Credibility prediction of first and second moments in the hierarchical model

WILLIAM S. JEWELL, Berkeley

Credibility Prediction of First and Second Moments in the Hierarchical Model

1 Introduction

The hierarchical model occupies an important place in Bayesian inference, since it shows how collateral data from a portfolio (cohort) of exchangeable risks can be used to improve predictions for any single risk in the portfolio. Unfortunately, an exact Bayesian formulation is difficult to apply in the general case, because analytic predictive distributions can only be obtained for the normal-normal-normal model with fixed variances (*Lindley/Smith*, 1972) and a heteroscedastic generalization (*Jewell*, 1987). Thus, an important step was the development of a one-dimensional credibility (linear least-squares) approximation to the predictive mean (*Taylor*, 1974; *Jewell*, 1975 b). In this paper, we use an idea of *Jewell/Schnieper* (1985) to find simultaneous approximations to both the first- and second-order predictive moments in the hierarchical model, by using multi-dimensional credibility theory. Because the resulting covariance matrix has special structure, the size of the necessary matrix inversions can be drastically reduced by introducing prototype forecasting formulae. We validate the method with exact results for variants of the normal-normal-normal model, then obtain approximate results for other models, with special attention to limiting behavior.

2 The Hierarchical Model

Consider an *individual risk* (labelled # 1), characterized by an unknown *risk parameter*, $\tilde{\theta}_1$, from which n_1 i.i.d. *observations*, $\mathscr{D}_1 = \{x_{1t}\}$ $(t = 1, 2, \ldots, n_1)$, are available; we wish to predict a *future observation*, say $\tilde{w}_1 = \tilde{x}_{1,n_1+1}$, of this risk. Given the *model density*, $p(x_{1t} \mid \theta_1)$, and the *prior parameter density*, $p(\theta_1)$, finding the *forecast density*, $p(w_1 \mid \mathscr{D})$, is then a simple exercise in Bayes law. For a variety of simple likelihoods and priors (*Jewell*, 1974, 1975 a), the predictive mean is a linear function of the data:

$$\mathscr{E}\{\tilde{w}_1 \mid \mathscr{D}_1\} = f_1(\mathscr{D}_1) = (1 - z_1)m + z_1 \frac{\sum x_{1t}}{n_1}, \tag{2.1}$$

with a mixing coefficient, called the *credibility factor*:

$$z_1 = \frac{n_1}{n_1 + (e/d)}, \tag{2.2}$$

and three required marginal (prior knowledge) moments:

$$m = \mathscr{E}\mathscr{E}\{\tilde{x}_{1t} \mid \tilde{\theta}_1\}; \qquad e = \mathscr{E}\mathscr{V}\{\tilde{x}_{1t} \mid \tilde{\theta}_1\}; \qquad d = \mathscr{V}\mathscr{E}\{\tilde{x}_{1t} \mid \tilde{\theta}_1\}. \tag{2.3}$$

The *credibility forecast*, $f_1(\mathscr{D}_1)$, is also a robust predictor for the true $\mathscr{E}\{\tilde{w}_1 \mid \mathscr{D}_1\}$ for *any* model or prior density, in the sense that it is also the *best linear least-squares approximation* (*Bühlmann*, 1967).

In many applications, there may be additional data, say $\{x_{it}\}$ $(i = 2, 3, \ldots, r)$ $(t = 1, 2, \ldots, n_i)$, available from $(r-1)$ "similar" *related risks*, characterized by *different* risk parameters, $(\tilde{\theta}_i)$, but with the *same* model density, $p(x \mid \theta)$. For example, in insurance we may have a portfolio of risks, presumed similar, *a priori*, as determined by some risk classification scheme. But, if the $(\tilde{\theta}_i)$ were i.i.d., then the *collateral data* would have no predictive value!

A convenient hypothesis that keeps the risks similar, yet introduces dependency in a natural way, is to assume that $\boldsymbol{\theta} = [\theta_1, \theta_2, \ldots, \theta_r]^T$ is composed of *exchangeable random variables*, with a joint prior density:

$$p(\theta_1, \theta_2, \ldots, \theta_r) = \int \prod_{i=1}^{r} p_0(\theta_i \mid \phi) p(\phi)\, d\phi, \tag{2.4}$$

where the common *conditional density*, $p_0(\cdot|\cdot)$, assures that the marginal densities, $p(\theta_i)$, have identical forms. The interpretation of this assumption is that the individual risk parameters now depend upon one or more unknown *portfolio hyperparameters*, ϕ. Of course, we must be willing to opine a *hyperprior density*, $p(\phi)$, over the set of all possible portfolios. In the insurance setting, ϕ represents simply the variation in similar portfolios between different insurance companies. The resulting three-level structure is called a *hierarchical model.*

Our goal is still to predict a future value of risk # 1, but now using the *total cohort data*, $\mathscr{D} = \{x_{it}\}$ $(i = 1, 2, \ldots, r)$ $(t = 1, 2, \ldots, n_i)$. *Jewell* (1975 b) finds the credibility approximation for the hierarchical predictive mean as the combination of two credibility-like forecasts:

$$\begin{aligned} \mathscr{E}\{\tilde{w}_1 \mid \mathscr{D}\} \approx f_1(\mathscr{D}) &= (1 - z_1) f_0(\mathscr{D}) + z_1 y_1 ; \\ f_0(\mathscr{D}) &= (1 - z_0) m + z_0 y_0 . \end{aligned} \tag{2.5}$$

The total cohort data is summarized in $r+1$ *linear sufficient statistics*:

$$y_i = \frac{\sum x_{it}}{n_i}\,; \qquad y_0 = \frac{\sum z_i y_i}{\sum z_j}\,; \tag{2.6}$$

and $r+1$ credibility factors for each risk and for the whole portfolio:

$$z_i = \frac{n_i}{n_i + (f/g)}\,; \qquad z_0 = \frac{\sum z_i}{\sum z_j + (g/h)}\,. \tag{2.7}$$

We will see that: $f_0(\mathcal{D})$ is an approximation to the predictive mean at the portfolio level, $\mathcal{E}\{\mathcal{E}\{\tilde{x}_{it} \mid \tilde{\phi}\} \mid \mathcal{D}\}$; m is the overall prior portfolio mean; and f, g, and h are new central moments averaged over all possible portfolios.

In 1986, the authors investigated the use of credibility to find the *second* moments of $(\tilde{w}_1 \mid \mathcal{D})$ of the hierarchical model, using an idea due to *Jewell/Schnieper* (1985) (for short, *J&S*); there, a *three*-dimensional model was used to find *simultaneous* linear approximations to $\mathcal{E}\{\tilde{x}_{n+1} \mid \mathcal{D}\}$, $\mathcal{E}\{\tilde{x}_{n+1}^2 \mid \mathcal{D}\}$, and $\mathcal{E}\{\tilde{x}_{n+1}\tilde{x}_{n+2} \mid \mathcal{D}\}$ for arbitrary model and prior densities, using the three corresponding sample statistics. The methodology to carry out this program was soon clear, but implementation immediately ran into complicated and messy problems of definition, notation, and computational efficiency. It has taken the intervening years and several articles (*Bühlmann/Jewell*, 1987; *Jewell*, 1987, 1988, 1989) to properly develop and test this approach.

For simplicity, we assume equal data lengths, $n_i = n$, for each risk i, and begin with a brief summary of credibility theory in several dimensions.

3 Summary of Multi-Dimensional Credibility Theory

Suppose we have one sample, $\mathcal{D} = \{\boldsymbol{y}\}$, of a random vector, $\tilde{\boldsymbol{y}}$, with which to predict another random vector, $\tilde{\boldsymbol{w}}$, through a linear form:

$$\mathcal{E}\{\tilde{\boldsymbol{w}} \mid \mathcal{D}\} \approx \boldsymbol{f}(\mathcal{D}) = \boldsymbol{a} + \boldsymbol{Z}\boldsymbol{y}\,. \tag{3.1}$$

adjusted to give the best least-squares fit. It can be shown (see *Jewell*, 1989) that the matrix $\boldsymbol{Z}$ is given by the *normal system* of equations:

$$\boldsymbol{Z} \cdot \boldsymbol{C} = \boldsymbol{R}\,; \qquad \boldsymbol{C} = \mathcal{C}\{\tilde{\boldsymbol{y}}; \tilde{\boldsymbol{y}}\}\,; \qquad \boldsymbol{R} = \mathcal{C}\{\tilde{\boldsymbol{w}}; \tilde{\boldsymbol{y}}\}\,. \tag{3.2}$$

$\boldsymbol{a}$ is then adjusted to make the forecast unbiased, $\boldsymbol{a} = \mathscr{E}\{\tilde{\boldsymbol{w}}\} - \boldsymbol{Z} \cdot \mathscr{E}\{\tilde{\boldsymbol{y}}\}$.[1]
In credibility prediction, $\tilde{\boldsymbol{w}}$ and $\tilde{\boldsymbol{y}}$ are usually of the same dimensionality and depend upon one or more unknown risk parameters $\tilde{\psi}$ so that:

$$\mathscr{E}\{\tilde{\boldsymbol{w}} \mid \psi\} = \mathscr{E}\{\tilde{\boldsymbol{y}} \mid \psi\} = \boldsymbol{m}(\psi)\,; \qquad \mathscr{C}\{\tilde{\boldsymbol{w}}; \tilde{\boldsymbol{y}} \mid \psi\} = \boldsymbol{0}\,. \tag{3.3}$$

