.98E+09	1.61E+09	8.04E+08	5.65E+10	6.49E+10
2003	29,825,844	3.37E+10	1.26E+11	6.69E+10	2.96E+10	2.96E+10	1.94E+10	3.03E+09	3.03E+09	1.64E+09	8.18E+08	5.75E+10	8.54E+10
2004	30,303,950	3.43E+10	1.28E+11	6.80E+10	3.01E+10	3.01E+10	1.97E+10	3.08E+09	3.08E+09	1.66E+09	8.32E+08	5.84E+10	1.17E+11
2005	31,737,838	3.59E+10	1.34E+11	7.12E+10	3.15E+10	3.15E+10	2.06E+10	3.22E+09	3.22E+09	1.74E+09	8.71E+08	6.12E+10	1.54E+11
2006	35,430,317	4.01E+10	1.50E+11	7.95E+10	3.52E+10	3.52E+10	2.30E+10	3.60E+09	3.60E+09	1.94E+09	9.72E+08	6.83E+10	2.51E+11
2007	37,271,114	4.22E+10	1.58E+11	8.36E+10	3.70E+10	3.70E+10	2.42E+10	3.78E+09	3.78E+09	2.05E+09	1.02E+09	7.19E+10	4.22E+11

AY	Adj Prem	Incurred Variance											IBNR
		12	24	36	48	60	72	84	96	108	120	@Ult	
1998	29,060,019	1.34E+11	2.54E+11	5.71E+10	5.15E+10	5.15E+10	1.54E+10	8.45E+09	8.45E+09	8.45E+09	5.81E+09	3.12E+10	3.12E+10
1999	29,301,751	1.35E+11	2.56E+11	5.76E+10	5.20E+10	5.20E+10	1.55E+10	8.52E+09	8.52E+09	8.52E+09	5.86E+09	3.14E+10	3.73E+10
2000	31,142,788	1.43E+11	2.72E+11	6.12E+10	5.52E+10	5.52E+10	1.65E+10	9.06E+09	9.06E+09	9.06E+09	6.23E+09	3.34E+10	4.87E+10
2001	30,375,837	1.40E+11	2.66E+11	5.97E+10	5.39E+10	5.39E+10	1.61E+10	8.83E+09	8.83E+09	8.83E+09	6.08E+09	3.26E+10	5.63E+10
2002	29,297,526	1.35E+11	2.56E+11	5.76E+10	5.19E+10	5.19E+10	1.55E+10	8.52E+09	8.52E+09	8.52E+09	5.86E+09	3.14E+10	6.28E+10
2003	29,825,844	1.37E+11	2.61E+11	5.86E+10	5.29E+10	5.29E+10	1.58E+10	8.67E+09	8.67E+09	8.67E+09	5.97E+09	3.20E+10	7.98E+10
2004	30,303,950	1.40E+11	2.65E+11	5.96E+10	5.37E+10	5.37E+10	1.61E+10	8.81E+09	8.81E+09	8.81E+09	6.06E+09	3.25E+10	1.35E+11
2005	31,737,838	1.46E+11	2.78E+11	6.24E+10	5.63E+10	5.63E+10	1.68E+10	9.23E+09	9.23E+09	9.23E+09	6.35E+09	3.40E+10	1.97E+11
2006	35,430,317	1.63E+11	3.10E+11	6.96E+10	6.28E+10	6.28E+10	1.88E+10	1.03E+10	1.03E+10	1.03E+10	7.09E+09	3.80E+10	2.90E+11
2007	37,271,114	1.72E+11	3.26E+11	7.33E+10	6.61E+10	6.61E+10	1.98E+10	1.08E+10	1.08E+10	1.08E+10	7.45E+09	4.00E+10	6.31E+11

