Statistical Models and Credibility

by Leigh J. Halliwell, FCAS, MAAA

Statistical Models and Credibility

Leigh J. Halliwell, FCAS, MAAA

Abstract

The theory of credibility is a cornerstone of actuarial science. Actuaries commonly use it, and with some pride regard it as their own invention, something which surpasses statistical theory and sets actuaries apart from statisticians. Nevertheless, the development of statistical models by statisticians and econometricians in the latter half of this century is very relevant to credibility theory; it can ground as well as generalize much of the theory, particularly the branch thereof known as least-squares credibility. It is the purpose of this paper to show how the theory and practice of credibility can benefit from statistical modeling.

The first half of the paper consists of eleven sections, notes, references, and twenty exhibits. The technical content is subdued, and readers may content themselves with this half. But the technically inclined are invited to study the six appendices (A through F) of the second half. Due to space limitations of the Call Paper Program, some of the appendices may be deleted. If this should happen, the deleted appendices can be obtained by calling the author at (201) 278-8860.

The author is grateful to Kenneth Kasner, FCAS, MAAA, for his thoughtful and kind review of the draft of this paper.

Mr. Halliwell is a Fellow of the Casualty Actuarial Society and a member of the American Academy of Actuaries. In August 1997 he became a consultant at the New York office of Milliman and Robertson. For two years prior to that he lived in Mexico City as the Regional Actuary of Latin America for the Zurich Insurance Group. And prior to that he was the Chief Actuary of the Louisiana Workers' Compensation Corporation, Baton Rouge, LA. His actuarial career began at the National Council on Compensation Insurance in Boca Raton, FL.

1. Introduction

Throughout the twentieth century actuaries have been practicing something that they call credibility. Although acknowledging some connections with statistics, especially with regard to Bayesian credibility, actuaries have tended to regard credibility as transcending statistics. This is illustrated in the historical sketch of the following section. But this paper will proceed to show that advances in statistical modeling during the latter half of this century legitimate and deepen typical uses of credibility. In order not to presume on the readers' knowledge of modern statistics, Sections 3, 4, and 5 will outline and illustrate the linear statistical model. The treatment of credibility *per se* will begin in Section 6, where we will show how to introduce prior (or non-sample) information into the statistical model. It is hoped that the reader will be persuaded that to express credibility in statistical terms is not only possible, but also advantageous. Six appendices at the end of the paper provide mathematical foundations for much of what is glossed over in the sections.

2. An Historical Perspective on Credibility

To Matthew Rodermund was entrusted the formidable task of writing the introduction to the textbook *Foundations of Casualty Actuarial Science*. The task was formidable because it demanded a engaging history of the casualty actuarial profession and a distillation of its essence. Rodermund states, "It is the concept of credibility that has been the casualty actuaries' most important and most enduring contribution to casualty actuarial science."

[11:3].* After recounting the accomplishments of actuaries in experience rating, retrospective rating, merit rating, ratemaking, and reserving – all with an eye on credibility, he asks, "Readers who have come this far may conclude from what they've read that casualty actuarial science is the study and application of the theory of credibility, and that's all. Is it all?" [11:19] An affirmative answer is implied. And almost thirty years earlier L. H. Longley-Cook, although more reserved than Rodermund, prefaced his famous monograph on credibility with the words "Credibility Theory is one of the cornerstones of actuarial science ..." [9:3]

The "Statement of Principles Regarding Property and Casualty Ratemaking," adopted by the Casualty Actuarial Society in 1988, defines credibility to be "a measure of the predictive value that an actuary attaches to a particular body of data."[1] Actuaries often speak equivalently of the "weight" given to a body of data. The language of *attaching* or *giving* credibility to data is suggestive of an important point made by Longley-Cook:

> ... the amount of credibility to be attached to a given body of data is not entirely an intrinsic property of the data. For example, there is always stated or implied in any measure of credibility the purpose to which data are to be used.
> ...
> Hence, we see that credibility is not a simple property of data which can be calculated by some mathematical formula ... [9: 4]

If credibility is not entirely intrinsic to the data, then it is at least partially extrinsic. In practice, credibility is largely, if not entirely, extrinsic to the data. And what is extrinsic to the data pertains to informed judgment; so it is fitting that Longley-Cook concluded his monograph as follows:

> It is perhaps necessary to stress that credibility procedures are not a substitute for informed judgment, but an aid thereto. Of necessity so many practical considerations must enter into

* In the '[*n*:*p*]' format '*n*' is the reference number and '*p*' gives the page number(s).

> any actuarial work that the student cannot substitute the blind application of a credibility formula for the careful consideration of all aspects of an actuarial problem. [9:25] (also quoted in [11:10f.])

Since the credibility of data is the predictive value or weight given to the data, the question arises what to do when the actuary judges the data not to have enough predictive value or weight. The answer is to weight the answer which is based on the data with an answer based on informed judgment; so it is natural for actuaries to speak of credibility-weighting the empirical answer with another source of information.

One great teacher and apologist of credibility was Arthur L. Bailey. Writing between 1945 and 1950, he claimed that certain credibility procedures conflicted with current statistical theory; in fact, statistical training could hinder someone from accepting these procedures:

> The basis for these credibility formulas has been a profound mystery to most people who have come in contact with them. The actuary finds them difficult to explain and, in some cases, even difficult to understand. Paradoxical as it may be, the more contact a person has had with statistical practices in other fields or the more training a person has had in the theory of mathematical statistics, the more difficult it has been to understand these credibility procedures or the validity of their application. [3:7]

Bailey listed as three offending credibility procedures (1) the use of prior hypotheses in estimation, (2) an estimation of groups together which is more accurate than estimating each group separately, and (3) estimating for an individual that belongs to a heterogeneous population [4:59f.]. Speaking from his own experience and with the ardor of a convert, he wrote:

> I personally entered the casualty insurance field from the completely unassociated field of statistical research in the banana business. The first year or so I spent proving to myself that all of the fancy actuarial procedures of the casualty business were mathematically unsound. They are unsound, if one is bound to accept the restrictions implied or specifically placed on the development of the classical statistical methods. Later on I realized that the hard-shelled underwriters were recognizing certain facts of life neglected by the statistical theorists. Now I am convinced that casualty insurance statisticians are a step ahead of those in most fields.

> This is because there has been a truly epistemological review of the basic conditions of which their statistics are measurements. I can only urge a similar review be made by statisticians in other fields. [4:61]

Bailey [3] sought to ground these procedures in what later became known as Bayesian analysis. No doubt, in his day statistical theory could not accommodate certain actuarial ideas. Therefore, he saw the actuarial profession as in "revolt," as for example when he wrote:

> Philosophers have recently discussed the credibilities to be given to various elements of knowledge, thus undermining the accepted philosophy of the statisticians. However, it appears to be only in the actuarial field that there has been an organized revolt against discarding all prior knowledge when an estimate is to be made using newly acquired data. [3:9f.]

But a revolt involving Bayesian analysis was soon to happen among the statisticians, as Allen Mayerson remarked in 1964:

> Statistical theory has now caught up with the actuary's problems. Starting with the 1954 book by Savage, and buttressed by the 1959 volume by Schlaifer and the 1961 book by Raiffa and Schlaifer, there has been, among probabilists and statisticians, an organized revolt against the classical approach and a trend toward the use of prior knowledge for statistical inference.
>
> ...
>
> The relationship between Bayes' theorem and credibility was first noticed by Arthur Bailey, who showed that the formula $ZA+(1-Z)B$ can be derived from Bayes' theorem ...
>
> ...
>
> It seems appropriate, in view of the growing interest among statisticians in the Bayesian point of view, to attempt to continue the work started 15 years ago by Bailey, and, using modern probability concepts, try to develop a theory of credibility which will bridge the gap that now separates the actuarial from the statistical world. [10:85f.]

Bayesian analysis has continued to be a popular basis of credibility theory. It plays a prominent role in Gary Venter's momentous chapter on credibility in the *Foundations* textbook [13]. But Bailey's seminal idea was a "greatest accuracy credibility" [2:20], of which Venter writes:

> The most well developed approach to greatest accuracy credibility is *least squares credibility*, which seeks to minimize the expected value of the square of the estimation error ...
>
> More recent statistical theory, Bayesian analysis for example, also addresses the use of data to update previous estimates, and this will be introduced later below. Credibility theory shares with Bayesian analysis the outlook toward data as strictly a source to update prior knowledge. Credibility, particularly least squares credibility is sometimes labeled Bayesian or empirical Bayesian for this reason. It also gives the same result as Bayesian analysis in some circumstances, although credibility theory can be developed within the frequentist view of probability ...
>
> Frequentist refers to an interpretation of probability as solely an expression of the relative frequency of events, in contrast to a subjectivist view which regards probability as a quantification of opinion. This latter view is a hallmark of Bayesian analysis. [13:384]

This quotation clearly indicates that Bayesian analysis is not the be-all and end-all of credibility theory. Rather, despite some similarities, greatest accuracy credibility is independent from Bayesian analysis, and especially from the on-going philosophical debate between the frequentists and the subjectivists. With all the limelight on Bayesian analysis, actuaries have not realized that statistical theory now has some non-Bayesian things to say about credibility. In particular, modern statistical modeling can accommodate the three "offending" credibility procedures mentioned above; moreover, it provides a richer world of ideas than the one-dimensional formula $ZA+(1-Z)B$.

3. An Overview of the Linear Statistical Model

In an earlier paper [7] the author treated the best linear unbiased estimation (BLUE) of the linear statistical model. That treatment was detailed and self-contained; so the author will assume it, rather than derive it. In Appendix C of that paper the author compared BLUE with Gary Venter's formulation of least-squares credibility [13:418], and concluded:

> 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. [7:335]

It is for the purpose of realizing that hope that the present paper is written.

The form of a linear[2] statistical model is $\mathbf{y} = X\beta + \mathbf{e}$, where $\mathrm{Var}[\mathbf{e}] = \Sigma = \sigma^2\Phi$. In this model $\mathbf{y}$ and $\mathbf{e}$ are $(t\times 1)$ vectors, X is a $(t\times k)$ matrix, β is a $(k\times 1)$ vector, and Σ and Φ are $(t\times t)$ matrices. The design matrix X is known, or posited; $\mathbf{y}$ is observed. Although the parameter vector β is not known, it is not random; an estimator of β is random, but β itself is not. What injects randomness into the vector $\mathbf{y}$ is the error term $\mathbf{e}$. $\mathbf{e}$ is not observable; however, $\mathrm{E}[\mathbf{e}] = 0_{(t\times 1)}$, and $\mathrm{Var}[\mathbf{e}]$ is known, or posited, at least to within a proportionality constant, i.e., $\mathrm{Var}[\mathbf{e}] \propto \Phi$. No assumption is made as to the probability distribution of $\mathbf{e}$.

Most presentations of the linear statistical model dwell on how to estimate β, but there is a wider approach. 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 so that the observed portion comes first. Similarly partition X and $\mathbf{e}$, so that the model looks like:

$$\begin{aligned}\mathbf{y}_1 &= X_1\beta + \mathbf{e}_1\\ \mathbf{y}_2 &= X_2\beta + \mathbf{e}_2\end{aligned} \quad , \text{ where } \mathrm{Var}\begin{bmatrix}\mathbf{e}_1\\ \mathbf{e}_2\end{bmatrix} = \Sigma = \begin{bmatrix}\Sigma_{11} & \Sigma_{12}\\ \Sigma_{21} & \Sigma_{22}\end{bmatrix} = \sigma^2\Phi = \sigma^2\begin{bmatrix}\Phi_{11} & \Phi_{12}\\ \Phi_{21} & \Phi_{22}\end{bmatrix}$$

Since variance matrices are symmetric (cf. [7:304] and [8:43]), $\Sigma_{21} = \Sigma_{12}'$ and $\Phi_{21} = \Phi_{12}'$.

Being unobserved, $\mathbf{y}_2$ contains missing values. The problem is 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 best; i.e., we want the best linear unbiased estimator (BLUE) of $\mathbf{y}_2$. In Appendix C of the earlier paper [7] it is shown that the BLUE of $\mathbf{y}_2$ is:

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

This is equivalent to:

$$\hat{\mathbf{y}}_2 = \mathrm{X}_2\hat{\beta} + \Phi_{21}\Phi_{11}^{-1}(\mathbf{y}_1 - \mathrm{X}_1\hat{\beta})$$
$$\mathrm{Var}[\mathbf{y}_2 - \hat{\mathbf{y}}_2] = \sigma^2(\Phi_{22} - \Phi_{21}\Phi_{11}^{-1}\Phi_{12}) + (\mathrm{X}_2 - \Phi_{21}\Phi_{11}^{-1}\mathrm{X}_1)\mathrm{Var}[\hat{\beta}](\mathrm{X}_2 - \Phi_{21}\Phi_{11}^{-1}\mathrm{X}_1)', \text{ where}$$
$$\hat{\beta} = (\mathrm{X}_1'\Phi_{11}^{-1}\mathrm{X}_1)^{-1}\mathrm{X}_1'\Phi_{11}^{-1}\mathbf{y}_1 \text{ and}$$
$$\mathrm{Var}[\hat{\beta}] = \sigma^2(\mathrm{X}_1'\Phi_{11}^{-1}\mathrm{X}_1)^{-1}$$

If σ^2 is not known, it can be unbiasedly estimated as $\hat{\sigma}^2 = \dfrac{\hat{\mathbf{e}}_1'\Phi_{11}^{-1}\hat{\mathbf{e}}_1}{t_1 - k}$, where $\hat{\mathbf{e}}_1 = \mathbf{y}_1 - \mathrm{X}_1\hat{\beta}$ [7:333f.].

What does it mean for $\hat{\mathbf{y}}_2$ to be best? As explained in Appendix A, of two competing linear unbiased estimators the best estimator is the one the variance of whose prediction error is smaller:

$$\mathrm{Var}[\mathbf{y}_2 - \hat{\mathbf{y}}_2] .\leq. \mathrm{Var}[\mathbf{y}_2 - \tilde{\mathbf{y}}_2], \text{ or}$$
$$0 .\leq. \mathrm{Var}[\mathbf{y}_2 - \tilde{\mathbf{y}}_2] - \mathrm{Var}[\mathbf{y}_2 - \hat{\mathbf{y}}_2]$$

This means that the right-hand side of the second inequality is a non-negative definite matrix. 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.

Before applying this overview to credibility, the next two sections will warm the reader up with two simple linear models. Prior to riding a horse it is wise to practice on ponies.

4. The Simplest Statistical Model (Example 1)

Suppose that we have seven non-covarying and identically distributed observations of a random variable: 6.164, 11.103, 9.663, 12.998, 10.329, 9.564, and 9.602. A simple model of the i^{th} observation ($i = 1, \ldots, 7$) is $y_i = \beta + e_i$, where $\mathrm{Var}[e_i] = \sigma^2$. The matrix formulation is:

$$\begin{bmatrix} 6.164 \\ 11.103 \\ 9.663 \\ 12.998 \\ 10.329 \\ 9.564 \\ 9.602 \end{bmatrix} = \mathbf{y} = X\beta + \mathbf{e} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \beta + \mathbf{e}$$

Since the observations are non-covarying and identically distributed, $\mathrm{Var}[\mathbf{e}] = \sigma^2 I_7$. In this simple example $\hat{\beta} = (X'X)^{-1}(X'\mathbf{y}) = \dfrac{\sum 1 * y_i}{\sum 1 * 1} = \bar{y} = 9.917$. So the parameter is the mean of the observations, and the estimator of σ^2 is the sample variance (= 4.240). One might react that this is like using a sledgehammer to crack a walnut: "Why go to all this trouble when the mean and the variance are the obvious solutions from the start?" The answer, however, deserves to be pondered: This model, the simplest of all, undergirds the mean and variance functions; these functions are in reality pre-packaged solutions of the simplest linear model.

Exhibits 1 and 2 present and solve this model. The seven observations are contained in $\mathbf{y}_1$. Since these observations are non-covarying, the off-diagonal elements of Φ_{11} are zero; since they are identically distributed, the diagonal elements of Φ_{11} are equal (ones). Thus, according to the formulas of the previous section (which are repeated in the exhibits), β and its variance may be estimated.

However, in this example we have chosen to estimate, or to predict, a certain (11×1) vector $\mathbf{y}_2 = X_2\beta + \mathbf{e}_2$. What $\mathbf{y}_2$ estimates is determined by X_2, Φ_{21}, and Φ_{22}. The first seven elements of $\mathbf{y}_2$ have the same variance as $\mathbf{e}_1$ and are perfectly correlated with $\mathbf{e}_1$. This means that as far as this statistical model is concerned, these seven elements are indiscernible from $\mathbf{e}_1$, and hence *are* $\mathbf{e}_1$. The eighth element of $\mathbf{y}_2$ models the constant 0. The ninth element models a new error term, i.e., an error term which has the same variance as $\mathbf{e}_1$ but does not covary with $\mathbf{e}_1$. The last two elements of $\mathbf{y}_2$ model β without an error term and with a new error term. Exhibit 2 derives the estimate of $\mathbf{y}_2$ and the variance of its prediction error.

5. Another Simple Statistical Model (Example 2)

Exhibits 3 and 4 concern a slightly less simple example. We have actual utility expenses for thirteen months (Sep95-Sep96). For each of these months there is a suitable utility index. We desire to estimate the expenses for the next three months (Oct96-Dec96), and are comfortable with 160, 162, and 168 as predictions of the utility index.

Many actuaries would simply rescale the last month's expenses. For example, Oct96 expenses are expected to be 2,192*(160/156.779) = 2,237. But this ignores the information from the earlier months. If one were to do a similar calculation for the other twelve months, one would then have thirteen estimates in need of combination. If this combination were performed correctly, one would be doing a statistical model in a roundabout manner.

Exhibit 4 tackles the problem directly. The observed expenses are equal to β times the utility index plus a error term. However, Φ_{11} is not of constant variance. It seems reasonable for the standard deviation of expenses to be proportional to the utility index (e.g., if prices were to double, the expense swings would double). This causes the variances of the expenses to be proportional to the squares of the utility indices, which squares are found along the diagonals of Φ_{11} and Φ_{22}. Each month's error is assumed not to covary with the other months' errors. In this exhibit β and $\mathbf{y}_2$ are estimated in accordance with the formulas already mentioned. One can also take linear combinations of $\mathbf{y}_2$ and of the variance of its prediction error. For example $\begin{bmatrix}1 & 1 & 1\end{bmatrix}\hat{\mathbf{y}}_2 = \begin{bmatrix}1 & 1 & 1\end{bmatrix}\begin{bmatrix}2{,}339\\ 2{,}368\\ 2{,}456\end{bmatrix} = \begin{bmatrix}7{,}163\end{bmatrix}$ is the estimated expense for the entire fourth quarter. Moreover, the variance of its prediction error is $\begin{bmatrix}1 & 1 & 1\end{bmatrix}\mathrm{Var}\left[\mathbf{y}_2 - \hat{\mathbf{y}}_2\right]\begin{bmatrix}1\\ 1\\ 1\end{bmatrix} = \begin{bmatrix}1 & 1 & 1\end{bmatrix}\begin{bmatrix}40672 & 2941 & 3050\\ 2941 & 41695 & 3089\\ 3050 & 3089 & 44841\end{bmatrix}\begin{bmatrix}1\\ 1\\ 1\end{bmatrix} = \begin{bmatrix}145370\end{bmatrix}$, for a standard deviation of 381.

6. A Simple Example of a Model with Prior Information (Example 3)

Now that we have warmed up on two simple models, let us see how to express credibility in a statistical model. We return again to the seven observations of Example 1 (Exhibit 1). The numbers 6.164, ... , 9.602 were actually generated as random numbers with mean 10 and variance 4. Therefore, the mean and variance estimates of 9.917 and 4.240 are close. Of course, if one knew the true parameters, they would not need to be estimated.

But suppose that prior to observation we believed (for whatever reason) that the mean is 11 and the variance is 3. Could we benefit from combining observation with our prior belief? (We will assume that the prior belief is well-founded, so that it is prior information, rather than prior *mis*information.) The answer is "Yes;" it is possible, even advisable, to combine prior information with observation.

One way of combining is Bayesian inference (Appendix B). But a simpler way is to treat the prior information *as if* it had been observed. Therefore, in Exhibit 5 the prior information is appended to the observations as an eighth row (separated from the genuine observations by a light line). In an earlier paper the author referred to prior information as quasi-observation [7:Section 6 and Appendix E]. Judge [8] refers to observation as sample information and to prior information as non-sample information. Combining the two is called mixed estimation [8:877]. Our formulation of this hybrid model, which differs only slightly from Judge's, is:

$$\begin{bmatrix}\mathbf{y}\\ \mathbf{r}\end{bmatrix}=\begin{bmatrix}X\\ R\end{bmatrix}\beta+\begin{bmatrix}\mathbf{e}\\ \mathbf{v}\end{bmatrix},\ \text{where Var}\begin{bmatrix}\mathbf{e}\\ \mathbf{v}\end{bmatrix}=\begin{bmatrix}\Sigma & \\ & V\end{bmatrix}$$

So the best linear unbiased estimator of β is:

$$\begin{aligned}\hat{\beta}&=\left(\begin{bmatrix}X\\ R\end{bmatrix}'\begin{bmatrix}\Sigma & \\ & V\end{bmatrix}^{-1}\begin{bmatrix}X\\ R\end{bmatrix}\right)^{-1}\begin{bmatrix}X\\ R\end{bmatrix}'\begin{bmatrix}\Sigma & \\ & V\end{bmatrix}^{-1}\begin{bmatrix}\mathbf{y}\\ \mathbf{r}\end{bmatrix}\\ &=\left(\begin{bmatrix}X' & R'\end{bmatrix}\begin{bmatrix}\Sigma^{-1} & \\ & V^{-1}\end{bmatrix}\begin{bmatrix}X\\ R\end{bmatrix}\right)^{-1}\begin{bmatrix}X' & R'\end{bmatrix}\begin{bmatrix}\Sigma^{-1} & \\ & V^{-1}\end{bmatrix}\begin{bmatrix}\mathbf{y}\\ \mathbf{r}\end{bmatrix}\\ &=\left(X'\Sigma^{-1}X+R'V^{-1}R\right)^{-1}\left(X'\Sigma^{-1}\mathbf{y}+R'V^{-1}\mathbf{r}\right)\end{aligned}$$

Certain properties of this estimator are explored in Appendices A and B. In particular, the estimator is a matrix-weighted average of more familiar estimators and has a smaller variance. These properties depend on the block diagonality of the hybrid variance matrix, i.e., that **e** and **v** do not covary. This is a natural assumption; however, the estimator can accommodate covariance if these properties are surrendered.

Exhibit 5 works out the mixed estimate of β as 10.099. This is equivalent to what actuaries would call a weighted-average of the data with the prior hypothesis, where the weight of the data, 0.832, results from the well-known $n/(n+k)$ formula. It is interesting, perhaps surprising, that the variance of the mixed estimator, 0.904, is less than both the variance from the unmixed model (4.240) and the variance of the prior hypothesis (3.000). This synergy of combination is analyzed in Appendix A.

One complicating detail of this model has to do with the variance matrix. Usually we specify the variance matrix not *absolutely*, but *relatively*, or to within a proportionality constant. In other words, in the model $\mathbf{y} = X\beta+\mathbf{e}$, where $\mathrm{Var}[\mathbf{e}] = \Sigma = \sigma^2\Phi$, the estimator of β is invariant to the scale of Σ. So we usually specify Φ, calculate the estimator of β, and then derive an estimate of σ^2. In the unusual event that V/σ^2 is known (or, V is known to within the same proportionality constant to within which Σ is known), then one can use the mixed estimator with the relative hybrid variance matrix. However, the usual case is that V is known absolutely and Σ is known relatively. In this case the author recommends that σ^2 be estimated in the unmixed model, and that the absolute matrix $\begin{bmatrix} \hat{\sigma}^2\Phi & \\ & V \end{bmatrix}$ be used in the mixed model. (This implies that one should solve the unmixed model as a prelude to solving the mixed.) This was done in Exhibit 5, where the 4.240 down the diagonal of Φ_{11} is the estimate of the σ^2 of Example 1. Using an estimate of the absolute variance for the absolute variance itself disturbs the optimality (the "bestness" of "best linear unbiased") of the estimator; however, statisticians and econometricians feel that this is a small price to pay for the benefit derived from combining observation with prior information. Moreover, the estimate of σ^2 in the mixed model (0.904 in Exhibit 5) will not significantly differ from 1 if the absolute variance matrix is correct. Therefore, one can assume the estimator of σ^2 in the mixed model to be a chi-square random variable with *df* degrees of freedom divided by *df* (i.e., a gamma random variable with mean 1 and variance $2/df$) and can perform a significance test. But seldom is there a problem, and this will not be mentioned again in the following examples.

7. A Statistical Model of Merit Rating (Example 4)

A simple method of merit rating a driver is to make the premium proportional to the expected number of accidents. This ignores differences of severity, e.g., driver A is half as likely to have an accident as driver B, but perhaps his accidents are likely to be twice as severe. However, as with experience rating in workers' compensation, it is natural to suppose that the insured has more control over whether an accident will happen than over how severe it will be. So we wish to estimate a driver's accident frequency, and the problem is to determine how much a driver's accident record should differentiate him from his peers.

Lester Dropkin paved the way for a Bayesian solution, viz., that every driver has his own accident frequency m, and that the number of his claims is Poisson distributed with mean m. Therefore, the probability of x claims is $\frac{m^x}{x!}e^{-m}$.[3] Moreover, the frequencies of the drivers of a certain class are gamma-distributed with parameters r and a [5:392f.]. So the probability density function of the ms is $\frac{a^r}{\Gamma(r)}e^{-am}m^{r-1}$, and the ms are distributed with mean r/a and variance r/a^2. As Dropkin showed [5:399], the claim count distribution of a driver randomly selected from the class is negative binomial with mean r/a and variance $\frac{r}{a}\frac{a+1}{a}$.

But the posterior density of a driver's one-period m given $x_1, \ldots, x_n$ accidents in n previous periods is proportional or equal to:

$$\begin{aligned}
&\propto \left(\prod \frac{m^{x_i}}{x_i!} e^{-m}\right) \frac{a^r}{\Gamma(r)} e^{-am} m^{r-1} \\
&\propto \left(\prod m^{x_i} e^{-m}\right) e^{-am} m^{r-1} \\
&\propto \left(m^{\sum x_i} e^{-nm}\right) e^{-am} m^{r-1} \\
&\propto e^{-(a+n)m} m^{\left(r+\sum x_i\right)-1} \\
&= \frac{(a+n)^{r+\sum x_i}}{\Gamma\left(r+\sum x_i\right)} e^{-(a+n)m} m^{\left(r+\sum x_i\right)-1}
\end{aligned}$$

This posterior density is gamma with parameters $r' = r + \sum x_i$ and $a' = a + n$. The posterior mean, to which the merit-rated premium should be proportional, is a weighted average of the prior mean (r/a) and the empirical mean (cf. also [10:99-101]):

$$\begin{aligned}
\frac{r'}{a'} &= \frac{r + \sum x_i}{a + n} \\
&= \frac{a\dfrac{r}{a} + n\dfrac{\sum x_i}{n}}{a + n} \\
&= \frac{a\dfrac{r}{a} + n\bar{x}}{a + n}
\end{aligned}$$

The same result is obtained from the following linear model:

$$\begin{bmatrix} x_1 \\ \vdots \\ x_n \\ r/a \end{bmatrix} = \begin{bmatrix} 1 \\ \vdots \\ 1 \\ 1 \end{bmatrix} \beta + \begin{bmatrix} e_1 \\ \vdots \\ e_n \\ v \end{bmatrix}, \text{ where } \operatorname{Var}\begin{bmatrix} e_1 \\ \vdots \\ e_n \\ v \end{bmatrix} = \begin{bmatrix} r/a & 0 & 0 & 0 \\ 0 & \ddots & 0 & 0 \\ 0 & 0 & r/a & 0 \\ 0 & 0 & 0 & r/a^2 \end{bmatrix}$$

Each x_i is explained as some mean value β plus an error, where the error is like a Poisson random variable (with parameter r/a) centered about zero. But the last row is a quasi-observation: it is as if β had been observed as r/a but obfuscated with an error whose variance is r/a^2. The mixed estimator is:

$$\hat{\beta} = \left(\begin{bmatrix} 1 \\ \vdots \\ 1 \\ 1 \end{bmatrix}' \begin{bmatrix} r/a & 0 & 0 & 0 \\ 0 & \ddots & 0 & 0 \\ 0 & 0 & r/a & 0 \\ 0 & 0 & 0 & r/a^2 \end{bmatrix}^{-1} \begin{bmatrix} 1 \\ \vdots \\ 1 \\ 1 \end{bmatrix} \right)^{-1} \begin{bmatrix} 1 \\ \vdots \\ 1 \\ 1 \end{bmatrix}' \begin{bmatrix} r/a & 0 & 0 & 0 \\ 0 & \ddots & 0 & 0 \\ 0 & 0 & r/a & 0 \\ 0 & 0 & 0 & r/a^2 \end{bmatrix}^{-1} \begin{bmatrix} x_1 \\ \vdots \\ x_n \\ r/a \end{bmatrix}$$

$$= \left(\begin{bmatrix} 1 & \cdots & 1 & 1 \end{bmatrix} \begin{bmatrix} a/r & 0 & 0 & 0 \\ 0 & \ddots & 0 & 0 \\ 0 & 0 & a/r & 0 \\ 0 & 0 & 0 & a^2/r \end{bmatrix} \begin{bmatrix} 1 \\ \vdots \\ 1 \\ 1 \end{bmatrix} \right)^{-1} \begin{bmatrix} 1 & \cdots & 1 & 1 \end{bmatrix} \begin{bmatrix} a/r & 0 & 0 & 0 \\ 0 & \ddots & 0 & 0 \\ 0 & 0 & a/r & 0 \\ 0 & 0 & 0 & a^2/r \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \\ r/a \end{bmatrix}$$

$$= \frac{\frac{a}{r}x_1 + \ldots + \frac{a}{r}x_n + \frac{a^2}{r}\frac{r}{a}}{\frac{a}{r} + \ldots + \frac{a}{r} + \frac{a^2}{r}}$$

$$= \frac{x_1 + \ldots + x_n + r}{1 + \ldots + 1 + a}$$

$$= \frac{r'}{a'}$$

The statistical model reaches the same conclusion without assuming a distributional form.