If this occurs, we say $\tilde{\boldsymbol{w}}$ and $\tilde{\boldsymbol{y}}$ are *mean-exchangeable random variables.*
Defining the marginal moments:

$$\begin{aligned} \boldsymbol{m} &= \mathscr{E}\{\boldsymbol{m}(\tilde{\psi})\}\,; & \boldsymbol{D} &= \mathscr{C}\{\boldsymbol{m}(\tilde{\psi})\,;\, \boldsymbol{m}(\tilde{\psi})\}\,; \\ \boldsymbol{C}(\psi) &= \mathscr{C}\{\tilde{\boldsymbol{y}}; \tilde{\boldsymbol{y}} \mid \psi\}\,; & \boldsymbol{B} &= \mathscr{E}\{\boldsymbol{C}(\tilde{\psi})\}\,, \end{aligned} \tag{3.4}$$

mean-exchangeability then implies:

$$\mathscr{E}\{\tilde{\boldsymbol{w}}\} = \mathscr{E}\{\tilde{\boldsymbol{y}}\} = \boldsymbol{m}\,; \quad \mathscr{C}\{\tilde{\boldsymbol{w}}; \tilde{\boldsymbol{y}}\} = \boldsymbol{D}\,; \quad \mathscr{C}\{\tilde{\boldsymbol{y}}; \tilde{\boldsymbol{y}}\} = \boldsymbol{B} + \boldsymbol{D}\,. \tag{3.5}$$

The linear predictor (3.1) becomes the *credibility forecast*:

$$\boldsymbol{f}(\mathscr{D}) = (\boldsymbol{I} - \boldsymbol{Z})\boldsymbol{m} + \boldsymbol{Z}\boldsymbol{y}\,, \tag{3.6}$$

with a square *credibility matrix* $\boldsymbol{Z}$ given by:

$$\boldsymbol{Z}(\boldsymbol{B} + \boldsymbol{D}) = \boldsymbol{D} \qquad \text{or} \qquad \boldsymbol{Z} = \boldsymbol{D}(\boldsymbol{B} + \boldsymbol{D})^{-1}\,. \tag{3.7}$$

For the hierarchical model, $\psi = (\theta; \phi)$, with model and prior densities having the special form (2.4). We take $\tilde{\boldsymbol{w}}$ to be a vector of first *and* second powers of future observations, and $\tilde{\boldsymbol{y}}$ as the corresponding first- *and* second-order statistics; this will retain mean-exchangeability. After defining $\tilde{\boldsymbol{w}}$ and $\tilde{\boldsymbol{y}}$ further, we will find $\boldsymbol{m}$, $\boldsymbol{B}$, and $\boldsymbol{D}$ in terms of underlying moments of $\tilde{\boldsymbol{x}}$, and then consider the problem of inverting (3.7).

4 Predictands and Predictors for the Basic Model

As in *J&S*, the use of moments about the origin provides the easiest way to develop the parameters (3.5). However, we will need to generalize the notation somewhat for the hierarchical case.

[1] We are using the vector and matrix expectation operator, $\mathscr{E}$, and the vector covariance-matrix operator, $\mathscr{C}\{\tilde{\boldsymbol{u}}; \tilde{\boldsymbol{v}}\} = \mathscr{E}\{\tilde{\boldsymbol{u}}\tilde{\boldsymbol{v}}^T\} - \mathscr{E}\{\tilde{\boldsymbol{u}}\} \cdot \mathscr{E}\{\tilde{\boldsymbol{v}}^T\}$, where $\tilde{\boldsymbol{u}}$ and $\tilde{\boldsymbol{v}}$ need not have the same dimensions.

The first- and second-order predictands of interest are:

$$\tilde{w}_i = \tilde{x}_{i,t} \; ; \quad \tilde{w}_{ii} = \tilde{x}_{i,t}^2 \; ; \quad \tilde{w}_{i\times i} = \tilde{x}_{i,t}\tilde{x}_{i,u} \; ; \quad \tilde{w}_{h*i} = \tilde{x}_{h,s}\tilde{x}_{i,t} \, , \tag{4.1}$$

where $h \neq i$, and s, t, and u are arbitrary time indices $\geq n+1$, except that $u \neq t$! This is because the special structure of the (stationary) hierarchical model gives different conditional second-order moments for the *same* risk when considering squares of observations (same time period) than when considering cross-products (different time periods); however, when considering *different* risks, the expected cross-product of observations is independent of the particular epochs chosen. This is reflected in the subscript notation, with double subscripts (ii) used for the square of observations from *any* time period, $(i \times i)$ denoting the product of observations from a *single* risk from *different* time periods, and $(h * i)$ denoting the product from two *different* risks at *any* future time period(s). From symmetry, only the $s = \frac{1}{2}r(r-1)$ cases with $h < i$ need to be considered. Forming the predictands into a vector:

$$\tilde{\boldsymbol{w}} = \left[\tilde{w}_i \mid \tilde{w}_{ii} \mid \tilde{w}_{i\times i} \mid \tilde{w}_{h*i}\right]^T , \qquad (h < i) \tag{4.2}$$

we see that there are $R = 3r + s = \frac{1}{2}r(r+5)$ total distinct predictands, with the special indexing scheme:

$$\mathcal{I} = \{(i) \mid (ii) \mid (i \times i) \mid (h * i)\} \, . \qquad (h < i)$$

We shall refer to the four different index types as groups a, b, c, and d.
For predictors in $\tilde{\boldsymbol{y}}$, we assume that $n \geq 2$ and use the corresponding "natural" sample moments:

$$y_i = \frac{1}{n}\sum_{t=1}^{n} x_{it} \; ; \qquad y_{ii} = \frac{1}{n}\sum_{t=1}^{n} x_{it}^2 \; ; \tag{4.3 A\&B}$$

$$y_{i\times i} = \frac{1}{n(n-1)} \underbrace{\sum_{t=1}^{n}\sum_{u=1}^{n} x_{it}x_{iu}}_{t\neq u} = \frac{2}{n(n-1)} \underbrace{\sum_{t=1}^{n}\sum_{u=1}^{n} x_{it}x_{iu}}_{t<u} \; ; \tag{4.3 C}$$

$$y_{h*i} = y_{i*h} = \frac{1}{n^2}\sum_{t=1}^{n}\sum_{u=1}^{n} x_{ht}x_{iu} = y_h y_i \, . \qquad (h \neq i) \, . \tag{4.3 D}$$

(One could also use $y_{i*i} = y_i^2$ in place of either y_{ii} or $y_{i\times i}$).

Assume $r \geq 2$. The number of predictors is also R, so that the same indexing scheme $\mathcal{J}$ can be used to define the vector $\tilde{\boldsymbol{y}}$. Note that, since $\tilde{\boldsymbol{w}}$ and $\tilde{\boldsymbol{y}}$ have the same conditional mean vector, $\boldsymbol{m}(\tilde{\boldsymbol{\theta}})$, and involve $\tilde{\boldsymbol{x}}_t$ from different time intervals, they are mean-exchangeable random variables. Finally, since the RHS of (3.7) is $\boldsymbol{D}$, we could define $\tilde{\boldsymbol{w}} = \boldsymbol{m}(\tilde{\boldsymbol{\theta}})$ and get the same result. Thus, a credibility formula can be thought of either as predicting a future value of an observable, or as providing a linear fit to its unknown mean value.

5 Conditional and Unconditional Moments

The next step is to define the various conditional moments that will appear in $\boldsymbol{m}(\boldsymbol{\theta})$ and $\boldsymbol{C}(\boldsymbol{\theta})$; as in *J&S*, we will need to find conditional moments up to order *four* from $p(\boldsymbol{x} \mid \boldsymbol{\theta})$, which we define as

$$m_k(\theta_i) = \mathscr{E}\{\tilde{x}_{it}^k \mid \theta_i\}\,; \qquad (k = 1, \ldots, 4) \quad (i = 1, \ldots, r) \tag{5.1}$$

any cross-risk or cross-temporal conditional moments factor into products of these functions (subscripts on moments refer to powers, not indices). In this notation, the conditional mean vector of both $\tilde{\boldsymbol{w}}$ and $\tilde{\boldsymbol{y}}$ is:

$$\boldsymbol{m}(\boldsymbol{\theta}) = \left[m_1(\theta_i) \mid m_2(\theta_i) \mid m_1^2(\theta_i) \mid m_1(\theta_h)m_1(\theta_i)\right]^T . \tag{5.2}$$

From the assumptions, the $\{m_k(\tilde{\theta}_i)\}$ are exchangeable r.v.s over the risk indices for each value of k. Given ϕ, we form *eleven* new conditional moments:

$$\begin{aligned}
&M_1(\phi) = \mathscr{E}\{m_1(\tilde{\theta}_i) \mid \phi\}\,; && && && (5.3)\\
&M_2(\phi) = \mathscr{E}\{m_2(\tilde{\theta}_i) \mid \phi\}\,; && M_{11}(\phi) = \mathscr{E}\{m_1^2(\tilde{\theta}_i) \mid \phi\}\,; && \\
&M_3(\phi) = \mathscr{E}\{m_3(\tilde{\theta}_i) \mid \phi\}\,; && M_{21}(\phi) = \mathscr{E}\{m_2(\tilde{\theta}_i)m_1(\tilde{\theta}_i) \mid \phi\}\,; && M_{111}(\phi) = \mathscr{E}\{m_1^3(\tilde{\theta}_i) \mid \phi\}\,;\\
&M_4(\phi) = \mathscr{E}\{m_4(\tilde{\theta}_i) \mid \phi\}\,; && M_{31}(\phi) = \mathscr{E}\{m_3(\tilde{\theta}_i)m_1(\tilde{\theta}_i) \mid \phi\}\,; && \\
&M_{22}(\phi) = \mathscr{E}\{m_2^2(\tilde{\theta}_i) \mid \phi\}\,; && M_{211}(\phi) = \mathscr{E}\{m_2(\bar{\theta}_i)m_1^2(\tilde{\theta}_i) \mid \phi\}\,; && M_{1111}(\phi) = \mathscr{E}\{m_1^4(\tilde{\theta}_i) \mid \phi\}\,,
\end{aligned}$$

where i can be *any* index. Cross-risk moments conditional on ϕ factor, i.e.,

$$\mathscr{E}\{m_1(\tilde{\theta}_i)m_1(\tilde{\theta}_j) \mid \phi\} = M_1^2(\phi) \quad \text{for any} \quad i \neq j\,.$$