Modeling Paid and Incurred Losses Together

Exhibit 7.1

Paid Linear Model

AY	Age	AdjPrem	y	@12	@24	@36	@48	@60	@72	@84	@96	@108	@120	@Ult	Σ
			X												
1998	12	29,060,019	4,651,588	29,060,019											3.29E+10
1998	24	29,060,019	4,933,554		29,060,019										1.23E+11
1998	36	29,060,019	3,021,114			29,060,019									6.52E+10
1998	48	29,060,019	1,488,504				29,060,019								2.88E+10
1998	60	29,060,019	1,173,282					29,060,019							2.88E+10
1998	72	29,060,019	806,542						29,060,019						1.89E+10
1998	84	29,060,019	594,058							29,060,019					2.95E+09
1998	96	29,060,019	438,193								29,060,019				2.95E+09
1998	108	29,060,019	316,106									29,060,019			1.59E+09
1998	120	29,060,019	316,058										29,060,019		7.97E+08
1999	12	29,301,751	4,211,880	29,301,751											3.31E+10
1999	24	29,301,751	5,420,600		29,301,751										1.24E+11
1999	36	29,301,751	3,118,015			29,301,751									6.57E+10
1999	48	29,301,751	1,868,494				29,301,751								2.91E+10
1999	60	29,301,751	1,018,079					29,301,751							2.91E+10
1999	72	29,301,751	584,906						29,301,751						1.91E+10
1999	84	29,301,751	531,983							29,301,751					2.98E+09
1999	96	29,301,751	446,822								29,301,751				2.98E+09
1999	108	29,301,751	356,478									29,301,751			1.61E+09
2000	12	31,142,788	4,553,584	31,142,788											3.52E+10
2000	24	31,142,788	5,812,588		31,142,788										1.32E+11
2000	36	31,142,788	3,342,985			31,142,788									6.99E+10
2000	48	31,142,788	1,870,185				31,142,788								3.09E+10
2000	60	31,142,788	1,144,950					31,142,788							3.09E+10
2000	72	31,142,788	640,842						31,142,788						2.03E+10
2000	84	31,142,788	595,970							31,142,788					3.16E+09
2000	96	31,142,788	471,781								31,142,788				3.16E+09
2001	12	30,375,837	4,556,995	30,375,837											3.44E+10
2001	24	30,375,837	5,786,328		30,375,837										1.28E+11
2001	36	30,375,837	3,418,250			30,375,837									6.81E+10
2001	48	30,375,837	1,858,209				30,375,837								3.01E+10
2001	60	30,375,837	738,292					30,375,837							3.01E+10
2001	72	30,375,837	442,905						30,375,837						1.98E+10
2001	84	30,375,837	488,139							30,375,837					3.08E+09
2002	12	29,297,526	4,262,115	29,297,526											3.31E+10
2002	24	29,297,526	5,263,681		29,297,526										1.24E+11
2002	36	29,297,526	3,002,075			29,297,526									6.57E+10
2002	48	29,297,526	1,649,991				29,297,526								2.91E+10
2002	60	29,297,526	1,106,736					29,297,526							2.91E+10
2002	72	29,297,526	614,683						29,297,526						1.91E+10
2003	12	29,825,844	4,274,440	29,825,844											3.37E+10
2003	24	29,825,844	5,177,285		29,825,844										1.26E+11
2003	36	29,825,844	2,938,488			29,825,844									6.69E+10
2003	48	29,825,844	1,747,993				29,825,844								2.96E+10
2003	60	29,825,844	1,145,332					29,825,844							2.96E+10
2004	12	30,303,950	4,624,395	30,303,950											3.43E+10
2004	24	30,303,950	5,174,240		30,303,950										1.28E+11
2004	36	30,303,950	2,674,991			30,303,950									6.80E+10
2004	48	30,303,950	1,660,882				30,303,950								3.01E+10
2005	12	31,737,838	4,865,363	31,737,838											3.59E+10
2005	24	31,737,838	5,081,513		31,737,838										1.34E+11
2005	36	31,737,838	2,842,925			31,737,838									7.12E+10
2006	12	35,430,317	5,130,174	35,430,317											4.01E+10
2006	24	35,430,317	5,593,828		35,430,317										1.50E+11
2007	12	37,271,114	5,211,936	37,271,114											4.22E+10
		Constraint	0	0.145	0.145	0.145	0.145	0.145	0.145	0.145	0.145	0.145	0.145	-0.855	
1998	Unpd	29,060,019											29,060,019		5.60E+10
1999	Unpd	29,301,751											29,301,751	29,301,751	5.73E+10
2000	Unpd	31,142,788											31,142,788	31,142,788	6.26E+10
2001	Unpd	30,375,837											30,375,837	30,375,837	6.42E+10
2002	Unpd	29,297,526											29,297,526	29,297,526	6.49E+10
2003	Unpd	29,825,844											29,825,844	29,825,844	8.54E+10
2004	Unpd	30,303,950											30,303,950	30,303,950	1.17E+11
2005	Unpd	31,737,838											31,737,838	31,737,838	1.54E+11
2006	Unpd	35,430,317											35,430,317	35,430,317	2.51E+11
2007	Unpd	37,271,114											37,271,114	37,271,114	4.22E+11