Exhibit 6 shows another example of merit rating. A driver had one accident in the second of three periods (years). The variance of his yearly accidents is assumed to be 0.0625 (standard deviation of 0.25). But there is prior information that drivers of this class are expected to have 0.25 claims per period with a variance of 0.0225 (standard deviation of 0.15). In Part A of the exhibit the three years are three one-year observations. But in Part B they are summarized into one three-year observation. The estimates are the same in both

parts, but their variances differ. This hints that summarization is attended with loss of information about prediction error variance. An amount of 1 over three years could mean 1/3 each year and no apparent variance. Or it could mean widely varying positive and negative amounts by year and an arbitrarily large variance. If actuaries wish to speak of variances, then they should know where to stop summarizing the data.

8. Stochastic and Exact Constraints (Example 5)

The prior information, or the quasi-observation, $\mathbf{r} = R\beta+\mathbf{v}$ is a stochastic constraint since $\mathbf{v}$ does not have to be zero. However, as $V = Var[\mathbf{v}]$ approaches a zero matrix, the constraint behaves more and more like the exact constraint $r = R\beta$. In an earlier paper [6:26] the author filled out a loss triangle by means of estimated pure premiums by payout year. But the pure premiums by year were exactly constrained so that the sum of the first seven of them (the pure premium of payments before 84 months) was 7.213. Exhibit 7 shows that the same result is obtained by adding a quasi-observation that this sum is 7.213 with a error whose variance is 10^{-15} relative to the variances of the observations.[4] Exhibit 8 shows how different the estimate is when the constraint is relaxed. (One should not suppose that the estimates of σ^2 in the two exhibits are equal; they differ by about six million.) Appendix C proves that the mixed model (stochastically constrained model) approaches the (non-stochastically) constrained model as V approaches zero.

9. Credibility and Random Effects (Example 6)

So far, credibility has been statistically modeled by adding quasi-observations to observations, i.e., by mixing non-sample with sample information. The non-sample information is aptly considered to be logically prior to, if not also temporally prior to, the sample information. It too may have been derived from a sample; but if so, its sample is a different sample. If the two samples are grouped into a grand statistical model, such as the first grand model of Appendix A, the submodels are naturally considered as non-simultaneous, or *temporally extensive* or *longitudinal*. For example, if we begin observing the pure premium of State X with the prior opinion that it is 0.10 with a standard deviation of 0.02, we opine thus because in the past we have observed the pure premiums of similar States A, B, ...

But credibility may also involve the simultaneous modeling of similar entities. Each entity has its own model, and the models are grouped into a grand model; however, the (sub)models are simultaneous, or *temporally intensive* or *latitudinal*. Example 6, which begins with Exhibit 9, will illustrate this concept. This example, taken from Venter [13:433], consists of six observations of a pure premium from each of nine states. If the pure premiums were unrelated, then one could do no better than to solve nine independent models (to take nine averages). If the pure premiums had to be equal, then one could do no better than to average the fifty-four observations. But an actuary would rightly feel that the truth lies in between these two extremes: the pure premiums of the states are neither unrelated nor identical. The pure premium of one state is related with those of the other

states, but it also has some identity of its own. A natural way of expressing this is to assume that the pure premiums deviate randomly from a common value, e.g., $\beta_i = \beta_0 + \mathbf{v}_i$. β_0 is the (fixed) effect common to all the states, and $\mathbf{v}_i$ is the (random) effect which differentiates State i from the other states. Each $\mathbf{v}_i$ is distributed with mean zero and some (known or unknown) variance V, and the $\mathbf{v}_i$s do not covary one with another. It is this assumption of being distributed that makes the effect random.

For the moment we will abstract from the example. In general we have n models, each of the form $\mathbf{y}_i = X_i\beta_i + \mathbf{e}_i$, where $\mathrm{Var}[\mathbf{e}_i] = \Sigma_i$ and the $\mathbf{e}_i$s do not covary. At this point we have n independent models. But now we introduce the random-effects linkage, viz., that $\beta_i = \beta_0 + \mathbf{v}_i$. Now each model becomes:

$$\begin{aligned}
\mathbf{y}_i &= X_i\beta_i + \mathbf{e}_i \\
&= X_i(\beta_0 + \mathbf{v}_i) + \mathbf{e}_i \\
&= X_i\beta_0 + (X_i\mathbf{v}_i + \mathbf{e}_i) \\
&= X_i\beta_0 + \tau_i, \\
\text{where } \mathrm{E}[\tau_i] &= X_i\mathrm{E}[\mathbf{v}_i] + \mathrm{E}[\mathbf{e}_i] \\
&= 0 \\
\text{and } \mathrm{Var}[\tau_i] &= X_i\mathrm{Var}[\mathbf{v}_i]X_i' + \mathrm{Var}[\mathbf{e}_i] \\
&= X_iVX_i' + \Sigma_i \\
&= T_i
\end{aligned}$$

The formula for $\mathrm{Var}[\tau_i]$ assumes that $\mathbf{v}_i$ and $\mathbf{e}_i$ do not covary. Moreover, since $\mathbf{v}_i$ and $\mathbf{e}_i$ do not covary across groups, the τ_is do not covary one with another. Thus we have the grand model in β_0: $\begin{bmatrix} \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_n \end{bmatrix} = \begin{bmatrix} X_1 \\ \vdots \\ X_n \end{bmatrix}\beta_0 + \begin{bmatrix} \tau_1 \\ \vdots \\ \tau_n \end{bmatrix}$, where $\mathrm{Var}\begin{bmatrix} \tau_1 \\ \vdots \\ \tau_n \end{bmatrix} = \begin{bmatrix} T_1 & & \\ & \ddots & \\ & & T_n \end{bmatrix}$. The solution of this model presents no difficulties, as long as V is known.[5] Hence, the only difficulty of this

form of credibility is to estimate V, the random-effects variance, if one has no prior information about it.

So the difficult task of Example 6 is to estimate the β,s and their common variance.[6] This involves first solving the model as if it were a fixed-effects model, as in Exhibit 10. The estimate of β in this exhibit contains the nine group means, which are carried over to Exhibit 11. The estimate of the grand mean β_0 is 0.563, and the variance of the group means about β_0 is 0.0662.[7] It would be a mistake to think that this represents the random-effects variance, because we have calculated the variance of the *estimates* of the β,s, rather than the variance of the β,s themselves. Unlike the β,s themselves, the estimates of the β,s are affected by the error terms, the **e**,s. So 0.0662 has two variance components, one from the **v**,s and one from the **e**,s, which is the reason for labeling it Var[**v**+**e**]. Back in Exhibit 10 the variance of β was estimated as if the model were a fixed-effects model. The variance of the grand parameter of a fixed-effects model must be ($k \times k$) block diagonal (here k equals one); and it is reasonable to attribute these variances to the **e**,s. Since the **v**,s and the **e**,s do not covary, one can estimate V by averaging the differences of these variances from Var[**v**+**e**]; thus V is estimated to be 0.0067.[8] Exhibit 11 goes on to show that we have derived the expected value of the process variance (EVPV) and the variance of the hypothetical means (VHM), which implies to an actuary that the credibility of each group is 10.1%. This will be checked at the end of the example.

But now we can estimate $\mathrm{Var}[X_i \mathbf{v}_i + \mathbf{e}_i] = \mathrm{Var}[\tau_i] = X_i V X_i' + \Sigma_i$, which the exhibit calls the Φ for each group. In Exhibit 12 the random-effects model is solved for the grand

parameter β_0, and the variance matrix of this model is block diagonal in Φ. But we are more interested in estimating the β_is (where $\beta_i = \beta_0 + v_i$) than we are in estimating β_0. So in Exhibit 13 we formulate $\mathbf{y}_2$ as an estimator of these β_is (we also leave an estimator of β_0 in its first row): $\mathbf{y}_2 = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \beta_0$, and the covariance matrix Φ_{21} takes the v_is into account. (See the discussion of the covariance matrix T_{21} in Appendix E for details.) So the estimate of $\mathbf{y}_2$ is obtained from the familiar formula $\hat{\mathbf{y}}_2 = X_2\hat{\beta}_0 + \Phi_{21}\Phi_{11}^{-1}\left(\mathbf{y}_1 - X_1\hat{\beta}_0\right)$. The exhibit illustrates that this estimator is equivalent to giving the fixed-effects estimators 10.1% credibility against the grand mean, as well as that the simple average of the $\hat{\beta}_i$s is $\hat{\beta}_0$. Appendix E backs up these specific illustrations with general proofs.

The results are the same as Venter's [13:432f.]. One might question whether anything has been gained by the setting up of a statistical model. Venter's discussion of credibility is hard enough for actuaries to understand; statistically modeling credibility may seem even harder. However, after developing some familiarity with best linear unbiased estimation, one will find it to be the more natural and more powerful way of handling credibility. Three reasons for its being more powerful are: 1) statistical modeling preserves two moments (the variance as well as the mean), 2) combinations of the parameter estimates can be estimated, and 3) it allows for multidimensional credibility. The third reason will be illustrated in the following trend model.

10. Random-Effects Credibility and Trend Modeling (Example 7)

A simulation of loss ratios for nine states over a six-year period is shown in Exhibit 14, and is graphed in Exhibit 15. Simulating nine states makes for a cluttered graph; however, as a practical matter, the reliable estimation of a random-effects variance requires enough data to distinguish the groups from one another. This requires a fair number of groups and/or a fair number of observations per group. The author knows of no rule as to what is a "fair" number, but a "fair" number of groups is probably not much less than the nine of this and the previous example. An upward trend is evident in the graph; but the states obviously have different slopes and intercepts. In fact, State G seems to have a negative slope.

Exhibit 16 solves the problem as a fixed-effects model, with the (18×1) $\hat{\beta}$ containing the (2×1) trend parameters of the nine states. The variance of the error matrix (Φ) is I_{54}, which simplifies the formulas. $\mathrm{Var}\left[\hat{\beta}\right]$ is diagonal in the same (2×2) block $\begin{bmatrix} 0.0011 & -0.0003 \\ -0.0003 & 0.0001 \end{bmatrix}$, which, as mentioned in the previous section, means that the model is balanced. The trend parameters vary widely by state, and State G is showing a negative slope. But will the negative slope be credible?

The random-effects variance is estimated in Exhibit 17. The mean state parameter is $\begin{bmatrix} 42.3\% \\ 2.7\% \end{bmatrix}$, and the individual states' parameters vary about it by $\begin{bmatrix} 0.0033 & -0.0005 \\ -0.0005 & 0.0006 \end{bmatrix}$. Removing the effect of the error term, we are left with the estimated random-effects

variance $V = \begin{bmatrix} 0.0022 & -0.0003 \\ -0.0003 & 0.0005 \end{bmatrix}$.[9] The variance of the intercept is greater than that of the slope (0.0022 versus 0.0005). Also, the intercept and the slope negatively covary (–0.0003, or a correlation coefficient of –27.0%). This is common in random-effects trend models: there seems to be a centroid through which every group pivots. So a higher than average intercept tends to pair with a lower than average slope, and *vice versa*. The exhibit then derives the Φ matrix for the random-effects model.

The random-effects trend model is set up and solved in Exhibit 18. From the previous section and Appendix E, and because of the balance, it comes as no surprise that the grand parameter $\hat{\beta}_0 = \begin{bmatrix} 42.3\% \\ 2.7\% \end{bmatrix}$, the simple average of the fixed-effects parameters. But we really want to estimate the states' trends, which are sums of the grand parameter and the random effects. This is accomplished in Exhibit 19, in which the most difficult concept is Φ_{21}. The blocks of this matrix represent how $\beta_i = \beta_0 + \mathbf{v}_i$ covaries with $\mathbf{y}_i = X_i\beta_i + \mathbf{e}_i$:

$$\begin{aligned}
\mathrm{Cov}[\beta_i, \mathbf{y}_i] &= \mathrm{Cov}[\beta_0 + \mathbf{v}_i, X_i(\beta_0 + \mathbf{v}_i) + \mathbf{e}_i] \\
&= \mathrm{Cov}[\beta_0 + \mathbf{v}_i, X_i\beta_0 + X_i\mathbf{v}_i + \mathbf{e}_i] \\
&= \mathrm{Cov}[\mathbf{v}_i, X_i\mathbf{v}_i + \mathbf{e}_i] \\
&= \mathrm{Cov}[\mathbf{v}_i, X_i\mathbf{v}_i] + \mathrm{Cov}[\mathbf{v}_i, \mathbf{e}_i] \\
&= \mathrm{Cov}[\mathbf{v}_i, X_i\mathbf{v}_i] \\
&= \mathrm{Cov}[\mathbf{v}_i, \mathbf{v}_i]X_i' \\
&= \mathrm{Var}[\mathbf{v}_i]X_i' \\
&= VX_i'
\end{aligned}$$

The usual formula for $\hat{\mathbf{y}}_2$ yields the random-effects trend parameters by state. State G remains with a negative slope, though less negative than before.

In Appendix E the relationship between the fixed-effects and random-effects estimators is explored. One result is the discovery of a $(k \times k)$ matrix Z such that:

$$\text{Random-effects } \hat{\beta}_i = \hat{\gamma}_i = Z\left(\text{Fixed-effects } \hat{\beta}_i\right) + (I_k - Z)\hat{\beta}_0$$

This is the k-dimensional extension of the well-known scalar (or 1-dimensional) credibility formula. Exhibit 20 expresses the random-effects estimates in this Z form. As remarked in Appendix A, a matrix-weighted average of two vectors is usually not collinear with the two vectors. But somewhat surprising is that occasionally the matrix-weighted average can fall outside the range of the two vectors. For example, the posterior slope of State A (5.9%) is outside the range of the prior and empirical slopes (2.7% and 5.7%). This happens also with the intercept of State F and with the slope of State H. Non-zero off-diagonal elements of Z make this possible.

11. Conclusion

Practice precedes theory and systematization. For example, the Egyptians were doing geometry for centuries before Euclid wrote the *Elements*. Euclid didn't discover Geometry; he didn't correct it; he may not even have contributed much in the way of new theorems. But he systematized it, made it rigorous, and enabled centuries of mathematicians to develop it further. So too, actuaries have been practicing beneficial things under the name of credibility largely in ignorance of statistical theory. But just as Euclid made Geometry better, so too the theory of statistical modeling makes credibility better.

How does credibility benefit? As mentioned at the ends of the Introduction and of Section 9, statistical modeling furnishes the actuary with the variances as well as with the means; and from this the actuary can work with combinations of estimates. But perhaps most important, statistical theory and modeling are as at home in n dimensions as in one. When the author began to study statistics and econometrics he erroneously believed that his linear algebra and multivariate calculus were sufficient for statistical work. As his background was typical for an actuary, he can "speak from his own experience and with the ardor of a convert" (as did Arthur Bailey in a quote of Section 2) that most of us actuaries, even the technically inclined, are Flatlanders as regards our statistical skills. As our problems become more complex, as well as the tools with which to solve them, this defect will become more grievous.

Bailey's three offending credibility procedures (cf. Section 2) were statistically ahead of their time. But times have changed, and now it is incumbent upon actuaries to keep up with the times. The examples of this paper show how these procedures are legitimated and generalized by current statistical theory. For the use of prior hypotheses in estimation see Examples 3, 4, and 5. For the estimation of groups together which is more accurate than estimating each separately see Examples 6 and 7. And the estimation of an individual that belongs to a heterogeneous population is in essence a disguised use of a prior hypothesis; but see especially Section 7. The appendices of the paper lay the theoretical groundwork for the examples, a groundwork from which credibility has much to gain.

Notes

[1] Longley-Cook's definition is similar: "The word credibility was originally introduced into actuarial science as a measure of the credence that the actuary believes should be attached to a particular body of experience for ratemaking purposes." [9:3] "Predictive value" in the CAS statement has a more precise meaning that Longley-Cook's noun "credence." One reason for not giving credence to data is that it is suspected of being erroneous. But in credibility theory the quality of the data is not at issue; it is supposed to be valid data. What is at issue is the value of the data for predicting.

[2] The title of this paper is "Statistical Models and Credibility," but only *linear* statistical models will be treated. In the earlier paper [7:325f.] the author argued that due to the multivariate Taylor's expansion, linearity is not much of a restriction. The interested reader can refer to Judge, who devotes a chapter of his book to non-linear statistical models [8:508-511].

[3] There is an easy way to derive the form of the Poisson distribution with parameter m. One need only to remember the Taylor series for e^m:

$$\begin{aligned} 1 &= e^m e^{-m} \\ &= \left(\sum_{x=0}^{\infty} \frac{m^x}{x!} \right) e^{-m} \\ &= \left(\sum_{x=0}^{\infty} \frac{m^x}{x!} e^{-m} \right) \\ &= \sum_{x=0}^{\infty} \text{Prob}(x) \end{aligned}$$

[4] Of course, the results are not really the same, only very close (to within the decimals shown in the exhibit). Reducing the variance of the quasi-observation still more will at some point run up against computational problems, and the results will stray. The author recommends that a tight stochastic constraint should not be substituted for an exact constraint.

[5] As in Section 6, either the Σ_is are known absolutely, in which case V must be known absolutely, or the Σ_is are known relatively, in which case V must be known to within the same proportionality constant to within which the Σ_is are known.

[6] Since in this example the β_is are the means of the groups (the hypothetical means), their common variance is what actuaries call the variance of the hypothetical means. But in general, the β_is are ($k \times 1$) parameters; so their common variance could be called the variance of the hypothetical parameters.

[7] Appendix D derives the formulas for the sample mean and variance of n identically distributed non-covarying ($k\times 1$) random vectors.

[8] The fixed-effects model is an instance of the first grand model of Appendix A, where it is proved that $\mathrm{Var}\left[\hat{\beta}\right] = \mathrm{Var}\begin{bmatrix}\hat{\beta}_1 \\ \vdots \\ \hat{\beta}_n\end{bmatrix} = \begin{bmatrix}\mathrm{Var}\left[\hat{\beta}_1\right] & & \\ & \ddots & \\ & & \mathrm{Var}\left[\hat{\beta}_n\right]\end{bmatrix} = \begin{bmatrix}\left(X_1'\Sigma_1^{-1}X_1\right)^{-1} & & \\ & \ddots & \\ & & \left(X_n'\Sigma_n^{-1}X_n\right)^{-1}\end{bmatrix}$.
When the n blocks of this matrix are equal, the submodels are equally influential in the determination of V. In this situation the corresponding random-effects model is said to be balanced. Both Examples 6 and 7 are balanced. The estimation of V by variance components is particularly suited to balanced models. The estimate of the V of an unbalanced model can be thrown off by the more volatile groups, and can easily end up not being non-negative definite. Nothing precludes positing V by prior information, and this recourse is the more recommended according as the model is the more unbalanced. Also, Appendix F mentions that V can be estimated by maximum likelihood, which despite its complexity is sometimes a useful alternative to variance components.

[9] Compare these estimates with the true values used in the simulation: $\beta_0 = \begin{bmatrix}40.0\% \\ 3.0\%\end{bmatrix}$ and $V = \begin{bmatrix}0.0025 & -0.0001875 \\ -0.0001875 & 0.000225\end{bmatrix}$ (so $\rho = -25\%$). And by generating bivariate normal random vectors with mean β_0 and variance V, the true β_is were:

$$\begin{bmatrix}\beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \beta_5 \\ \beta_6 \\ \beta_7 \\ \beta_8 \\ \beta_9\end{bmatrix} = \begin{bmatrix}47.7\% \\ 5.4\% \\ 41.5\% \\ 3.8\% \\ 47.3\% \\ 3.1\% \\ 39.3\% \\ 0.8\% \\ 44.0\% \\ 5.1\% \\ 37.1\% \\ 2.9\% \\ 46.1\% \\ -0.8\% \\ 38.8\% \\ 1.5\% \\ 38.1\% \\ 3.8\%\end{bmatrix}$$

Normal random variables with a standard deviation of 4.0% were added to the resulting trend lines to form the fifty-four loss ratios of Exhibit 14.

References

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

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

3. Bailey, Arthur, "Credibility Procedures, Laplace's Generalization of Bayes' Rule, and the Combination of Collateral Knowledge with Observed Data," *PCAS* XXXVII, 1950, 7-23.

4. Bailey, Arthur, "Discussion of 'An Introduction to Credibility Theory by L. H. Longley-Cook,'" *PCAS* L, 1963, 59-61. Posthumously reprinted from the Journal of the American Teachers of Insurance (1950), 17-24.

5. Dropkin, Lester, "Some Considerations on Automobile Rating Systems Utilizing Individual Driving Records," *PCAS* LXXIV, 1987, 391-405. Reprinted from *PCAS* XLVI, 1959.

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

7. Halliwell, Leigh J., "Conjoint Prediction of Paid and Incurred Losses," 1997 Loss Reserving Discussion Papers, Casualty Actuarial Society, 1997, 241-379.

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

9. Longley-Cook, L.H., "An Introduction to Credibility Theory," Casualty Actuarial Society, 1962.

10. Mayerson, Allen L. "A Bayesian View of Credibility," *PCAS* LI, 1964, 85-104.

11. Rodermund, Matthew, "Introduction," Foundations of Casualty Actuarial Science, Casualty Actuarial Society, 1990, 1-24.

12. SAS/STAT® Software: Changes and Enhancements through Release 6.12, Cary, NC, SAS Institute Inc., 1997.

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

Exhibit 1

Example 1: The Simplest Statistical Model

$y = X\beta + e$, where $Var[e] = \sigma^2\Phi$

y_1	X_1	Φ_{11}	Φ_{12}	$y_1 - X_1\hat{\beta}$
6.164	1	1 0 0 0 0 0 0	1 0 0 0 0 0 0 0 0 0 0	-3.754
11.103	1	0 1 0 0 0 0 0	0 1 0 0 0 0 0 0 0 0 0	1.185
9.663	1	0 0 1 0 0 0 0	0 0 1 0 0 0 0 0 0 0 0	-0.255
12.998	1	0 0 0 1 0 0 0	0 0 0 1 0 0 0 0 0 0 0	3.080
10.329	1	0 0 0 0 1 0 0	0 0 0 0 1 0 0 0 0 0 0	0.411
9.564	1	0 0 0 0 0 1 0	0 0 0 0 0 1 0 0 0 0 0	-0.353
9.602	1	0 0 0 0 0 0 1	0 0 0 0 0 0 1 0 0 0 0	-0.315

y_2	X_2	Φ_{21}	Φ_{22}
-	0	1 0 0 0 0 0 0	1 0 0 0 0 0 0 0 0 0 0
-	0	0 1 0 0 0 0 0	0 1 0 0 0 0 0 0 0 0 0
-	0	0 0 1 0 0 0 0	0 0 1 0 0 0 0 0 0 0 0
-	0	0 0 0 1 0 0 0	0 0 0 1 0 0 0 0 0 0 0
-	0	0 0 0 0 1 0 0	0 0 0 0 1 0 0 0 0 0 0
-	0	0 0 0 0 0 1 0	0 0 0 0 0 1 0 0 0 0 0
-	0	0 0 0 0 0 0 1	0 0 0 0 0 0 1 0 0 0 0
-	0	0 0 0 0 0 0 0	0 0 0 0 0 0 0 0 0 0 0
-	0	0 0 0 0 0 0 0	0 0 0 0 0 0 0 1 0 0 0
-	1	0 0 0 0 0 0 0	0 0 0 0 0 0 0 0 0 0 0
-	1	0 0 0 0 0 0 0	0 0 0 0 0 0 0 0 0 0 1

$X_1'\Phi_{11}^{-1}y_1$	$X_1'\Phi_{11}^{-1}X_1$	$X_1'\Phi_{11}^{-1}$	$df = t_1 - k$
69.421	7	1 1 1 1 1 1 1	6

$\hat{\beta}$	$(X_1'\Phi_{11}^{-1}X_1)^{-1}$	$\hat{\sigma}^2$
9.917	0.143	4.240

$Var[\hat{\beta}]$

0.606

$$\hat{\beta} = \left(X_1'\Phi_{11}^{-1}X_1\right)^{-1} X_1'\Phi_{11}^{-1}y_1$$

$$Var[\hat{\beta}] = \hat{\sigma}^2\left(X_1'\Phi_{11}^{-1}X_1\right)^{-1}$$

$$\hat{\sigma}^2 = \left(y_1 - X_1\hat{\beta}\right)'\Phi_{11}^{-1}\left(y_1 - X_1\hat{\beta}\right)/df$$

Exhibit 2

Example 1. The Simplest Statistical Model (Cont'd)

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

1
1
1
1
1
1
1
0
0
0
0

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

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

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

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

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

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

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

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

$\hat{\mathbf{y}}_2$

-3.754
1.185
-0.255
3.080
0.411
-0.353
-0.315
0.000
0.000
9.917
9.917

$\mathrm{Var}[\mathbf{y}_2-\hat{\mathbf{y}}_2]$

0.606	0.606	0.606	0.606	0.606	0.606	0.606	0	0	-0.606	-0.606
0.606	0.606	0.606	0.606	0.606	0.606	0.606	0	0	-0.606	-0.606
0.606	0.606	0.606	0.606	0.606	0.606	0.606	0	0	-0.606	-0.606
0.606	0.606	0.606	0.606	0.606	0.606	0.606	0	0	-0.606	-0.606
0.606	0.606	0.606	0.606	0.606	0.606	0.606	0	0	-0.606	-0.606
0.606	0.606	0.606	0.606	0.606	0.606	0.606	0	0	-0.606	-0.606
0.606	0.606	0.606	0.606	0.606	0.606	0.606	0	0	-0.606	-0.606
0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	4.240	0	0
-0.606	-0.606	-0.606	-0.606	-0.606	-0.606	-0.606	0	0	0.606	0.606
-0.606	-0.606	-0.606	-0.606	-0.606	-0.606	-0.606	0	0	0.606	4.846

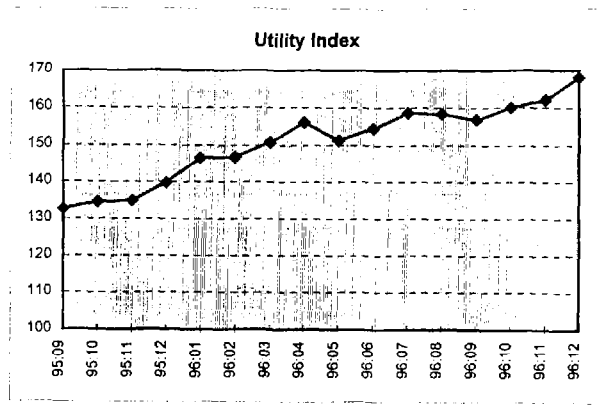
$$\hat{\mathbf{y}}_2 = X_2\hat{\beta} + \Phi_{21}\Phi_{11}^{-1}(\mathbf{y}_1 - X_1\hat{\beta})$$

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

Exhibit 3

Example 2: Expense Model

Month	Index	Expense
95:09	132.545	1,714
95:10	134.440	1,804
95:11	134.820	1,862
95:12	139.690	2,265
96:01	146.572	2,553
96:02	146.745	2,170
96:03	150.687	2,315
96:04	155.983	2,217
96:05	151.240	2,279
96:06	154.417	2,293
96:07	158.616	2,171
96:08	158.302	2,263
96:09	156.779	2,192
96:10	160	
96:11	162	
96:12	168	



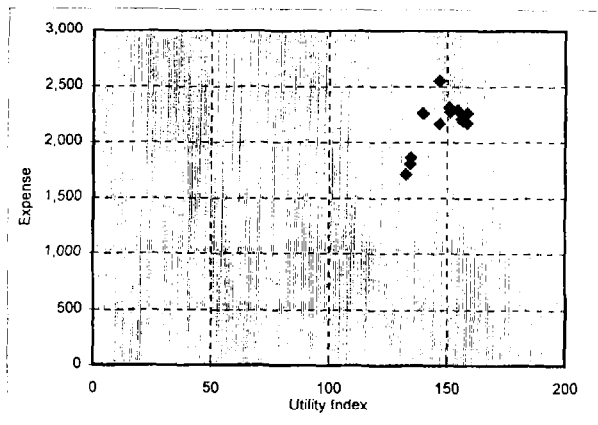


Exhibit 4

Example 2: Expense Model (Cont'd)