In this new notation, the classical credibility model can be considered as one in which ϕ is fixed, with parameters (2.3) that are:

$$\begin{aligned} m &= M_1(\phi)\,; \\ e &= M_2(\phi) - M_{11}(\phi) \geq 0\,; \\ d &= M_{11}(\phi) - M_1^2(\phi) \geq 0\,. \end{aligned} \tag{5.4}$$

In the second-moment *J&S* extension, all 11 moments (5.3) are used. However, in the hierarchical model, the portfolio hyperparameter is an unknown quantity. So, by taking averages over $\tilde{\phi}$ of all possible combinations of the conditional moments, we finally obtain the 24 *unconditional moments*:

$$\begin{aligned} &\text{(a)}\quad \text{First order:} && M(1); \\ &\text{(b)}\quad \text{Second order:} && M(2),\ M(11),\ M(1;1); \\ &\text{(c)}\quad \text{Third order:} && M(3),\ M(21),\ M(111),\ M(2;1),\ M(11;1),\ M(1;1;1); \\ &\text{(d)}\quad \text{Fourth order:} && M(4),\ M(31),\ M(22),\ M(211),\ M(1111),\ M(3;1), \\ & && M(21;1),\ M(111;1),\ M(2;2),\ M(2;11),\ M(11;11), \\ & && M(2;1;1),\ M(11;1;1),\ M(1;1;1;1). \end{aligned} \tag{5.5}$$

Here the semicolons in the arguments separate products and powers of the eleven conditional moments, that is, terms arising from different risks. This makes the notation easy to read in terms of the basic r.v.s., (5.1), or (5.3); for example, for any three *distinct* risk indices h, i, and j:

$$\begin{aligned} &\mathscr{E}\{\tilde{x}_{ht}\tilde{x}_{hu}\tilde{x}_{iv}\tilde{x}_{jw}\} \\ &= \left\{ \begin{array}{ll} \mathscr{E}\{m_2(\tilde{\theta}_h)m_1(\tilde{\theta}_i)m_1(\tilde{\theta}_j)\} = \mathscr{E}\{M_2(\tilde{\phi})M_1^2(\tilde{\phi})\} = M(2;1;1) & (t=u) \\ \mathscr{E}\{m_1^2(\tilde{\theta}_h)m_1(\tilde{\theta}_i)m_1(\tilde{\theta}_j)\} = \mathscr{E}\{M_{11}(\tilde{\phi})M_1^2(\tilde{\phi})\} = M(11;1;1) & (t\neq u) \end{array} \right\}, \end{aligned}$$

for *all* values of the time indices v and w. So the general hierarchical portfolio model requires only *24 hyper-hyperparameters*, for any $r \geq 2$!
For the predictive means (2.5), we need only four of these moments:

$$\begin{aligned} m &= M(1)\,; & f &= M(2) - M(11) \geq 0\,; \\ g &= M(11) - M(1;1) \geq 0\,; & h &= M(1;1) - M(1)^2 \geq 0\,. \end{aligned} \tag{5.6}$$

6 Calculation of Model Hyperparameters

We now calculate $\boldsymbol{m}$, $\boldsymbol{D}$, and $\boldsymbol{B}$ in terms of the moments (5.5).

6.1 *Mean Vector*

From (5.2), the R-dimensional mean vector is:

$$\boldsymbol{m} = [M(1) \mid M(2) \mid M(11) \mid M(1;1)]^T , \tag{6.1}$$

where the values are extended as appropriate within each block.

6.2 *Predictand–Predictor Covariance Matrix*

Since both $\tilde{\boldsymbol{w}}$ and $\tilde{\boldsymbol{y}}$ are partitioned by $\mathcal{J}$, we subdivide the $R \times R$ covariance matrices into 16 submatrices, as shown in Figure 1, using subscripts $aa, ab, \ldots$, etc., to refer to the different blocks.
Because every covariance matrix is symmetric, the diagonal blocks are symmetric, and there are obvious transposes between the rectangular blocks. For the hierarchical model, it turns out that all square submatrices are symmetrical; in fact there are many structural simplifications, which we now describe. Rather than give specific formulae for each block, we shall express our results in terms of the patterns shown in Figure 2.
For the matrix $\boldsymbol{D} = \mathcal{E}\{\boldsymbol{m}(\tilde{\boldsymbol{\theta}})\boldsymbol{m}(\tilde{\boldsymbol{\theta}})^T\} - \boldsymbol{m}\boldsymbol{m}^T$, we use (5.2) to find that the values for each of the nine $r \times r$ square submatrices are constant over the pattern shown in Figure 2 (a), with one value, call it h, in the unshaded cells, and a larger value, call it $g + h$, on the diagonal. In terms of the marginal moments, we find for the different blocks:

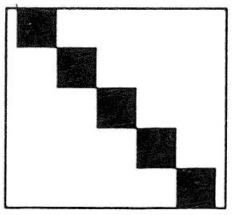
Block	g	h
aa	$M(11) - M(1;1)$	$M(1;1) - M^2(1)$
ab, ba	$M(21) - M(2;1)$	$M(2;1) - M(2)M(1)$
ac, ca	$M(111) - M(11;1)$	$M(11;1) - M(11)M(1)$
bb	$M(22) - M(2;2)$	$M(2;2) - M^2(2)$
bc, cb	$M(211) - M(2;11)$	$M(2;11) - M(2)M(11)$
cc	$M(1111) - M(11;11)$	$M(11;11) - M^2(11)$

(6.2)

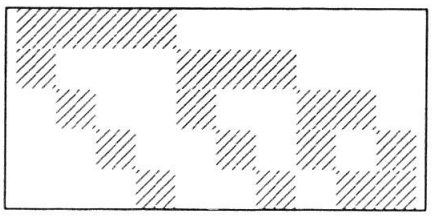
Figure 1.
Partitioning and indexing of covariance matrices.

	(j)	(jj)	(j×j)	(j*k)
(i)	*aa*	*ab*	*ac*	*ad*
(ii)	*ba*	*bb*	*bc*	*bd*
(i×i)	*ca*	*cb*	*cc*	*cd*
(h*i)	*da*	*db*	*dc*	*dd*

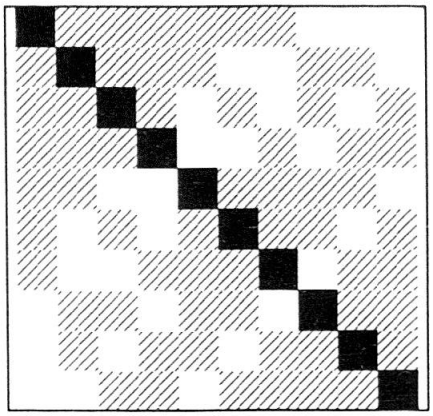
Figure 2.
Special structures of covariance submatrices ($r = 5$, $s = 10$).
(a) $r \times r$ submatrices *aa*, *ab*, *ac*, *ba*, *bb*, *bc*, *ca*, *cb*, and *cc*.



(b) $r \times s$ submatrices *ad*, *bd*, and *cd* (and transposes *da*, *db*, and *dc*).



(c) $s \times s$ submatrix dd.



For the six $r \times s$ rectangular submatrices in $\boldsymbol{D}$ there are no diagonals, but now a pattern of constant values like that shown in Figure 2 (b) emerges where a row index coincides with one of the column indices. Let us call the common value in each unshaded cell h, and the common value in each shaded cell by $\gamma + h$. The new central moments are:

Block	γ	h
ad, da	$M(11;1) - M(1;1;1)$	$M(1;1;1) - M(1;1)M(1)$
bd, db	$M(21;1) - M(2;1;1)$	$M(2;1;1) - M(2)M(1;1)$
cd, dc	$M(111;1) - M(11;1;1)$	$M(11;1;1) - M(11)M(1;1)$

(6.3)

The pattern of common values for the $s \times s$ block dd is shown in Figure 2 (c). As before, we call the unshaded cell values h and the diagonal values $g + h$. However, we now have shaded cells where one row index coincides with one of the column indices, with a new common value we will call $\gamma + h$. We find:

$$\begin{aligned} g_{dd} &= M(11;11) - M(1;1;1;1)\,; \\ \gamma_{dd} &= M(11;1;1) - M(1;1;1;1)\,; \\ h_{dd} &= M(1;1;1;1) - M^2(1;1)\,. \end{aligned} \tag{6.4}$$

In summary, there are 21 distinct central moments, g, h, and γ, needed to compute $\boldsymbol{D}$, independent of either r or n!

Notice that $\boldsymbol{D}$ appears twice in (3.7): once as part of the total covariance matrix, and once as the RHS. If a different predictand $\tilde{\boldsymbol{w}}$ than (4.1) is sought, then a different RHS must be used in:

$$\boldsymbol{Z}(\boldsymbol{B} + \boldsymbol{D}) = \mathscr{C}\{\tilde{\boldsymbol{w}}; \tilde{\boldsymbol{y}}\} = \boldsymbol{R}\,, \tag{6.5}$$

say, where a non-square $\boldsymbol{R}$ would induce a similarly-shaped $\boldsymbol{Z}$. For instance, suppose we wish to predict first and second moments at the *portfolio* level. Then:

$$\boldsymbol{w} = \boldsymbol{M}(\boldsymbol{\phi}) = [M_1(\phi) \mid M_2(\phi) \mid M_{11}(\phi) \mid M_1^2(\phi)]^T \tag{6.6}$$

and we find that the $4 \times s$ matrix $\boldsymbol{R}$ consists *only* of the 10 central moments h, arranged in a constant-by-blocks pattern. We will later interpret this result, and show how it fits into the general solution to our model.