Modeling Paid and Incurred Losses Together

Exhibit 7.2

Case\_Incurred Linear Model

AY	Age	AdjPrem	y	@12 X	@24	@36	@48	@60	@72	@84	@96	@108	@120	@Ult	Σ
1998	12	29,060,019	10,440,449	29,060,019											1.34E+11
1998	24	29,060,019	4,086,220		29,060,019										2.54E+11
1998	36	29,060,019	1,688,495			29,060,019									5.71E+10
1998	48	29,060,019	1,044,239				29,060,019								5.15E+10
1998	60	29,060,019	851,747					29,060,019							5.15E+10
1998	72	29,060,019	616,672						29,060,019						1.54E+10
1998	84	29,060,019	420,021							29,060,019					8.45E+09
1998	96	29,060,019	321,247								29,060,019				8.45E+09
1998	108	29,060,019	71,684									29,060,019			8.45E+09
1998	120	29,060,019	224,296										29,060,019		5.81E+09
1999	12	29,301,751	10,104,076	29,301,751											1.35E+11
1999	24	29,301,751	4,262,241		29,301,751										2.56E+11
1999	36	29,301,751	2,008,640			29,301,751									5.76E+10
1999	48	29,301,751	1,266,374				29,301,751								5.20E+10
1999	60	29,301,751	765,915					29,301,751							5.20E+10
1999	72	29,301,751	450,157						29,301,751						1.55E+10
1999	84	29,301,751	479,312							29,301,751					8.52E+09
1999	96	29,301,751	211,550								29,301,751				8.52E+09
1999	108	29,301,751	264,671									29,301,751			8.52E+09
2000	12	31,142,788	10,614,330	31,142,788											1.43E+11
2000	24	31,142,788	5,087,335		31,142,788										2.72E+11
2000	36	31,142,788	2,000,022			31,142,788									6.12E+10
2000	48	31,142,788	1,142,896				31,142,788								5.52E+10
2000	60	31,142,788	824,094					31,142,788							5.52E+10
2000	72	31,142,788	425,884						31,142,788						1.65E+10
2000	84	31,142,788	307,470							31,142,788					9.06E+09
2000	96	31,142,788	337,415								31,142,788				9.06E+09
2001	12	30,375,837	11,104,926	30,375,837											1.40E+11
2001	24	30,375,837	4,741,998		30,375,837										2.66E+11
2001	36	30,375,837	2,116,895			30,375,837									5.97E+10
2001	48	30,375,837	969,052				30,375,837								5.39E+10
2001	60	30,375,837	246,184					30,375,837							5.39E+10
2001	72	30,375,837	305,161						30,375,837						1.61E+10
2001	84	30,375,837	327,231							30,375,837					8.83E+09
2002	12	29,297,526	10,379,583	29,297,526											1.35E+11
2002	24	29,297,526	4,729,077		29,297,526										2.56E+11
2002	36	29,297,526	1,886,096			29,297,526									5.76E+10
2002	48	29,297,526	691,197				29,297,526								5.19E+10
2002	60	29,297,526	548,041					29,297,526							5.19E+10
2002	72	29,297,526	355,685						29,297,526						1.55E+10
2003	12	29,825,844	10,932,703	29,825,844											1.37E+11
2003	24	29,825,844	4,391,717		29,825,844										2.61E+11
2003	36	29,825,844	1,574,142			29,825,844									5.86E+10
2003	48	29,825,844	797,113				29,825,844								5.29E+10
2003	60	29,825,844	618,675					29,825,844							5.29E+10
2004	12	30,303,950	11,239,343	30,303,950											1.40E+11
2004	24	30,303,950	4,081,055		30,303,950										2.65E+11
2004	36	30,303,950	1,522,852			30,303,950									5.96E+10
2004	48	30,303,950	788,922				30,303,950								5.37E+10
2005	12	31,737,838	11,978,411	31,737,838											1.46E+11
2005	24	31,737,838	3,653,908		31,737,838										2.78E+11
2005	36	31,737,838	1,588,938			31,737,838									6.24E+10
2006	12	35,430,317	12,468,437	35,430,317											1.63E+11
2006	24	35,430,317	4,353,742		35,430,317										3.10E+11
2007	12	37,271,114	12,931,177	37,271,114											1.72E+11
		Constraint	0	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	-0.95	
1998	IBNR	29,060,019										29,060,019			3.12E+10
1999	IBNR	29,301,751										29,301,751			3.73E+10
2000	IBNR	31,142,788									31,142,788	31,142,788	31,142,788		4.87E+10
2001	IBNR	30,375,837									30,375,837	30,375,837	30,375,837		5.63E+10
2002	IBNR	29,297,526									29,297,526	29,297,526	29,297,526		6.28E+10
2003	IBNR	29,825,844									29,825,844	29,825,844	29,825,844		7.98E+10
2004	IBNR	30,303,950									30,303,950	30,303,950	30,303,950		1.35E+11
2005	IBNR	31,737,838									31,737,838	31,737,838	31,737,838		1.97E+11
2006	IBNR	35,430,317									35,430,317	35,430,317	35,430,317		2.90E+11
2007	IBNR	37,271,114									37,271,114	37,271,114	37,271,114		6.31E+11









