

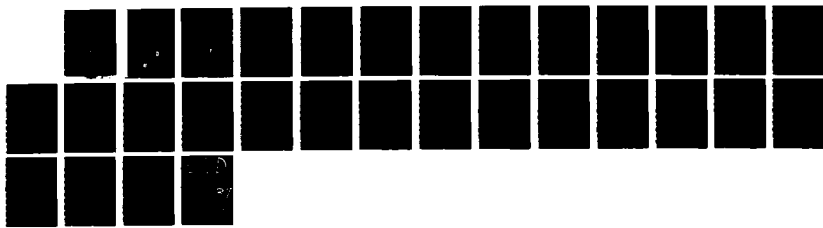
AD-A184 256

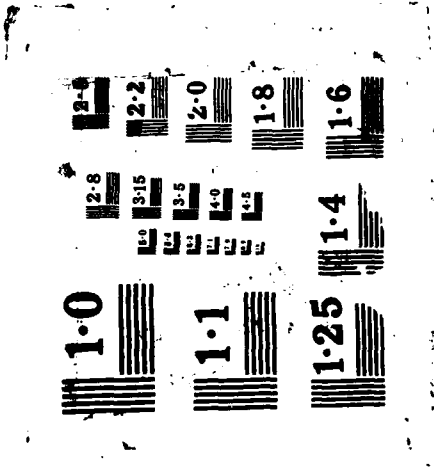
A HETEROSCEDASTIC HIERARCHICAL MODEL (U) CALIFORNIA UNIV 1/1  
BERKELEY OPERATIONS RESEARCH CENTER W S JEWELL APR 87  
ORC-87-11 AFORR-81-8122

UNCLASSIFIED

F/G 12/4

NL





AD-A184 256

# OPERATIONS RESEARCH CENTER

A HETEROSCEDASTIC HIERARCHICAL MODEL\*\*\*

by

William S. Jewell\*

ORC 87-11

April 1987

DTIC  
DTIC  
ELECTE  
SEP 08 1987  
S D

UNIVERSITY OF CALIFORNIA



BERKELEY

DISTRIBUTION STATEMENTS  
Approved for public release  
Distribution unlimited

87 9 3 091

11

A HETEROSCEDASTIC HIERARCHICAL MODEL\*\*\*

by

William S. Jewell\*

ORC 87-11

April 1987

DTIC  
ELECTE  
SEP 08 1987  
D

\*Department of Industrial Engineering and Operations Research, California University of California, Berkeley, California 94720.

\*\*\*This research has been partially supported by the Air Force Office of Scientific Research (AFSC), USAF, under Grant AFOSR-81-0122 with the University of California. Reproduction in whole or in part is permitted for any purpose of the United States Government.

Approved for  
Distribution

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

AD-A134 25

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION <b>UNCLASSIFIED</b>		1b. RESTRICTIVE MARKINGS	
2a. SECURITY CLASSIFICATION AUTHORITY NA		3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for Public Release; Distribution Unlimited	
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE NA			
4. PERFORMING ORGANIZATION REPORT NUMBER(S) 87-11		5. MONITORING ORGANIZATION REPORT NUMBER(S)	
6a. NAME OF PERFORMING ORGANIZATION University of Calif., Berkeley	6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION AFOSR/NM	
6c. ADDRESS (City, State and ZIP Code) Engineering Systems Research Center 3115 Etcheverry Hall, Univ. of Calif. Berk. Berkeley, California 94720		7b. ADDRESS (City, State and ZIP Code) Bldg. 410 Bolling AFB, DC 20332-6448	
8a. NAME OF FUNDING/SPONSORING ORGANIZATION AFOSR	8b. OFFICE SYMBOL (If applicable) NM	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER	
8c. ADDRESS (City, State and ZIP Code) Bldg. 410 Bolling AFB, DC		10. SOURCE OF FUNDING NOS.	
		PROGRAM ELEMENT NO. 6.1102F	TASK NO. 2304
			WORK UNIT NO. leave blk
11. TITLE (Include Security Classification) A Heteroscedastic Hierarchical Model			
12. PERSONAL AUTHOR(S) William S. Jewell			
13a. TYPE OF REPORT technical	13b. TIME COVERED FROM _____ TO _____	14. DATE OF REPORT (Yr., Mo., Day)	15. PAGE COUNT
16. SUPPLEMENTARY NOTATION Keywords: Hierarchical models, collateral data, Bayesian prediction, heteroscedastic variances, credibility formulae.			
17. COSATI CODES		18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number) see #16	
FIELD	GROUP		
SUB. GR.			
XXXXXXXXXXXXXX			
19. ABSTRACT (Continue on reverse if necessary and identify by block number) see report			
20. DISTRIBUTION/AVAILABILITY OF ABSTRACT UNCLASSIFIED/UNLIMITED <input checked="" type="checkbox"/> SAME AS RPT. <input type="checkbox"/> DTIC USERS <input type="checkbox"/>		21. ABSTRACT SECURITY CLASSIFICATION UNCLASSIFIED	
22a. NAME OF RESPONSIBLE INDIVIDUAL Major Brian Woodruff		22b. TELEPHONE NUMBER (202) 767-5027	22c. OFFICE SYMBOL AFOSR/NM

ABSTRACT

Hierarchical models are important in Bayesian prediction because they enable the use of collateral data from related risks with exchangeable parameters. The classical normal-normal-normal model with random means show clearly how the linear predictive mean for a single risk is improved by the availability of cohort data. However, this model has the disadvantage that the predictive density is *homoscedastic*, that is, the posterior variance depends only on the design (number of risks and number of samples). In most applications, one would assume that the variance also depended upon the data values.