6.3 *Predictor Covariance Matrix*

Calculation of $\boldsymbol{B}$ is simplified because $\boldsymbol{C}(\boldsymbol{\theta}) = \mathscr{E}\{\tilde{\boldsymbol{y}}\tilde{\boldsymbol{y}}^T \mid \boldsymbol{\theta}\} - \boldsymbol{m}(\boldsymbol{\theta})\boldsymbol{m}(\boldsymbol{\theta})^T$ depends directly upon the individual risk parameters, which are independent, given ϕ. Thus $\boldsymbol{B} = \mathscr{E}\{\mathscr{E}\{\boldsymbol{C}(\tilde{\boldsymbol{\theta}}) \mid \tilde{\phi}\}\}$ is *zero* whenever the row indices are distinct from the column indices, that is, in the unshaded cells of Figure 2. As $\boldsymbol{B}$ depends upon n, we set $\boldsymbol{B} = \frac{1}{n}\boldsymbol{E}(n)$, where $\boldsymbol{E}(n)$ is a weak function of n, independent of it except in blocks cc and dd.

For the eight $r \times r$ square submatrices, excluding cc, only the diagonal cells have a non-zero common value, call it f. Then:

Block	f
aa	$M(2) - M(11)$
ab, ba	$M(3) - M(21)$
ac, ca	$2[M(21) - M(111)]$
bb	$M(4) - M(22)$
bc, cb	$2[M(31) - M(211)]$

(6.7)

In block cc, the common value on the diagonal varies with n:

$$\begin{aligned} f_{cc}(n) &= \left[f_{cc} + \frac{1}{n-1}\tau_{cc}\right]; \\ f_{cc} &= 4[M(211) - M(1111)]; \\ \tau_{cc} &= 2[M(22) - 2M(211) + M(1111)]. \end{aligned} \tag{6.8}$$

A similar result was found previously in *J&S*.

For the six $r \times s$ rectangular blocks, non-zero values are obtained only in the non-shaded cells of Figure 2 (b), with a common value we shall call φ (not to be confused with the portfolio parameter ϕ). The values by block are:

Block	φ
ad, da	$M(2;1) - M(11;1)$
bd, db	$M(3;1) - M(21;1)$
cd, dc	$2[M(21;1) - M(111;1)]$

(6.9)

Finally, as might be expected, the common diagonal value in $\boldsymbol{E}^{dd}(n)$ is also a weak function of n:

$$\begin{aligned} f_{dd}(n) &= \left[f_{dd} + \frac{1}{n} \tau_{dd} \right] ; \\ f_{dd} &= 2[M(2;11) - M(11;11)] ; \\ \tau_{dd} &= M(2;2) - 2M(2;11) + M(11;11) . \end{aligned} \tag{6.10}$$

and the shaded cells in Figure 2 (c) have common value:

$$\varphi_{dd} = M(2;1;1) - M(11;1;1) . \tag{6.11}$$

The variation with n of the transient term in block dd is somewhat different than that in block cc because of the different number of samples involved.
In summary, the calculation of $\boldsymbol{E}(n)$ requires 13 central moments, which, when combined with those in $\boldsymbol{D}$ and $\boldsymbol{m}$, means that 28 values need to be computed from the 24 moments in (5.5); most of them are, in fact, non-negative. Rather than finding $\boldsymbol{Z}$ through the inversion of the $R \times R$ matrix $\boldsymbol{C}$, we now develop methods that will exploit the special structure just found.

7 Diagonal Block Solutions

In our original paper, we showed how to use the special structure in Figure 2 to rederive the basic hierarchical formula (2.5), by predicting $\tilde{\boldsymbol{w}}^a$ from just $\tilde{\boldsymbol{y}}^a$, using only the diagonal square submatrices $\boldsymbol{B}^{aa}$ and $\boldsymbol{D}^{aa}$. In the interest of brevity, we omit discussion of this and similar independent block solutions. Full details are available from the author.

8 The Reduced Prediction Problem

8.1 *Some Matrix Results*

We now gloss over several years labor to simplify solutions and take an intuitive leap that will greatly simplify our problem. Consider the identical and complete patterns of constant values in the $R \times R$ matrices $\boldsymbol{B}$ and $\boldsymbol{D}$, as formed from Figure 2. Surprisingly, this overall pattern is *invariant under multiplication and inversion*! Further details can be found in *Jewell* (1988). For our problem, this means that:

(a) $(\boldsymbol{B}+\boldsymbol{D})^{-1}$ will have the same pattern as $\boldsymbol{B}$ and $\boldsymbol{D}$; and

(b) $\boldsymbol{Z}=\boldsymbol{R}(\boldsymbol{B}+\boldsymbol{D})^{-1}$ will also have the same pattern if $\boldsymbol{R}$ does!

Thus, our computational problem reduces to finding the constant values associated with each pattern, that is, to solving a small linear system whose dimension does not vary with the number of risks, r!

8.2 *Sufficient Statistics and Prototype Formulae*

Assume for the moment that we are not interested in cross-moment predictions of $\tilde{x}_{h,n+1}\tilde{x}_{i,n+1}$ or $m_1(\tilde{\theta}_h)m_1(\tilde{\theta}_i)$ $(h \neq i)$. Considering the patterns in $\boldsymbol{Z}$ and in the product $\boldsymbol{Zy}$, one can show that only *eight* summary predictands are needed! For each risk i, we can use the previously defined statistics y_i, y_{ii}, and y_{i*i}, and, for the portfolio, we obtain y_0, y_{00}, and $y_{0\times 0}$. Two new statistics are needed:

$$\begin{aligned} y_{i*0} &= \frac{1}{r-1}\sum_{h\neq i} y_{h*i}\,; \\ y_{0*0} &= \frac{1}{s}\sum_{h<i}\sum y_{h*i} = \frac{1}{r(r-1)}\sum_{h\neq i}\sum y_{h*i}\,, \end{aligned} \tag{8.1}$$

which, to be distinct, now require $r \geq 3$. (The asterisks in y_{i*0} and y_{0*0} do not have the same meaning as in y_{h*i}). It follows that *any* f_i or f_{ii} or $f_{i\times i}$ will be a linear combination of only *eight* statistics:

$$\tilde{\boldsymbol{y}} = [\tilde{y}_i\,, \tilde{y}_{ii}\,, \tilde{y}_{i\times i}\,, \tilde{y}_{i*0} \mid \tilde{y}_0\,, \tilde{y}_{00}\,, \tilde{y}_{0\times 0}, \tilde{y}_{0*0}]^T\,. \tag{8.2}$$

For symmetry, we select the eight-vector of predictands:

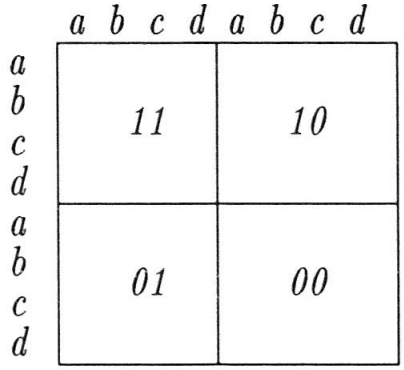
$$\tilde{w} = \Big[m_1(\tilde{\theta}_i), m_2(\tilde{\theta}_i), m_1^2(\tilde{\theta}_i), m_1(\tilde{\theta}_i)M_1(\tilde{\phi}) \mid M_1(\tilde{\phi}), M_2(\tilde{\phi}), M_{11}(\tilde{\phi}), M_1^2(\tilde{\phi})\Big]^T . \tag{8.3}$$

Although this new $\tilde{w}$ has the same prior mean as $\tilde{y}$, they are not mean-exchangeable r.v.s., so that the general form (6.8), $\boldsymbol{Z} = \boldsymbol{R}(\boldsymbol{B} + \boldsymbol{D})^{-1}$, must be used in the reduced problem after finding the appropriate 8×8 RHS matrix, $\boldsymbol{R}$.

So the original $R \times R$ problem now reduces to one in which: $\boldsymbol{B}$, $\boldsymbol{D}$, the new $\boldsymbol{R}$, and $\boldsymbol{Z}$ will all be 8×8 matrices, whatever the value of r; the new $\boldsymbol{m}$ will be an 8-vector; and the values of $\boldsymbol{Z}$ and $\boldsymbol{m}$ will not depend in any way upon the index i! In other words, the 8×8 version of (3.6) for the reduced formulation will be a *prototype prediction formula* suitable for *any* individual risk i!! This being the case, let us henceforth, for simplicity, set $i = 1$.

Our task is now to express the new moments, $\boldsymbol{m}$, $\boldsymbol{B}$, $\boldsymbol{D}$, and $\boldsymbol{R}$, in terms of the central moments found in Section 6. In place of the partitioning of Figure 1, the two groups of four variables in (8.2) (8.3) will be labelled 1 and 0, respectively, and, within each group, the variables will be labelled by a, b, c, and d, corresponding to the original block labels. This leads to the partitioning of the reduced matrices, $\boldsymbol{B}$, $\boldsymbol{D}$, $\boldsymbol{R}$, and $\boldsymbol{Z}$, shown in Figure 3.

Figure 3.
Block structure for reduced 8×8 matrices.