$\mathbf{y} = X\beta + \mathbf{e}$, where $\mathrm{Var}[\mathbf{e}] = \sigma^2\Phi$

y_1	X_1	Φ_{11}													Φ_{12}			$y_1 - X_1\hat{\beta}$
1,714	132.545	17568	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-224
1,804	134.440	0	18074	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-161
1,862	134.820	0	0	18176	0	0	0	0	0	0	0	0	0	0	0	0	0	-109
2,265	139.690	0	0	0	19513	0	0	0	0	0	0	0	0	0	0	0	0	223
2,553	146.572	0	0	0	0	21483	0	0	0	0	0	0	0	0	0	0	0	410
2,170	146.745	0	0	0	0	0	21534	0	0	0	0	0	0	0	0	0	0	25
2,315	150.687	0	0	0	0	0	0	22707	0	0	0	0	0	0	0	0	0	112
2,217	155.983	0	0	0	0	0	0	0	24331	0	0	0	0	0	0	0	0	-63
2,279	151.240	0	0	0	0	0	0	0	0	22874	0	0	0	0	0	0	0	68
2,293	154.417	0	0	0	0	0	0	0	0	0	23845	0	0	0	0	0	0	36
2,171	158.616	0	0	0	0	0	0	0	0	0	0	25159	0	0	0	0	0	-148
2,263	158.302	0	0	0	0	0	0	0	0	0	0	0	25060	0	0	0	0	-51
2,192	156.779	0	0	0	0	0	0	0	0	0	0	0	0	24580	0	0	0	-100

y_2	X_2	Φ_{21}													Φ_{22}		
-	160	0	0	0	0	0	0	0	0	0	0	0	0	0	25600	0	0
-	162	0	0	0	0	0	0	0	0	0	0	0	0	0	0	26244	0
-	168	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	28224

$X_1'\Phi_{11}^{-1}y_1$	$X_1'\Phi_{11}^{-1}X_1$	$X_1'\Phi_{11}^{-1}$													$df = t_1 - k$
190.039	13	0.008	0.007	0.007	0.007	0.007	0.007	0.007	0.006	0.007	0.006	0.006	0.006	0.006	12

$\hat{\beta}$	$(X_1'\Phi_{11}^{-1}X_1)^{-1}$	$\hat{\sigma}^2$
14.618	0.077	1.475

$\mathrm{Var}[\hat{\beta}]$
0.113

$\hat{y}_2$
2,339
2,368
2,456

$\mathrm{Var}[y_2 - \hat{y}_2]$		
40672	2941	3050
2941	41695	3089
3050	3089	44841

Exhibit 5

Example 3: Example 1 with Prior Information

$\mathbf{y} = X\beta + \mathbf{e}$, where $\mathrm{Var}[\mathbf{e}] = \sigma^2\Phi$

$\mathbf{y}_1$	X_1	Φ_{11}								$\mathbf{y}_1 - X_1\hat{\beta}$
6.164	1	4.240	0	0	0	0	0	0	0	-3.936
11.103	1	0	4.240	0	0	0	0	0	0	1.004
9.663	1	0	0	4.240	0	0	0	0	0	-0.437
12.998	1	0	0	0	4.240	0	0	0	0	2.898
10.329	1	0	0	0	0	4.240	0	0	0	0.229
9.564	1	0	0	0	0	0	4.240	0	0	-0.535
9.602	1	0	0	0	0	0	0	4.240	0	-0.497
11.000	1	0	0	0	0	0	0	0	3.000	0.901

$X_1'\Phi_{11}^{-1}\mathbf{y}_1$	$X_1'\Phi_{11}^{-1}X_1$	$X_1'\Phi_{11}^{-1}$								$df = t_1 - k$
20.038	1.984155	0.236	0.236	0.236	0.236	0.236	0.236	0.236	0.333	7

$\hat{\beta}$	$(X_1'\Phi_{11}^{-1}X_1)^{-1}$	$\hat{\sigma}^2$
10.099	0.504	0.904

$\mathrm{Var}[\hat{\beta}]$
0.455

Credibility-Weighed Estimate		
Prior	11.000	0.168
Empirical	9.917	Z = 0.832
Posterior	10.099	1.000

Credibility Weight	
EVPV (σ^2)	4.240
VHM (τ^2)	3.000
k = EVPV/VHM	1.413
n	7
$Z = n/(n+k)$	0.832

Exhibit 6

Example 4: Merit Rating

A. Separate Observations

y_1	X_1	Φ_{11}				Φ_{21}	$y_1 - X_1\hat{\beta}$
0	1	0.0625	0	0	0	0	-0.293
1	1	0	0.0625	0	0	0	0.707
0	1	0	0	0.0625	0	0	-0.293
0.25	1	0	0	0	0.0225	0	-0.043

y_2	X_2	Φ_{12}				Φ_{22}	
-	1	0	0	0	0	0.0625	

$X_1'\Phi_{11}^{-1}y_1$	$X_1'\Phi_{11}^{-1}X_1$	$X_1'\Phi_{11}^{-1}$				$df = t_1 - k$
27.111	92.44444	16	16	16	44.444	3

$\hat{\beta}$	$(X_1'\Phi_{11}^{-1}X_1)^{-1}$	$\hat{\sigma}^2$
0.293	0.011	3.609

$Var[\hat{\beta}]$
0.039

$\hat{y}_2$	$Var[y_2 - \hat{y}_2]$
0.293	0.265

B. Summarized Observations

y_1	X_1	Φ_{11}		Φ_{21}	$y_1 - X_1\hat{\beta}$
1	3	0.1875	0	0	0.120
0.25	1	0	0.0225	0	-0.043

y_2	X_2	Φ_{12}		Φ_{22}
-	1	0	0	0.0625

$X_1'\Phi_{11}^{-1}y_1$	$X_1'\Phi_{11}^{-1}X_1$	$X_1'\Phi_{11}^{-1}$		$df = t_1 - k$
27.111	92.44444	16	44.444	1

$\hat{\beta}$	$(X_1'\Phi_{11}^{-1}X_1)^{-1}$	$\hat{\sigma}^2$
0.293	0.011	0.160

$Var[\hat{\beta}]$
0.002

$\hat{y}_2$	$Var[y_2 - \hat{y}_2]$
0.293	0.012

Exhibit 7

Example 5: Stochastic Constraint for Exact Constraint

Year	Age	y	X							Φ	y − Xβ̂
1988	12	268,354	131332							1	32,538
1988	24	166,572		131332						1	-88,469
1988	36	32,329			131332						-133,541
1988	48	53,610				131332					-59,763
1988	60	8,124					131332				-63,069
1988	72	16,924						131332			-44,391
1988	84	39,109							131332		-7,552
1989	12	246,981	141672								-5,244
1989	24	359,380		141672							84,237
1989	36	229,016			141672						50,087
1989	48	89,539				141672					-52,760
1989	60	118,635					141672				41,837
1989	72	100,292						141672			34,150
1990	12	203,178	141677								-49,056
1990	24	375,768		141677							100,615
1990	36	276,617			141677						97,682
1990	48	74,912				141677					-47,392
1990	60	86,428					141677				9,627
1991	12	395,630	142578								141,793
1991	24	260,643		142578							-16,259
1991	36	167,709			142578						-12,364
1991	48	270,692				142578					147,611
1992	12	207,698	143286								-47,399
1992	24	174,615		143286							-103,661
1992	36	162,640			143286						-18,327
1993	12	167,681	138262								-78,472
1993	24	280,178		138262							11,659
1994	12	215,740	121858							1	-1,208
		7.213	1	1	1	1	1	1	1	1E-15	9.92E-07

$X'\Phi^{-1}y$	$X'\Phi^{-1}X$						
7.21E+15	1E+15	1E+15	1E+15	1E+15	1E+15	1E+15	1E+15
7.21E+15	1E+15	1E+15	1E+15	1E+15	1E+15	1E+15	1E+15
7.21E+15	1E+15	1E+15	1E+15	1E+15	1E+15	1E+15	1E+15
7.21E+15	1E+15	1E+15	1E+15	1E+15	1E+15	1E+15	1E+15
7.21E+15	1E+15	1E+15	1E+15	1E+15	1E+15	1E+15	1E+15
7.21E+15	1E+15	1E+15	1E+15	1E+15	1E+15	1E+15	1E+15
7.21E+15	1E+15	1E+15	1E+15	1E+15	1E+15	1E+15	1E+15

$X'\Phi^{-1}$			
1E+05		1E+05	1E+15
	1E+05		1E+15
			1E+15
		…	1E+15
			1E+15
			1E+15
			1E+15

$df = t-k$
22

$\hat{\beta}$	$(X'\Phi^{-1}X)^{-1}$						
1.780	7.16E-12	-4.6E-13	-5.4E-13	-6.9E-13	-9.3E-13	-1.4E-12	-3.1E-12
1.942	-4.6E-13	8.01E-12	-6.1E-13	-7.8E-13	-1.1E-12	-1.6E-12	-3.5E-12
1.263	-5.4E-13	-6.1E-13	9.45E-12	-9.3E-13	-1.3E-12	-1.9E-12	-4.2E-12
0.863	-6.9E-13	-7.8E-13	-9.3E-13	1.17E-11	-1.6E-12	-2.4E-12	-5.3E-12
0.542	-9.3E-13	-1.1E-12	-1.3E-12	-1.6E-12	1.53E-11	-3.3E-12	-7.1E-12
0.467	-1.4E-12	-1.6E-12	-1.9E-12	-2.4E-12	-3.3E-12	2.17E-11	-1.1E-11
0.355	-3.1E-12	-3.5E-12	-4.2E-12	-5.3E-12	-7.1E-12	-1.1E-11	3.42E-11
7.213							

$\hat{\sigma}^2$
6.27E+09

$Var[\hat{\beta}]$

0.04490	-0.00286	-0.00342	-0.00432	-0.00585	-0.00899	-0.01946
-0.00286	0.05021	-0.00385	-0.00488	-0.00659	-0.01013	-0.02192
-0.00342	-0.00385	0.05924	-0.00581	-0.00787	-0.01210	-0.02619
-0.00432	-0.00486	-0.00581	0.07335	-0.00995	-0.01530	-0.03310
-0.00585	-0.00659	-0.00787	-0.00995	0.09581	-0.02072	-0.04483
-0.00899	-0.01013	-0.01210	-0.01530	-0.02072	0.13619	-0.06894
-0.01946	-0.02192	-0.02619	-0.03310	-0.04483	-0.06894	0.21445

Exhibit 8

Example 5 Stochastic Constraint for Exact Constraint (Cont'd)

Year	Age	y	X							Φ							y - Xβ
1988	12	266,354	131332							1							33,523
1988	24	166,572		131332							1						-87,379
1988	36	32,329			131332							1					-132,215
1988	48	53,610				131332							1				-58,087
1988	60	8,124					131332							1			-60,799
1988	72	16,924						131332							1		-40,900
1988	84	39,109							131332							1	-0
1989	12	246,981	141672														-4,181
1989	24	359,380		141672													85,434
1989	36	229,016			141672												51,517
1989	48	69,539				141672											-50,952
1989	60	118,635					141672										44,285
1989	72	100,292						141672									37,915
1990	12	203,178	141677														-47,993
1990	24	375,768		141677													101,813
1990	36	276,617			141677												99,112
1990	48	74,912				141677											-45,584
1990	60	86,428					141677										12,076
1991	12	395,630	142578														142,862
1991	24	260,643		142578													-15,054
1991	36	167,709			142578												-10,925
1991	48	270,892				142578											149,430
1992	12	207,698	143286														-46,324
1992	24	174,615		143286													-102,450
1992	36	162,640			143286												-16,880
1993	12	167,681	138262														-77,435
1993	24	280,178		138262													12,827
1994	12	215,740	121858														-294
		7,213	1	1	1	1	1	1	1							1	0 140171

$X'\Phi^{-1}y$

2 34E+11
2 27E+11
1 23E+11
6 61E+10
3 01E+10
1 64E+10
5 14E+09

$X'\Phi^{-1}X$

1.32E+11	1	1	1	1	1	1
1	1 17E+11	1	1	1	1	1
1	1	9 83E+10	1	1	1	1
1	1	1	7 77E+10	1	1	1
1	1	1	1	5 74E+10	1	1
1	1	1	1	1	3 73E+10	1
1	1	1	1	1	1	1.72E+10

$X'\Phi^{-1}$

1E+05	1E+05	1
1E+05		1
		1
		1
		1
		1
		1

$df = t\text{-}k$

22

$\hat{\beta}$

1.773
1 934
1 253
0 850
0 525
0 440
0 298
7 073

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

7 56E-12	-6 4E-23	-7 7E-23	-9 7E-23	-1 3E-22	-2E-22	-4 4E-22
-6 4E-23	8 52E-12	-8 7E-23	-1 1E-22	-1 5E-22	-2 3E-22	-4 9E-22
-7 7E-23	-8 7E-23	1 02E-11	-1 3E-22	-1 8E-22	-2 7E-22	-5 9E-22
-9 7E-23	-1 1E-22	-1 3E-22	1 29E-11	-2 2E-22	-3 4E-22	-7 5E-22
-1 3E-22	-1 5E-22	-1 8E-22	-2 2E-22	1 74E-11	-4 7E-22	-1E-21
-2E-22	-2 3E-22	-2 7E-22	-3 4E-22	-4 7E-22	2 68E-11	-1 6E-21
-4 4E-22	-4 9E-22	-5 9E-22	-7 5E-22	-1E-21	-1 6E-21	5 8E-11

$\hat{\sigma}^2$

6 27E+09

$Var[\hat{\beta}]$

0 04739	0.00000	0 00000	0 00000	0 00000	0 00000	0 00000
0 00000	0 05338	0 00000	0 00000	0 00000	0 00000	0.00000
0 00000	0 00000	0 06377	0 00000	0.00000	0 00000	0 00000
0.00000	0 00000	0 00000	0 08061	0.00000	0 00000	0 00000
0.00000	0 00000	0 00000	0 00000	0.10917	0 00000	0 00000
0.00000	0 00000	0.00000	0 00000	0.00000	0.16789	0.00000
0 00000	0 00000	0 00000	0.00000	0 00000	0 00000	0.36325

Exhibit 9

Example 6: The Simplest Random-Effects Model

Preliminary Fixed-Effects Model

	y	X								$y - X\hat{\beta}$
1	0.430	1								-0.371
1	0.375	1								-0.426
1	2.341	1								1.541
1	0.175	1								-0.626
1	1.016	1								0.216
1	0.466	1								-0.335
2	0.247		1							-0.553
2	1.587		1							0.787
2	1.939		1							1.139
2	0.712		1							-0.088
2	0.054		1							-0.746
2	0.261		1							-0.539
3	0.661			1						0.242
3	0.237			1						-0.182
3	0.063			1						-0.356
3	0.250			1						-0.169
3	0.602			1						0.183
3	0.700			1						0.281
4	0.182				1					0.043
4	0.351				1					0.212
4	0.011				1					-0.129
4	0.022				1					-0.118
4	0.019				1					-0.121
4	0.252				1					0.113
5	0.311					1				-0.504
5	0.664					1				-0.151
5	1.002					1				0.188
5	0.038					1				-0.777
5	0.370					1				-0.445
5	2.502					1				1.688
6	0.301						1			-0.316
6	0.253						1			-0.364
6	0.044						1			-0.573
6	0.109						1			-0.508
6	2.105						1			1.488
6	0.891						1			0.274
7	0.219							1		-0.495
7	1.186							1		0.472
7	0.431							1		-0.283
7	1.405							1		0.691
7	0.241							1		-0.473
7	0.804							1		0.090
8	0.002								1	-0.204
8	0.058								1	-0.148
8	0.235								1	0.029
8	0.018								1	-0.188
8	0.713								1	0.507
8	0.208								1	0.002
9	0.796									0.242
9	0.260									-0.294
9	0.932									0.378
9	0.857									0.303
9	0.129									-0.425
9	0.349									-0.205

Exhibit 10

Example 6: The Simplest Random-Effects Model (Cont'd)

Solution of Fixed Effects

X'y	X'X								
4.803	6	0	0	0	0	0	0	0	0
4.800	0	6	0	0	0	0	0	0	0
2.513	0	0	6	0	0	0	0	0	0
0.837	0	0	0	6	0	0	0	0	0
4.887	0	0	0	0	6	0	0	0	0
3.703	0	0	0	0	0	6	0	0	0
4.286	0	0	0	0	0	0	6	0	0
1.234	0	0	0	0	0	0	0	6	0
3.323	0	0	0	0	0	0	0	0	6

$\hat{\beta}$	$(X'X)^{-1}$								
0.801	0.16667	0	0	0	0	0	0	0	0
0.800	0	0.16667	0	0	0	0	0	0	0
0.419	0	0	0.16667	0	0	0	0	0	0
0.140	0	0	0	0.16667	0	0	0	0	0
0.815	0	0	0	0	0.16667	0	0	0	0
0.617	0	0	0	0	0	0.16667	0	0	0
0.714	0	0	0	0	0	0	0.16667	0	0
0.206	0	0	0	0	0	0	0	0.16667	0
0.554	0	0	0	0	0	0	0	0	0.16667

t	54
k	9
$df = t - k$	45
$\hat{\sigma}^2$	0.357

$Var[\hat{\beta}]$

0.0595	0	0	0	0	0	0	0	0
0	0.0595	0	0	0	0	0	0	0
0	0	0.0595	0	0	0	0	0	0
0	0	0	0.0595	0	0	0	0	0
0	0	0	0	0.0595	0	0	0	0
0	0	0	0	0	0.0595	0	0	0
0	0	0	0	0	0	0.0595	0	0
0	0	0	0	0	0	0	0.0595	0
0	0	0	0	0	0	0	0	0.0595

Exhibit 11

Example 6: The Simplest Random-Effects Model (Cont'd)

Estimation of the Variance of the Random Effects

$\hat{\beta}$	$\hat{\beta} - \hat{\beta}_0$	$(\hat{\beta} - \hat{\beta}_0)(\hat{\beta} - \hat{\beta}_0)'$	Var[**v**+**e**]	– Var[**e**]	= Var[**v**]
0.801	0.238	0.0565	0.0662	0.0595	0.0067
0.800	0.237	0.0563	0.0662	0.0595	0.0067
0.419	-0.144	0.0207	0.0662	0.0595	0.0067
0.140	-0.423	0.1791	0.0662	0.0595	0.0067
0.815	0.252	0.0634	0.0662	0.0595	0.0067
0.617	0.054	0.003	0.0662	0.0595	0.0067
0.714	0.152	0.023	0.0662	0.0595	0.0067
0.206	-0.357	0.1275	0.0662	0.0595	0.0067
0.554	-0.009	8E-05	0.0662	0.0595	0.0067

	$\hat{\beta}_0$	Var[**v**+**e**]	V = Var[**v**]
Mean:	0.563	0.0662	0.0067

Credibility Weight	
EVPV (σ^2)	0.3570
VHM (V)	0.0067
k = EVPV/VHM	53.332
n	6
$Z = n/(n+k)$	10.1%

Φ for each group

0.3637	0.0067	0.0067	0.0067	0.0067	0.0067
0.0067	0.3637	0.0067	0.0067	0.0067	0.0067
0.0067	0.0067	0.3637	0.0067	0.0067	0.0067
0.0067	0.0067	0.0067	0.3637	0.0067	0.0067
0.0067	0.0067	0.0067	0.0067	0.3637	0.0067
0.0067	0.0067	0.0067	0.0067	0.0067	0.3637

Φ^{-1} for each group

2.754	-0.047	-0.047	-0.047	-0.047	-0.047
-0.047	2.754	-0.047	-0.047	-0.047	-0.047
-0.047	-0.047	2.754	-0.047	-0.047	-0.047
-0.047	-0.047	-0.047	2.754	-0.047	-0.047
-0.047	-0.047	-0.047	-0.047	2.754	-0.047
-0.047	-0.047	-0.047	-0.047	-0.047	2.754

Exhibit 12

Example 6. The Simplest Random-Effects Model (Cont'd)

Estimation of the General Mean β_0

	y_i	X_i	$y_i - X_i\hat{\beta}_0$	Φ_{11}^{-1}							
1	0.430	1	-0.133	2.753812	-0.04721	-0.04721	-0.04721	-0.04721	-0.04721	0	0
1	0.375	1	-0.188	-0.04721	2.753812	-0.04721	-0.04721	-0.04721	-0.04721	0	0
1	2.341	1	1.778	-0.04721	-0.04721	2.753812	-0.04721	-0.04721	-0.04721	0	0
1	0.175	1	-0.388	-0.04721	-0.04721	-0.04721	2.753812	-0.04721	-0.04721	0	0
1	1.016	1	0.453	-0.04721	-0.04721	-0.04721	-0.04721	2.753812	-0.04721	0	0
1	0.466	1	-0.097	-0.04721	-0.04721	-0.04721	-0.04721	-0.04721	2.753812	0	0
2	0.247	1	-0.316	0	0	0	0	0	0	2.753812	-0.04721
2	1.587	1	1.024	0	0	0	0	0	0	-0.04721	2.753812
2	1.939	1	1.376	0	0	0	0	0	0	-0.04721	-0.04721
2	0.712	1	0.149	0	0	0	0	0	0	-0.04721	-0.04721
2	0.054	1	-0.509	0	0	0	0	0	0	-0.04721	-0.04721
2	0.261	1	-0.302	0	0	0	0	0	0	-0.04721	-0.04721
3	0.661	1	0.098	0	0	0	0	0	0	0	0
3	0.237	1	-0.326	0	0	0	0	0	0	0	0
3	0.063	1	-0.500	0	0	0	0	0	0	0	0
3	0.250	1	-0.313	0	0	0	0	0	0	0	0
3	0.602	1	0.039	0	0	0	0	0	0	0	0
3	0.700	1	0.137	0	0	0	0	0	0	0	0
4	0.182	1	-0.381	0	0	0	0	0	0	0	0
4	0.351	1	-0.212	0	0	0	0	0	0	0	0
4	0.011	1	-0.552	0	0	0	0	0	0	0	0
4	0.022	1	-0.541	0	0	0	0	0	0	0	0
4	0.019	1	-0.544	0	0	0	0	0	0	0	0
4	0.252	1	-0.311	0	0	0	0	0	0	0	0
5	0.311	1	-0.252	0	0	0	0	0	0	0	0
5	0.664	1	0.101	0	0	0	0	0	0	0	0
5	1.002	1	0.439	0	0	0	0	0	0	0	0
5	0.038	1	-0.525	0	0	0	0	0	0	0	0
5	0.370	1	-0.193	0	0	0	0	0	0	0	0
5	2.502	1	1.939	0	0	0	0	0	0	0	0
6	0.301	1	-0.262	0	0	0	0	0	0	0	0
6	0.253	1	-0.310	0	0	0	0	0	0	0	0
6	0.044	1	-0.519	0	0	0	0	0	0	0	0
6	0.109	1	-0.454	0	0	0	0	0	0	0	0
6	2.105	1	1.542	0	0	0	0	0	0	0	0
6	0.891	1	0.328	0	0	0	0	0	0	0	0
7	0.219	1	-0.344	0	0	0	0	0	0	0	0
7	1.186	1	0.623	0	0	0	0	0	0	0	0
7	0.431	1	-0.132	0	0	0	0	0	0	0	0
7	1.405	1	0.842	0	0	0	0	0	0	0	0
7	0.241	1	-0.322	0	0	0	0	0	0	0	0
7	0.804	1	0.241	0	0	0	0	0	0	0	0
8	0.002	1	-0.561	0	0	0	0	0	0	0	0
8	0.058	1	-0.505	0	0	0	0	0	0	0	0
8	0.235	1	-0.328	0	0	0	0	0	0	0	0
8	0.018	1	-0.545	0	0	0	0	0	0	0	0
8	0.713	1	0.150	0	0	0	0	0	0	0	0
8	0.208	1	-0.355	0	0	0	0	0	0	0	0
9	0.796	1	0.233	0	0	0	0	0	0	0	0
9	0.260	1	-0.303	0	0	0	0	0	0	0	0
9	0.932	1	0.369	0	0	0	0	0	0	0	0
9	0.857	1	0.294	0	0	0	0	0	0	0	0
9	0.129	1	-0.434	0	0	0	0	0	0	0	0
9	0.349	1	-0.214	0	0	0	0	0	0	0	0

$X_i'\Phi_{11}^{-1}y_i$	$X_i'\Phi_{11}^{-1}X_i$	df	$X_i'\Phi_{11}^{-1}$							
76.505	135.96	53	2.517766	2.517766	2.517766	2.517766	2.517766	2.517766	2.517766	2.517766

$\hat{\beta}_0$	$(X_i'\Phi_{11}^{-1}X_i)^{-1}$	$\hat{\sigma}^2$
0.563	0.0074	1.000

$\text{Var}[\hat{\beta}_0]$

0.0074

Exhibit 13

Example 6: The Simplest Random-Effects Model (Cont'd)

Estimation of the Group Means

	y_2
	-
1	-
2	-
3	-
4	-
5	-
6	-
7	-
8	-
9	-

X_2
1
1
1
1
1
1
1
1
1
1

Φ_{21}

	0	0	0	0	0	0	0	0
1	0.0067	0.0067	0.0067	0.0067	0.0067	0.0067	0	0
2	0	0	0	0	0	0	0.0067	0.0067
3	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	0

	$\hat{y}_2$
	0.563
1	0.587
2	0.587
3	0.548
4	0.520
5	0.588
6	0.568
7	0.578
8	0.527
9	0.562

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

	0	0	0	0	0	0	0	0
1	0.016854	0.016854	0.016854	0.016854	0.016854	0.016854	0	0
2	0	0	0	0	0	0	0.016854	0.016854
3	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	0

Credibility-Weighed Estimates					
Group	1 - Z	Prior	Z	Empirical	Posterior
1	0.899	0.563	0.101	0.801	0.587
2		0.563		0.800	0.587
3		0.563		0.419	0.548
4		0.563		0.140	0.520
5		0.563		0.815	0.588
6		0.563		0.617	0.568
7		0.563		0.714	0.578
8		0.563		0.206	0.527
9		0.563		0.554	0.562
Unweighted Mean					0.563

Exhibit 14

Example 7: Random-Effects Trend Model

Preliminary Fixed-Effects Model

State	Year	Loss Ratio y	X																								$y - X\hat{\beta}$
A	1	54.3%	1	1																							1.8%
A	2	57.2%	1	2																							-1.1%
A	3	64.6%	1	3																							0.6%
A	4	67.6%	1	4																							-2.2%
A	5	73.5%	1	5																							-2.0%
A	6	84.1%	1	6																							2.9%
B	1	44.2%			1	1																					0.6%
B	2	48.6%			1	2																					1.0%
B	3	54.8%			1	3																					3.2%
B	4	48.2%			1	4																					-7.4%
B	5	57.7%			1	5																					-1.9%
B	6	68.2%			1	6																					4.5%
C	1	53.9%					1	1																			0.1%
C	2	57.0%					1	2																			1.9%
C	3	54.8%					1	3																			-1.8%
C	4	59.9%					1	4																			2.0%
C	5	52.7%					1	5																			-6.6%
C	6	65.2%					1	6																			4.5%
D	1	41.8%							1	1																	-2.0%
D	2	45.2%							1	2																	1.2%
D	3	45.1%							1	3																	0.8%
D	4	48.4%							1	4																	1.9%
D	5	43.9%							1	5																	-0.9%
D	6	44.0%							1	6																	-1.0%
E	1	46.3%									1	1															1.5%
E	2	48.6%									1	2															-2.4%
E	3	57.6%									1	3															0.5%
E	4	63.3%									1	4															0.1%
E	5	69.6%									1	5															0.3%
E	6	75.4%									1	6															-0.1%
F	1	46.9%											1	1													3.6%
F	2	38.4%											1	2													-6.3%
F	3	48.1%											1	3													1.9%
F	4	46.0%											1	4													-1.6%
F	5	53.4%											1	5													4.4%
F	6	48.2%											1	6													-2.1%
G	1	45.7%													1	1											-0.6%
G	2	44.5%													1	2											-0.8%
G	3	44.2%													1	3											-0.2%
G	4	46.7%													1	4											3.2%
G	5	43.2%													1	5											0.6%
G	6	39.5%													1	6											-2.2%
H	1	38.2%															1	1									0.8%
H	2	42.0%															1	2									2.3%
H	3	36.8%															1	3									-5.4%
H	4	46.1%															1	4									1.4%
H	5	47.3%															1	5									0.2%
H	6	50.2%															1	6									0.7%
I	1	43.1%																	1	1							3.2%
I	2	44.8%																	1	2							0.8%
I	3	47.3%																	1	3							-0.9%
I	4	46.4%																	1	4							-6.0%
I	5	52.5%																	1	5							-4.1%
I	6	67.9%																	1	6							7.1%

Exhibit 15

Example 7: Random-Effects Trend Model (Cont'd)