Modeling Paid and Incurred Losses Together

Exhibit 10

Summary of Linear Models

AY	EarnPrem	Paid	CaseIncd	Joint Paid-Incurred						Paid		Case-Incurred	
				Unpaid	Std Dev	IBNR	Std Dev	Ultimate	=	Ultimate	Std Dev	Ultimate	Std Dev
1998	23,278,084	17,738,999	19,765,070	2,999,915	± 142,339	973,844	± 142,339	20,738,914	TRUE	20,663,109	± 239,110	20,780,207	± 177,147
1999	21,555,421	17,557,257	19,812,936	3,398,227	± 158,471	1,142,547	± 158,471	20,955,484	TRUE	20,824,378	± 244,026	21,062,679	± 209,846
2000	23,495,444	18,432,885	20,739,446	3,804,647	± 177,631	1,498,084	± 177,630	22,237,532	TRUE	22,264,182	± 257,624	22,247,195	± 248,293
2001	25,864,065	17,289,118	19,811,447	4,227,840	± 186,625	1,705,511	± 186,625	21,516,958	TRUE	21,486,528	± 262,929	21,577,395	± 268,263
2002	29,134,414	15,899,281	18,589,679	4,662,500	± 192,377	1,972,102	± 192,377	20,561,781	TRUE	20,487,825	± 265,436	20,667,840	± 282,302
2003	32,391,860	15,283,538	18,314,350	5,421,105	± 220,945	2,390,298	± 220,945	20,704,643	TRUE	20,572,601	± 310,529	20,860,557	± 317,408
2004	36,533,278	14,134,508	17,632,172	6,568,927	± 272,535	3,071,266	± 272,536	20,703,435	TRUE	20,579,410	± 367,486	20,871,758	± 408,036
2005	39,208,849	12,789,801	17,221,257	8,692,524	± 320,964	4,261,067	± 320,964	21,482,325	TRUE	21,381,119	± 424,534	21,630,036	± 491,580
2006	42,065,555	10,724,002	16,822,179	13,144,841	± 403,064	7,046,664	± 403,064	23,868,843	TRUE	23,895,269	± 547,487	23,858,441	± 597,042
2007	40,220,014	5,211,936	12,931,177	20,392,908	± 549,291	12,673,667	± 549,291	25,604,844	TRUE	25,571,140	± 709,931	25,642,727	± 867,816
Total	313,746,984	145,061,325	181,639,713	73,313,433	± 1,196,054	36,735,050	± 1,196,055	218,374,758	TRUE	217,725,562	± 1,572,984	219,198,836	± 1,866,883

## APPENDIX A

### The Effect of Covariance in a Simple Model

This appendix will present variations of an elementary model so as to illustrate the effect of covariance. The top part of Exhibit A.1 shows the basic form (Model 1). Eight quantities are observed (Obs 1–8), and two predictions are desired (Pred 1–2). The model for each  $y$ , whether observed or predicted, is  $y_i = x_i\beta + e_i$ . Except for Pred 2, all  $x$  and  $\phi$  values are one; Pred 2 is like a doubling of Pred 1, which makes its variance relativity four. Zeroes in the variance structure are not shown. This is a heteroskedastic model (homoskedastic in the observed part). The formulæ of the linear statistical model were explained in Section 2.

In the middle of the exhibit is shown the estimate of the parameter:  $\hat{\beta} = 97.875 \pm 2.074$ . Those unfamiliar with this formulation of the linear model should at least recognize that Model 1 is equivalent to the simple average. At the bottom of the exhibit are various diagnostics (left side) and weighted sums of squares and crossproducts (“SSCP”, right side). The diagonal elements of the  $3 \times 3$  SSCP matrix must satisfy the equation  $d_1 = d_2 + d_3$ , and the fit accounts for 99.7% of  $d_1$ . Eight (unrelated) observations and one parameter make for seven degrees of freedom, and the estimate of the variance scale is  $\hat{\sigma}^2 = 240.875/7 = 34.411$ .

The estimator of  $\beta$  linearly depends on the error terms of the observations. Because the error terms of the predictions do not covary with those of the observations, nether do they covary with  $\hat{\beta}$ . Hence:

$$\begin{aligned} \text{Var}[\text{Pred}] &= \begin{bmatrix} 1 \\ 2 \end{bmatrix} \text{Var}[\hat{\beta}] \begin{bmatrix} 1 & 2 \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 0 & 4 \end{bmatrix} \hat{\sigma}^2 \\ &= \begin{bmatrix} 4.301 & 8.603 \\ 8.603 & 17.205 \end{bmatrix} + \begin{bmatrix} 34.411 & 0 \\ 0 & 137.643 \end{bmatrix} \\ &= \begin{bmatrix} 38.712 & 8.603 \\ 8.603 & 154.848 \end{bmatrix} \end{aligned}$$

(We explain the model in this manner for the benefit of those unfamiliar with matrix algebra and multivariate statistics; nevertheless, we encourage them to study these fields until they become natural. See references.) We conclude Model 1 by saying that Pred 2 is *like* Pred 1, but on twice the scale. The covariance between the two predictions is solely due to their reliance on the estimate of  $\beta$ ; it has to do with parameter variance or

uncertainty.