8.3 *Reduced Model Hyperparameters*

The eight-vector of prior means is similar to (6.1):

$$\boldsymbol{m} = \left[M(1)\,M(2)\,M(11)\,M(1;1) \mid M(1)\,M(2)\,M(11)\,M(1;1)\right]^T. \tag{8.4}$$

The 8×8 total covariance matrix, $\boldsymbol{C} = \boldsymbol{B} + \boldsymbol{D}$, has only two different forms in the four blocks of Figure 3. In block *11*, position *aa*:

$$C_{aa}^{11} = \frac{1}{n} f_{aa} + g_{aa} + h_{aa}, \tag{8.5}$$

and similarly for subscripts *ab*, *ac*, *ba*, *bb*, *bc*, *ca*, *cb*, and *cc*; in position *cc*, $f_{cc}(n)$ must be used. In the six border locations corresponding to the rectangular blocks in Figure 1:

$$C_{ad}^{11} = \frac{1}{n} \varphi_{ad} + \gamma_{ad} + h_{ad}, \tag{8.6}$$

and similarly for *bd*, *cd*, *da*, *db*, and *dc*. (Note that all 15 of these values are invariant with *r*). The 16th position, corresponding to the old $s \times s$ block, is the most complicated, with:

$$C_{dd}^{11} = \left[\frac{1}{r-1}\right]\left[\frac{1}{n} f_{dd}(n) + g_{dd}\right] + \left[\frac{r-2}{r-1}\right]\left[\frac{1}{n}\varphi_{dd} + \gamma_{dd}\right] + h_{dd} \tag{8.7}$$

The remaining three blocks are *identical*, $\boldsymbol{C}^{10} = \boldsymbol{C}^{01} = \boldsymbol{C}^{00}$, with:

$$C_{aa}^{10} = C_{aa}^{01} = C_{aa}^{00} = \frac{1}{r}\left[\frac{1}{n} f_{aa} + g_{aa}\right] + h_{aa}, \tag{8.8}$$

and similarly for the eight other NW-corner locations in each block, as in (8.5) above (remember to use $f_{cc}(n)$ in cell *cc*). For the six border locations corresponding to (8.6), we have the form:

$$C_{ad}^{00} = \frac{2}{r}\left[\frac{1}{n}\varphi_{ad} + \gamma_{ad}\right] + h_{ad}, \tag{8.9}$$

noting especially the factor of 2. Finally, for the SE-corner location *dd*, we have the complex form:

$$C_{dd}^{00} = \frac{1}{s}\left[\frac{1}{n} f_{dd}(n) + g_{dd}\right] + 2\left[\frac{r-2}{s}\right]\left[\frac{1}{n}\varphi_{dd} + \gamma_{dd}\right] + h_{dd}. \tag{8.10}$$

Turning to the RHS matrix, $\boldsymbol{R}^{11}$ has the simple form:

$$R_{aa}^{11} = g_{aa} + h_{aa}, \tag{8.11}$$

and similarly in the eight other NW-corner locations. Then:

$$R_{ad}^{11} = \gamma_{ad} + h_{ad}, \tag{8.12}$$

and likewise in the five other border cells, *and* also for R_{dd}^{11}!
The submatrix $\boldsymbol{R}^{10}$ is the only asymmetric 4×4 block, with:

$$R_{aa}^{10} = \frac{1}{r} g_{aa} + h_{aa}, \tag{8.13}$$

and similarly in the eight other NW-corner cells. Then:

$$R_{ad}^{10} = \frac{2}{r} \gamma_{ad} + h_{ad}, \tag{8.14}$$

and likewise in *bd*, *cd*, and the SE corner location *dd*. However:

$$R_{da}^{10} = \frac{1}{r} \gamma_{ad} + h_{ad}, \tag{8.15}$$

and similarly in the other lower border locations *db* and *dc*. Note the different coefficients 2 and 1 in (8.14) and (8.15)! These are *not* misprints, but reflect an underlying asymmetry in the reduced predictors and predictands. The unchanged subscripts on the central moments γ and h remind us that they are still symmetric w.r.t. their subscripts.
The remaining two blocks, $\boldsymbol{R}^{01}$ and $\boldsymbol{R}^{00}$, are easy to describe:

$$R_{aa}^{01} = R_{aa}^{00} = h_{aa}, \tag{8.16}$$

and similarly for *all 15* remaining cells in both blocks!
Figure 4 summarizes the final block structure of $\boldsymbol{C}$ and $\boldsymbol{R}$. Only $\boldsymbol{R}^{10}$ is asymmetric.

Figure 4.
Final block structure in reduced matrices.

$$\mathbf{C} = \begin{array}{|c|c|}\hline \mathbf{C}^{11} & \mathbf{C}^{00} \\ \hline \mathbf{C}^{00} & \mathbf{C}^{00} \\ \hline \end{array} \quad ; \quad \mathbf{R} = \begin{array}{|c|c|}\hline \mathbf{R}^{11} & \mathbf{R}^{10} \\ \hline \mathbf{R}^{00} & \mathbf{R}^{00} \\ \hline \end{array}$$

9 Final Simplification of Credibility Prediction

We can further simplify the 8×8 formulation by using the special structure just found. Partitioning $\boldsymbol{Z}$ as in Figure 3, we find:

$$\begin{aligned} \boldsymbol{Z}^{11} &= (\boldsymbol{R}^{11} - \boldsymbol{R}^{10})(\boldsymbol{C}^{11} - \boldsymbol{C}^{00})^{-1} ; \\ \boldsymbol{Z}^{10} &= \boldsymbol{R}^{10}(\boldsymbol{C}^{00})^{-1} - \boldsymbol{Z}^{11} ; \\ \boldsymbol{Z}^{01} &= \mathbf{0} ; \qquad\qquad \boldsymbol{Z}^{00} = \boldsymbol{R}^{00}(\boldsymbol{C}^{00})^{-1} , \end{aligned} \tag{9.1}$$

which reduces the original 8×8 inversion to just *two* 4×4 inversions! Furthermore, by partitioning $\boldsymbol{y}$ into $[\boldsymbol{y}^1 \mid \boldsymbol{y}^0]^T$, and similarly with $\boldsymbol{f} = [\boldsymbol{f}^1 \mid \boldsymbol{f}^0]^T$ and $\boldsymbol{m} = [\boldsymbol{m}^0 \mid \boldsymbol{m}^0]^T$, we obtain the final two prototype first- and second-moment prediction formulae at the individual and collective levels:

$$\begin{aligned} \boldsymbol{f}^1(\mathscr{D}) &= (\boldsymbol{I} - \boldsymbol{Z}^{10} - \boldsymbol{Z}^{11})\boldsymbol{m}^0 + \boldsymbol{Z}^{10}\boldsymbol{y}^0 + \boldsymbol{Z}^{11}\boldsymbol{y}^1 , \\ \boldsymbol{f}^0(\mathscr{D}) &= (\boldsymbol{I} - \boldsymbol{Z}^{00})\boldsymbol{m}^0 + \boldsymbol{Z}^{00}\boldsymbol{y}^0 . \end{aligned} \tag{9.2}$$

Comparing with the first-moment hierarchical prediction (2.5), we see that $\boldsymbol{f}^0(\mathscr{D})$ is in the correct form, but that $\boldsymbol{f}^1(\mathscr{D})$ does not quite have the same simple form. Further investigation shows that this could only be true if $\boldsymbol{Z}^{10} = (\boldsymbol{I} - \boldsymbol{Z}^{11})\boldsymbol{Z}^{00}$; in fact, the numerical match is never close unless r is very large, or one throws away d-type terms that use y_{1*0} and y_{0*0} in the prediction. We think (9.2) is simple enough!

10 Cross-Moment Prediction

To forecast cross-moments, say $\widetilde{w}_{1*2} = \widetilde{x}_{1,n+1}\widetilde{x}_{2,n+1}$ or $m_1(\widetilde{\theta}_1)m_1(\widetilde{\theta}_2)$, we must expand (8.2) and (8.3) to contain predictors and predictands of both risks # 1

and # 2 (as prototypes). The dimension of the problem now jumps from 8 to 13, with, for instance:

$$\tilde{\boldsymbol{y}} = \Big[\tilde{y}_1, \tilde{y}_{11}, \tilde{y}_{1\times 1}, \tilde{y}_{1*0}; \; \tilde{y}_2, \tilde{y}_{22}, \tilde{y}_{2\times 2}, \tilde{y}_{2*0}; \; \tilde{y}_{1\times 2} \mid \tilde{y}_0, \tilde{y}_{00}, \tilde{y}_{0\times 0}, \tilde{y}_{0*0}\Big]^T, \tag{10.1}$$

and similarly for $\tilde{\boldsymbol{w}}$. Many of the coefficients in the 13×13 covariance matrices duplicate those already found, but there are new cross-risk moments to be found, which we leave to the reader. The final credibility formulae are similar to (9.2), and require one 9×9 and one 4×4 inversion.

11 General Asymptotic Results

In contrast to (2.5), it is difficult to say much about asymptotic results for the general hierarchical model as the number of samples, n, gets very large. Terms in f and φ drop out of $\boldsymbol{C}$, but otherwise no simplification occurs in the structure of $\boldsymbol{Z}$. In other words, neither y_i nor y_0 achieves "full credibility" for the mean forecasts, f_i and f_0, in the general case. We shall see below, however, that this can occur in special cases.

As r, the number of risks, increases, $\boldsymbol{C}^{00}$, $\boldsymbol{R}^{10}$, and $\boldsymbol{R}^{00}$ approach the same 4×4 matrix of h coefficients, call it $\boldsymbol{H}$. Assuming $\boldsymbol{H}^{-1}$ exists, then:

$$\begin{aligned} \boldsymbol{f}^0(\mathscr{D}) &= \boldsymbol{y}^0 ; \\ \boldsymbol{f}^1(\mathscr{D}) &= (\boldsymbol{I} - \boldsymbol{Z}^{11})\boldsymbol{y}^0 + \boldsymbol{Z}^{11}\boldsymbol{y}^1 ; \\ \boldsymbol{Z}^{11} &= (\boldsymbol{R}^{11} - \boldsymbol{H})(\boldsymbol{C}^{11} - \boldsymbol{H})^{-1} , \end{aligned} \tag{11.1}$$

giving full credibility at the portfolio level for the four statistics in $\boldsymbol{y}^0$, a satisfying result. However, in the examples below, $|\boldsymbol{H}| = 0$, and (11.1) is not true for all components.

12 Numerical Examples

The first two numerical examples are for models in which all of the first and second moment forecasts are exact; the third example is similar, but is known to be analytically intractable, even for the predictive mean. Other computational experience is then summarized.