Loss Ratios by State and Year

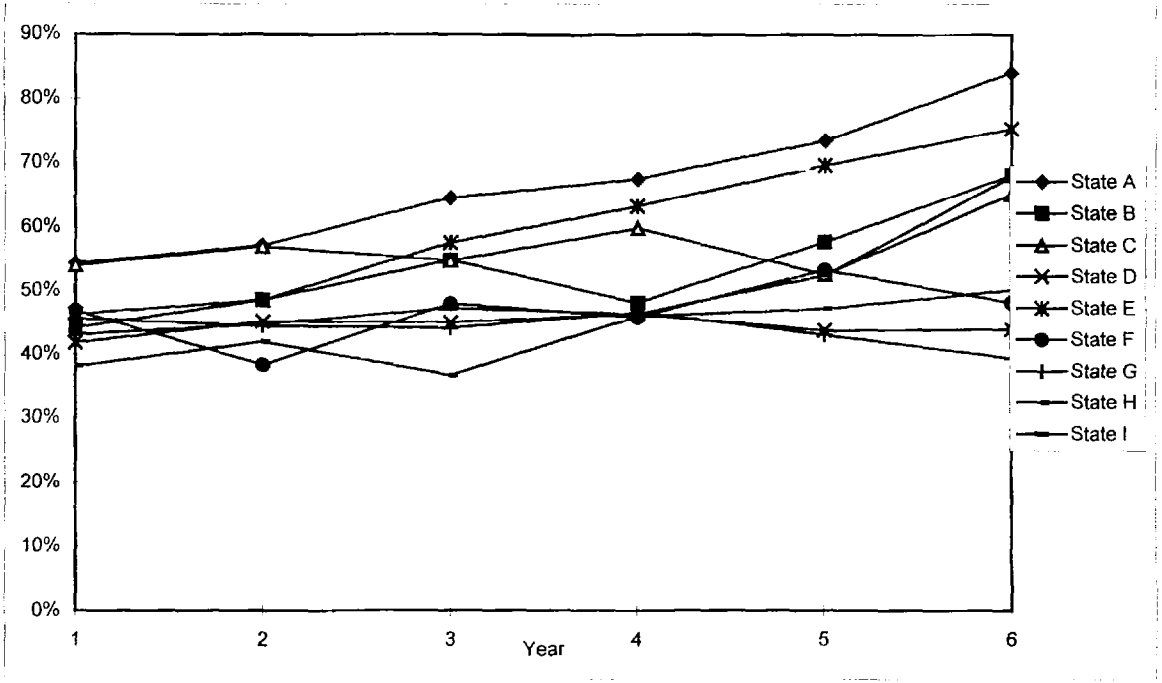


Exhibit 16

Example 7 Random-Effects Trend Model (Cont'd)

Solution of Fixed Effects

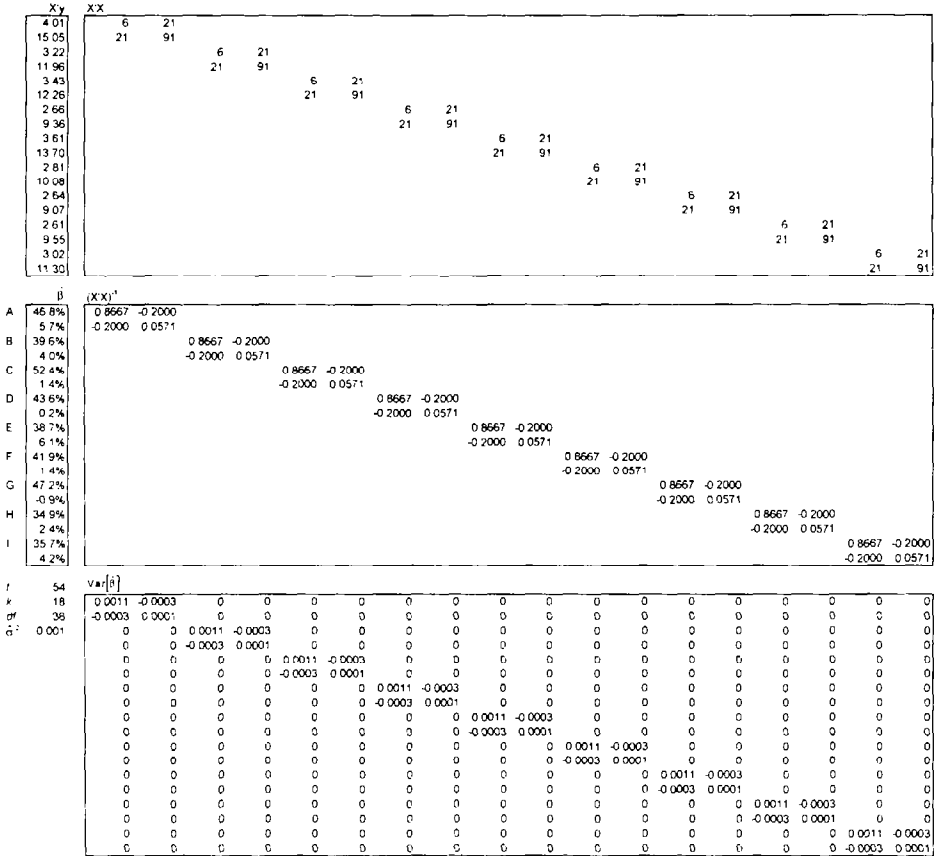


Exhibit 17

Example 7: Random-Effects Trend Model (Cont'd)

Estimation of the Variance of the Random Effects

	$\hat{\beta}$	$\hat{\beta} - \hat{\beta}_o$	$(\hat{\beta} - \hat{\beta}_o)(\hat{\beta} - \hat{\beta}_o)'$		Var[v+e]		– Var[e]		= Var[v]	
A	46.8%	4.5%	0.0020	0.0013	0.0033	-0.0005	0.0011	-0.0003	0.0022	-0.0003
	5.7%	3.0%	0.0013	0.0009	-0.0005	0.0006	-0.0003	0.0001	-0.0003	0.0005
B	39.6%	-2.7%	0.0007	-0.0003	0.0033	-0.0005	0.0011	-0.0003	0.0022	-0.0003
	4.0%	1.3%	-0.0003	0.0002	-0.0005	0.0006	-0.0003	0.0001	-0.0003	0.0005
C	52.4%	10.1%	0.0102	-0.0014	0.0033	-0.0005	0.0011	-0.0003	0.0022	-0.0003
	1.4%	-1.3%	-0.0014	0.0002	-0.0005	0.0006	-0.0003	0.0001	-0.0003	0.0005
D	43.6%	1.3%	0.0002	-0.0003	0.0033	-0.0005	0.0011	-0.0003	0.0022	-0.0003
	0.2%	-2.5%	-0.0003	0.0006	-0.0005	0.0006	-0.0003	0.0001	-0.0003	0.0005
E	38.7%	-3.6%	0.0013	-0.0012	0.0033	-0.0005	0.0011	-0.0003	0.0022	-0.0003
	6.1%	3.4%	-0.0012	0.0011	-0.0005	0.0006	-0.0003	0.0001	-0.0003	0.0005
F	41.9%	-0.4%	0.0000	0.0001	0.0033	-0.0005	0.0011	-0.0003	0.0022	-0.0003
	1.4%	-1.3%	0.0001	0.0002	-0.0005	0.0006	-0.0003	0.0001	-0.0003	0.0005
G	47.2%	4.9%	0.0024	-0.0018	0.0033	-0.0005	0.0011	-0.0003	0.0022	-0.0003
	-0.9%	-3.7%	-0.0018	0.0013	-0.0005	0.0006	-0.0003	0.0001	-0.0003	0.0005
H	34.9%	-7.4%	0.0055	0.0002	0.0033	-0.0005	0.0011	-0.0003	0.0022	-0.0003
	2.4%	-0.3%	0.0002	0.0000	-0.0005	0.0006	-0.0003	0.0001	-0.0003	0.0005
I	35.7%	-6.6%	0.0043	-0.0010	0.0033	-0.0005	0.0011	-0.0003	0.0022	-0.0003
	4.2%	1.4%	-0.0010	0.0002	-0.0005	0.0006	-0.0003	0.0001	-0.0003	0.0005

	$\hat{\beta}_o$		Var[v+e]	
Mean	42.3%		0.0033	-0.0005
	2.7%		-0.0005	0.0006

V = Var[v]	
0.0022	-0.0003
-0.0003	0.0005

$\left(\Phi_i = X_i V X_i' + \hat{\sigma}^2 I_6\right)$

Φ for each group

0.0034	0.0024	0.0026	0.0029	0.0031	0.0033
0.0024	0.0044	0.0039	0.0046	0.0054	0.0062
0.0026	0.0039	0.0064	0.0064	0.0077	0.0090
0.0029	0.0046	0.0064	0.0095	0.0100	0.0118
0.0031	0.0054	0.0077	0.0100	0.0136	0.0147
0.0033	0.0062	0.0090	0.0118	0.0147	0.0188

Φ^{-1} for each group

509.28	-204.33	-134.11	-63.89	6.34	76.56
-204.33	618.15	-127.04	-88.39	-49.75	-11.10
-134.11	-127.04	663.87	-112.90	-105.83	-98.76
-63.89	-88.39	-112.90	646.43	-161.91	-186.42
6.34	-49.75	-105.83	-161.91	565.84	-274.08
76.56	-11.10	-98.76	-186.42	-274.08	422.10

Exhibit 18

Example 7: Random-Effects Trend Model (Cont'd)

Estimation of the General Parameter β_0

State	Year	y_i	X_i		$y_i - X_i\hat{\beta}$	Φ_{ii}^{-1}						
A	1	54.3%	1	1	9.2%	509.28	-204.33	-134.11	-63.89	6.34	76.56	0
A	2	57.2%	1	2	9.4%	-204.33	618.15	-127.04	-88.39	-49.75	-11.10	0
A	3	64.6%	1	3	14.1%	-134.11	-127.04	663.87	-112.90	-105.83	-98.76	0
A	4	67.6%	1	4	14.3%	-63.89	-88.39	-112.90	646.43	-161.91	-186.42	0
A	5	73.5%	1	5	17.5%	6.34	-49.75	-105.83	-161.91	585.84	-274.08	0
A	6	84.1%	1	6	25.4%	76.56	-11.10	-98.76	-186.42	-274.08	422.10	0
B	1	44.2%	1	1	-0.8%	0	0	0	0	0	0	509.28
B	2	48.6%	1	2	0.8%	0	0	0	0	0	0	-204.33
B	3	54.8%	1	3	4.3%	0	0	0	0	0	0	-134.11
B	4	48.2%	1	4	-5.0%	0	0	0	0	0	0	-63.89
B	5	57.7%	1	5	1.8%	0	0	0	0	0	0	6.34
B	6	68.2%	1	6	9.5%	0	0	0	0	0	0	76.56
C	1	53.9%	1	1	8.9%	0	0	0	0	0	0	0
C	2	57.0%	1	2	9.2%	0	0	0	0	0	0	0
C	3	54.8%	1	3	4.2%	0	0	0	0	0	0	0
C	4	59.9%	1	4	6.7%	0	0	0	0	0	0	0
C	5	52.7%	1	5	-3.3%	0	0	0	0	0	0	0
C	6	65.2%	1	6	6.5%	0	0	0	0	0	0	0
D	1	41.8%	1	1	-3.2%	0	0	0	0	0	0	0
D	2	45.2%	1	2	-2.6%	0	0	0	0	0	0	0
D	3	45.1%	1	3	-5.4%	0	0	0	0	0	0	0
D	4	46.4%	1	4	-6.8%	0	0	0	0	0	0	0
D	5	43.9%	1	5	-12.1%	0	0	0	0	0	0	0
D	6	44.0%	1	6	-14.7%	0	0	0	0	0	0	0
E	1	46.3%	1	1	1.3%	0	0	0	0	0	0	0
E	2	48.6%	1	2	0.8%	0	0	0	0	0	0	0
E	3	57.6%	1	3	7.1%	0	0	0	0	0	0	0
E	4	63.3%	1	4	10.1%	0	0	0	0	0	0	0
E	5	69.6%	1	5	13.7%	0	0	0	0	0	0	0
E	6	75.4%	1	6	16.7%	0	0	0	0	0	0	0
F	1	46.9%	1	1	1.8%	0	0	0	0	0	0	0
F	2	38.4%	1	2	-9.3%	0	0	0	0	0	0	0
F	3	48.1%	1	3	-2.5%	0	0	0	0	0	0	0
F	4	46.0%	1	4	-7.3%	0	0	0	0	0	0	0
F	5	53.4%	1	5	-2.6%	0	0	0	0	0	0	0
F	6	48.2%	1	6	-10.5%	0	0	0	0	0	0	0
G	1	45.7%	1	1	0.6%	0	0	0	0	0	0	0
G	2	44.5%	1	2	-3.3%	0	0	0	0	0	0	0
G	3	44.2%	1	3	-6.3%	0	0	0	0	0	0	0
G	4	48.7%	1	4	-6.5%	0	0	0	0	0	0	0
G	5	43.2%	1	5	-12.8%	0	0	0	0	0	0	0
G	6	39.5%	1	6	-19.2%	0	0	0	0	0	0	0
H	1	38.2%	1	1	-6.9%	0	0	0	0	0	0	0
H	2	42.0%	1	2	-5.7%	0	0	0	0	0	0	0
H	3	36.8%	1	3	-13.7%	0	0	0	0	0	0	0
H	4	46.1%	1	4	-7.2%	0	0	0	0	0	0	0
H	5	47.3%	1	5	-8.7%	0	0	0	0	0	0	0
H	6	50.2%	1	6	-8.5%	0	0	0	0	0	0	0
I	1	43.1%	1	1	-2.0%	0	0	0	0	0	0	0
I	2	44.8%	1	2	-2.9%	0	0	0	0	0	0	0
I	3	47.3%	1	3	-3.2%	0	0	0	0	0	0	0
I	4	48.4%	1	4	-6.8%	0	0	0	0	0	0	0
I	5	52.5%	1	5	-3.5%	0	0	0	0	0	0	0
I	6	67.9%	1	6	9.2%	0	0	0	0	0	0	0

$X_i'\Phi_{ii}^{-1}y_i$	$X_i'\Phi_{ii}^{-1}X_i$		df	$X_i'\Phi_{ii}^{-1}$						
1429.4	3190	2925	52	189.853	137.541	85.228	32.915	-19.398	-71.711	189.853
1724.4	2925	17823		-66.221	-18.064	30.092	78.248	126.404	174.560	-66.221

$\hat{\beta}_0$	$(X_i'\Phi_{ii}^{-1}X_i)^{-1}$		$\hat{\sigma}^2$
42.3%	0.0004	-6E-05	1.000
2.7%	-6E-05	7E-05	

Exhibit 19

Example 7: Random-Effects Trend Model (Cont'd)

Estimation of the Group Parameters

Estimations	
State	y_2
A	-
	-
B	-
	-
C	-
	-
D	-
	-
E	-
	-
F	-
	-
G	-
	-
H	-
	-
I	-
	-

X_2

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

Φ_{21}

0.0019	0.0016	0.0013	0.0011	0.0008	0.0005	0
0.0002	0.0008	0.0013	0.0018	0.0023	0.0028	0
0	0	0	0	0	0	0.0019
0	0	0	0	0	0	0.0002
0	0	0	0	0	0	0
0	0	0	0	0	0	0
0	0	0	0	0	0	0
0	0	0	0	0	0	0
0	0	0	0	0	0	0
0	0	0	0	0	0	0
0	0	0	0	0	0	0
0	0	0	0	0	0	0
0	0	0	0	0	0	0
0	0	0	0	0	0	0
0	0	0	0	0	0	0
0	0	0	0	0	0	0
0	0	0	0	0	0	0
0	0	0	0	0	0	0

$$\hat{y}_2 = X_2\hat{\beta}_0 + \Phi_{21}\Phi_{11}^{-1}\left(y_1 - X_1\hat{\beta}_0\right)$$

A	45.8%
	5.9%
B	40.6%
	3.8%
C	49.1%
	2.1%
D	42.8%
	0.5%
E	40.3%
	5.7%
F	41.8%
	1.5%
G	45.2%
	-0.4%
H	37.1%
	2.0%
I	38.0%
	3.6%

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

0.4399	0.31	0.1801	0.0502	-0.0796	-0.2095	0
-0.0896	-0.0493	-0.009	0.0313	0.0715	0.1118	0
0	0	0	0	0	0	0.4399
0	0	0	0	0	0	-0.0896
0	0	0	0	0	0	0
0	0	0	0	0	0	0
0	0	0	0	0	0	0
0	0	0	0	0	0	0
0	0	0	0	0	0	0
0	0	0	0	0	0	0
0	0	0	0	0	0	0
0	0	0	0	0	0	0
0	0	0	0	0	0	0
0	0	0	0	0	0	0
0	0	0	0	0	0	0
0	0	0	0	0	0	0
0	0	0	0	0	0	0
0	0	0	0	0	0	0

Exhibit 20

Example 7: Random-Effects Trend Model (Cont'd)

Credibility-Weighted Estimates

X_i

1	1
1	2
1	3
1	4
1	5
1	6

$T_i^{-1} = (X_i V X_i' + \Sigma_i)^{-1}$

509.28	-204.33	-134.11	-63.89	6.34	76.56
-204.33	618.15	-127.04	-88.39	-49.75	-11.10
-134.11	-127.04	663.87	-112.90	-105.83	-98.76
-63.89	-88.39	-112.90	646.43	-161.91	-186.42
6.34	-49.75	-105.83	-161.91	565.84	-274.08
76.56	-11.10	-98.76	-186.42	-274.08	422.10

$X_i' T_i^{-1} X_i$

354.43	325.02
325.02	1980.29

$Z_i = V X_i' T_i^{-1} X_i$

0.691	0.146
0.067	0.939

State (i)	$I_2 - Z_i$		Prior	Z_i		Empirical	Posterior
A	0.309	-0.146	42.3%	0.691	0.146	46.8%	45.8%
	-0.067	0.061	2.7%	0.067	0.939	5.7%	5.9%
B			42.3%			39.6%	40.6%
			2.7%			4.0%	3.8%
C			42.3%			52.4%	49.1%
			2.7%			1.4%	2.1%
D			42.3%			43.6%	42.8%
			2.7%			0.2%	0.5%
E			42.3%			38.7%	40.3%
			2.7%			6.1%	5.7%
F			42.3%			41.9%	41.8%
			2.7%			1.4%	1.5%
G			42.3%			47.2%	45.2%
			2.7%			-0.9%	-0.4%
H			42.3%			34.9%	37.1%
			2.7%			2.4%	2.0%
I			42.3%			35.7%	38.0%
			2.7%			4.2%	3.6%
Unweighted Mean							42.3%
							2.7%

Appendix A

Groups of Statistical Models

The basis of credibility is a grand statistical model which is a group of statistical submodels. Suppose that we have n linear models of the form $\mathbf{y}_i = \mathrm{X}_i\beta_i + \mathbf{e}_i$, where $\mathrm{Var}[\mathbf{e}_i] = \Sigma_i$, for $i = 1, \ldots, n$. As to the dimensions of the matrices, $\mathbf{y}_i$ and $\mathbf{e}_i$ are $(t_i \times 1)$, X_i is $(t_i \times k_i)$, β_i is $(k_i \times 1)$, and Σ_i is $(t_i \times t_i)$. We assume that each Σ_i is non-singular and that each X_i is of full column rank, i.e., $\mathrm{rank}(\mathrm{X}_i) = k_i$. These assumptions ensure that each $\mathrm{X}_i'\Sigma_i^{-1}\mathrm{X}_i$ is non-singular. The best linear unbiased estimator [7:Appendix C] of each β_i is $\hat{\beta}_i = \left(\mathrm{X}_i'\Sigma_i^{-1}\mathrm{X}_i\right)^{-1}\mathrm{X}_i'\Sigma_i^{-1}\mathbf{y}_i$, and $\mathrm{Var}[\hat{\beta}_i] = \left(\mathrm{X}_i'\Sigma_i^{-1}\mathrm{X}_i\right)^{-1}$.

Let $t = t_1 + \ldots + t_n$, and $k = k_1 + \ldots + k_n$. The first model of models is as follows:

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

$$\begin{bmatrix} \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_n \end{bmatrix} = \begin{bmatrix} \mathrm{X}_1 & & \\ & \ddots & \\ & & \mathrm{X}_n \end{bmatrix}\begin{bmatrix} \beta_1 \\ \vdots \\ \beta_n \end{bmatrix} + \begin{bmatrix} \mathbf{e}_1 \\ \vdots \\ \mathbf{e}_n \end{bmatrix},$$

$$\text{where } \mathrm{Var}[\mathbf{e}] = \mathrm{Var}\begin{bmatrix} \mathbf{e}_1 \\ \vdots \\ \mathbf{e}_n \end{bmatrix} = \begin{bmatrix} \Sigma_1 & & \\ & \ddots & \\ & & \Sigma_n \end{bmatrix} = \Sigma_{(t\times t)}.$$

The best linear unbiased estimator of β is:

$$\hat{\beta} = \left(X'\Sigma^{-1}X\right)^{-1}X'\Sigma^{-1}\mathbf{y}$$

$$= \left(\begin{bmatrix} X_1 & & \\ & \ddots & \\ & & X_n \end{bmatrix}' \begin{bmatrix} \Sigma_1 & & \\ & \ddots & \\ & & \Sigma_n \end{bmatrix}^{-1} \begin{bmatrix} X_1 & & \\ & \ddots & \\ & & X_n \end{bmatrix}\right)^{-1} \begin{bmatrix} X_1 & & \\ & \ddots & \\ & & X_n \end{bmatrix}' \begin{bmatrix} \Sigma_1 & & \\ & \ddots & \\ & & \Sigma_n \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_n \end{bmatrix}$$

$$= \left(\begin{bmatrix} X_1' & & \\ & \ddots & \\ & & X_n' \end{bmatrix} \begin{bmatrix} \Sigma_1^{-1} & & \\ & \ddots & \\ & & \Sigma_n^{-1} \end{bmatrix} \begin{bmatrix} X_1 & & \\ & \ddots & \\ & & X_n \end{bmatrix}\right)^{-1} \begin{bmatrix} X_1' & & \\ & \ddots & \\ & & X_n' \end{bmatrix} \begin{bmatrix} \Sigma_1^{-1} & & \\ & \ddots & \\ & & \Sigma_n^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_n \end{bmatrix}$$

$$= \left(\begin{bmatrix} X_1'\Sigma_1^{-1}X_1 & & \\ & \ddots & \\ & & X_n'\Sigma_n^{-1}X_n \end{bmatrix}\right)^{-1} \begin{bmatrix} X_1'\Sigma_1^{-1}\mathbf{y}_1 \\ \vdots \\ X_n'\Sigma_n^{-1}\mathbf{y}_n \end{bmatrix}$$

$$= \begin{bmatrix} \left(X_1'\Sigma_1^{-1}X_1\right)^{-1} & & \\ & \ddots & \\ & & \left(X_n'\Sigma_n^{-1}X_n\right)^{-1} \end{bmatrix} \begin{bmatrix} X_1'\Sigma_1^{-1}\mathbf{y}_1 \\ \vdots \\ X_n'\Sigma_n^{-1}\mathbf{y}_n \end{bmatrix}$$

$$= \begin{bmatrix} \left(X_1'\Sigma_1^{-1}X_1\right)^{-1}X_1'\Sigma_1^{-1}\mathbf{y}_1 \\ \vdots \\ \left(X_n'\Sigma_n^{-1}X_n\right)^{-1}X_n'\Sigma_n^{-1}\mathbf{y}_n \end{bmatrix}$$

$$= \begin{bmatrix} \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_n \end{bmatrix}$$

Also, $\mathrm{Var}\left[\hat{\beta}\right] = \left(X'\Sigma^{-1}X\right)^{-1} = \begin{bmatrix} \left(X_1'\Sigma_1^{-1}X_1\right)^{-1} & & \\ & \ddots & \\ & & \left(X_n'\Sigma_n^{-1}X_n\right)^{-1} \end{bmatrix} = \begin{bmatrix} \mathrm{Var}\left[\hat{\beta}_1\right] & & \\ & \ddots & \\ & & \mathrm{Var}\left[\hat{\beta}_n\right] \end{bmatrix}$. In this grand model the submodels appear together, but they are unrelated.

But this leads us to a second model of models, the one that forms the basis of credibility. Instead of n models and n betas, let there be n models and one beta. In this model $k = k_1 = \ldots = k_n$:

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

$$\begin{bmatrix}\mathbf{y}_1\\ \vdots \\ \mathbf{y}_n\end{bmatrix} = \begin{bmatrix}\mathbf{X}_1\\ \vdots \\ \mathbf{X}_n\end{bmatrix}\beta + \begin{bmatrix}\mathbf{e}_1\\ \vdots \\ \mathbf{e}_n\end{bmatrix},$$

$$\text{where } \mathrm{Var}[\mathbf{e}] = \mathrm{Var}\begin{bmatrix}\mathbf{e}_1\\ \vdots \\ \mathbf{e}_n\end{bmatrix} = \begin{bmatrix}\Sigma_1 & & \\ & \ddots & \\ & & \Sigma_n\end{bmatrix} = \Sigma_{(t\times t)}.$$

The estimator of β is:

$$\hat{\beta} = \left(\mathbf{X}'\Sigma^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'\Sigma^{-1}\mathbf{y}$$

$$= \left(\begin{bmatrix}\mathbf{X}_1\\ \vdots \\ \mathbf{X}_n\end{bmatrix}'\begin{bmatrix}\Sigma_1 & & \\ & \ddots & \\ & & \Sigma_n\end{bmatrix}^{-1}\begin{bmatrix}\mathbf{X}_1\\ \vdots \\ \mathbf{X}_n\end{bmatrix}\right)^{-1}\begin{bmatrix}\mathbf{X}_1\\ \vdots \\ \mathbf{X}_n\end{bmatrix}'\begin{bmatrix}\Sigma_1 & & \\ & \ddots & \\ & & \Sigma_n\end{bmatrix}^{-1}\begin{bmatrix}\mathbf{y}_1\\ \vdots \\ \mathbf{y}_n\end{bmatrix}$$

$$= \left(\begin{bmatrix}\mathbf{X}_1' & \cdots & \mathbf{X}_n'\end{bmatrix}\begin{bmatrix}\Sigma_1^{-1} & & \\ & \ddots & \\ & & \Sigma_n^{-1}\end{bmatrix}\begin{bmatrix}\mathbf{X}_1\\ \vdots \\ \mathbf{X}_n\end{bmatrix}\right)^{-1}\begin{bmatrix}\mathbf{X}_1' & \cdots & \mathbf{X}_n'\end{bmatrix}\begin{bmatrix}\Sigma_1^{-1} & & \\ & \ddots & \\ & & \Sigma_n^{-1}\end{bmatrix}\begin{bmatrix}\mathbf{y}_1\\ \vdots \\ \mathbf{y}_n\end{bmatrix}$$

$$= \left(\mathbf{X}_1'\Sigma_1^{-1}\mathbf{X}_1 + \ldots + \mathbf{X}_n'\Sigma_n^{-1}\mathbf{X}_n\right)^{-1}\left(\mathbf{X}_1'\Sigma_1^{-1}\mathbf{y}_1 + \ldots + \mathbf{X}_n'\Sigma_n^{-1}\mathbf{y}_n\right)$$

$$= \left(\mathbf{X}_1'\Sigma_1^{-1}\mathbf{X}_1 + \ldots + \mathbf{X}_n'\Sigma_n^{-1}\mathbf{X}_n\right)^{-1}\left(\mathbf{X}_1'\Sigma_1^{-1}\mathbf{X}_1\left(\mathbf{X}_1'\Sigma_1^{-1}\mathbf{X}_1\right)^{-1}\mathbf{X}_1'\Sigma_1^{-1}\mathbf{y}_1 + \ldots + \mathbf{X}_n'\Sigma_n^{-1}\mathbf{X}_n\left(\mathbf{X}_n'\Sigma_n^{-1}\mathbf{X}_n\right)^{-1}\mathbf{X}_n'\Sigma_n^{-1}\mathbf{y}_n\right)$$

$$= \left(\mathrm{Var}^{-1}\left[\hat{\beta}_1\right] + \ldots + \mathrm{Var}^{-1}\left[\hat{\beta}_n\right]\right)^{-1}\left(\mathrm{Var}^{-1}\left[\hat{\beta}_1\right]\hat{\beta}_1 + \ldots + \mathrm{Var}^{-1}\left[\hat{\beta}_n\right]\hat{\beta}_n\right)$$

$$= \left(\mathrm{Var}\left[\hat{\beta}\right]\right)\left(\mathrm{Var}^{-1}\left[\hat{\beta}_1\right]\hat{\beta}_1 + \ldots + \mathrm{Var}^{-1}\left[\hat{\beta}_n\right]\hat{\beta}_n\right)$$

The estimator of this grand model is a matrix-weighted average of the estimators of the submodels. The weights themselves, which are $(k\times k)$ matrices, are the inverses of the variances of the estimators. This is a k-dimensional form of the well-known rule that non-covarying estimates of the same parameter are best averaged according to weights inversely proportional to their variances. Judge [8:287] notes that a matrix-weighted average of two vectors need not be collinear with the two vectors, unlike a scalar-weighted average, which must be collinear.