Model 2 of Exhibit A.2 is identical to Model 1 except that the variance relativity of Obs 4 is zero, rather than one. The column ‘Constrnd’ signals a variance degeneracy with a ‘1’ for this observation. Whether this be unrealistic, it is at least instructive as a limiting case. The model says that in one observation we were able to see  $\beta$  without the obfuscation of an error term. So  $\hat{\beta} = 96 \pm 0$ , plain and simple. The error terms readjust, and according to the formulation of our software<sup>9</sup> there are seven (stochastic) observations and zero estimated parameters, which again makes for seven degrees of freedom. In this model it is easier to see that Pred 2 is like Pred 1, but on twice the scale. If, in addition, the variance of any other observation were set to zero, the resulting two equations would be inconsistent.

Things become very interesting with Model 3 (Exhibit A.3), in which Obs 2 and Obs 6 covary, as well as Obs 7 and Pred 2. Both covariances imply perfect positive correlations. In the latter case, we show non-negative definiteness by:

$$\text{Var} \begin{bmatrix} 2 & -1 \\ \text{Obs 7} \\ \text{Pred 2} \end{bmatrix} = \begin{bmatrix} 2 & -1 \\ 1 & 2 \\ 2 & 4 \end{bmatrix} \begin{bmatrix} 2 \\ -1 \end{bmatrix} = 0$$

So the error term of Pred 2 equals (with probability one) the error term of Obs 7. It is not *like* twice the error term of Obs 7; rather, it is the *same* as twice the error term. And because its dependent variable is twice that of Obs 7, Pred 2 is not *like* twice Obs 7; it *is* twice Obs 7. Without even estimating  $\beta$  we know Pred 2 to be  $2 \cdot 90 = 180 \pm 0$ . Covariance makes the difference between like and same.

But within the observations themselves is a variance degeneracy. Though all the observation error terms are *alike*, the error term of Obs 7 *is* that of Obs 3. The ‘Cnstrnd’ column indicates with ‘0.5’ the dependency of the two observations. In effect, it says that the same observation is written twice, and should be counted once. Thus instead of eight observations, there are really seven, which with one parameter estimated makes for six degrees of freedom. When covariance is properly considered, one cannot create information *ex nihilo*, i.e., by repeating the same observation. One cannot fool the model

---

<sup>9</sup> It would take us too far afield to detail how our software solves the linear statistical model when the variance of the observations  $\Sigma_{11}$  is not positive definite. But briefly, it eigen-decomposes the observations, and treats the once-transformed rows whose eigenvalues are zero as constraints on  $\beta$ . Then it transforms constrained  $\beta$ -space into a lower-dimensional, unconstrained  $\gamma$ -space. The twice-transformed model (cf. “TTy1” in SSCP, where “TT” stands for “twice-transformed”) is solved, and transformed back. It is a theorem that the solution of a linear model is invariant to any invertible, or one-to-one, transformation of the observations, i.e.,  $A\mathbf{y}_1 = A\mathbf{X}_1\beta + A\mathbf{e}_1$ , for any nonsingular  $A$ . But one must not forget to transform the covariances:  $\Sigma_{12} \rightarrow A\Sigma_{12}$  and  $\Sigma_{21} \rightarrow \Sigma_{21}A'$ .



even by repeating a linear combination of observations, for  $Var \begin{bmatrix} \mathbf{y}_1 \\ A\mathbf{y}_1 \end{bmatrix} = \begin{bmatrix} \Sigma_{11} & \Sigma_{11}A' \\ A\Sigma_{11} & A\Sigma_{11}A' \end{bmatrix}$  contains no more information than  $\Sigma_{11}$  contains. Note that although Obs 2 and Obs 6 are redundant, they are consistent. If they were not both 93, they would be inconsistent equations.

Lastly, Model 4 in Exhibit A.4 is a mixture of Models 2 and 3. The reason for changing Obs 6 from 93 to 94 will soon appear. Similarly to Model 3:

$$Var \begin{bmatrix} 2 & -1 \end{bmatrix} \begin{bmatrix} \text{Obs 2} \\ \text{Obs 6} \end{bmatrix} = \begin{bmatrix} 2 & -1 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix} \begin{bmatrix} 2 \\ -1 \end{bmatrix} = 0,$$

i.e.,  $\text{Prob}[2e_2 - 1e_6 = 0] = 1$ , or simply,  $e_6 = 2e_2$ . But here Obs 2 and Obs 6 are not redundant. Nevertheless, they are equivalent to three equations in three variables:

$$\begin{aligned} 93 &= \beta + e_2 \\ 94 &= \beta + e_6, \\ e_6 &= 2e_2 \end{aligned}$$

whose solution is  $\beta = 92, e_2 = 1, e_6 = 2$ . In Model 2  $\beta$  was like a gem lying on the surface; here we had to pan a little for it. So  $\hat{\beta} = 92 \pm 0$ . If we had not changed Obs 6, the equations still would have been consistent, but the error terms would have both been zero – a much less interesting result. The ‘Cnstrnd’ column indicates the variance degeneracy between the two observations; but their counting as one observation is apportioned inversely according to their 1:4 variance relativities.