12.1 Example A – Homoscedastic Normal-Normal-Normal Model

In this model, developed by *Lindley/Smith* (1972) in a slightly more general setting, the densities at all levels are normal with known variances:

$$\begin{aligned} (\tilde{x}_{it} \mid \theta_i) &\sim \mathcal{N}ormal\ (\theta_i, f)\ ; \\ (\tilde{\theta}_i \mid \phi) &\sim \mathcal{N}ormal\ (\phi, g)\ ; \\ \tilde{\phi} &= \mathcal{N}ormal\ (m, h)\,. \end{aligned} \tag{12.1}$$

In this case, *Jewell* (1975 b) showed that the linear mean hierarchical forecasts (2.5) for $f_1(\mathcal{D})$ and $f_0(\mathcal{D})$ are exact, using the same m, f, g, h. However, it is easy to see that the second moment forecasts are *homoscedastic*, that is, all updated variances and covariances depend only upon the sampling design parameters, (n, r), and not upon the sample values in $\mathcal{D}$. This means that our second-moment forecasts are basically uninteresting, for example:

$$\begin{aligned} f_{11}(\mathcal{D}) &= f + (1 - z_1)g + (1 - z_0)(1 - z_1)^2 h + f_1^2(\mathcal{D})\ ; \\ f_{00}(\mathcal{D}) &= f + g + (1 - z_0)h + f_0^2(\mathcal{D})\,, \end{aligned} \tag{12.2}$$

and with similar data dependence for $f_{1\times 1}$, $f_{0\times 0}$, and f_{0*0}; f_{1*0} is a hybrid, and depends linearly on $y_1 f_0$ and f_0^2.

For the calculations, we took $m = 1$, $f = 4$, $g = 0.4$, and $h = 0.04$, reflecting increasing certainty about higher-level values (which we believe is the usual case in insurance). This gives time constants of $n_0 = r_0 = 10$ in the usual credibility factors z_1 and z_0. The resulting central moments are shown in Table I (a). Note the large number of values that are duplicate or stand in constant ratio.

Table I.
Central moments for three numerical examples

(a) Example A

	f	g	h	φ	γ	τ
aa	4.0	0.4	0.04	---	---	---
ab	8.0	0.8	0.08	---	---	---
ac	8.0	0.8	0.08	---	---	---
bb	55.04	1.984	0.1632	---	---	---
bc	23.04	1.984	0.1632	---	---	---
cc	23.04	1.984	0.1632	---	---	32.0
ad	---	---	0.08	4.0	0.4	---
bd	---	---	0.1632	8.32	0.832	---
cd	---	---	0.1632	8.32	0.832	---
dd	11.52	0.992	0.1632	4.16	0.416	16.0

(b) Example B

	f	g	h	φ	γ	τ
aa	4.0	0.4	0.04	---	---	---
ab	8.0	0.8	0.08	---	---	---
ac	8.0	0.8	0.08	---	---	---
bb	74.56	2.176	10.0216	---	---	---
bc	26.56	2.176	1.1416	---	---	---
cc	26.56	2.176	0.2616	---	---	48.0
ad	---	---	0.08	4.0	0.4	---
bd	---	---	0.2536	8.48	0.848	---
cd	---	---	0.1736	8.48	0.848	---
dd	13.28	1.088	0.1656	4.24	0.424	24.0

(c) Example C

	f	g	h	φ	γ	τ
aa	4.0	0.4	0.04	---	---	---
ab	8.0	0.8	0.08	---	---	---
ac	8.0	0.8	0.08	---	---	---
bb	55.04	2.144	0.2432	---	---	---
bc	23.04	2.144	0.2432	---	---	---
cc	23.04	2.144	0.2432	---	---	32.0
ad	---	---	0.08	4.0	0.4	---
bd	---	---	0.1632	8.32	0.832	---
cd	---	---	0.1632	8.32	0.832	---
dd	11.52	1.072	0.1632	4.16	0.416	16.0

Table II shows the credibility matrix, $\boldsymbol{Z}$, for selected values of n and r. For instance, with $n = 10$, $r = 5$, rows # 1 and # 5 show that f_1 and f_0 are just (2.5), with $z_1 = 0.5$ and $z_0 = 0.2$.

Table II.
Numerical values of Z for example A

0.5	0.0	0.0	0.0	0.1	0.0	0.0	0.0
0.4	0.027	0.243	0.08	0.08	0.0002	0.0018	0.008
0.4	0.027	0.243	0.08	0.08	0.0002	0.0018	0.008
0.4	0.002	0.018	0.08	0.16	0.0004	0.0036	0.016
0	0	0	0	0.2	0.0	0.0	0.0
0			0	0.32	0.0008	0.0072	0.032
0	n = 10	r = 5	0	0.32	0.0008	0.0072	0.032
0			0	0.32	0.0008	0.0072	0.032

0.8333	0.0	0.0	0.0	0.0490	0.0	0.0	0.0
0.1961	0.0142	0.6966	0.0654	0.0115	0.0000	0.0005	0.0019
0.1961	0.0142	0.6966	0.0654	0.0115	0.0000	0.0005	0.0019
0.5882	0.0010	0.0480	0.1961	0.0692	0.0001	0.0028	0.0115
0	0	0	0	0.2941	0.0	0.0	0.0
0			0	0.4152	0.0004	0.0170	0.0692
0	n = 50	r = 5	0	0.4152	0.0004	0.0170	0.0692
0			0	0.4152	0.0004	0.0170	0.0692

1.0	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	1.0	0.0	0.0	0.0	0.0	0.0
0.6667	0.0	0.0667	0.2667	0.0	0.0	0.0	0.0
0	0	0	0	0.3333	0.0	0.0	0.0
0			0	0.4444	0.0	0.0222	0.0889
0	$n \to \infty$	r = 5	0	0.4444	0.0	0.0222	0.0889
0			0	0.4444	0.0	0.0222	0.0889

0.5	0.0	0.0	0.0	0.5	0.0	0.0	0.0
0.0	0.025	0.225	0.5	0.0	0.0	0.0	0.25
0.0	0.025	0.225	0.5	0.0	0.0	0.0	0.25
0.0	0.0	0.	0.5	0.0	0.0	0.0	0.5
0	0	0	0	1.0	0.0	0.0	0.0
0			0	0.0	0.0	0.0	1.0
0	n = 10	$r \to \infty$	0	0.0	0.0	0.0	1.0
0			0	0.0	0.0	0.0	1.0

1.0	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	1.0	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	1.0	0.0	0.0	0.0
0			0	0.0	0.0	0.0	1.0
0	$n \to \infty$	$r \to \infty$	0	0.0	0.0	0.0	1.0
0			0	0.0	0.0	0.0	1.0

Now, it is always true that:

$$
\begin{aligned}
y_i^2 &= \left[\frac{1}{n}\right] y_{ii} + \left[\frac{n-1}{n}\right] y_{i\times i}\,; \\
y_0^2 &= \left[\frac{1}{r}\right] y_{00} + \left[\frac{r-1}{r}\right] y_{0*0}\,; \quad \text{and} \\
y_i y_0 &= \left[\frac{1}{r}\right] y_i^2 + \left[\frac{r-1}{r}\right] y_{i*0}\,.
\end{aligned}
\tag{12.3}
$$

So, if f_1 and f_0 obey (2.5) exactly, we can then expand $f_1^2(\mathscr{D})$, $f_0^2(\mathscr{D})$, and $y_1 f_0(\mathscr{D})$ solely in terms of the eight statistics in $\boldsymbol{y}$! This leads to the duplicate coefficients in the remaining rows of $\boldsymbol{Z}$, as explained above.
As $n \to \infty$, the exact nature of the forecasts leads to $f_1 = 1 \cdot y_1$, as expected, and to $z_0 = (r/(r + r_0))$; the other limits, such as $f_{11} = f_{1\times 1} = y_{11}$, can be explained in terms of the limiting behavior described in *J&S*. As $r \to \infty$, the limiting $\boldsymbol{Z}$ is not $\boldsymbol{I}$, because $|\boldsymbol{H}| = 0$.

12.2 Example B – Jewell's Heteroscedastic Model

We now consider ways to make model A heteroscedastic, that is, to have predictive (co-)variances that depend upon the data values; in the normal-normal-normal model, this clearly requires that one or more of the basic variances in (12.1) be considered as unknown, *a priori*. *Jewell* (1987) showed that if the densities in (12.1) are taken as *conditional* on (f, g, h), with $p(f, g, h)$ a given joint prior, then the key to retaining analytic solvability was to *fix* the credibility factors z_1 and z_0 by *linking* all three variances through the relations:

$$
\tilde{f} = n_0 \tilde{g} = n_0 r_0 \tilde{h}\,, \tag{12.4}
$$

thus fixing the time constants n_0 and r_0. The essential work is then to find $\mathscr{E}\{\tilde{f} \mid \mathscr{D}\}$, etc. now needed in the generalized forecasts like (12.2).
If we choose a $\mathscr{G}amma$ (α, β) prior for $(\tilde{f})^{-1}$, we obtain credibility-type formulae for the updated (co-)variance(s)! The main result is:

$$
\begin{aligned}
\mathscr{E}\{\tilde{f} \mid \mathscr{D}\} &= (1 - z_*)\mathscr{E}\{\tilde{f}\} + z_* \boldsymbol{B}(\mathscr{D})\,; \\
\mathscr{E}\{\tilde{f}\} &= \frac{\beta}{\alpha - 1}\,; \\
z_* &= \frac{nr}{nr + 2(\alpha - 1)}\,,
\end{aligned}
\tag{12.5}
$$

with a linked-variance statistic that, in its most pleasing form, can be written:

$$\begin{aligned} \boldsymbol{B}(\mathscr{D}) = {} & (1 - z_0)(1 - z_1)(y_{00} - m)^2 \\ & + \left[1 - \frac{1}{n}\left[z_1 + \frac{1}{r}(1 - z)\right]\right][y_{00} - y_{0\times 0}] \\ & + (1 - z_1)\left(1 - \frac{1}{r}\right)[y_{0\times 0} - y_{0*0}]\,. \end{aligned} \tag{12.6}$$

Other forms are given in *Jewell* (1987). Note that the "volume" of effective data in the learning-curve z_* is (nr), and that, with fixed $\mathscr{E}\{\tilde{f}\}$, the effective time constant, $2(\alpha - 1)$, is larger, the more certain we are about $\tilde{f}$, *a priori*. The forecasts of (12.2) are easily generalized to this model by setting $g = f/n_0$ and $h = f/n_0 r_0$ and using (12.5) to uncondition on $\tilde{f}$. So the only change in f_{11}, $f_{1\times 1}$, and f_{1*0} will be in the coefficients from $\boldsymbol{y}^0$.