A third model of models looks like the first, but has a general variance matrix:

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

$$\begin{bmatrix} \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_n \end{bmatrix} = \begin{bmatrix} \mathrm{X}_1 & & \\ & \ddots & \\ & & \mathrm{X}_n \end{bmatrix} \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_n \end{bmatrix} + \begin{bmatrix} \mathbf{e}_1 \\ \vdots \\ \mathbf{e}_n \end{bmatrix},$$

$$\text{where } \mathrm{Var}[\mathbf{e}] = \mathrm{Var}\begin{bmatrix} \mathbf{e}_1 \\ \vdots \\ \mathbf{e}_n \end{bmatrix} = \begin{bmatrix} \Sigma_{11} & \cdots & \Sigma_{1n} \\ \vdots & \ddots & \vdots \\ \Sigma_{n1} & \cdots & \Sigma_{nn} \end{bmatrix} = \Sigma_{(t\times t)}$$

Because we are accustomed to regarding the variance matrix as block diagonal, as in the first grand model, the submodels are *seemingly* unrelated. Models of seemingly unrelated models are discussed in [7:Appendix H] and [8:444-466].

The remainder of this appendix will be devoted to proving that the estimator of the second grand model is better than the estimators of its submodels. Both the proof itself and the precise meaning of 'better' require a discussion of non-negative definite and positive definite matrices. In an earlier paper [7:Appendix A] the author discussed such matrices, and developed many basic theorems concerning them (cf. also [1:459-461] and [8:960f]). This discussion dovetails with that of the earlier paper, and anything simply asserted here will be found proven there.

Let A and B be square matrices of the same dimension, say $(n\times n)$, and let x be an $(n\times 1)$ vector. The (1×1) matrices x'Ax and x'Bx are called quadratic forms in x (Judge [8:959]). Let '~' stand for one of the five following comparison relations among the real numbers:

'<', '≤', '=', '≥', and '>'. What might 'A ~ B' mean? In the case of equality, we know that 'A = B' means that corresponding elements are A and B are equal (elementwise equality). So it would be natural to define 'A ~ B' as elementwise '~', as is already the case with '='.

But there is another very useful definition: A ~ B if and only if for every non-zero x, $\{x'Ax\}_{11} \sim \{x'Bx\}_{11}$. (Of course, a zero x will result in equality.) The operator $\{\bullet\}_{ij}$ yields the ij^{th} element of the matrix inside the brackets, which is a scalar result. Being (1×1) matrices, x'Ax and x'Bx have only one element; thus, $\{\bullet\}_{11}$ makes quadratic forms comparable on a scalar basis. According to this definition, A ~ B depends on the *matrices* A and B, rather than on the *elements* of A and B. But the matrices must be reduced to the *definite* level of (1×1) quadratic forms in order to invite comparison. If '~' in the first sense is elementwise comparison, we might say that '~' in the second sense is definite comparison, perhaps distinguishing it with dots '.~.'. Therefore, A .~. B if and only if for every non-zero x, $\{x'Ax\}_{11} \sim \{x'Bx\}_{11}$.

Let C be an (n×n) matrix. Obviously, if for all non-zero x, $\{x'Ax\}_{11} \sim \{x'Bx\}_{11}$, and for all non-zero x, $\{x'Bx\}_{11} \sim \{x'Cx\}_{11}$, then for all non-zero x, $\{x'Ax\}_{11} \sim \{x'Cx\}_{11}$. So the five definite comparisons are transitive. Also, adding or subtracting the same amount from both sides of a scalar comparison does not affect the comparison. Hence,

$$
\begin{aligned}
A \mathrel{.\sim.} B &\Leftrightarrow \forall x \neq 0,\ x'Ax \mathrel{.\sim.} x'Bx \\
&\Leftrightarrow \forall x \neq 0,\ (x'Ax - x'Bx) \mathrel{.\sim.} (x'Bx - x'Bx) \\
&\Leftrightarrow \forall x \neq 0,\ x'(A - B)x \mathrel{.\sim.} x'(B - B)x \\
&\Leftrightarrow \forall x \neq 0,\ x'(A - B)x \mathrel{.\sim.} x'0_{(n\times n)}x \\
&\Leftrightarrow (A - B) \mathrel{.\sim.} 0
\end{aligned}
$$

So A compares definitely with B as (A – B) compares definitely with the zero matrix. Similarly, multiplying or dividing both sides of a scalar comparison by a positive scalar does not affect the comparison; so if $k > 0$, then kA .~. kB.

As for inequalities, if A .≤. [.≥.] B, then B .≥. [.≤.] A. And if A .≤. B and B .≤. A, then A .=. B. So far, definite comparisons behave like scalar comparisons. But the scalar comparison '$a \leq b$' is equivalent to '($a < b$) or ($a = b$)'. It is different with the definite comparison: 'A .≤. B' means 'for all x, $\{x'Ax\}_{11} \leq \{x'Bx\}_{11}$'. It is possible that for some values of x the relation is '<' and for other values it is '='. Thus 'A .≤. B' is not equivalent to '(A .<. B) or (A .=. B)'. One must be cautious in handling the compound comparisons '.≤.' and '.≥.'; for instance, it is tempting but fallacious to argue that if A .≤. B and not (A .=. B), then A .<. B. In a similar vein, according to the law of trichotomy, for any two scalars a and b, ($a < b$) or ($a = b$) or ($a > b$). But it is *not* true that for any two ($n\times n$) matrices A and B, (A .<. B) or (A .=. B) or (A .>. B).

As for equalities, since every (1×1) matrix is symmetric, x'Ax = (x'Ax)' = x'A'x. So, for all non-zero x, $\{x'Ax\}_{11} = \{x'A'x\}_{11}$, implying that A .=. A' and that (A – A') .=. 0. Moreover, if A .=. 0, then A' = –A (skew symmetry). For if A .=. 0, then for all non-zero x, $\{x'Ax\}_{11} =$

0. But $\{x'Ax\}_{11} = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}x_i x_j$. If one of the xs (say x_i) equals one and the rest are zero, then $\{x'Ax\}_{11} = a_{ii}x_i^2 = a_{ii} = 0$. And if two of the xs (say x_i and x_j) equal one and the rest are zero, then $\{x'Ax\}_{11} = a_{ii}x_i^2 + a_{ij}x_i x_j + a_{ji}x_j x_i + a_{jj}x_j^2 = a_{ii} + a_{ij} + a_{ji} + a_{jj} = a_{ij} + a_{ji} = 0$. For all k and l, $a_{ik} = -a_{ki}$, which makes A skew symmetric. Conversely, if A is skew symmetric then A .=. (A + A)/2 .=. (A – (–A))/2 .=. (A – A')/2 .=. 0/2 .=. 0. Therefore, A .=. 0 if and only if A is skew symmetric. Moreover, if A is symmetric and A .=. 0, then A is both symmetric (A' = A) and skew symmetric (A' = –A), which implies that A = –A = 0. Finally, if A and B are symmetric and A .=. B, then A – B is both symmetric and .=. 0. Hence, A – B = 0; so A = B.

A matrix A is *non-negative definite* [*positive definite*] if and only if A is symmetric and A .≥. [.>.] 0. Obviously, if A is positive definite then it is non-negative definite, but not necessarily *vice versa*. It is a theorem that if A is a non-negative definite matrix, then A is positive definite if and only if A^{-1} exists (or A is non-singular). Another theorem is that A is non-negative definite if and only if there exists a square matrix W, such that A = WW'. Such a W is sometimes called a square root matrix of A. If A is positive definite, then it is non-singular and every square root matrix of it must be non-singular. In such circumstances, $A^{-1} = (WW')^{-1} = (W')^{-1}(W)^{-1} = (W^{-1})'(W^{-1})$, which is non-negative definite. But since A^{-1} is non-singular, it must also be positive definite. Therefore, if A is positive definite, then so too is A^{-1}.

If $\mathbf{x}$ is an $(n\times 1)$ random vector with $\mathrm{Var}[\mathbf{x}] = \Sigma$, and A is an $(m\times n)$ non-stochastic matrix, then $A\mathbf{x}$ is an $(m\times 1)$ random vector with $\mathrm{Var}[A\mathbf{x}] = A\Sigma A'$. If A is $(1\times n)$, then $A\mathbf{x}$ is a (1×1) random vector, whose element must be non-negative. Hence, a variance matrix, which must be symmetric, must also be $.\geq. 0$; otherwise some non-zero linear combination of the elements of the random vector would imply a scalar random variable with a negative variance. In other words, every variance matrix is non-negative definite.

We would like to compare two $(n\times n)$ variance matrices $\mathrm{Var}[\mathbf{x}_1] = \Sigma_1$ and $\mathrm{Var}[\mathbf{x}_2] = \Sigma_2$. If $\Sigma_1 .<. [.\leq.] \Sigma_2$, then the variance of every non-zero linear combination of $\mathbf{x}_1$ is less than [less than or equal to] the variance of the same linear combination of $\mathbf{x}_2$. If Σ_1 and Σ_2 are the variance matrices of two estimates of an unknown parameter and $\Sigma_1 .<. \Sigma_2$, then Σ_1 is the better estimate. If $\Sigma_1 .\leq. \Sigma_2$, then Σ_1 may not be better; however, it is at least as good. But if, in addition, $\Sigma_1 \neq \Sigma_2$, then not $(\Sigma_1 .=. \Sigma_2)$ and Σ_1 is again better.

So, turning back to the second grand model, we will prove that $\forall i, \mathrm{Var}[\hat{\beta}].<.\mathrm{Var}[\hat{\beta}_i]$, which is equivalent to $\mathrm{Var}[\hat{\beta}_i] - \mathrm{Var}[\hat{\beta}].>.0$. This too is equivalent to the statement that $\mathrm{Var}[\hat{\beta}_i] - \mathrm{Var}[\hat{\beta}]$ is positive definite.

As above, $\mathrm{Var}[\hat{\beta}] = \left(\mathrm{Var}^{-1}[\hat{\beta}_1]+\ldots+\mathrm{Var}^{-1}[\hat{\beta}_n]\right)^{-1}$, or $\mathrm{Var}^{-1}[\hat{\beta}] = \mathrm{Var}^{-1}[\hat{\beta}_1]+\ldots+\mathrm{Var}^{-1}[\hat{\beta}_n]$. Being a variance matrix, each $\mathrm{Var}^{-1}[\hat{\beta}_i]$ is non-negative definite. And being non-singular, each is positive definite. Therefore, $\mathrm{Var}^{-1}[\hat{\beta}] - \mathrm{Var}^{-1}[\hat{\beta}_i]$ is positive definite. So there exists a non-

singular ($k \times k$) matrix W such that $\mathrm{Var}^{-1}\left[\hat{\beta}\right] = \mathrm{Var}^{-1}\left[\hat{\beta}_t\right] + WW' = \mathrm{Var}^{-1}\left[\hat{\beta}_t\right] + WI_k W'$, or

$\mathrm{Var}\left[\hat{\beta}\right] = \left(\mathrm{Var}^{-1}\left[\hat{\beta}_t\right] + WI_k W'\right)^{-1}$.

Now it is a theorem that if $D^{-1} + CA^{-1}B$ exists and is non-singular, then $(A + BDC)^{-1} = A^{-1} - A^{-1}B(D^{-1} + CA^{-1}B)^{-1}CA^{-1}$. As the first part of the proof:

$$\begin{aligned}
(A + BDC)\left(A^{-1} - A^{-1}B\left(D^{-1} + CA^{-1}B\right)^{-1}CA^{-1}\right) &= AA^{-1} - AA^{-1}B\left(D^{-1} + CA^{-1}B\right)^{-1}CA^{-1} \\
&\quad + BDCA^{-1} - BDCA^{-1}B\left(D^{-1} + CA^{-1}B\right)^{-1}CA^{-1} \\
&= I - B\left(D^{-1} + CA^{-1}B\right)^{-1}CA^{-1} \\
&\quad + BDCA^{-1} - BDCA^{-1}B\left(D^{-1} + CA^{-1}B\right)^{-1}CA^{-1} \\
&= I - BDD^{-1}\left(D^{-1} + CA^{-1}B\right)^{-1}CA^{-1} \\
&\quad + BDCA^{-1} - BDCA^{-1}B\left(D^{-1} + CA^{-1}B\right)^{-1}CA^{-1} \\
&= I + BDCA^{-1} \\
&\quad - BD\left(D^{-1} + CA^{-1}B\right)\left(D^{-1} + CA^{-1}B\right)^{-1}CA^{-1} \\
&= I + BDCA^{-1} - BDCA^{-1} \\
&= I
\end{aligned}$$

Reversing the order of the multiplication is the second and final part of the proof:

$$\begin{aligned}
\left(A^{-1} - A^{-1}B\left(D^{-1} + CA^{-1}B\right)^{-1}CA^{-1}\right)(A + BDC) &= A^{-1}A + A^{-1}BDC \\
&\quad - A^{-1}B\left(D^{-1} + CA^{-1}B\right)^{-1}CA^{-1}A - A^{-1}B\left(D^{-1} + CA^{-1}B\right)^{-1}CA^{-1}BDC \\
&= I + A^{-1}BDC \\
&\quad - A^{-1}B\left(D^{-1} + CA^{-1}B\right)^{-1}C - A^{-1}B\left(D^{-1} + CA^{-1}B\right)^{-1}CA^{-1}BDC \\
&= I + A^{-1}B\left(D^{-1} + CA^{-1}B\right)^{-1}\left(D^{-1} + CA^{-1}B\right)DC \\
&\quad - A^{-1}B\left(D^{-1} + CA^{-1}B\right)^{-1}C - A^{-1}B\left(D^{-1} + CA^{-1}B\right)^{-1}CA^{-1}BDC \\
&= I + A^{-1}B\left(D^{-1} + CA^{-1}B\right)^{-1}\left(D^{-1}\right)DC - A^{-1}B\left(D^{-1} + CA^{-1}B\right)^{-1}C \\
&= I + A^{-1}B\left(D^{-1} + CA^{-1}B\right)^{-1}C - A^{-1}B\left(D^{-1} + CA^{-1}B\right)^{-1}C \\
&= I
\end{aligned}$$

Therefore, using this theorem:

$$
\begin{aligned}
\mathrm{Var}[\hat{\beta}] &= \left(\mathrm{Var}^{-1}[\hat{\beta}_i] + WI_k W'\right)^{-1} \\
&= \left(\mathrm{Var}^{-1}[\hat{\beta}_i]\right)^{-1} - \left(\mathrm{Var}^{-1}[\hat{\beta}_i]\right)^{-1} W\left(I_k^{-1} + W'\left(\mathrm{Var}^{-1}[\hat{\beta}_i]\right)^{-1} W\right)^{-1} W'\left(\mathrm{Var}^{-1}[\hat{\beta}_i]\right)^{-1} \\
&= \mathrm{Var}[\hat{\beta}_i] - \mathrm{Var}[\hat{\beta}_i]W\left(I_k + W'\mathrm{Var}[\hat{\beta}_i]W\right)^{-1} W'\mathrm{Var}[\hat{\beta}_i] \\
&= \mathrm{Var}[\hat{\beta}_i] - \left(\mathrm{Var}[\hat{\beta}_i]W\right)\left(I_k + W'\mathrm{Var}[\hat{\beta}_i]W\right)^{-1}\left(\mathrm{Var}[\hat{\beta}_i]W\right)' \\
&.<. \mathrm{Var}[\hat{\beta}_i]
\end{aligned}
$$

Some explanation is in order: $\mathrm{Var}[\hat{\beta}_i]$ is positive definite and thus factorable as, say, UU'. So $I_k + W'\mathrm{Var}[\hat{\beta}_i]W = I_k + W'UU'W = I_k + W'U(W'U)'$. The identity matrix is positive definite, and to it is added a non-negative definite matrix. Hence, $I_k + W'\mathrm{Var}[\hat{\beta}_i]W$ is positive definite. It follows that $\left(I_k + W'\mathrm{Var}[\hat{\beta}_i]W\right)^{-1}$ is positive definite and factorable as, say, VV', where V is non-singular. Thus, $\left(\mathrm{Var}[\hat{\beta}_i]W\right)\left(I_k + W'\mathrm{Var}[\hat{\beta}_i]W\right)^{-1}\left(\mathrm{Var}[\hat{\beta}_i]W\right)'$ is factorable as $\left(\mathrm{Var}[\hat{\beta}_i]W\right)VV'\left(\mathrm{Var}[\hat{\beta}_i]W\right)' = \left(\mathrm{Var}[\hat{\beta}_i]WV\right)\left(\mathrm{Var}[\hat{\beta}_i]WV\right)'$. This is a non-negative definite matrix. But inasmuch as the square root matrix consists of the product of three non-singular matrices, the root itself is non-singular, and so too is the root times its transpose. Therefore, $\left(\mathrm{Var}[\hat{\beta}_i]W\right)VV'\left(\mathrm{Var}[\hat{\beta}_i]W\right)' = \left(\mathrm{Var}[\hat{\beta}_i]WV\right)\left(\mathrm{Var}[\hat{\beta}_i]WV\right)'$ is positive definite, and so $\mathrm{Var}[\hat{\beta}].<.\mathrm{Var}[\hat{\beta}_i]$.

So the grand model is better than every submodel. But an even more powerful statement can be made. Consider a partial grand model, consisting of some, but not all, of the submodels. Let $\tilde{\beta}$ be the estimator of the partial model. Then

$\mathrm{Var}^{-1}\left[\hat{\beta}\right] = \mathrm{Var}^{-1}\left[\tilde{\beta}\right] + \ldots + \sum_i \mathrm{Var}^{-1}\left[\hat{\beta}_i\right]$, where the subscript i is ranges over the submodels left out of the partial grand model. Then, by similar reasoning, $\mathrm{Var}^{-1}\left[\hat{\beta}\right] < \mathrm{Var}^{-1}\left[\tilde{\beta}\right]$. This goes to show that the more submodels, the better the estimate.

$\mathrm{Var}^{-1}\left[\hat{\beta}\right] = \mathrm{Var}^{-1}\left[\hat{\beta}_1\right] + \ldots + \mathrm{Var}^{-1}\left[\hat{\beta}_n\right]$ is called a harmonic sum. It is a k-dimensional equation. But there is an interesting 1-dimensional analogue in electricity, which may help the reader to understand the meaning of the statement 'the more submodels, the better the estimate'. A group of n resistors in parallel, whose resistances are $r_1, \ldots, r_n$, has an overall resistance R such that $\frac{1}{R} = \frac{1}{r_1} + \cdots + \frac{1}{r_n}$ (a harmonic sum). Every extra resistor added in parallel allows a little more current to flow through group, which in effect reduces the overall resistance. If the extra resistor is of high resistance (almost an insulator), then the reduction is small; if it is of low resistance (almost a short), then the reduction is great. The variance of an additional submodel is like the resistance of an additional resistor: when the variance is high, the extra group provides little additional information, so the reduction of variance of the estimate of the grand model is small (but a reduction nonetheless). When the variance is low, the extra group provides much additional information, with a great reduction of overall variance. Of course, the assumption implicit throughout is that each submodel is an appropriate model; otherwise, information could be created *ex nihilo*.

The case of a grand model in which some $\mathrm{X}_i'\Sigma_i^{-1}\mathrm{X}_i$ may be singular deserves a discussion. We will consider a model with only two submodels, in which $\mathrm{Var}[\hat{\beta}_1] = \left(\mathrm{X}_1'\Sigma_1^{-1}\mathrm{X}_1\right)^{-1}$

exists, but $X_2'\Sigma_2^{-1}X_2$ may be singular. In this case, the second submodel, though informative, may not be sufficiently informative for a unique estimate of β. For example, if β were (2×1), the second submodel might be a non-sample judgment that the first element of β has a mean of 1 and a variance of 2:

$$\begin{aligned}\mathbf{y}_2 &= X_2\beta + \mathbf{e}_2\\ [1] &= [1 \quad 0]\beta + \mathbf{e}_2,\end{aligned}$$

where $\mathrm{Var}[\mathbf{e}_2] = \Sigma_2 = [2]$. In this example, β cannot be uniquely estimated because $X_2'\Sigma_2^{-1}X_2 = \begin{bmatrix}1\\0\end{bmatrix}[2]^{-1}[1 \quad 0] = \begin{bmatrix}1/2 & 0\\ 0 & 0\end{bmatrix}$, which is singular.

$X_2'\Sigma_2^{-1}X_2$ is non-negative definite; therefore, for all non-zero u, $\{u'(X_2'\Sigma_2^{-1}X_2)u\}_{11} \geq 0$. We will define a set Z, possibly empty, of all non-zero u such that $\{u'(X_2'\Sigma_2^{-1}X_2)u\}_{11} = 0$. But $u'(X_2'\Sigma_2^{-1}X_2)u = (X_2u)'\Sigma_2^{-1}(X_2u)$. Since Σ_2^{-1} is positive definite, $\{(X_2u)'\Sigma_2^{-1}(X_2u)\}_{11} = 0$ if and only if $X_2u = 0$. Therefore, $u \in Z$ if and only if u is non-zero and $X_2u = 0$. Recall that X_2 is $(t_2\times k)$. At the beginning of the appendix it was assumed that $\mathrm{rank}(X_2) = k$, but now we will relax this assumption. Let $\mathrm{rank}(X_2) = j \leq k$. Then the set of all u such that $X_2u = 0$ is a $(k - j)$-dimensional linear subspace of k-space. Z is this subspace less the zero vector (so if $\mathrm{rank}(X_2) = j = k$, then Z is empty and $X_2'\Sigma_2^{-1}X_2$ is positive definite).

Therefore, in the grand model:

$$
\begin{aligned}
\mathrm{Var}\left[\hat{\beta}\right] &= \left(\mathrm{Var}^{-1}\left[\hat{\beta}_1\right] + X_2'\Sigma_2^{-1}X_2\right)^{-1} \\
&= \left(\mathrm{Var}^{-1}\left[\hat{\beta}_1\right]\right)^{-1} - \left(\mathrm{Var}^{-1}\left[\hat{\beta}_1\right]\right)^{-1} X_2'\left(\left(\Sigma_2^{-1}\right)^{-1} + X_2\left(\mathrm{Var}^{-1}\hat{\beta}_1\right)^{-1}X_2'\right)^{-1} X_2\left(\mathrm{Var}^{-1}\left[\hat{\beta}_1\right]\right)^{-1} \\
&= \mathrm{Var}\left[\hat{\beta}_1\right] - \mathrm{Var}\left[\hat{\beta}_1\right]X_2'\left(\Sigma_2 + X_2\mathrm{Var}\left[\hat{\beta}_1\right]X_2'\right)^{-1}X_2\mathrm{Var}\left[\hat{\beta}_1\right] \\
&= \mathrm{Var}\left[\hat{\beta}_1\right] - \left(X_2\mathrm{Var}\left[\hat{\beta}_1\right]\right)'\left(\Sigma_2 + X_2\mathrm{Var}\left[\hat{\beta}_1\right]X_2'\right)^{-1}\left(X_2\mathrm{Var}\left[\hat{\beta}_1\right]\right) \\
&\leq \mathrm{Var}\left[\hat{\beta}_1\right]
\end{aligned}
$$

The inequality follows from the fact that $\left(X_2\mathrm{Var}\left[\hat{\beta}_1\right]\right)'\left(\Sigma_2 + X_2\mathrm{Var}\left[\hat{\beta}_1\right]X_2'\right)^{-1}\left(X_2\mathrm{Var}\left[\hat{\beta}_1\right]\right)$ is non-negative definite. But strict inequality, which represents an efficiency gain, depends on $\left\{u'\left(X_2\mathrm{Var}\left[\hat{\beta}_1\right]\right)'\left(\Sigma_2 + X_2\mathrm{Var}\left[\hat{\beta}_1\right]X_2'\right)^{-1}\left(X_2\mathrm{Var}\left[\hat{\beta}_1\right]\right)u\right\}_{11} > 0$. Since $\left(\Sigma_2 + X_2\mathrm{Var}\left[\hat{\beta}_1\right]X_2'\right)^{-1}$ is positive definite, strict inequality is thwarted only when $X_2\mathrm{Var}\left[\hat{\beta}_1\right]u = 0$. And $X_2\mathrm{Var}\left[\hat{\beta}_1\right]u = 0$ if and only if $\mathrm{Var}\left[\hat{\beta}_1\right]u \in Z$. Since $\mathrm{Var}\left[\hat{\beta}_1\right]$ is non-singular, there exists a $(k-j)$-dimensional subspace of k-space, Z^*, formed by premultiplying each member of Z by $\mathrm{Var}^{-1}\left[\hat{\beta}_1\right]$. When $\mathrm{rank}(X_2) = j \leq k$, $\mathrm{Var}\left[\hat{\beta}\right] < \mathrm{Var}\left[\hat{\beta}_1\right]$ except in the $(k-j)$-dimensional subspace Z^*, within which $\mathrm{Var}\left[\hat{\beta}\right] = \mathrm{Var}\left[\hat{\beta}_1\right]$ (so $\mathrm{Var}\left[\hat{\beta}\right] = \mathrm{Var}\left[\hat{\beta}_1\right]$).

Appendix B

A Bayesian Interpretation of Prior Information

Consider the model $\mathbf{y}_{(t\times 1)} = \mathbf{X}_{(t\times k)}\beta_{(k\times 1)} + \mathbf{e}_{(t\times 1)}$. What makes this model Bayesian is that β is stochastic. Let us assume that β is multivariate normal with mean β_0 and variance V, i.e., $\beta \sim N(\beta_0, V)$. We will assume also that $\mathbf{e} \sim N(0, \Sigma)$, and that $\mathbf{e}$ is independent of β. Being variance matrices, Σ and V must be non-negative definite. But we will further assume that both matrices are positive definite, which implies that their inverses exist and that their determinants are positive.

The probability density function of $\beta \sim N(\beta_0, V)$ is [8:49f.]:

$$f_{\beta}(\beta) = (2\pi)^{-\frac{k}{2}}|V|^{-\frac{1}{2}}e^{-\frac{1}{2}(\beta-\beta_0)'V^{-1}(\beta-\beta_0)} \propto e^{-\frac{1}{2}(\beta-\beta_0)'V^{-1}(\beta-\beta_0)}$$

As for the random vector $\mathbf{y}$ given that $\beta = \beta$, or $\mathbf{y}|\beta = \beta$:

$$\begin{aligned}\mathbf{y}|\beta = \beta &= (X\beta + \mathbf{e})|\beta = \beta \\ &= X\beta + (\mathbf{e}|\beta = \beta) \\ &= X\beta + \mathbf{e}\end{aligned}$$

The last equation follows from $\mathbf{e}$'s being independent of β. Hence, $\mathbf{y}|\beta = \beta \sim N(X\beta, \Sigma)$; so its probability density function is:

$$f_{\mathbf{y}|\beta=\beta}(y) = (2\pi)^{-\frac{t}{2}}|\Sigma|^{-\frac{1}{2}}e^{-\frac{1}{2}(y-X\beta)'\Sigma^{-1}(y-X\beta)} \propto e^{-\frac{1}{2}(y-X\beta)'\Sigma^{-1}(y-X\beta)}$$

Therefore, according to Bayes' theorem, the probability density function of $\beta|\mathbf{y} = y$ is:

$$f_{\beta|y=y}(\beta) = \frac{f_{y|\beta=\beta}(y) f_{\beta}(\beta)}{f_{y}(y)} \propto f_{y|\beta=\beta}(y) f_{\beta}(\beta)$$

$$\propto \left(e^{-\frac{1}{2}(y-X\beta)' \Sigma^{-1}(y-X\beta)} \right) \left(e^{-\frac{1}{2}(\beta-\beta_0)' V^{-1}(\beta-\beta_0)} \right)$$

$$\propto e^{-\frac{1}{2}\left[(y-X\beta)' \Sigma^{-1}(y-X\beta)+(\beta-\beta_0)' V^{-1}(\beta-\beta_0)\right]}$$

We will now expand the exponent of this density function:

$$\begin{aligned}
(y-X\beta)' \Sigma^{-1}(y-X\beta)+(\beta-\beta_0)' V^{-1}(\beta-\beta_0) &= y'\Sigma^{-1}y-(X\beta)' \Sigma^{-1}y-y'\Sigma^{-1}(X\beta)+(X\beta)' \Sigma^{-1}(X\beta) \\
&\quad +\beta' V^{-1}\beta-\beta_0' V^{-1}\beta-\beta' V^{-1}\beta_0+\beta_0' V^{-1}\beta_0 \\
&= y'\Sigma^{-1}y-\beta'X'\Sigma^{-1}y-y'\Sigma^{-1}X\beta+\beta'X'\Sigma^{-1}X\beta \\
&\quad +\beta' V^{-1}\beta-\beta_0' V^{-1}\beta-\beta' V^{-1}\beta_0+\beta_0' V^{-1}\beta_0 \\
&= -\beta'X'\Sigma^{-1}y-y'\Sigma^{-1}X\beta+\beta'X'\Sigma^{-1}X\beta \\
&\quad +\beta' V^{-1}\beta-\beta_0' V^{-1}\beta-\beta' V^{-1}\beta_0+c
\end{aligned}$$

The two terms of the expansion which did not involve β were absorbed into the term c, which will be a catch-all for all terms not involving β.