Granted, the behavior of Models 2–4 depends on perfect correlation, and is akin to imagining relativistic effects at the speed of light. As long as  $\Sigma_{11}$  has no variance degeneracy, i.e., is positive definite, the observations consist of  $t_1$  consistent equations in  $t_1 + k$  variables. They comprise a system of equations that can be solved only probabilistically. However, these limiting cases confirm the conservation of information. Just as there is no magic, just illusion, so too only by trickery can someone produce information out of nothing.

Consequently, to include in a joint paid-incurred model tautologous observations for completely observed exposure periods, correctly accounting for covariance, will furnish no additional information. For it’s simply a linear combination of old information. Moreover, for numerical-analytic reasons it’s dangerous, since the software must decide when small eigenvalues should be treated as zeroes. Without these redundant equations the joint model will be of full rank.

Modeling Paid and Incurred Losses Together

Exhibit A.1

Model 1

ID	y	X	$\Phi$									
Obs 1	106	1	1									
Obs 2	93	1		1								
Obs 3	99	1			1							
Obs 4	96	1				1						
Obs 5	105	1					1					
Obs 6	93	1						1				
Obs 7	90	1							1			
Obs 8	101	1								1		
Pred 1		1									1	
Pred 2		2										4

ID	y2hat	StdPrdErr	VarPrdErr	VarPrdErr
Pred 1	97.875	6.222	38.712	8.603
Pred 2	195.750	12.444	8.603	154.848

Betahat	StdBeta	VarBeta
97.875	2.074	4.301

ID	y1	Fitted	Resid	Student	Cnstrnd
Obs 1	106	97.875	8.125	1.48	0
Obs 2	93	97.875	-4.875	-0.89	0
Obs 3	99	97.875	1.125	0.21	0
Obs 4	96	97.875	-1.875	-0.34	0
Obs 5	105	97.875	7.125	1.30	0
Obs 6	93	97.875	-4.875	-0.89	0
Obs 7	90	97.875	-7.875	-1.44	0
Obs 8	101	97.875	3.125	0.57	0

SSCP	TTy1	Fitted	Resid
TTy1	76877	76636.13	240.875
Fitted	76636.13	76636.13	0
Resid	240.875	0	240.875
rhosq	100.0%	99.7%	0.3%
df	8	1	7
s2hat	9609.625	76636.13	34.41071
s2sel			<b>34.41071</b>

Exhibit A.2

Model 2

ID	y	X	$\Phi$									
Obs 1	106	1	1									
Obs 2	93	1		1								
Obs 3	99	1			1							
Obs 4	96	1				0						
Obs 5	105	1					1					
Obs 6	93	1						1				
Obs 7	90	1							1			
Obs 8	101	1								1		
Pred 1		1									1	
Pred 2		2										4

ID	y2hat	StdPrdErr	VarPrdErr	VarPrdErr
Pred 1	96	6.199	38.429	0
Pred 2	192	12.398	0	153.714

Betahat	StdBeta	VarBeta
96	0	0

ID	y1	Fitted	Resid	Student	Cnstrnd
Obs 1	106	96	10	1.61	0
Obs 2	93	96	-3	-0.48	0
Obs 3	99	96	3	0.48	0
Obs 4	96	96	0	0.00	1
Obs 5	105	96	9	1.45	0
Obs 6	93	96	-3	-0.48	0
Obs 7	90	96	-6	-0.97	0
Obs 8	101	96	5	0.81	0

SSCP	TTy1	Fitted	Resid
TTy1	269	0	269
Fitted	0	0	0
Resid	269	0	269
rhosq	100.0%	0.0%	100.0%
df	7	0	7
s2hat	38.42857	0	38.42857
s2sel			<b>38.42857</b>

Modeling Paid and Incurred Losses Together

Exhibit A.3

Model 3

ID	y	X	$\Phi$							
Obs 1	106	1	1							
Obs 2	93	1		1				1		
Obs 3	99	1			1					
Obs 4	96	1				1				
Obs 5	105	1					1			
Obs 6	93	1		1				1		
Obs 7	90	1							1	2
Obs 8	101	1							1	
Pred 1		1								1
Pred 2		2							2	4

ID	y2hat	StdPrdErr	VarPrdErr	VarPrdErr
Pred 1	98.571	6.380	40.707	0
Pred 2	180	0	0	0