For the numerical calculations, we have set $m = 1$, $\mathscr{E}\{\tilde{f}\} = 4.0$, and $n_0 = r_0 = 10$, so that the mean values of the variances are comparable with example A. Heteroscedasity is introduced by setting $\mathscr{V}\{\tilde{f}\} = 8.0$, giving $\mathscr{V}\{\tilde{g}\} = 0.08$ and $\mathscr{V}\{\tilde{h}\} = 0.0008$, and similarly for the covariances. The resulting central moments are shown in Table I (b). The credibility matrices in Table III, calculated for the same values of n and r as Table II, show the same results for f_1 and f_0, as well as the same coefficients in the rest of $\boldsymbol{Z}^{11}$, as expected. However, negative and/or larger-than-unity coefficients now appear because of the complexity of the the statistic $\boldsymbol{B}(\mathscr{D})$. Of particular interest is the fact that, as $n \to \infty$, $f_{11} \to y_{11} + (y_{00} - y_{0\times 0})$, with $f_{1\times 1} \to y_{1\times 1}$, as before. $|\boldsymbol{H}| = 0$ for this model, too, but here, as $r \to \infty$, f_{00} and $f_{0\times 0}$ both use all the second-moment statistics from $\boldsymbol{y}^0$ in their exact, linked predictions.

12.3 Example C – Berger's Heteroscedastic Model

Berger (1985: Section 4.6) analyzes a normal-normal-normal hierarchical model that, in our notation, has f and h fixed, but with $\tilde{g}$ having an arbitrary prior. He essentially develops (2.5) and the covariance form of (12.2), conditional on g, gives the likelihood $p(\mathscr{D} \mid g)$, and then proposes using several one-dimensional numerical integrations to uncondition on $\tilde{g}$. Because of the complexity of $p(\mathscr{D} \mid g)$, and the way in which g enters $z_1(g)$ and $z_0(g)$, we see that not even the predictive means can be linear functions of our statistics. Nevertheless, because Berger's model is "close" to those analyzed above, we feel intuitively that the credibility approximation should be "good enough" for most purposes.

Table III.
Numerical values of $\boldsymbol{Z}$ for example B

0.5	0.0	0.0	0.0	0.1	0.0	0.0	0.0
0.4	0.027	0.243	0.08	-0.6714	0.8906	-0.4378	-0.0671
0.4	0.027	0.243	0.08	0.0429	0.0442	-0.0199	0.0043
0.4	0.002	0.018	0.08	0.1571	0.0038	0.0019	0.0157
0	0	0	0	0.2	0.0	0.0	0.0
0	n = 10	r = 5	0	-0.4714	0.9386	-0.4558	-0.0471
0			0	0.2429	0.0922	-0.0379	0.0243
0			0	0.3143	0.0076	0.0039	0.0314

0.8333	0.0	0.0	0.0	0.0490	0.0	0.0	0.0
0.1961	0.0142	0.6966	0.0654	-0.2221	0.9763	-0.8200	-0.0370
0.1961	0.0142	0.6966	0.0654	0.0077	0.0162	-0.0131	0.0013
0.5882	0.0010	0.0480	0.1961	0.0689	0.0012	0.0019	0.0115
0	0	0	0	0.2941	0.0	0.0	0.0
0	n = 50	r = 5	0	0.1609	1.0632	-0.8763	0.0268
0			0	0.3906	0.1031	-0.0694	0.0651
0			0	0.4136	0.0071	0.0113	0.0689

1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.0	0.0	1.0	0.0	0.0	1.0	-1.0	0.0
0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0
0.6667	0.0	0.0667	0.2667	0.0	0.0	0.0	0.0
0	0	0	0	0.3333	0.0	0.0	0.0
0	$n \to \infty$	r = 5	0	0.4444	1.1067	-1.0844	0.0889
0			0	0.4444	0.1067	-0.0844	0.0889
0			0	0.4444	0.0067	0.0156	0.0889

0.5	0.0	0.0	0.0	0.5	0.0	0.0	0.0
0.0	0.025	0.225	0.5	0.0	0.9975	-0.4725	-0.275
0.0	0.025	0.225	0.5	0.0	0.0475	-0.0225	0.225
0.0	0.0	0.0	0.5	0.0	0.0	0.0	0.5
0	0	0	0	1.0	0.0	0.0	0.0
0	n = 10	$r \to \infty$	0	0.0	1.045	-0.495	0.45
0			0	0.0	0.095	-0.045	0.95
0			0	0.0	0.0	0.0	1.0

1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.0	0.0	1.0	0.0	0.0	1.0	-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	1.0	0.0	0.0	0.0	0.0
0	0	0	0	1.0	0.0	0.0	0.0
0	$n \to \infty$	$r \to \infty$	0	0.0	1.1	-1.1	1.0
0			0	0.0	0.1	-0.1	1.0
0			0	0.0	0.0	0.0	1.0

Table IV.
Numerical values of $\boldsymbol{Z}$ for example C

0.5	0.0	0.0	0.0	0.1	0.0	0.0	0.0
0.2727	0.0346	0.3109	0.0485	-0.0793	0.0137	0.1232	-0.0276
0.2727	0.0346	0.3109	0.0485	-0.0793	0.0137	0.1232	-0.0276
0.4091	0.0018	0.0164	0.0727	0.1752	-0.0001	-0.0005	0.0175
0	0	0	0	0.2	0.0	0.0	0.0
0	n = 10	r = 5	0	0.0340	0.0212	0.1906	-0.0288
0			0	0.0340	0.0212	0.1906	-0.0288
0			0	0.3281	0.0006	0.0053	0.0301

0.8333	0.0	0.0	0.0	0.0490	0.0	0.0	0.0
0.0919	0.0156	0.7633	0.0227	-0.0513	0.0017	0.0815	-0.0227
0.0919	0.0156	0.7633	0.0227	-0.0513	0.0017	0.0815	-0.0227
0.6365	0.0008	0.0386	0.1575	0.1180	-0.0002	-0.0113	0.0257
0	0	0	0	0.2941	0.0	0.0	0.0
0	n = 50	r = 5	0	-0.0165	0.0076	0.3724	-0.0777
0			0	-0.0165	0.0076	0.3724	-0.0777
0			0	0.4476	0.0002	0.0091	0.0610

1.0	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	1.0	0.0	0.0	0.0	0.0	0.0
0.75	0.0	0.05	0.20	0.0661	0.0	-0.0190	0.0276
0	0	0	0	0.3333	0.0	0.0	0.0
0	$n \to \infty$	r = 5	0	-0.0230	0.0	0.4414	-0.0966
0			0	-0.0230	0.0	0.4414	-0.0966
0			0	0.4943	0.0	0.0103	0.0759

0.5	0.0	0.0	0.0	0.5	0.0	0.0	0.0
0.0	0.0333	0.3	0.3333	0.0	0.0667	0.6	-0.3333
0.0	0.0333	0.3	0.3333	0.0	0.0667	0.6	-0.3333
0.0	0.0	0.0	0.5	0.0	0.0	0.0	0.5
0	0	0	0	1.0	0.0	0.0	0.0
0	n = 10	$r \to \infty$	0	0.0	0.1	0.9	0.0
0			0	0.0	0.1	0.9	0.0
0			0	0.0	0.0	0.0	1.0

1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.0	0.0	1.0	0.0	0.0	1.0	-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	1.0	0.0	0.0	0.0	0.0
0	0	0	0	1.0	0.0	0.0	0.0
0	$n \to \infty$	$r \to \infty$	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

To maintain similarity, our numerical trials set $m = 1$, $f = 4$, and $h = 0.04$, as in example A, and $\mathscr{V}\{\tilde{g}\} = 0.08$, as in example B; all covariances and the other variances are zero. Table I (c) shows that the central moments are closer to those of A than those of B; in fact, only seven moments in the g and h columns differ from those of A!

Table IV shows the numerical results for $\boldsymbol{Z}$, for the same values of n and r used previously. The first surprise is that the *approximations* $f_1(\mathscr{D})$ and $f_0(\mathscr{D})$ for $\mathscr{E}\{\tilde{\theta} \mid \mathscr{D}\}$ and $\mathscr{E}\{\tilde{\phi} \mid \mathscr{D}\}$, respectively, are still the simple formulae of (2.5)! (discussion below). On the other hand, the other coefficients of $\boldsymbol{Z}^{11}$ are different than those of either example A or B. The other coefficients for $\boldsymbol{y}^0$ have negative signs in different locations than those of example B, and are generally smaller in magnitude, closer to those of A.

Asymptotically, there are few surprises, although, as $r \to \infty$, both f_{11} and $f_{1\times 1}$ have the fixed term $0.9[M(2) - M(11)]$, so that neither estimate is ever fully credible. On the other hand, if both n and r are large, $\boldsymbol{Z} \to \boldsymbol{I}$, probably because of the underlying non-linearity of the exact predictions.

Overall, the fact the credibility coefficients of example C are "between" those of A and B, for the most part, gives us additional confidence that they are robust approximations to the true predictive formulae.

12.4 Other Computational Experience

Is the behavior of examples A, B, and C typical of general normal-normal-normal models with arbitrary priors, p (f, g, h)? (Note that credibility approximations only use the means and covariances of these r.v.s). Based upon computational experience with a variety of cases, we can say empirically that f_1 always uses only y_1 and y_0, and that f_0 only depends upon y_0. A proof of this approximation simplicity seems difficult, but it must be related to the assumption of normality at all levels.