Next we will perform a multivariate "completion of the square" with respect to β. To do this we must recognize that since V is positive definite, V^{-1} exists and is positive definite. It is similar with Σ, so $X'\Sigma^{-1}X$ exists and is non-negative definite. This implies that $X'\Sigma^{-1}X + V^{-1}$ is positive definite. Therefore, there exists a nonsingular $(k \times k)$ matrix W such that $W'W = X'\Sigma^{-1}X + V^{-1}$. So we continue:

$$
\begin{aligned}
(y-X\beta)'\Sigma^{-1}(y-X\beta)+(\beta-\beta_0)'V^{-1}(\beta-\beta_0) &= -\beta'X'\Sigma^{-1}y - y'\Sigma^{-1}X\beta + \beta'X'\Sigma^{-1}X\beta \\
&\quad + \beta'V^{-1}\beta - \beta_0'V^{-1}\beta - \beta'V^{-1}\beta_0 + c \\
&= \beta'(X'\Sigma^{-1}X + V^{-1})\beta - \beta'(X'\Sigma^{-1}y + V^{-1}\beta_0) \\
&\quad - (y'\Sigma^{-1}X + \beta_0'V^{-1})\beta + c \\
&= \beta'(X'\Sigma^{-1}X + V^{-1})\beta - \beta'(X'\Sigma^{-1}y + V^{-1}\beta_0) \\
&\quad - (X'\Sigma^{-1}y + V^{-1}\beta_0)'\beta + c \\
&= \beta'W'W\beta - \beta'W'(W')^{-1}(X'\Sigma^{-1}y + V^{-1}\beta_0) \\
&\quad - (X'\Sigma^{-1}y + V^{-1}\beta_0)'W^{-1}W\beta + c \\
&= (W\beta)'W\beta - (W\beta)'(W^{-1})'(X'\Sigma^{-1}y + V^{-1}\beta_0) \\
&\quad - (X'\Sigma^{-1}y + V^{-1}\beta_0)'W^{-1}(W\beta) + c \\
&= (W\beta)'W\beta - (W\beta)'(W^{-1})'(X'\Sigma^{-1}y + V^{-1}\beta_0) \\
&\quad - ((W^{-1})'(X'\Sigma^{-1}y + V^{-1}\beta_0))'(W\beta) + c \\
&= (W\beta)'W\beta - (W\beta)'(W^{-1})'(X'\Sigma^{-1}y + V^{-1}\beta_0) \\
&\quad - ((W^{-1})'(X'\Sigma^{-1}y + V^{-1}\beta_0))'(W\beta) \\
&\quad + ((W^{-1})'(X'\Sigma^{-1}y + V^{-1}\beta_0))'((W^{-1})'(X'\Sigma^{-1}y + V^{-1}\beta_0)) \\
&\quad - ((W^{-1})'(X'\Sigma^{-1}y + V^{-1}\beta_0))'((W^{-1})'(X'\Sigma^{-1}y + V^{-1}\beta_0)) + c \\
&= (W\beta - (W^{-1})'(X'\Sigma^{-1}y + V^{-1}\beta_0))'(W\beta - (W^{-1})'(X'\Sigma^{-1}y + V^{-1}\beta_0)) \\
&\quad - ((W^{-1})'(X'\Sigma^{-1}y + V^{-1}\beta_0))'((W^{-1})'(X'\Sigma^{-1}y + V^{-1}\beta_0)) + c \\
&= (W\beta - (W^{-1})'(X'\Sigma^{-1}y + V^{-1}\beta_0))'(W\beta - (W^{-1})'(X'\Sigma^{-1}y + V^{-1}\beta_0)) + c
\end{aligned}
$$

In the last equation a term not involving β has been absorbed into the catch-all term c. Now we can simplify:

$$
\begin{aligned}
(y-X\beta)'\Sigma^{-1}(y-X\beta)+(\beta-\beta_0)'V^{-1}(\beta-\beta_0) &= (W\beta - (W^{-1})'(X'\Sigma^{-1}y + V^{-1}\beta_0))'(W\beta - (W^{-1})'(X'\Sigma^{-1}y + V^{-1}\beta_0)) + c \\
&= (W\beta - WW^{-1}(W^{-1})'(X'\Sigma^{-1}y + V^{-1}\beta_0))'(W\beta - WW^{-1}(W^{-1})'(X'\Sigma^{-1}y + V^{-1}\beta_0)) + c \\
&= (W\beta - W(W'W)^{-1}(X'\Sigma^{-1}y + V^{-1}\beta_0))'(W\beta - W(W'W)^{-1}(X'\Sigma^{-1}y + V^{-1}\beta_0)) + c \\
&= (\beta - (W'W)^{-1}(X'\Sigma^{-1}y + V^{-1}\beta_0))'W'W(\beta - (W'W)^{-1}(X'\Sigma^{-1}y + V^{-1}\beta_0)) + c
\end{aligned}
$$

Therefore, the probability density function of $\beta|\mathbf{y} = y$ is:

$$
\begin{aligned}
f_{\beta|y=y}(\beta) &\propto e^{-\frac{1}{2}\left[(y-X\beta)'\Sigma^{-1}(y-X\beta)+(\beta-\beta_0)'V^{-1}(\beta-\beta_0)\right]} \\
&\propto e^{-\frac{1}{2}\left(\left[\beta-(W'W)^{-1}\left(X'\Sigma^{-1}y+V^{-1}\beta_0\right)\right]'W'W\left[\beta-(W'W)^{-1}\left(X'\Sigma^{-1}y+V^{-1}\beta_0\right)\right]+c\right)} \\
&\propto e^{-\frac{1}{2}\left[\beta-(W'W)^{-1}\left(X'\Sigma^{-1}y+V^{-1}\beta_0\right)\right]'W'W\left[\beta-(W'W)^{-1}\left(X'\Sigma^{-1}y+V^{-1}\beta_0\right)\right]}e^{-\frac{1}{2}c} \\
&\propto e^{-\frac{1}{2}\left[\beta-(W'W)^{-1}\left(X'\Sigma^{-1}y+V^{-1}\beta_0\right)\right]'W'W\left[\beta-(W'W)^{-1}\left(X'\Sigma^{-1}y+V^{-1}\beta_0\right)\right]} \\
&\propto e^{-\frac{1}{2}\left[\beta-(W'W)^{-1}\left(X'\Sigma^{-1}y+V^{-1}\beta_0\right)\right]'\left((W'W)^{-1}\right)^{-1}\left[\beta-(W'W)^{-1}\left(X'\Sigma^{-1}y+V^{-1}\beta_0\right)\right]}
\end{aligned}
$$

The term having c in the exponent was absorbed into the proportionality, since c does not depend on β. This function is proportional to the probability density function of a normal random vector whose mean is $(W'W)^{-1}\left(X'\Sigma^{-1}y + V^{-1}\beta_0\right)$ and whose variance is $(W'W)^{-1}$; therefore, $\beta|\mathbf{y} = y \sim N\left((W'W)^{-1}\left(X'\Sigma^{-1}y + V^{-1}\beta_0\right), (W'W)^{-1}\right)$, or:

$$\beta|\mathbf{y} = y \sim N\left(\left(X'\Sigma^{-1}X + V^{-1}\right)^{-1}\left(X'\Sigma^{-1}y + V^{-1}\beta_0\right), \left(X'\Sigma^{-1}X + V^{-1}\right)^{-1}\right)$$

This is the same result as that obtained from the mixed linear statistical model:

$$\begin{bmatrix}\mathbf{y}\\ \beta_0\end{bmatrix} = \begin{bmatrix}X\\ I\end{bmatrix}\beta + \begin{bmatrix}\mathbf{e}\\ \mathbf{v}\end{bmatrix}, \text{ where } \mathrm{Var}\begin{bmatrix}\mathbf{e}\\ \mathbf{v}\end{bmatrix} = \begin{bmatrix}\Sigma & \\ & V\end{bmatrix},$$

which mixes the sample information $\mathbf{y} = X\beta + \mathbf{e}$ with the non-sample information $\beta_0 = \beta + \mathbf{v}$. Because the results are the same, Judge says that estimating the β of such a model, i.e., mixed estimation, is a "quasi-Bayesian approach" [8:877].

It may seem when R is not an identity matrix that the mixed model $\begin{bmatrix}\mathbf{y}\\ \mathbf{r}\end{bmatrix}=\begin{bmatrix}X\\ R\end{bmatrix}\beta+\begin{bmatrix}\mathbf{e}\\ \mathbf{v}\end{bmatrix}$, where $\mathrm{Var}\begin{bmatrix}\mathbf{e}\\ \mathbf{v}\end{bmatrix}=\begin{bmatrix}\Sigma & \\ & V\end{bmatrix}$, has no Bayesian interpretation. However, if R is of full row rank (which is not a restrictive condition), there exists an S such that $\begin{bmatrix}R\\ S\end{bmatrix}=Q_{(k\times k)}$ is non-singular. Add to the non-sample information thus:

$$\begin{bmatrix}\mathbf{y}\\ \mathbf{r}\\ \mathbf{s}\end{bmatrix}=\begin{bmatrix}X\\ R\\ S\end{bmatrix}\beta+\begin{bmatrix}\mathbf{e}\\ \mathbf{v}_1\\ \mathbf{v}_2\end{bmatrix},\ \text{where Var}\begin{bmatrix}\mathbf{e}\\ \mathbf{v}_1\\ \mathbf{v}_2\end{bmatrix}=\begin{bmatrix}\Sigma & & \\ & V_1 & \\ & & V_2\end{bmatrix}$$

Letting $\gamma = Q\beta$, a one-to-one transformation because $\beta = Q^{-1}\gamma$, we can transform:

$$\begin{aligned}\begin{bmatrix}\mathbf{y}\\ \begin{bmatrix}\mathbf{r}\\ \mathbf{s}\end{bmatrix}\end{bmatrix}&=\begin{bmatrix}X\\ \begin{bmatrix}R\\ S\end{bmatrix}\end{bmatrix}\beta+\begin{bmatrix}\mathbf{e}\\ \begin{bmatrix}\mathbf{v}_1\\ \mathbf{v}_2\end{bmatrix}\end{bmatrix}\\ &=\begin{bmatrix}X\\ Q\end{bmatrix}\beta+\begin{bmatrix}\mathbf{e}\\ \begin{bmatrix}\mathbf{v}_1\\ \mathbf{v}_2\end{bmatrix}\end{bmatrix}\\ &=\begin{bmatrix}XQ^{-1}\\ I_k\end{bmatrix}Q\beta+\begin{bmatrix}\mathbf{e}\\ \begin{bmatrix}\mathbf{v}_1\\ \mathbf{v}_2\end{bmatrix}\end{bmatrix}\\ &=\begin{bmatrix}XQ^{-1}\\ I_k\end{bmatrix}\gamma+\begin{bmatrix}\mathbf{e}\\ \begin{bmatrix}\mathbf{v}_1\\ \mathbf{v}_2\end{bmatrix}\end{bmatrix},\ \text{where Var}\begin{bmatrix}\mathbf{e}\\ \begin{bmatrix}\mathbf{v}_1\\ \mathbf{v}_2\end{bmatrix}\end{bmatrix}=\begin{bmatrix}\Sigma & \\ & \begin{bmatrix}V_1 & \\ & V_2\end{bmatrix}\end{bmatrix}\end{aligned}$$

The transformed model does admit of a Bayesian interpretation. Both the mixed estimator and the Bayesian estimator are the same:

$$\hat{\gamma} = \left(\left(XQ^{-1}\right)' \Sigma^{-1} \left(XQ^{-1}\right) + \begin{bmatrix} V_1 & \\ & V_2 \end{bmatrix}^{-1} \right)^{-1} \left(\left(XQ^{-1}\right)' \Sigma^{-1} \mathbf{y} + \begin{bmatrix} V_1 & \\ & V_2 \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{r} \\ \mathbf{s} \end{bmatrix} \right)$$

$$= \left(Q'^{-1} X' \Sigma^{-1} X Q^{-1} + \begin{bmatrix} V_1^{-1} & \\ & V_2^{-1} \end{bmatrix} \right)^{-1} \left(Q'^{-1} X' \Sigma^{-1} \mathbf{y} + \begin{bmatrix} V_1^{-1} & \\ & V_2^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{r} \\ \mathbf{s} \end{bmatrix} \right)$$

$$= \left(Q'^{-1} X' \Sigma^{-1} X Q^{-1} + Q'^{-1} Q' \begin{bmatrix} V_1^{-1} & \\ & V_2^{-1} \end{bmatrix} Q Q^{-1} \right)^{-1} \left(Q'^{-1} X' \Sigma^{-1} \mathbf{y} + \begin{bmatrix} V_1^{-1} & \\ & V_2^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{r} \\ \mathbf{s} \end{bmatrix} \right)$$

$$= Q \left(X' \Sigma^{-1} X + Q' \begin{bmatrix} V_1^{-1} & \\ & V_2^{-1} \end{bmatrix} Q \right)^{-1} Q' \left(Q'^{-1} X' \Sigma^{-1} \mathbf{y} + \begin{bmatrix} V_1^{-1} & \\ & V_2^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{r} \\ \mathbf{s} \end{bmatrix} \right)$$

$$= Q \left(X' \Sigma^{-1} X + Q' \begin{bmatrix} V_1^{-1} & \\ & V_2^{-1} \end{bmatrix} Q \right)^{-1} \left(X' \Sigma^{-1} \mathbf{y} + Q' \begin{bmatrix} V_1^{-1} & \\ & V_2^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{r} \\ \mathbf{s} \end{bmatrix} \right)$$

Therefore:

$$\hat{\beta} = Q^{-1} \hat{\gamma}$$

$$= \left(X' \Sigma^{-1} X + Q' \begin{bmatrix} V_1^{-1} & \\ & V_2^{-1} \end{bmatrix} Q \right)^{-1} \left(X' \Sigma^{-1} \mathbf{y} + Q' \begin{bmatrix} V_1^{-1} & \\ & V_2^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{r} \\ \mathbf{s} \end{bmatrix} \right)$$

$$= \left(X' \Sigma^{-1} X + \begin{bmatrix} R' & S' \end{bmatrix} \begin{bmatrix} V_1^{-1} & \\ & V_2^{-1} \end{bmatrix} \begin{bmatrix} R \\ S \end{bmatrix} \right)^{-1} \left(X' \Sigma^{-1} \mathbf{y} + \begin{bmatrix} R' & S' \end{bmatrix} \begin{bmatrix} V_1^{-1} & \\ & V_2^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{r} \\ \mathbf{s} \end{bmatrix} \right)$$

$$= \left(X' \Sigma^{-1} X + R' V_1^{-1} R + S' V_2^{-1} S \right)^{-1} \left(X' \Sigma^{-1} \mathbf{y} + R' V_1^{-1} \mathbf{r} + S' V_2^{-1} \mathbf{s} \right)$$

However, this model has extraneous non-sample information. But if V_2, the variance of the extraneous non-sample information, is allowed to approach infinity, this extraneous information will have no effect. Hence:

$$\lim_{V_2 \to \infty} \hat{\beta} = \lim_{V_2^{-1} \to 0} \hat{\beta}$$

$$= \lim_{V_2^{-1} \to 0} \left(X' \Sigma^{-1} X + R' V_1^{-1} R + S' V_2^{-1} S \right)^{-1} \left(X' \Sigma^{-1} \mathbf{y} + R' V_1^{-1} \mathbf{r} + S' V_2^{-1} \mathbf{s} \right)$$

$$= \left(X' \Sigma^{-1} X + R' V_1^{-1} R \right)^{-1} \left(X' \Sigma^{-1} \mathbf{y} + R' V_1^{-1} \mathbf{r} \right)$$

Thus, in general, a Bayesian formulation, suitably transformed and taken to a limit, can be made equivalent to the mixed model.

Appendix C

The Limiting Behavior of a Stochastic Constraint

The model $\begin{bmatrix}\mathbf{y}\\ \mathbf{r}\end{bmatrix} = \begin{bmatrix}X\\ R\end{bmatrix}\beta + \begin{bmatrix}\mathbf{e}\\ \mathbf{v}\end{bmatrix}$, where $\mathrm{Var}\begin{bmatrix}\mathbf{e}\\ \mathbf{v}\end{bmatrix} = \begin{bmatrix}\Sigma & \\ & V\end{bmatrix}$, contains the stochastic constraint $\mathbf{r} = R\beta + \mathbf{v}$. The constraint loosens as $\mathrm{Var}[\mathbf{v}] = V$ increases, and tightens as it decreases. In the limit, as V approaches 0, the constraint is non-stochastic, or absolute. The problem of estimating β in the model $\mathbf{y} = X\beta + \mathbf{e}$ subject to the non-stochastic constraint that $R\beta = r$ has been solved by many authors, e.g., [1:20-23], [6:35-42], and [8:235-240]. In this appendix we will demonstrate that that the same solution obtains from a stochastically constrained model as the variance of the constraint approaches zero.

Consider the model $\begin{bmatrix}\mathbf{y}_{(t\times 1)}\\ \mathbf{r}_{(j\times 1)}\end{bmatrix} = \begin{bmatrix}X_{(t\times k)}\\ R_{(j\times k)}\end{bmatrix}\beta_{(k\times 1)} + \begin{bmatrix}\mathbf{e}_{(t\times 1)}\\ \mathbf{v}_{(j\times 1)}\end{bmatrix}$, where $\mathrm{Var}\begin{bmatrix}\mathbf{e}\\ \mathbf{v}\end{bmatrix} = \begin{bmatrix}\Sigma_{(t\times t)} & \\ & V_{(k\times k)}\end{bmatrix}$. We will assume that both Σ and V are positive definite, so that their inverses exist. Also, assume that R is of full row rank, i.e., $\mathrm{rank}(R) = j$. This means that the j constraints on β contain no redundancy. We will also assume that the $(k\times k)$ matrix $X'\Sigma^{-1}X$ has an inverse. Normally this is guaranteed by assuming that X is of full column rank. From these assumptions it follows that the $(j\times j)$ matrix $R(X'\Sigma^{-1}X)^{-1}R'$ has an inverse, which inverse we will call $H = (R(X'\Sigma^{-1}X)^{-1}R')^{-1}$.

The best linear unbiased estimator of β, sometimes in this context called the mixed estimator ([1:25] and [8:877]), is:

$$\hat{\beta} = \left(\begin{bmatrix} X \\ R \end{bmatrix}' \begin{bmatrix} \Sigma & \\ & V \end{bmatrix}^{-1} \begin{bmatrix} X \\ R \end{bmatrix} \right)^{-1} \begin{bmatrix} X \\ R \end{bmatrix}' \begin{bmatrix} \Sigma & \\ & V \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{y} \\ \mathbf{r} \end{bmatrix}$$

$$= \left(\begin{bmatrix} X' & R' \end{bmatrix} \begin{bmatrix} \Sigma^{-1} & \\ & V^{-1} \end{bmatrix} \begin{bmatrix} X \\ R \end{bmatrix} \right)^{-1} \begin{bmatrix} X' & R' \end{bmatrix} \begin{bmatrix} \Sigma^{-1} & \\ & V^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{y} \\ \mathbf{r} \end{bmatrix}$$

$$= \left(X'\Sigma^{-1}X + R'V^{-1}R \right)^{-1} \left(X'\Sigma^{-1}\mathbf{y} + R'V^{-1}\mathbf{r} \right)$$

The expectation of the estimator is β (hence unbiased), and the variance thereof is $\left(X'\Sigma^{-1}X + R'V^{-1}R \right)^{-1}$. Therefore, $\hat{\beta} = \mathrm{Var}\left[\hat{\beta}\right]\left(X'\Sigma^{-1}\mathbf{y} + R'V^{-1}\mathbf{r} \right)$. Evaluating this expression as V approaches 0 is complicated due to the fact that as V approaches 0, V^{-1} approaches infinity. Thus, $\hat{\beta} = \mathrm{Var}\left[\hat{\beta}\right]\left(X'\Sigma^{-1}\mathbf{y} + R'V^{-1}\mathbf{r} \right) \to (\infty)^{-1}\infty$, an indeterminate form. The trick is to transform the expression so as to remove V^{-1}.

In Appendix A we proved that $(A + BDC)^{-1} = A^{-1} - A^{-1}B(D^{-1} + CA^{-1}B)^{-1}CA^{-1}$, provided that the inverses exist. We can apply this theorem to the variance of the estimator:

$$\mathrm{Var}\left[\hat{\beta}\right] = \left(X'\Sigma^{-1}X + R'V^{-1}R \right)^{-1}$$

$$= \left(X'\Sigma^{-1}X \right)^{-1} - \left(X'\Sigma^{-1}X \right)^{-1} R' \left(\left(V^{-1} \right)^{-1} + R\left(X'\Sigma^{-1}X \right)^{-1} R' \right)^{-1} R\left(X'\Sigma^{-1}X \right)^{-1}$$

$$= \left(X'\Sigma^{-1}X \right)^{-1} - \left(X'\Sigma^{-1}X \right)^{-1} R' \left(V + R\left(X'\Sigma^{-1}X \right)^{-1} R' \right)^{-1} R\left(X'\Sigma^{-1}X \right)^{-1}$$

Because of the assumptions, all the inverses exist; in particular, $V + R(X'\Sigma^{-1}X)^{-1}R'$ is the sum of positive definite V and non-negative definite $R(X'\Sigma^{-1}X)^{-1}R'$. Therefore, it is positive definite, and hence non-singular. This expression has no V^{-1}, so:

$$\lim_{V \to 0} \mathrm{Var}\left[\hat{\beta}\right] = \left(X'\Sigma^{-1}X \right)^{-1} - \left(X'\Sigma^{-1}X \right)^{-1} R' \left(R\left(X'\Sigma^{-1}X \right)^{-1} R' \right)^{-1} R\left(X'\Sigma^{-1}X \right)^{-1}$$

$$= \left(X'\Sigma^{-1}X \right)^{-1} - \left(X'\Sigma^{-1}X \right)^{-1} R'HR\left(X'\Sigma^{-1}X \right)^{-1}$$

Also worth noting is:

$$\begin{aligned}
\lim_{V\to 0} R\,\mathrm{Var}\left[\hat{\beta}\right] &= R \lim_{V\to 0} \mathrm{Var}\left[\hat{\beta}\right] \\
&= R\left(\left(X'\Sigma^{-1}X\right)^{-1} - \left(X'\Sigma^{-1}X\right)^{-1}R'\left(R\left(X'\Sigma^{-1}X\right)^{-1}R'\right)^{-1}R\left(X'\Sigma^{-1}X\right)^{-1}\right) \\
&= R\left(X'\Sigma^{-1}X\right)^{-1} - R\left(X'\Sigma^{-1}X\right)^{-1}R'HR\left(X'\Sigma^{-1}X\right)^{-1} \\
&= R\left(X'\Sigma^{-1}X\right)^{-1} - H^{-1}HR\left(X'\Sigma^{-1}X\right)^{-1} \\
&= R\left(X'\Sigma^{-1}X\right)^{-1} - R\left(X'\Sigma^{-1}X\right)^{-1} \\
&= 0
\end{aligned}$$

Now we are ready to remove the remaining V^{-1} from the estimator:

$$\begin{aligned}
\hat{\beta} &= \mathrm{Var}\left[\hat{\beta}\right]\left(X'\Sigma^{-1}\mathbf{y} + R'V^{-1}\mathbf{r}\right) \\
&= \mathrm{Var}\left[\hat{\beta}\right]X'\Sigma^{-1}\mathbf{y} + \mathrm{Var}\left[\hat{\beta}\right]R'V^{-1}\mathbf{r} \\
&= \mathrm{Var}\left[\hat{\beta}\right]X'\Sigma^{-1}\mathbf{y} + \left(\left(X'\Sigma^{-1}X\right)^{-1} - \left(X'\Sigma^{-1}X\right)^{-1}R'\left(V + R\left(X'\Sigma^{-1}X\right)^{-1}R'\right)^{-1}R\left(X'\Sigma^{-1}X\right)^{-1}\right)R'V^{-1}\mathbf{r} \\
&= \mathrm{Var}\left[\hat{\beta}\right]X'\Sigma^{-1}\mathbf{y} + \left(X'\Sigma^{-1}X\right)^{-1}R'V^{-1}\mathbf{r} \\
&\quad - \left(X'\Sigma^{-1}X\right)^{-1}R'\left(V + R\left(X'\Sigma^{-1}X\right)^{-1}R'\right)^{-1}R\left(X'\Sigma^{-1}X\right)^{-1}R'V^{-1}\mathbf{r} \\
&= \mathrm{Var}\left[\hat{\beta}\right]X'\Sigma^{-1}\mathbf{y} + \left(X'\Sigma^{-1}X\right)^{-1}R'\left(V + R\left(X'\Sigma^{-1}X\right)^{-1}R'\right)^{-1}\left(V + R\left(X'\Sigma^{-1}X\right)^{-1}R'\right)V^{-1}\mathbf{r} \\
&\quad - \left(X'\Sigma^{-1}X\right)^{-1}R'\left(V + R\left(X'\Sigma^{-1}X\right)^{-1}R'\right)^{-1}R\left(X'\Sigma^{-1}X\right)^{-1}R'V^{-1}\mathbf{r} \\
&= \mathrm{Var}\left[\hat{\beta}\right]X'\Sigma^{-1}\mathbf{y} + \left(X'\Sigma^{-1}X\right)^{-1}R'\left(V + R\left(X'\Sigma^{-1}X\right)^{-1}R'\right)^{-1}VV^{-1}\mathbf{r} \\
&= \mathrm{Var}\left[\hat{\beta}\right]X'\Sigma^{-1}\mathbf{y} + \left(X'\Sigma^{-1}X\right)^{-1}R'\left(V + R\left(X'\Sigma^{-1}X\right)^{-1}R'\right)^{-1}\mathbf{r}
\end{aligned}$$

Therefore:

$$\begin{aligned}
\lim_{V\to 0}\hat{\beta} &= \lim_{V\to 0}\left(\mathrm{Var}\left[\hat{\beta}\right]X'\Sigma^{-1}\mathbf{y} + \left(X'\Sigma^{-1}X\right)^{-1}R'\left(V + R\left(X'\Sigma^{-1}X\right)^{-1}R'\right)^{-1}\mathbf{r}\right) \\
&= \lim_{V\to 0}\left(\mathrm{Var}\left[\hat{\beta}\right]X'\Sigma^{-1}\mathbf{y}\right) + \lim_{V\to 0}\left(X'\Sigma^{-1}X\right)^{-1}R'\left(V + R\left(X'\Sigma^{-1}X\right)^{-1}R'\right)^{-1}\mathbf{r} \\
&= \left(\lim_{V\to 0}\mathrm{Var}\left[\hat{\beta}\right]\right)X'\Sigma^{-1}\mathbf{y} + \left(X'\Sigma^{-1}X\right)^{-1}R'\left(R\left(X'\Sigma^{-1}X\right)^{-1}R'\right)^{-1}\mathbf{r} \\
&= \left(\lim_{V\to 0}\mathrm{Var}\left[\hat{\beta}\right]\right)X'\Sigma^{-1}\mathbf{y} + \left(X'\Sigma^{-1}X\right)^{-1}R'H\mathbf{r} \\
&= \left(\left(X'\Sigma^{-1}X\right)^{-1} - \left(X'\Sigma^{-1}X\right)^{-1}R'HR\left(X'\Sigma^{-1}X\right)^{-1}\right)X'\Sigma^{-1}\mathbf{y} + \left(X'\Sigma^{-1}X\right)^{-1}R'H\mathbf{r} \\
&= \left(I_k - \left(X'\Sigma^{-1}X\right)^{-1}R'HR\right)\left(X'\Sigma^{-1}X\right)^{-1}X'\Sigma^{-1}\mathbf{y} + \left(X'\Sigma^{-1}X\right)^{-1}R'H\mathbf{r}
\end{aligned}$$

In the limit the non-stochastic constraint is satisfied:

$$\begin{aligned}
\lim_{V\to 0}R\hat{\beta} &= \lim_{V\to 0}\left(R\left(\mathrm{Var}\left[\hat{\beta}\right]X'\Sigma^{-1}\mathbf{y} + \left(X'\Sigma^{-1}X\right)^{-1}R'\left(V + R\left(X'\Sigma^{-1}X\right)^{-1}R'\right)^{-1}\mathbf{r}\right)\right) \\
&= \lim_{V\to 0}\left(R\,\mathrm{Var}\left[\hat{\beta}\right]X'\Sigma^{-1}\mathbf{y}\right) + \lim_{V\to 0}R\left(X'\Sigma^{-1}X\right)^{-1}R'\left(V + R\left(X'\Sigma^{-1}X\right)^{-1}R'\right)^{-1}\mathbf{r} \\
&= \left(\lim_{V\to 0}R\,\mathrm{Var}\left[\hat{\beta}\right]\right)X'\Sigma^{-1}\mathbf{y} + R\left(X'\Sigma^{-1}X\right)^{-1}R'\left(R\left(X'\Sigma^{-1}X\right)^{-1}R'\right)^{-1}\mathbf{r} \\
&= (0)X'\Sigma^{-1}\mathbf{y} + I_r\mathbf{r} \\
&= \mathbf{r}
\end{aligned}$$

In an earlier paper [6:35f.] the author derived the formula for the non-stochastically constrained estimator $\beta^* = \left(I_k - (X'X)^{-1}R'HR\right)(X'X)^{-1}X'\mathbf{y} + (X'X)^{-1}R'H\mathbf{r}$. We see that the formulas are identical except for the presence of Σ^{-1} in the middle of X'X and X'**y**. (Remember too that H contains an X'X.) But the earlier paper simplistically assumed $\mathrm{Var}[\mathbf{e}] = \Sigma$ to be some scalar multiple of an identity matrix, i.e., $\sigma^2 I$, [6:35]. The general model can be reduced to the simpler model by a transformation [8:329f.]: If $\mathrm{Var}[\mathbf{e}] = \Sigma = \sigma^2\Phi$, where Φ is positive definite, then Φ^{-1} is also positive definite and there exists a non-singular W such that $\Phi^{-1} = W'W$ (*cf.* Appendix A). Transform the general model by

premultiplying it by W (a one-to-one transformation): $W\mathbf{y} = WX\beta + W\mathbf{e}$, where $Var[W\mathbf{e}] = WVar[\mathbf{e}]W' = W\sigma^2\Phi W' = \sigma^2 W(W'W)^{-1}W' = \sigma^2 WW^{-1}(W')^{-1}W' = \sigma^2 I_t$. The transformed model $(W\mathbf{y}) = (WX)\beta + (W\mathbf{e})$ has the scalar-identity variance of the simpler model, so the term corresponding to X'X is $(WX)'(WX) = X'W'WX = X'\Phi^{-1}X$. Similarly, the term corresponding to X'**y** is $(WX)'(W\mathbf{y}) = X'\Phi^{-1}\mathbf{y}$. The formula for β^* is so constructed as to be invariant to the scale of Φ; hence, Φ can be replaced by Σ with the result:

$$\beta^* = \left(I_k - \left(X'\Sigma^{-1}X\right)^{-1}R'HR\right)\left(X'\Sigma^{-1}X\right)^{-1}X'\Sigma^{-1}\mathbf{y} + \left(X'\Sigma^{-1}X\right)^{-1}R'Hr,$$

where $H = (R(X'\Sigma^{-1}X)^{-1}R')^{-1}$. Therefore, we have demonstrated that the non-stochastically constrained model is a limiting case of the stochastically constrained model.