Betahat	StdBeta	VarBeta
98.571	2.256	5.088

ID	y1	Fitted	Resid	Student	Cnstrnd
Obs 1	106	98.571	7.429	1.34	0
Obs 2	93	98.571	-5.571	-1.01	0.5
Obs 3	99	98.571	0.429	0.08	0
Obs 4	96	98.571	-2.571	-0.47	0
Obs 5	105	98.571	6.429	1.16	0
Obs 6	93	98.571	-5.571	-1.01	0.5
Obs 7	90	98.571	-8.571	-1.55	0
Obs 8	101	98.571	2.429	0.44	0

SSCP	TTy1	Fitted	Resid
TTy1	68228	68014.29	213.7143
Fitted	68014.29	68014.29	0
Resid	213.7143	0	213.7143
rhosq	100.0%	99.7%	0.3%
df	7	1	6
s2hat	9746.857	68014.29	35.61905
s2sel			<b>35.61905</b>

Exhibit A.4

Model 4

ID	y	X	$\Phi$								
Obs 1	106	1	1								
Obs 2	93	1		1				2			
Obs 3	99	1			1						
Obs 4	96	1				1					
Obs 5	105	1					1				
Obs 6	94	1		2				4			
Obs 7	90	1							1		
Obs 8	101	1								1	
Pred 1		1								1	
Pred 2		2									4

ID	y2hat	StdPrdErr	VarPrdErr	VarPrdErr
Pred 1	92	8.586	73.714	0
Pred 2	184	17.171	0	294.857

Betahat	StdBeta	VarBeta
92	0	0

ID	y1	Fitted	Resid	Student	Cnstrnd
Obs 1	106	92	14	1.63	0
Obs 2	93	92	1	0.12	0.8
Obs 3	99	92	7	0.82	0
Obs 4	96	92	4	0.47	0
Obs 5	105	92	13	1.51	0
Obs 6	94	92	2	0.12	0.2
Obs 7	90	92	-2	-0.23	0
Obs 8	101	92	9	1.05	0

SSCP	TTY1	Fitted	Resid
TTY1	516	0	516
Fitted	0	0	0
Resid	516	0	516
rhosq	100.0%	0.0%	100.0%
df	7	0	7
s2hat	73.71429	0	73.71429
s2sel			<b>73.71429</b>

## APPENDIX B

### Correlation Constraints among Three Random Variables

Our solution to the joint model involved the addition of tautologous observations which covary with certain of the paid and incurred observations. Often the loss observations are not inter-correlated. According to statistical and econometric terminology (e.g., Judge [1988], Chapter 9), the variance structure of such observations is homo- or heteroskedastic, as opposed to autocorrelated. In our simple example they were homoskedastic; in the Workers' Compensation example they were heteroskedastic. If  $z$  be a tautologous observation that involves loss observation  $x$  (so that  $Cov[z, x] \neq 0$ ; in our models,  $Cov[z, x] = \pm Var[x]$ ), and  $x$  does not covary with any other loss observation, then we may assume that  $z$  does not *secondarily* covary with any other loss observation. But in general, for  $z$  to covary with  $x$  and for  $x$  to covary with  $y$  places a transitive tendency for  $z$  to covary with  $y$ . Ignoring secondary covariance may lead one to create models whose variance structure is not non-negative definite. Such models would be defective, because a variance structure is legitimate if and only if it is non-negative definite.

So our task here is to solve an interesting problem: If the correlation between two random variables is  $\rho$ , what are legitimate values of  $x$  and  $y$ , the correlations of a third random variable with the first two? Mathematically expressed, for what values of  $x$  and  $y$  is  $\begin{bmatrix} \rho & 1 & y \\ 1 & \rho & x \\ y & x & 1 \end{bmatrix}$  non-negative definite? To express the problem as correlation is simpler and no less general than to express it as covariance.

First, because the correlation coefficient is bounded,  $-1 \leq \rho, x, y \leq 1$ . Hence, regardless of  $\rho$ , allowable pairs  $(x, y)$  must be on or within the square whose four corners are  $(\pm 1, \pm 1)$ . And second, it is a theorem of matrix algebra that a matrix  $\Sigma$  is non-negative definite if and only if it has a "square root," i.e., a real-valued matrix  $W$  such that  $\Sigma = WW'$ . The Cholesky decomposition yields a suitable square  $W$  that is lower-triangular (i.e., zero above the main diagonal).<sup>10</sup>

The Cholesky decomposition of the correlation matrix is:

$$\begin{bmatrix} 1 & \rho & x \\ \rho & 1 & y \\ x & y & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ \rho & \sqrt{1-\rho^2} & 0 \\ x & a & b \end{bmatrix} \begin{bmatrix} 1 & \rho & x \\ 0 & \sqrt{1-\rho^2} & a \\ 0 & 0 & b \end{bmatrix},$$