On the other hand, it is easy to construct normal models with $|\boldsymbol{H}| \neq 0$, so that, as $r \to \infty$, $\boldsymbol{f}^0 = \boldsymbol{y}^0$ and $\boldsymbol{f}^1 = (\boldsymbol{I} - \boldsymbol{Z}^{11})\boldsymbol{y}^0 + \boldsymbol{Z}^{11}\boldsymbol{y}^1$. We leave the exact conditions for the reader to discover, but if, for instance: $\mathscr{E}\{\tilde{h}\} = 0$; or $\mathscr{V}\{\tilde{f}\} = 0$; or $\mathscr{V}\{\tilde{g}\} = 0$; or all correlations among $(\tilde{f}, \tilde{g}, \tilde{h})$ are unity, as in example B, then $|\boldsymbol{H}| = 0$ and we obtain different limiting results. But setting only $\mathscr{V}\{\tilde{g}\} = 0$ or making $(\tilde{f}, \tilde{g}, \tilde{h})$ independent will lead to full credibility at the portfolio level as the number of risks becomes very large.

To obtain approximation formulae where f_1 depends upon more than y_1 and y_0, one must turn to non-normal model densities. From *J&S* we know that the non-hierarchical Gamma-Exponential gives exact forecasts for f_1 in terms of

y_1, and f_{11} and $f_{1\times 1}$ in terms of y_1, y_{11}, and $y_{1\times 1}$. Now let us make the scale parameter of the Gamma a random variable, with its own Gamma hyperprior density (thus giving a Gamma-Gamma-Exponential hierarchical model). The exact predictive formulae are already analytically intractable, so our f_1 and f_0 can only be linear approximations. Yet we have found empirically that f_1 still depends only upon y_1, but now upon *all* of $\boldsymbol{y}^0$; the latter is due solely to the complexity of f_0, so we get a formula for f_1 like the first line of (2.5)! Now, as $n \to \infty$, the 3×3 NW corner of $\boldsymbol{Z}^{11}$ behaves like the Gamma-Exponential example in *J&S*, which is some consolation. $\boldsymbol{H}^{-1}$ always exists in this model (in non-degenerate cases), so that, as $r \to \infty$, we have full credibility at the portfolio level.

The Gamma-Poisson is another (non-hierarchical) model that has exact first- and second-moment credibility forecasts. Creating a hierarchical model by again making the Gamma scale parameter into a Gamma r.v., we obtain different results than those described above. Now the approximation f_1 depends upon *all* components of $\boldsymbol{y}^1$ and $\boldsymbol{y}^0$! Moreover, we obtain the surprising empirical result that, as $n \to \infty$, the 3×3 NW corner of $\boldsymbol{Z}^{11}$ now approaches $\boldsymbol{I}$, rather than the corresponding limit in *J&S*! $\boldsymbol{H}^{-1}$ always exists in this model, so $\boldsymbol{Z}^{00} \to \boldsymbol{I}$ as the number of risks increases without limit, although the rate of convergence is very slow, compared with previous models. There seems to be much more variability in Poisson model prediction.

It is dangerous to extrapolate this experience to other models, since each case requires developing the appropriate 24 analytic moments to look for simplifications. However, the pathologies observed above seem to be due to the fact that densities at all three levels are members of the quadratic variance family of Morris and are natural conjugate between levels. With different densities, one should obtain "regular" asymptotic behavior. The reader is cautioned against choosing the 24 moments at random for experimental trials, as the moments must obey many Schwarzian inequalities based upon applying $\mathscr{E}\{\tilde{y}^4\} \geq [\mathscr{E}\{\tilde{y}^2\}]^2 \geq [\mathscr{E}\{\tilde{y}\}]^4$ at all levels; assuming positive random variables and/or special densities adds further constraints.

Special thanks are due to M. Lin, who assisted in the complicated details of program development, computation, and validation for these examples.

13 Conclusion

We have seen that the hierarchical credibility model reduces from a large least-squares formulation to a pair of 4×4 problems, one of which finds the portfolio-level $\boldsymbol{f}^0(\mathscr{D})$, and the other the individual-level $\boldsymbol{f}^1(\mathscr{D})$. The

methodology also agrees numerically with those simple hierarchical models for which analytic forms are known, and gives reasonable approximations for "nearby" models for which exact results are not available.
The use of the common exponential-family modeling densities greatly simplifies parameter estimation, but raises interesting questions about the asymptotic behavior of these formulae when n and/or r are large, since "full credibility", using the natural estimators, may not be attained. And we have not at all considered whether a subset of our statistics might give almost as good approximations. Thus, there remains a great deal of practical exploration of the methodology, which we must leave to the future.

14 Dedication

Some scientific ideas are like complex wines – they need substantial blending and ageing before they can be properly appreciated. So it is with the topic of this paper.
The basic ideas and the formulae of Section 6 were developed during a visit of HB to Berkeley in Spring, 1986. However, extensive "cellaring" was necessary to simplify notation, to understand and exploit the special structure of the coefficients, to find a heteroscedastic test model, and, above all, to develop and test the many computer programs used for the numerical calculations.
On this special occasion of his sixtieth birthday, the cellarmaster would like to dedicate the long-overdue bottling of this work to Hans Bühlmann, with the hope that he will find it to his taste.

William S. Jewell
Engineering Systems Research Center
University of California, Berkeley
Berkeley, Ca 94720
USA

References

Bühlmann, H. (1967): Experience rating and credibility. ASTIN Bulletin *4*, 199 – 207.

Bühlmann, H./Jewell, W. S. (1987): Hierarchical credibility revisited. Bulletin of the Association of Swiss Actuaries *87*, 35 – 54.

Jewell, W. S. (1974 a): Credible Means are exact Bayesian for simple exponential families. ASTIN Bulletin *8*, 79 – 90.

Jewell, W. S. (1974 b): Exact multidimensional credibility. Bulletin of the Association of Swiss Actuaries *74*, 193 – 214.

Jewell, W. S. (1975 a): Regularity conditions for exact credibility. ASTIN Bulletin *8*, 336 – 341.

Jewell, W. S. (1975 b): The use of collateral data in credibility theory: a hierarchical model. Giornale dell'Istituto Italiano degli Attuari *38*, 1 – 16.

Jewell, W. S. (1983): Enriched multinormal priors revisited. Journal of Econometrics *23*, 5 – 35.

Jewell, W. S./Schnieper, R. (1985): Credibility approximations for Bayesian prediction of second moments. ASTIN Bulletin *15*, 103 – 121.

Jewell, W. S. (1987): A heteroscedastic hierarchical model. Bayesian Statistics 3, Proceedings of the Third Valencia International Meeting, 1 – 5 June 1987. Clarendon Press, Oxford, 657 – 663.

Jewell, W. S. (1988): Covariance matrix patterns invariant under multiplication and inversion. Engineering Systems Research Center Report ESRC 88 – 13, University of California, Berkeley.

Jewell, W. S. (1989): A general framework for credibility prediction of multidimensional first and second moments. Insurance: Mathematics and Economics *8*, 1 – 10.

Lindley, D. V./Smith, A. F. M. (1972): Bayes estimates for the linear model. Journal of the Royal Statistical Society B *34*, 1 – 41.

Norberg, R. (1986): Hierarchical credibility: analysis of a random effect linear model with nested classification. Scandinavian Actuarial Journal, 204 – 222.

Sundt, B. (1979): A hierarchical credibility regression model. Scandinavian Actuarial Journal, 107 – 113.

Taylor, G. C. (1979): Credibility analysis of general hierarchical models. Scandinavian Actuarial Journal, 1 – 12.

Taylor, G. C. (1984): Experience rating with credibility adjustment of the manual premium. ASTIN Bulletin *7*, 323 – 336.

Summary

Multidimensional credibility provides simultaneous linear least-squares approximations to the first and second predictive moments of the hierarchical model, using natural first- and second-order statistics formed from all risks in the portfolio. This gives a very large, special-structure covariance matrix, which would be very difficult to invert for a typical portfolio. Matrix arguments show that the formulation can be reduced to small, fixed-dimension matrices, giving prototypical credibility formulae for use with any number of risks. Numerical examples are given for the normal-normal-normal model, for which exact Bayesian results can be obtained in some cases, and for other models.

Zusammenfassung

Mehrdimensionale Kredibilitätstheorie liefert lineare Approximationen (mittels der Methode der kleinsten Quadrate) gleichzeitig für das erste und das zweite Moment des hierarchischen Modells, wobei natürliche Statistiken erster und zweiter Ordnung – basierend auf allen Risiken eines Portefeuilles – benutzt werden. Dies führt zu einer sehr grossen, speziell strukturierten Kovarianzmatrix, deren Invertierung sehr schwierig sein dürfte. Mit Hilfe der Matrizenrechnung wird der Formalismus reduziert auf kleine Matrizen fester Dimension. Numerische Beispiele werden erläutert für das "normal-normal-normal"-Modell, für welches in gewissen Fällen exakte Bayesianische Resultate hergeleitet werden können, sowie auch für andere Modelle.

Résumé

La crédibilité multidimensionnelle livre des approximations par moindre carrés simultanées et linéaires pour les premier et second moments a priori du modèle hiérarchique, et cela sur la base des statistiques naturelles de premier et de second ordres formées par l'ensemble des risques du portefeuille. Ce fait fournit une matrice de covariance très étendue et de structure spéciale, qu'il serait très difficile d'inverser dans le cas d'un portefeuille typique. Une étude montre que la formulation peut être réduite au cas de matrices de petite taille et de dimension fixe comportant des formules de crédibilité prototypiques pour un usage en présence d'un nombre quelconque de risques. Des exemples numériques suivent dans le cas du modèle normal-normal-normal, pour lequel il est possible d'obtenir des résultats bayesiens exacts dans quelques cas de figure, ainsi qu'avec d'autres modèles.