Amemiya [1:25f.] performs a similar demonstration, but with the simplistic assumption that $Var\begin{bmatrix}\mathbf{e}\\ \mathbf{v}\end{bmatrix} = \sigma^2\begin{bmatrix}I_t & \\ & \frac{1}{\lambda^2}I_t\end{bmatrix}$. (His notation is different, but this is in effect his reasoning.) The limiting case results from letting λ^2 approach infinity. Our demonstration is more powerful, since it allows Var[**v**] to approach zero in any manner, not just as a shrinking scalar multiple of an identity matrix.

Appendix D

Estimating the Mean and the Variance of a Multivariate Random Sample

The variance of the error term of a linear statistical model is usually assumed to be known to within a proportionality constant, i.e., $\mathrm{Var}[\mathbf{e}] \propto \Phi$. But in the case of a multivariate random sample the entire variance can be estimated. We start with n $(k\times 1)$ random vectors $\mathbf{y}_1, \ldots, \mathbf{y}_n$ which are randomly sampled from a population of unknown mean and variance, μ and Σ. According to the definition of variance ([7:Appendix A] and [8:43]),

$$\Sigma = \mathrm{Var}[\mathbf{y}_i] = \mathrm{E}\left[(\mathbf{y}_i - \mu)(\mathbf{y}_i - \mu)'\right].$$

The mean and the variance will be estimated from the linear model:

$$\begin{bmatrix}\mathbf{y}_1\\ \vdots \\ \mathbf{y}_n\end{bmatrix}_{(nk\times 1)} = \begin{bmatrix}\mathrm{I}_k\\ \vdots \\ \mathrm{I}_k\end{bmatrix}_{(nk\times k)} \mu_{(k\times 1)} + \begin{bmatrix}\mathbf{e}_1\\ \vdots \\ \mathbf{e}_n\end{bmatrix}_{(nk\times 1)}, \text{ where } \mathrm{Var}\begin{bmatrix}\mathbf{e}_1\\ \vdots \\ \mathbf{e}_n\end{bmatrix} = \begin{bmatrix}\Sigma & & \\ & \ddots & \\ & & \Sigma\end{bmatrix}_{(nk\times nk)}$$

The variance matrix is block diagonal in Σ, because random sampling implies independent, identically distributed trials. The best linear unbiased estimator of μ happens not to depend on the unknown Σ:

$$\hat{\mu} = \left(\begin{bmatrix} I_k \\ \vdots \\ I_k \end{bmatrix}' \begin{bmatrix} \Sigma & & \\ & \ddots & \\ & & \Sigma \end{bmatrix}^{-1} \begin{bmatrix} I_k \\ \vdots \\ I_k \end{bmatrix} \right)^{-1} \begin{bmatrix} I_k \\ \vdots \\ I_k \end{bmatrix}' \begin{bmatrix} \Sigma & & \\ & \ddots & \\ & & \Sigma \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_n \end{bmatrix}$$

$$= \left(\begin{bmatrix} I_k & \cdots & I_k \end{bmatrix} \begin{bmatrix} \Sigma^{-1} & & \\ & \ddots & \\ & & \Sigma^{-1} \end{bmatrix} \begin{bmatrix} I_k \\ \vdots \\ I_k \end{bmatrix} \right)^{-1} \begin{bmatrix} I_k & \cdots & I_k \end{bmatrix} \begin{bmatrix} \Sigma^{-1} & & \\ & \ddots & \\ & & \Sigma^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_n \end{bmatrix}$$

$$= \left(\Sigma^{-1} + \ldots + \Sigma^{-1} \right)^{-1} \left(\Sigma^{-1}\mathbf{y}_1 + \ldots + \Sigma^{-1}\mathbf{y}_n \right)$$

$$= \left(n\Sigma^{-1} \right)^{-1} \left(\Sigma^{-1}\mathbf{y}_1 + \ldots + \Sigma^{-1}\mathbf{y}_n \right)$$

$$= \frac{1}{n}\Sigma \left(\Sigma^{-1}\mathbf{y}_1 + \ldots + \Sigma^{-1}\mathbf{y}_n \right)$$

$$= \frac{1}{n}\left(\mathbf{y}_1 + \ldots + \mathbf{y}_n \right)$$

(That the true Σ might be singular does not impugn the validity of the estimator.) Since the estimator is unbiased, $\mathrm{E}[\hat{\mu}] = \mu$. The variance is:

$$\mathrm{Var}[\hat{\mu}] = \mathrm{E}\left[(\hat{\mu} - \mu)(\hat{\mu} - \mu)' \right] = \left(\begin{bmatrix} I_k \\ \vdots \\ I_k \end{bmatrix}' \begin{bmatrix} \Sigma & & \\ & \ddots & \\ & & \Sigma \end{bmatrix}^{-1} \begin{bmatrix} I_k \\ \vdots \\ I_k \end{bmatrix} \right)^{-1} = \frac{1}{n}\Sigma$$

For future reference it is noted here that $\sum_1^n (\mathbf{y}_i - \hat{\mu}) = 0$ and $\sum_1^n (\mathbf{y}_i - \mu) = \sum_1^n (\hat{\mu} - \mu)$.

Now consider the function $\Psi(\nu) = \sum_1^n (\mathbf{y}_i - \nu)(\mathbf{y}_i - \nu)'$. This function can be minimized:

$$
\begin{aligned}
\Psi(\nu) &= \sum_{1}^{n} (\mathbf{y}_i - \nu)(\mathbf{y}_i - \nu)' \\
&= \sum_{1}^{n} \left((\mathbf{y}_i - \hat{\mu}) - (\nu - \hat{\mu})\right)\left((\mathbf{y}_i - \hat{\mu}) - (\nu - \hat{\mu})\right) \\
&= \sum_{1}^{n} (\mathbf{y}_i - \hat{\mu})(\mathbf{y}_i - \hat{\mu})' - \sum_{1}^{n} (\mathbf{y}_i - \hat{\mu})(\nu - \hat{\mu})' - \sum_{1}^{n} (\nu - \hat{\mu})(\mathbf{y}_i - \hat{\mu})' + \sum_{1}^{n} (\nu - \hat{\mu})(\nu - \hat{\mu})' \\
&= \Psi(\hat{\mu}) - \sum_{1}^{n} (\mathbf{y}_i - \hat{\mu})(\nu - \hat{\mu})' - \sum_{1}^{n} (\nu - \hat{\mu})(\mathbf{y}_i - \hat{\mu})' + \sum_{1}^{n} (\nu - \hat{\mu})(\nu - \hat{\mu})' \\
&= \Psi(\hat{\mu}) - \left(\sum_{1}^{n} (\mathbf{y}_i - \hat{\mu})\right)(\nu - \hat{\mu})' - (\nu - \hat{\mu})\left(\sum_{1}^{n} (\mathbf{y}_i - \hat{\mu})'\right) + \sum_{1}^{n} (\nu - \hat{\mu})(\nu - \hat{\mu})' \\
&= \Psi(\hat{\mu}) - (0)(\nu - \hat{\mu})' - (\nu - \hat{\mu})(0) + \sum_{1}^{n} (\nu - \hat{\mu})(\nu - \hat{\mu})' \\
&= \Psi(\hat{\mu}) + \sum_{1}^{n} (\nu - \hat{\mu})(\nu - \hat{\mu})' \\
&= \Psi(\hat{\mu}) + n(\nu - \hat{\mu})(\nu - \hat{\mu})' \\
&\geq \Psi(\hat{\mu})
\end{aligned}
$$

The matrix inequality (cf. Appendix A) holds because $n(\nu - \hat{\mu})(\nu - \hat{\mu})'$ is non-negative definite, with equality obtaining if and only if $\nu = \hat{\mu}$. Due to the existence and uniqueness of this minimum, we could have defined $\hat{\mu}$ as the minimizing argument of Ψ, rather than as the best linear unbiased estimator of the model above.

The minimum of Ψ is:

$$
\begin{aligned}
\Psi(\hat{\mu}) &= \sum_{1}^{n}(\mathbf{y}_i - \hat{\mu})(\mathbf{y}_i - \hat{\mu})' \\
&= \sum_{1}^{n}\left((\mathbf{y}_i - \mu) - (\hat{\mu} - \mu)\right)\left((\mathbf{y}_i - \mu) - (\hat{\mu} - \mu)\right)' \\
&= \sum_{1}^{n}(\mathbf{y}_i - \mu)(\mathbf{y}_i - \mu)' - \sum_{1}^{n}(\mathbf{y}_i - \mu)(\hat{\mu} - \mu)' - \sum_{1}^{n}(\hat{\mu} - \mu)(\mathbf{y}_i - \mu)' + \sum_{1}^{n}(\hat{\mu} - \mu)(\hat{\mu} - \mu)' \\
&= \sum_{1}^{n}(\mathbf{y}_i - \mu)(\mathbf{y}_i - \mu)' - \left(\sum_{1}^{n}(\mathbf{y}_i - \mu)\right)(\hat{\mu} - \mu)' - (\hat{\mu} - \mu)\left(\sum_{1}^{n}(\mathbf{y}_i - \mu)'\right) + \sum_{1}^{n}(\hat{\mu} - \mu)(\hat{\mu} - \mu)' \\
&= \sum_{1}^{n}(\mathbf{y}_i - \mu)(\mathbf{y}_i - \mu)' - \left(\sum_{1}^{n}(\hat{\mu} - \mu)\right)(\hat{\mu} - \mu)' - (\hat{\mu} - \mu)\left(\sum_{1}^{n}(\hat{\mu} - \mu)'\right) + \sum_{1}^{n}(\hat{\mu} - \mu)(\hat{\mu} - \mu)' \\
&= \sum_{1}^{n}(\mathbf{y}_i - \mu)(\mathbf{y}_i - \mu)' - \sum_{1}^{n}(\hat{\mu} - \mu)(\hat{\mu} - \mu)' - \sum_{1}^{n}(\hat{\mu} - \mu)(\hat{\mu} - \mu)' + \sum_{1}^{n}(\hat{\mu} - \mu)(\hat{\mu} - \mu)' \\
&= \sum_{1}^{n}(\mathbf{y}_i - \mu)(\mathbf{y}_i - \mu)' - \sum_{1}^{n}(\hat{\mu} - \mu)(\hat{\mu} - \mu)' \\
&= \sum_{1}^{n}(\mathbf{y}_i - \mu)(\mathbf{y}_i - \mu)' - n(\hat{\mu} - \mu)(\hat{\mu} - \mu)'
\end{aligned}
$$

But the importance of this minimum lies in its expected value:

$$
\begin{aligned}
\mathrm{E}\left[\Psi(\hat{\mu})\right] &= \mathrm{E}\left[\sum_{1}^{n}(\mathbf{y}_i - \mu)(\mathbf{y}_i - \mu)' - n(\hat{\mu} - \mu)(\hat{\mu} - \mu)'\right] \\
&= \mathrm{E}\left[\sum_{1}^{n}(\mathbf{y}_i - \mu)(\mathbf{y}_i - \mu)'\right] - \mathrm{E}\left[n(\hat{\mu} - \mu)(\hat{\mu} - \mu)'\right] \\
&= \sum_{1}^{n}\mathrm{E}\left[(\mathbf{y}_i - \mu)(\mathbf{y}_i - \mu)'\right] - n\mathrm{E}\left[(\hat{\mu} - \mu)(\hat{\mu} - \mu)'\right] \\
&= \sum_{1}^{n}\mathrm{Var}\left[\mathbf{y}_i\right] - n\mathrm{Var}\left[\hat{\mu}\right] \\
&= \sum_{1}^{n}\Sigma - n\left(\frac{1}{n}\Sigma\right) \\
&= n\Sigma - \Sigma \\
&= (n-1)\Sigma
\end{aligned}
$$

Therefore, $\hat{\Sigma} = \dfrac{1}{n-1}\Psi(\hat{\mu}) = \dfrac{1}{n-1}\sum_{1}^{n}(\mathbf{y}_i - \hat{\mu})(\mathbf{y}_i - \hat{\mu})'$ is an unbiased estimator of Σ.

Appendix E

Credibility and the Random-Effects Model

Appendix A introduced groups of statistical models. The first model consisted of n linear models of the form $\mathbf{y}_i = \mathrm{X}_i\beta_i + \mathbf{e}_i$, where $\mathrm{Var}[\mathbf{e}_i] = \Sigma_i$, for $i = 1, \ldots, n$. $\mathbf{y}_i$ and $\mathbf{e}_i$ were $(t_i \times 1)$, X_i was $(t_i \times k_i)$, β_i is $(k_i \times 1)$, and Σ_i was $(t_i \times t_i)$. Each Σ_i was non-singular, and each X_i was of full column rank, i.e., $\mathrm{rank}(\mathrm{X}_i) = k_i$, which ensured that each $(\mathrm{X}_i'\Sigma_i^{-1}\mathrm{X}_i)^{-1}$ existed. These specifications will be adopted here, but with the additional specification that all the k_is are equal: $k = k_1 = \ldots = k_n$. The model then appears as:

$$\begin{bmatrix} \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_n \end{bmatrix} = \begin{bmatrix} \mathrm{X}_1 & & \\ & \ddots & \\ & & \mathrm{X}_n \end{bmatrix} \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_n \end{bmatrix} + \begin{bmatrix} \mathbf{e}_1 \\ \vdots \\ \mathbf{e}_n \end{bmatrix}$$

$$\text{where } \mathrm{Var}\begin{bmatrix} \mathbf{e}_1 \\ \vdots \\ \mathbf{e}_n \end{bmatrix} = \begin{bmatrix} \Sigma_1 & & \\ & \ddots & \\ & & \Sigma_n \end{bmatrix}$$

As shown in Appendix A, the best linear unbiased estimator of β is:

$$\hat{\beta} = \begin{bmatrix} \hat{\beta}_1 \\ \\ \hat{\beta}_n \end{bmatrix} = \begin{bmatrix} \left(\mathrm{X}_1'\Sigma_1^{-1}\mathrm{X}_1\right)^{-1}\mathrm{X}_1'\Sigma_1^{-1}\mathbf{y}_1 \\ \vdots \\ \left(\mathrm{X}_n'\Sigma_n^{-1}\mathrm{X}_n\right)^{-1}\mathrm{X}_n'\Sigma_n^{-1}\mathbf{y}_n \end{bmatrix} = \begin{bmatrix} \mathrm{Var}\left[\hat{\beta}_1\right]\mathrm{X}_1'\Sigma_1^{-1}\mathbf{y}_1 \\ \vdots \\ \mathrm{Var}\left[\hat{\beta}_n\right]\mathrm{X}_n'\Sigma_n^{-1}\mathbf{y}_n \end{bmatrix}$$

This is called a fixed-effects model because every submodel is given its own β_i.

The second model of models was like the first, but with the constraint that all the β_is be equal: $\beta_0 = \beta_1 = \ldots = \beta_n$. The model then appears as:

$$\begin{bmatrix}\mathbf{y}_1\\ \vdots\\ \mathbf{y}_n\end{bmatrix}=\begin{bmatrix}\mathbf{X}_1\\ \vdots\\ \mathbf{X}_n\end{bmatrix}\beta_0+\begin{bmatrix}\mathbf{e}_1\\ \vdots\\ \mathbf{e}_n\end{bmatrix}$$

$$\text{where } \mathrm{Var}\begin{bmatrix}\mathbf{e}_1\\ \vdots\\ \mathbf{e}_n\end{bmatrix}=\begin{bmatrix}\Sigma_1 & & \\ & \ddots & \\ & & \Sigma_n\end{bmatrix}$$

Again, as shown in Appendix A, the best linear unbiased estimator of β_0 is:

$$\begin{aligned}\hat{\beta}_0 &= \left(\mathbf{X}_1'\Sigma_1^{-1}\mathbf{X}_1+\ldots+\mathbf{X}_n'\Sigma_n^{-1}\mathbf{X}_n\right)^{-1}\left(\mathbf{X}_1'\Sigma_1^{-1}\mathbf{y}_1+\ldots+\mathbf{X}_n'\Sigma_n^{-1}\mathbf{y}_n\right)\\ &= \left(\mathrm{Var}^{-1}\left[\hat{\beta}_1\right]+\ldots+\mathrm{Var}^{-1}\left[\hat{\beta}_n\right]\right)^{-1}\left(\mathrm{Var}^{-1}\left[\hat{\beta}_1\right]\hat{\beta}_1+\ldots+\mathrm{Var}^{-1}\left[\hat{\beta}_n\right]\hat{\beta}_n\right)\\ &= \left(\mathrm{Var}\left[\hat{\beta}_0\right]\right)\left(\mathrm{Var}^{-1}\left[\hat{\beta}_1\right]\hat{\beta}_1+\ldots+\mathrm{Var}^{-1}\left[\hat{\beta}_n\right]\hat{\beta}_n\right)\end{aligned}$$

This too is a fixed-effects model, but with only one fixed effect.

Now an attractive basis of a credibility model is the belief that the parameters (here γ_is) constitute a random sample from a distribution of mean γ_0 and variance V. So $\gamma_i = \gamma_0 + \mathbf{v}_i$, where $E[\mathbf{v}_i] = 0$, $\mathrm{Var}[\mathbf{v}_i] = V$, and the $\mathbf{v}_i$s do not covary either with each other or with the $\mathbf{e}_i$s. This transforms the fixed-effects model into the random-effects model (with estimations):

$$\begin{bmatrix}\mathbf{y}_1\\ \vdots\\ \mathbf{y}_n\\ \gamma_1\\ \vdots\\ \gamma_n\end{bmatrix}=\begin{bmatrix}\mathbf{X}_1 & & \\ & \ddots & \\ & & \mathbf{X}_n\\ \mathbf{I}_k & \cdots\cdots & \\ & \ddots & \\ & & \mathbf{I}_k\end{bmatrix}\begin{bmatrix}\gamma_0+\mathbf{v}_1\\ \vdots\\ \gamma_0+\mathbf{v}_n\end{bmatrix}+\begin{bmatrix}\mathbf{e}_1\\ \vdots\\ \mathbf{e}_n\\ 0\\ \vdots\\ 0\end{bmatrix}$$

$$=\begin{bmatrix}\mathbf{X}_1\\ \vdots\\ \mathbf{X}_n\\ \mathbf{I}_k\\ \vdots\\ \mathbf{I}_k\end{bmatrix}\gamma_0+\begin{bmatrix}\mathbf{X}_1\mathbf{v}_1+\mathbf{e}_1\\ \vdots\\ \mathbf{X}_n\mathbf{v}_n+\mathbf{e}_n\\ \mathbf{v}_1\\ \vdots\\ \mathbf{v}_n\end{bmatrix}$$

Instead of each submodel having its own fixed effect γ_i, there is one fixed-effect (γ_0) for the whole model and each submodel has its own random effect $\mathbf{v}_i$. Therefore, the estimations of this random-effects model have a non-zero error term, and the variance matrix is:

$$\mathrm{Var}\begin{bmatrix} \mathbf{X}_1\mathbf{v}_1 + \mathbf{e}_1 \\ \vdots \\ \mathbf{X}_n\mathbf{v}_n + \mathbf{e}_n \\ \mathbf{v}_1 \\ \vdots \\ \mathbf{v}_n \end{bmatrix} = \left[\begin{array}{ccc:ccc} \mathbf{X}_1\mathbf{V}\mathbf{X}_1' + \Sigma_1 & & & \mathbf{X}_1\mathbf{V} & & \\ & \ddots & & & \ddots & \\ & & \mathbf{X}_n\mathbf{V}\mathbf{X}_n' + \Sigma_n & & & \mathbf{X}_n\mathbf{V} \\ \hdashline \mathbf{V}\mathbf{X}_1' & & & \mathbf{V} & & \\ & \ddots & & & \ddots & \\ & & \mathbf{V}\mathbf{X}_n' & & & \mathbf{V} \end{array}\right] = \begin{bmatrix} \mathbf{T}_{11} & \mathbf{T}_{12} \\ \mathbf{T}_{21} & \mathbf{T}_{22} \end{bmatrix}$$

The best linear unbiased estimations are:

$$\begin{bmatrix} \hat{\gamma}_1 \\ \vdots \\ \hat{\gamma}_n \end{bmatrix} = \begin{bmatrix} \mathbf{I}_k \\ \vdots \\ \mathbf{I}_k \end{bmatrix} \hat{\gamma}_0 + \mathbf{T}_{21}\mathbf{T}_{11}^{-1}\left(\begin{bmatrix} \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_n \end{bmatrix} - \begin{bmatrix} \mathbf{X}_1 \\ \vdots \\ \mathbf{X}_n \end{bmatrix} \hat{\gamma}_0 \right),$$

$$\text{where } \hat{\gamma}_0 = \left(\begin{bmatrix} \mathbf{X}_1 \\ \vdots \\ \mathbf{X}_n \end{bmatrix}' \mathbf{T}_{11}^{-1} \begin{bmatrix} \mathbf{X}_1 \\ \vdots \\ \mathbf{X}_n \end{bmatrix} \right)^{-1} \begin{bmatrix} \mathbf{X}_1 \\ \vdots \\ \mathbf{X}_n \end{bmatrix}' \mathbf{T}_{11}^{-1} \begin{bmatrix} \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_n \end{bmatrix}$$

$$= \mathrm{Var}\left[\hat{\gamma}_0\right] \begin{bmatrix} \mathbf{X}_1 \\ \vdots \\ \mathbf{X}_n \end{bmatrix}' \mathbf{T}_{11}^{-1} \begin{bmatrix} \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_n \end{bmatrix}$$

Let us define the n blocks of $\mathbf{T}_{11}$ as $\mathbf{T}_{11_i} = \mathbf{X}_i\mathbf{V}\mathbf{X}_i' + \Sigma_i$. $\mathbf{T}_{11_i}$ is $(t_i \times t_i)$ and positive definite. We will also use the shorthand expression of '$\mathbf{T}_i$' for '$\mathbf{T}_{11_i}$'. Then the estimations may be written as:

$$\begin{bmatrix}\hat{\gamma}_1\\ \vdots\\ \hat{\gamma}_n\end{bmatrix} = \begin{bmatrix}I_k\\ \vdots\\ I_k\end{bmatrix}\hat{\gamma}_0 + T_{21}T_{11}^{-1}\left(\begin{bmatrix}\mathbf{y}_1\\ \vdots\\ \mathbf{y}_n\end{bmatrix} - \begin{bmatrix}X_1\\ \vdots\\ X_n\end{bmatrix}\hat{\gamma}_0\right)$$

$$= \begin{bmatrix}\hat{\gamma}_0\\ \vdots\\ \hat{\gamma}_0\end{bmatrix} + \begin{bmatrix}VX_1' & & \\ & \ddots & \\ & & VX_n'\end{bmatrix}\begin{bmatrix}T_1 & & \\ & \ddots & \\ & & T_n\end{bmatrix}^{-1}\begin{bmatrix}\mathbf{y}_1 - X_1\hat{\gamma}_0\\ \vdots\\ \mathbf{y}_n - X_n\hat{\gamma}_0\end{bmatrix}$$

$$= \begin{bmatrix}\hat{\gamma}_0\\ \vdots\\ \hat{\gamma}_0\end{bmatrix} + \begin{bmatrix}VX_1' & & \\ & \ddots & \\ & & VX_n'\end{bmatrix}\begin{bmatrix}T_1^{-1} & & \\ & \ddots & \\ & & T_n^{-1}\end{bmatrix}\begin{bmatrix}\mathbf{y}_1 - X_1\hat{\gamma}_0\\ \vdots\\ \mathbf{y}_n - X_n\hat{\gamma}_0\end{bmatrix}$$

$$= \begin{bmatrix}\hat{\gamma}_0 + VX_1'T_1^{-1}(\mathbf{y}_1 - X_1\hat{\gamma}_0)\\ \vdots\\ \hat{\gamma}_0 + VX_n'T_n^{-1}(\mathbf{y}_n - X_n\hat{\gamma}_0)\end{bmatrix},$$

where $\hat{\gamma}_0 = \left(\begin{bmatrix}X_1\\ \vdots\\ X_n\end{bmatrix}'\begin{bmatrix}T_1 & & \\ & \ddots & \\ & & T_n\end{bmatrix}^{-1}\begin{bmatrix}X_1\\ \vdots\\ X_n\end{bmatrix}\right)^{-1}\begin{bmatrix}X_1\\ \vdots\\ X_n\end{bmatrix}'\begin{bmatrix}T_1 & & \\ & \ddots & \\ & & T_n\end{bmatrix}^{-1}\begin{bmatrix}\mathbf{y}_1\\ \vdots\\ \mathbf{y}_n\end{bmatrix}$

$$= \left(\begin{bmatrix}X_1' & \cdots & X_n'\end{bmatrix}\begin{bmatrix}T_1^{-1} & & \\ & \ddots & \\ & & T_n^{-1}\end{bmatrix}\begin{bmatrix}X_1\\ \vdots\\ X_n\end{bmatrix}\right)^{-1}\begin{bmatrix}X_1' & \cdots & X_n'\end{bmatrix}\begin{bmatrix}T_1^{-1} & & \\ & \ddots & \\ & & T_n^{-1}\end{bmatrix}\begin{bmatrix}\mathbf{y}_1\\ \vdots\\ \mathbf{y}_n\end{bmatrix}$$

$$= \left(X_1'T_1^{-1}X_1 + \cdots + X_n'T_n^{-1}X_n\right)^{-1}\left(X_1'T_1^{-1}\mathbf{y}_1 + \cdots + X_n'T_n^{-1}\mathbf{y}_n\right)$$

$$= \mathrm{Var}[\hat{\gamma}_0]\left(X_1'T_1^{-1}\mathbf{y}_1 + \cdots + X_n'T_n^{-1}\mathbf{y}_n\right)$$

The penultimate expression for $\hat{\gamma}_0$ looks like the expression for $\hat{\beta}_0$ except that it contains terms with T_i^{-1} instead of terms with Σ_i^{-1}. But this small difference has great effects, which must be investigated. As a beginning, borrowing a theorem from Appendix A, viz., that $(A + BDC)^{-1} = A^{-1} - A^{-1}B(D^{-1} + CA^{-1}B)^{-1}CA^{-1}$, we have:

$$
\begin{aligned}
\mathbf{T}_i^{-1} &= \left(\boldsymbol{\Sigma}_i + \mathbf{X}_i \mathbf{V} \mathbf{X}_i'\right)^{-1} \\
&= \boldsymbol{\Sigma}_i^{-1} - \boldsymbol{\Sigma}_i^{-1} \mathbf{X}_i \left(\mathbf{V}^{-1} + \mathbf{X}_i' \boldsymbol{\Sigma}_i^{-1} \mathbf{X}_i\right)^{-1} \mathbf{X}_i' \boldsymbol{\Sigma}_i^{-1}
\end{aligned}
$$