<sup>10</sup> Cf. Halliwell [1997; Appendix A], Healy [1986; 54f], and Judge [1988; 961].

where:

$$y = \rho x + a\sqrt{1-\rho^2}$$

$$1 = x^2 + a^2 + b^2$$

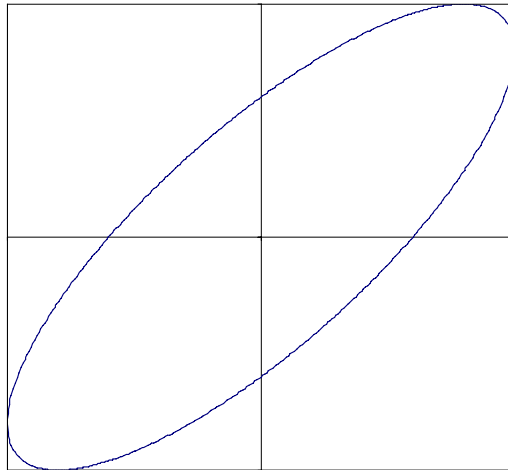
Hence, the correlation matrix is non-negative definite if and only if real values of  $a$  and  $b$  exist that solve these two last equations (in which  $\rho$ ,  $x$ , and  $y$  are given).

Of course,  $\sqrt{1-\rho^2} \geq 0$ . If  $\sqrt{1-\rho^2} = 0$ ,  $y$  must equal  $\rho x$ , and  $a$  may be any real number. Setting  $a$  to zero in this case gives the most leeway for  $b$  to be real. Hence, a Cholesky decomposition exists if and only if  $b^2 \geq 0$ . With this information we derive the inequality:

$$\begin{aligned} x^2 - 2\rho xy + y^2 &= x^2 - \rho^2 x^2 + y^2 - 2\rho xy + \rho^2 x^2 \\ &= x^2(1-\rho^2) + (y-\rho x)^2 \\ &= x^2(1-\rho^2) + (a\sqrt{1-\rho^2})^2 \\ &= (x^2 + a^2)(1-\rho^2) \\ &= (x^2 + a^2 + 0)(1-\rho^2) \\ &\leq (x^2 + a^2 + b^2)(1-\rho^2) \\ &= 1 \cdot (1-\rho^2) \end{aligned}$$

Therefore, the correlation matrix is non-negative definite if and only if  $x^2 - 2\rho xy + y^2 \leq (1-\rho^2)$ .

Legitimate  $(x, y)$  points satisfy the equation  $x^2 - 2\rho xy + y^2 \leq (1-\rho^2)$ . The region of these legitimate points is symmetric about the two lines  $y = \pm x$ . In fact, it is an ellipse whose axes are on those lines, its half lengths along the lines being  $\sqrt{1 \pm \rho}$  respectively. Here is a graph of the ellipse when  $\rho = 0.8$ :



Interior points of the ellipse, i.e.,  $x^2 - 2\rho xy + y^2 < (1 - \rho^2)$ , produce positive definite matrices; boundary points indicate a linear dependence among the three random variables. If  $\rho = 0$ , the ellipse becomes the unit circle. In the case that  $\rho = \pm 1$ , the ellipse degenerates into the respective lines  $y = \pm x$ , as expected. The origin is always a legitimate candidate for  $(x, y)$ , since for any two random variables there exists a third uncorrelated with either one of them.

The area of the ellipse is  $\pi\sqrt{1 - \rho^2}$ . That this is maximized for  $\rho = 0$  means that one may accommodate new random variables most freely into a universe of uncorrelated random variables. We can integrate the area of ellipse( $\rho$ ) over  $\rho$ :  $\int \pi\sqrt{1 - \rho^2} d\rho = \pi \frac{\pi}{2} = \frac{\pi^2}{2}$ . From this we conclude that the probability of constituting a legitimate correlation structure by randomly sampling  $\rho$ ,  $x$ , and  $y$  from a Uniform $[-1, 1]$  distribution is  $\frac{\pi^2}{2} / 2^3 = \left(\frac{\pi}{4}\right)^2 \approx 61.7\%$ .<sup>11</sup>

---

<sup>11</sup> The legitimacy equation has the three-way symmetric form  $x^2 + y^2 + z^2 - 2xyz \leq 1$ , as well as the

determinant form  $\begin{vmatrix} 1 & x & y \\ x & 1 & z \\ y & z & 1 \end{vmatrix} \geq 0$ . However, non-negative (positive) definiteness means more than a non-

negative (positive) determinant. A symmetric  $n \times n$  matrix is non-negative (positive) definite if and only if all its subdeterminants, of which there are  $2^n - 1$ , are  $\geq 0$  ( $> 0$ ). However, as a test of definiteness this is much less efficient than the Cholesky decomposition.