Then $\mathbf{X}_i' \mathbf{T}_i^{-1} \mathbf{X}_i = \mathbf{X}_i' \boldsymbol{\Sigma}_i^{-1} \mathbf{X}_i - \mathbf{X}_i' \boldsymbol{\Sigma}_i^{-1} \mathbf{X}_i \left(\mathbf{V}^{-1} + \mathbf{X}_i' \boldsymbol{\Sigma}_i^{-1} \mathbf{X}_i\right)^{-1} \mathbf{X}_i' \boldsymbol{\Sigma}_i^{-1} \mathbf{X}_i$. Moreover:

$$
\begin{aligned}
\mathbf{X}_i' \mathbf{T}_i^{-1} \mathbf{y}_i &= \mathbf{X}_i' \boldsymbol{\Sigma}_i^{-1} \mathbf{y}_i - \mathbf{X}_i' \boldsymbol{\Sigma}_i^{-1} \mathbf{X}_i \left(\mathbf{V}^{-1} + \mathbf{X}_i' \boldsymbol{\Sigma}_i^{-1} \mathbf{X}_i\right)^{-1} \mathbf{X}_i' \boldsymbol{\Sigma}_i^{-1} \mathbf{y}_i \\
&= \mathbf{X}_i' \boldsymbol{\Sigma}_i^{-1} \mathbf{X}_i \left(\mathbf{X}_i' \boldsymbol{\Sigma}_i^{-1} \mathbf{X}_i\right)^{-1} \mathbf{X}_i' \boldsymbol{\Sigma}_i^{-1} \mathbf{y}_i - \mathbf{X}_i' \boldsymbol{\Sigma}_i^{-1} \mathbf{X}_i \left(\mathbf{V}^{-1} + \mathbf{X}_i' \boldsymbol{\Sigma}_i^{-1} \mathbf{X}_i\right)^{-1} \mathbf{X}_i' \boldsymbol{\Sigma}_i^{-1} \mathbf{X}_i \left(\mathbf{X}_i' \boldsymbol{\Sigma}_i^{-1} \mathbf{X}_i\right)^{-1} \mathbf{X}_i' \boldsymbol{\Sigma}_i^{-1} \mathbf{y}_i \\
&= \left(\mathbf{X}_i' \boldsymbol{\Sigma}_i^{-1} \mathbf{X}_i - \mathbf{X}_i' \boldsymbol{\Sigma}_i^{-1} \mathbf{X}_i \left(\mathbf{V}^{-1} + \mathbf{X}_i' \boldsymbol{\Sigma}_i^{-1} \mathbf{X}_i\right)^{-1} \mathbf{X}_i' \boldsymbol{\Sigma}_i^{-1} \mathbf{X}_i\right) \left(\mathbf{X}_i' \boldsymbol{\Sigma}_i^{-1} \mathbf{X}_i\right)^{-1} \mathbf{X}_i' \boldsymbol{\Sigma}_i^{-1} \mathbf{y}_i \\
&= \mathbf{X}_i' \mathbf{T}_i^{-1} \mathbf{X}_i \left(\mathbf{X}_i' \boldsymbol{\Sigma}_i^{-1} \mathbf{X}_i\right)^{-1} \mathbf{X}_i' \boldsymbol{\Sigma}_i^{-1} \mathbf{y}_i \\
&= \mathbf{X}_i' \mathbf{T}_i^{-1} \mathbf{X}_i \hat{\boldsymbol{\beta}}_i
\end{aligned}
$$

And finally:

$$
\begin{aligned}
\hat{\gamma}_0 &= \left(\mathbf{X}_1' \mathbf{T}_1^{-1} \mathbf{X}_1 + \cdots + \mathbf{X}_n' \mathbf{T}_n^{-1} \mathbf{X}_n\right)^{-1} \left(\mathbf{X}_1' \mathbf{T}_1^{-1} \mathbf{y}_1 + \cdots + \mathbf{X}_n' \mathbf{T}_n^{-1} \mathbf{y}_n\right) \\
&= \left(\mathbf{X}_1' \mathbf{T}_1^{-1} \mathbf{X}_1 + \cdots + \mathbf{X}_n' \mathbf{T}_n^{-1} \mathbf{X}_n\right)^{-1} \left(\mathbf{X}_1' \mathbf{T}_1^{-1} \mathbf{X}_1 \hat{\boldsymbol{\beta}}_1 + \cdots + \mathbf{X}_n' \mathbf{T}_n^{-1} \mathbf{X}_n \hat{\boldsymbol{\beta}}_n\right) \\
&= \mathrm{Var}\left[\hat{\gamma}_0\right] \left(\mathbf{X}_1' \mathbf{T}_1^{-1} \mathbf{X}_1 \hat{\boldsymbol{\beta}}_1 + \cdots + \mathbf{X}_n' \mathbf{T}_n^{-1} \mathbf{X}_n \hat{\boldsymbol{\beta}}_n\right)
\end{aligned}
$$

Therefore, the estimator of the grand parameter of this credibility model (γ_0) is like the estimator of the grand parameter of the non-credibility model (β_0) in that both are weighted averages of the estimators of the fixed-effects model (the $\hat{\boldsymbol{\beta}}_i$s). The difference is that the weights of the credibility model are $\mathbf{X}_i' \mathbf{T}_i^{-1} \mathbf{X}_i$, whereas those of the non-credibility model are $\mathbf{X}_i' \boldsymbol{\Sigma}_i^{-1} \mathbf{X}_i$.

There is a danger of using the fixed-effects estimators $\hat{\boldsymbol{\beta}}_i = \left(\mathbf{X}_i' \boldsymbol{\Sigma}_i^{-1} \mathbf{X}_i\right)^{-1} \mathbf{X}_i' \boldsymbol{\Sigma}_i^{-1} \mathbf{y}_i$ in this random-effects model. Whichever model is assumed:

$$\begin{aligned}\mathrm{Var}\left[\hat{\beta}_i\right] &= \mathrm{Var}\left[\left(X_i'\Sigma_i^{-1}X_i\right)^{-1}X_i'\Sigma_i^{-1}\mathbf{y}_i\right] \\ &= \left(X_i'\Sigma_i^{-1}X_i\right)^{-1}X_i'\Sigma_i^{-1}\mathrm{Var}\left[\mathbf{y}_i\right]\Sigma_i^{-1}X_i\left(X_i'\Sigma_i^{-1}X_i\right)^{-1}\end{aligned}$$

However, under the fixed-effects model:

$$\begin{aligned}\mathrm{Var}\left[\hat{\beta}_i\right] &= \left(X_i'\Sigma_i^{-1}X_i\right)^{-1}X_i'\Sigma_i^{-1}\mathrm{Var}\left[\mathbf{y}_i\right]\Sigma_i^{-1}X_i\left(X_i'\Sigma_i^{-1}X_i\right)^{-1} \\ &= \left(X_i'\Sigma_i^{-1}X_i\right)^{-1}X_i'\Sigma_i^{-1}\Sigma_i\Sigma_i^{-1}X_i\left(X_i'\Sigma_i^{-1}X_i\right)^{-1} \\ &= \left(X_i'\Sigma_i^{-1}X_i\right)^{-1}X_i'\Sigma_i^{-1}X_i\left(X_i'\Sigma_i^{-1}X_i\right)^{-1} \\ &= \left(X_i'\Sigma_i^{-1}X_i\right)^{-1}\end{aligned}$$

But under the random-effects model:

$$\begin{aligned}\mathrm{Var}\left[\hat{\beta}_i\right] &= \left(X_i'\Sigma_i^{-1}X_i\right)^{-1}X_i'\Sigma_i^{-1}\mathrm{Var}\left[\mathbf{y}_i\right]\Sigma_i^{-1}X_i\left(X_i'\Sigma_i^{-1}X_i\right)^{-1} \\ &= \left(X_i'\Sigma_i^{-1}X_i\right)^{-1}X_i'\Sigma_i^{-1}T_i\Sigma_i^{-1}X_i\left(X_i'\Sigma_i^{-1}X_i\right)^{-1} \\ &= \left(X_i'\Sigma_i^{-1}X_i\right)^{-1}X_i'\Sigma_i^{-1}\left(X_iVX_i'+\Sigma_i\right)\Sigma_i^{-1}X_i\left(X_i'\Sigma_i^{-1}X_i\right)^{-1} \\ &= \left(X_i'\Sigma_i^{-1}X_i\right)^{-1}X_i'\Sigma_i^{-1}X_iVX_i'\Sigma_i^{-1}X_i\left(X_i'\Sigma_i^{-1}X_i\right)^{-1}+\left(X_i'\Sigma_i^{-1}X_i\right)^{-1}X_i'\Sigma_i^{-1}\Sigma_i\Sigma_i^{-1}X_i\left(X_i'\Sigma_i^{-1}X_i\right)^{-1} \\ &= V+\left(X_i'\Sigma_i^{-1}X_i\right)^{-1}X_i'\Sigma_i^{-1}X_i\left(X_i'\Sigma_i^{-1}X_i\right)^{-1} \\ &= V+\left(X_i'\Sigma_i^{-1}X_i\right)^{-1}\end{aligned}$$

Further manipulation (again, using the theorem from Appendix A cited above) yields:

$$\begin{aligned}\mathrm{Var}\left[\hat{\beta}_i\right] &= V+\left(X_i'\Sigma_i^{-1}X_i\right)^{-1} \\ &= \left(\left(\left(X_i'\Sigma_i^{-1}X_i\right)^{-1}+V\right)^{-1}\right)^{-1} \\ &= \left(X_i'\Sigma_i^{-1}X_i-X_i'\Sigma_i^{-1}X_i\left(V^{-1}+X_i'\Sigma_i^{-1}X_i\right)^{-1}X_i'\Sigma_i^{-1}X_i\right)^{-1} \\ &= \left(X_i'T_i^{-1}X_i\right)^{-1}\end{aligned}$$

So, the variance of this estimator under the random-effects model differs from that under the fixed-effects model either by the addition of V to the latter or by the substitution of T for Σ in the latter. With the use of the correct variance, the true formula for $\mathrm{Var}\left[\hat{\gamma}_0\right]$ results:

$$
\begin{aligned}
\mathrm{Var}[\hat{\gamma}_0] &= \mathrm{Var}\left[\left(X_1'T_1^{-1}X_1+\cdots+X_n'T_n^{-1}X_n\right)^{-1}\left(X_1'T_1^{-1}X_1\hat{\beta}_1+\cdots+X_n'T_n^{-1}X_n\hat{\beta}_n\right)\right] \\
&= \left(X_1'T_1^{-1}X_1+\cdots+X_n'T_n^{-1}X_n\right)^{-1}\mathrm{Var}\left[X_1'T_1^{-1}X_1\hat{\beta}_1+\cdots+X_n'T_n^{-1}X_n\hat{\beta}_n\right] \\
&\quad \left(X_1'T_1^{-1}X_1+\cdots+X_n'T_n^{-1}X_n\right)^{-1} \\
&= \left(X_1'T_1^{-1}X_1+\cdots+X_n'T_n^{-1}X_n\right)^{-1}\left(X_1'T_1^{-1}X_1\mathrm{Var}\left[\hat{\beta}_1\right]X_1'T_1^{-1}X_1+\cdots+X_n'T_n^{-1}X_n\mathrm{Var}\left[\hat{\beta}_n\right]X_n'T_n^{-1}X_n\right) \\
&\quad \left(X_1'T_1^{-1}X_1+\cdots+X_n'T_n^{-1}X_n\right)^{-1} \\
&= \left(X_1'T_1^{-1}X_1+\cdots+X_n'T_n^{-1}X_n\right)^{-1}\left(X_1'T_1^{-1}X_1\left(X_1'T_1^{-1}X_1\right)^{-1}X_1'T_1^{-1}X_1+\cdots+X_n'T_n^{-1}X_n\left(X_n'T_n^{-1}X_n\right)^{-1}X_n'T_n^{-1}X_n\right) \\
&\quad \left(X_1'T_1^{-1}X_1+\cdots+X_n'T_n^{-1}X_n\right)^{-1} \\
&= \left(X_1'T_1^{-1}X_1+\cdots+X_n'T_n^{-1}X_n\right)^{-1}\left(X_1'T_1^{-1}X_1+\cdots+X_n'T_n^{-1}X_n\right) \\
&\quad \left(X_1'T_1^{-1}X_1+\cdots+X_n'T_n^{-1}X_n\right)^{-1} \\
&= \left(X_1'T_1^{-1}X_1+\cdots+X_n'T_n^{-1}X_n\right)^{-1}
\end{aligned}
$$

Limiting cases for V of the random-effects model are important. The first limiting case, as $V \to 0$, is simple. As $V \to 0$, the $\mathbf{v}_i$s $\to$ zero vectors, and model approaches:

$$
\begin{bmatrix} \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_n \\ \gamma_1 \\ \vdots \\ \gamma_n \end{bmatrix} = \begin{bmatrix} X_1 & & \\ & \ddots & \\ & & X_n \\ I_k & & \\ & \ddots & \\ & & I_k \end{bmatrix} \begin{bmatrix} \gamma_0 + 0 \\ \vdots \\ \gamma_0 + 0 \end{bmatrix} + \begin{bmatrix} \mathbf{e}_1 \\ \vdots \\ \mathbf{e}_n \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \begin{bmatrix} X_1 \\ \vdots \\ X_n \\ I_k \\ \vdots \\ I_k \end{bmatrix} \gamma_0 + \begin{bmatrix} \mathbf{e}_1 \\ \vdots \\ \mathbf{e}_n \\ 0 \\ \vdots \\ 0 \end{bmatrix}
$$

But this is the fixed-effects model with the constraint that all the γ_is be equal: $\gamma_0 = \gamma_1 = \ldots = \gamma_n$. And as $V \to 0$, $T_i \to \Sigma_i$, and:

$$
\begin{aligned}
\lim_{V\to 0}\hat{\gamma}_0 &= \lim_{V\to 0}\left(X_1'T_1^{-1}X_1+\cdots+X_n'T_n^{-1}X_n\right)^{-1}\left(X_1'T_1^{-1}\mathbf{y}_1+\cdots+X_n'T_n^{-1}\mathbf{y}_n\right) \\
&= \left(X_1'\Sigma_1^{-1}X_1+\cdots+X_n'\Sigma_n^{-1}X_n\right)^{-1}\left(X_1'\Sigma_1^{-1}\mathbf{y}_1+\cdots+X_n'\Sigma_n^{-1}\mathbf{y}_n\right) \\
&= \mathrm{Var}\left[\hat{\beta}_0\right]\left(X_1'\Sigma_1^{-1}\mathbf{y}_1+\cdots+X_n'\Sigma_n^{-1}\mathbf{y}_n\right) \\
&= \hat{\beta}_0
\end{aligned}
$$

In typical actuarial parlance, the submodels of this case have no credibility.

The opposite limiting case is for V to approach infinity. But the more precise meaning is that $V^{-1} \to 0$. It is best to see what happens to $\hat{\gamma}_0$ in this case. Inasmuch as $\hat{\gamma}_0 = \left(X_1'T_1^{-1}X_1 + \cdots + X_n'T_n^{-1}X_n\right)^{-1}\left(X_1'T_1^{-1}X_1\hat{\beta}_1 + \cdots + X_n'T_n^{-1}X_n\hat{\beta}_n\right)$ and $X_i'T_i^{-1}X_i \to 0$ as $V^{-1} \to 0$, the limit is the indeterminate form $0^{-1}0$. However, since $V + \left(X_i'\Sigma_i^{-1}X_i\right)^{-1} = \left(X_i'T_i^{-1}X_i\right)^{-1}$:

$$\begin{aligned} X_i'T_i^{-1}X_i &= \left(V + \left(X_i'\Sigma_i^{-1}X_i\right)^{-1}\right)^{-1} \\ &= \left(\left(I_k + \left(X_i'\Sigma_i^{-1}X_i\right)^{-1}V^{-1}\right)V\right)^{-1} \\ &= V^{-1}\left(I_k + \left(X_i'\Sigma_i^{-1}X_i\right)^{-1}V^{-1}\right)^{-1} \\ &= V^{-1}U_i \end{aligned}$$

Therefore:

$$\begin{aligned} \hat{\gamma}_0 &= \left(V^{-1}U_1 + \cdots + V^{-1}U_n\right)^{-1}\left(V^{-1}U_1\hat{\beta}_1 + \cdots + V^{-1}U_n\hat{\beta}_n\right) \\ &= \left(V^{-1}\left(U_1 + \cdots + U_n\right)\right)^{-1}\left(V^{-1}\left(U_1\hat{\beta}_1 + \cdots + U_n\hat{\beta}_n\right)\right) \\ &= \left(U_1 + \cdots + U_n\right)^{-1}VV^{-1}\left(U_1\hat{\beta}_1 + \cdots + U_n\hat{\beta}_n\right) \\ &= \left(U_1 + \cdots + U_n\right)^{-1}\left(U_1\hat{\beta}_1 + \cdots + U_n\hat{\beta}_n\right) \end{aligned}$$

But $\lim_{V^{-1}\to 0} U_i = \lim_{V^{-1}\to 0}\left(I_k + \left(X_i'\Sigma_i^{-1}X_i\right)^{-1}V^{-1}\right)^{-1} = \lim_{V^{-1}\to 0}\left(I_k + \left(X_i'\Sigma_i^{-1}X_i\right)^{-1}0\right)^{-1} = I_k$. Hence:

$$\begin{aligned} \lim_{V^{-1}\to 0}\hat{\gamma}_0 &= \lim_{V^{-1}\to 0}\left(U_1 + \cdots + U_n\right)^{-1}\left(U_1\hat{\beta}_1 + \cdots + U_n\hat{\beta}_n\right) \\ &= \left(I_k + \cdots + I_k\right)^{-1}\left(I_k\hat{\beta}_1 + \cdots + I_k\hat{\beta}_n\right) \\ &= \frac{1}{n}\left(\hat{\beta}_1 + \cdots + \hat{\beta}_n\right) \end{aligned}$$

In actuarial parlance, the submodels of this case have full credibility. Therefore, it makes sense here for $\hat{\gamma}_0$ to be the simple average of the fixed-effects estimators.

We turn now to the credibility estimators, and elaborate the formula:

$$\begin{bmatrix} \hat{\gamma}_1 \\ \vdots \\ \hat{\gamma}_n \end{bmatrix} = \begin{bmatrix} \hat{\gamma}_0 + V X_1' T_1^{-1} (\mathbf{y}_1 - X_1 \hat{\gamma}_0) \\ \vdots \\ \hat{\gamma}_0 + V X_n' T_n^{-1} (\mathbf{y}_n - X_n \hat{\gamma}_0) \end{bmatrix}$$

$$= \begin{bmatrix} \hat{\gamma}_0 + V X_1' T_1^{-1} \mathbf{y}_1 - V X_1' T_1^{-1} X_1 \hat{\gamma}_0 \\ \vdots \\ \hat{\gamma}_0 + V X_n' T_n^{-1} \mathbf{y}_n - V X_n' T_n^{-1} X_n \hat{\gamma}_0 \end{bmatrix}$$

$$= \begin{bmatrix} \hat{\gamma}_0 + V X_1' T_1^{-1} X_1 \hat{\beta}_1 - V X_1' T_1^{-1} X_1 \hat{\gamma}_0 \\ \vdots \\ \hat{\gamma}_0 + V X_n' T_n^{-1} X_n \hat{\beta}_n - V X_n' T_n^{-1} X_n \hat{\gamma}_0 \end{bmatrix}$$

$$= \begin{bmatrix} V X_1' T_1^{-1} X_1 \hat{\beta}_1 + (I_k - V X_1' T_1^{-1} X_1) \hat{\gamma}_0 \\ \vdots \\ V X_n' T_n^{-1} X_n \hat{\beta}_n + (I_k - V X_n' T_n^{-1} X_n) \hat{\gamma}_0 \end{bmatrix}$$

$$= \begin{bmatrix} Z_1 \hat{\beta}_1 + (I_k - Z_1) \hat{\gamma}_0 \\ \vdots \\ Z_n \hat{\beta}_n + (I_k - Z_n) \hat{\gamma}_0 \end{bmatrix}$$

So the credibility estimators are matrix-weighted averages of the fixed-effects estimators and the estimator of the grand parameter. This is a *k*-dimensional generalization of what actuaries call credibility weighting. (See the remark in Appendix A on how matrix-weighted averages differ from scalar-weighted averages.)

In the first limiting case, $\lim_{V \to 0} Z_i = \lim_{V \to 0} V X_i T_i^{-1} X_i' = 0 X_i \Sigma_i^{-1} X_i' = 0$; so, $\lim_{V \to 0} \hat{\gamma}_i = \hat{\gamma}_0$. This is to say that as the submodels lose credibility, random-effects model approaches the constrained fixed-effects model, or the one-fixed-effect model. In the opposite limiting case, $\lim_{V^{-1} \to 0} Z_i = \lim_{V^{-1} \to 0} V X_i T_i^{-1} X_i' = \lim_{V^{-1} \to 0} V V^{-1} U_i = \lim_{V^{-1} \to 0} U_i = I_k$; so, $\lim_{V^{-1} \to 0} \hat{\gamma}_i = \hat{\beta}_i$. This means

that as the submodels gain credibility, the random-effects model approaches the *n*-fixed-effects model.

In the linear model $\mathbf{y} = X\beta + \mathbf{e}$, where $\mathrm{Var}[\mathbf{e}] = \Sigma$, $\hat{\beta} = \left(X'\Sigma^{-1}X\right)^{-1}X'\Sigma^{-1}\mathbf{y}$. Therefore:

$$\begin{aligned} X'\Sigma^{-1}\left(\mathbf{y} - X\hat{\beta}\right) &= X'\Sigma^{-1}\mathbf{y} - X'\Sigma^{-1}X\hat{\beta} \\ &= X'\Sigma^{-1}\mathbf{y} - X'\Sigma^{-1}X\left(X'\Sigma^{-1}X\right)^{-1}X'\Sigma^{-1}\mathbf{y} \\ &= X'\Sigma^{-1}\mathbf{y} - X'\Sigma^{-1}\mathbf{y} \\ &= 0 \end{aligned}$$

We will apply this identity in the following derivation:

$$
\begin{aligned}
\begin{bmatrix} I_k & \cdots & I_k \end{bmatrix}\begin{bmatrix} \hat{\gamma}_1 \\ \vdots \\ \hat{\gamma}_n \end{bmatrix} &= \begin{bmatrix} I_k & \cdots & I_k \end{bmatrix}\left(\begin{bmatrix} \hat{\gamma}_0 \\ \vdots \\ \hat{\gamma}_0 \end{bmatrix} + \begin{bmatrix} VX_1' & & \\ & \ddots & \\ & & VX_n' \end{bmatrix}\begin{bmatrix} T_1 & & \\ & \ddots & \\ & & T_n \end{bmatrix}^{-1}\begin{bmatrix} \mathbf{y}_1 - X_1\hat{\gamma}_0 \\ \vdots \\ \mathbf{y}_n - X_n\hat{\gamma}_0 \end{bmatrix}\right) \\
&= \begin{bmatrix} I_k & \cdots & I_k \end{bmatrix}\begin{bmatrix} \hat{\gamma}_0 \\ \vdots \\ \hat{\gamma}_0 \end{bmatrix} \\
&\quad + \begin{bmatrix} I_k & \cdots & I_k \end{bmatrix}\begin{bmatrix} VX_1' & & \\ & \ddots & \\ & & VX_n' \end{bmatrix}\begin{bmatrix} T_1 & & \\ & \ddots & \\ & & T_n \end{bmatrix}^{-1}\begin{bmatrix} \mathbf{y}_1 - X_1\hat{\gamma}_0 \\ \vdots \\ \mathbf{y}_n - X_n\hat{\gamma}_0 \end{bmatrix} \\
&= \begin{bmatrix} I_k & \cdots & I_k \end{bmatrix}\begin{bmatrix} \hat{\gamma}_0 \\ \vdots \\ \hat{\gamma}_0 \end{bmatrix} + \begin{bmatrix} VX_1' & \cdots & VX_n' \end{bmatrix}\begin{bmatrix} T_1 & & \\ & \ddots & \\ & & T_n \end{bmatrix}^{-1}\begin{bmatrix} \mathbf{y}_1 - X_1\hat{\gamma}_0 \\ \vdots \\ \mathbf{y}_n - X_n\hat{\gamma}_0 \end{bmatrix} \\
&= \begin{bmatrix} I_k & \cdots & I_k \end{bmatrix}\begin{bmatrix} \hat{\gamma}_0 \\ \vdots \\ \hat{\gamma}_0 \end{bmatrix} + V\begin{bmatrix} X_1' & \cdots & X_n' \end{bmatrix}\begin{bmatrix} T_1 & & \\ & \ddots & \\ & & T_n \end{bmatrix}^{-1}\begin{bmatrix} \mathbf{y}_1 - X_1\hat{\gamma}_0 \\ \vdots \\ \mathbf{y}_n - X_n\hat{\gamma}_0 \end{bmatrix} \\
&= \begin{bmatrix} I_k & \cdots & I_k \end{bmatrix}\begin{bmatrix} \hat{\gamma}_0 \\ \vdots \\ \hat{\gamma}_0 \end{bmatrix} + VX'T^{-1}\left(\mathbf{y} - X\hat{\gamma}_0\right) \\
&= \begin{bmatrix} I_k & \cdots & I_k \end{bmatrix}\begin{bmatrix} \hat{\gamma}_0 \\ \vdots \\ \hat{\gamma}_0 \end{bmatrix} + V(0) \\
&= \begin{bmatrix} I_k & \cdots & I_k \end{bmatrix}\begin{bmatrix} \hat{\gamma}_0 \\ \vdots \\ \hat{\gamma}_0 \end{bmatrix}
\end{aligned}
$$

This shows that the simple average of the credibility estimators equals the grand parameter. Earlier we saw that in the fixed effects model, $\hat{\beta}_0$ was a weighted average of the $\hat{\beta}_i$s, the weights being proportional to the inverses of the variances of the $\hat{\beta}_i$s. This average is aristocratic in that the better $\hat{\beta}_i$s (i.e., those with the smaller variances) receive more weight. But in the random effects model, $\hat{\gamma}_0$ is a simple average of the $\hat{\gamma}_i$s. This suggests

an interpretation of credibility: credibility democratizes submodels. After a credibility adjustment, every submodel is entitled to one vote in determining the grand parameter. Of course, the weaker submodels are adjusted more vigorously.

Appendix F

A SAS® Procedure for Credibility Problems

According to Appendix E, many credibility problems can be expressed as random-effects statistical models. There is a SAS® procedure, PROC MIXED, which is very versatile with random-effects models. This procedure formulates the model as [12:575f., 634] $\mathbf{y} = X\beta + Z\gamma + \varepsilon$, where $E\begin{bmatrix}\gamma \\ \varepsilon\end{bmatrix} = \begin{bmatrix}0 \\ 0\end{bmatrix}$, and $Var\begin{bmatrix}\gamma \\ \varepsilon\end{bmatrix} = \begin{bmatrix}G & 0 \\ 0 & R\end{bmatrix}$. $Z\gamma + \varepsilon$ is the total error term, with a mean of 0 and a variance of $V = ZGZ' + R$. We know that the best linear unbiased estimator of β is $(X'V^{-1}X)^{-1}X'V^{-1}\mathbf{y}$. To estimate γ, we would use the estimator equation $\gamma = 0\beta + \gamma$; so $\hat{\gamma} = 0\hat{\beta} + Cov[\gamma,\ Z\gamma + \varepsilon]V^{-1}\left(\mathbf{y} - X\hat{\beta}\right) = GZ'V^{-1}\left(\mathbf{y} - X\hat{\beta}\right)$ [12:641]. But the most powerful feature of this procedure is that the variance matrices may be specified with an unknown parameter vector, viz., $G(\theta)$ and $R(\theta)$. The procedure will estimate θ, whether by variance components or by maximum likelihood [12:588, 639f.]. This model is more general than the random-effects examples of this paper; and estimating θ is a more general problem than estimating the random-effects variance of those examples. The following code succinctly solves the problem posed by Gary Venter [13] and treated as Example 5 of this paper:

```
/** This SAS program uses PROC MIXED to solve the problem on page 433 **/
/** of Gary Venter's "Credibility," Foundations of Casualty Actuarial **/
/** Science, Casualty Actuarial Society, 1990.                        **/

data data1;
   input risk year1-year6;
   cards;
1 0.430 0.375 2.341 0.175 1.016 0.466
2 0.247 1.587 1.939 0.712 0.054 0.261
3 0.661 0.237 0.063 0.250 0.602 0.700
```

```
4 0.182 0.351 0.011 0.022 0.019 0.252
5 0.311 0.664 1.002 0.038 0.370 2.502
6 0.301 0.253 0.044 0.109 2.105 0.891
7 0.219 1.186 0.431 1.405 0.241 0.804
8 0.002 0.058 0.235 0.018 0.713 0.208
9 0.796 0.260 0.932 0.857 0.129 0.349
;

proc transpose data=data1 out=data1 (rename=(_name_=time col1=x));
   by risk;

proc mixed data=data1;
   class risk;
   model x= /p s;
   random intercept /g s subject=risk;
run;
```

Once the time is invested to learn how to use routines like PROC MIXED, many complicated problems can be solved easily and quickly. However, it is possible to go overboard and to pose problems that are so complicated that one might unknowingly misuse the software. In such cases, a wrong answer may go undetected because intuition has been overwhelmed by the complexity.