One can, of course, change the variances at each level into random parameters, but this modifies the predictive/mean formulae and leads to messy results in general. In the course of examining approximations to predictive/variances, the author has found an extended normal model with variances that are quadratic in the data, and with the additional advantage that the linear mean formulae are unchanged.

Keywords: Hierarchical models, collateral data, Bayesian prediction, heteroscedastic variances, credibility formulae.



Decision For	
NTIS CRA&I	<input checked="" type="checkbox"/>
BTIC TAB	<input type="checkbox"/>
Unannounced	<input type="checkbox"/>
Justification	
By	
Distribution/	
Availability Codes	
Dist	Avail and/or Special
A-1	

## 1. Introduction

Consider an individual risk (#1), characterized by an unknown risk parameter,  $\tilde{\theta}_1$ , from which  $n_1$  i.i.d. sample observations,

$\mathcal{D}_1 = \{x_{1t} : (t=1,2,\dots,n_1)\}$ , are available; the problem is to predict a future observation, say  $\tilde{w}_1 = \tilde{x}_{1,n_1+1}$ . Given the model density,

$p(x_{1t}|\theta_1)$ , and prior density,  $p(\theta_1)$ , finding the forecast density,

$p(w_1|\mathcal{D})$ , is then a simple exercise in Bayes' law.

For a variety of simple likelihoods and priors (Jewell [1974] [1975a]), the forecast mean turns out to be a linear function of the data:

$$(1.1) \quad \mathcal{E}\{\tilde{w}_1|\mathcal{D}_1\} = f_1(\mathcal{D}_1) = (1 - z_1)m + z_1(\sum x_{1t}/n_1) .$$

where the mixing coefficient,

$$(1.2) \quad z_1 = n_1/(n_1 + (e/d)) .$$

is called the *credibility factor*, and the three required marginal moments are:

$$(1.3) \quad m = \mathcal{E}\{\tilde{x}_{1t}|\tilde{\theta}_1\} \quad ; \quad e = \mathcal{E}\mathcal{V}\{\tilde{x}_{1t}|\tilde{\theta}_1\} \quad ; \quad d = \mathcal{V}\mathcal{E}\{\tilde{x}_{1t}|\tilde{\theta}_1\} .$$

The *credibility formula*,  $f_1(\mathcal{D}_1)$ , has an obvious interpretation as a mixture of the prior mean with the sample mean, according to a learning curve,  $z_1$ , which tends towards unity with increasing sample size at a rate governed by a time constant,  $(e/d)$ .  $f_1(\mathcal{D}_1)$  is a robust formula in the sense that it is also the best linear least-squares fit to the true  $\mathcal{E}\{\tilde{w}_1|\mathcal{D}_1\}$  for arbitrary  $p(x_{1t}|\theta_1)$  and  $p(\theta_1)$  (Bühlmann [1967]).

In many applications, the number of samples from risk #1 will be small, but there may be additional information available from related risks, that is,  $(\tilde{x}_{1t})$  characterized by a different risk parameter,  $\tilde{\theta}_i$ , but by the same form of model density,  $p(x|\theta)$  ( $i=2,3,\dots,r$ ). For example, in insurance we may have a portfolio of risks which, a priori, are similar in nature, as measured by some risk classification scheme. Or, in health statistics we may have a cohort of apparently similar lives, with the same ages, heritage, diet, etc. Of course, if these risk parameters were independent, with the same density  $p(\theta)$ , then the collateral data would have no predictive value.

However, if we assume that  $(\tilde{\theta}_1, \tilde{\theta}_2, \dots, \tilde{\theta}_r)$  are exchangeable, we are able to keep the assumed similar nature of the related risks as well as to introduce correlation between the risks in a natural way; we do this by adding another unknown parameter,  $\tilde{\varphi}$ , which characterizes additional uncertainty at the portfolio or cohort level. In other words, the prior  $p(\theta)$  is changed to a conditional prior  $p(\theta|\varphi)$ , and we assume a hyperprior density,  $p(\varphi)$ , is known, from which we can calculate the exchangeable joint prior:

$$(1.4) \quad p(\theta_1, \theta_2, \dots, \theta_r) = \int \prod p(\theta_i | \varphi) p(\varphi) d\varphi .$$

The resulting three-level structure is called a hierarchical model.

Unfortunately, the number of cases in which explicit analytic results can be obtained from hierarchical models is quite limited. As far as we know, the only case in which the predictive individual mean is linear in the individual and collateral data is in the normal-normal-normal linear model of Lindley & Smith [1972].

To simplify notation, we shall consider the important special case of this model:

$$\begin{aligned}
 (\tilde{x}_{it} | \mu_i, \varphi) &\sim (\tilde{x}_{it} | \mu_i) \sim No(\mu_i; f) ; \\
 &\hspace{15em} (i=1,2,\dots,r) \\
 (1.5) \quad (\mu_i | \varphi) &\sim No(\varphi; g) ; \\
 &\hspace{15em} (t=1,2,\dots) \\
 (\varphi) &\sim No(m; h)
 \end{aligned}$$

As before, we predict a future value of the risk #1, but now using the total cohort data,  $\mathcal{D} = \{x_{it} ; (i=1,2,\dots,r)(t=1,2,\dots,n_i)\}$ . It will be apparent later that the predictive mean can now be written as the combination of two credibility-like forecasts:

$$\begin{aligned}
 (1.6) \quad \xi\{\tilde{w}_1 | \mathcal{D}\} &= f_1(\mathcal{D}) = (1 - z_1) f_0(\mathcal{D}) + z_1 y_1 ; \\
 f_0(\mathcal{D}) &= (1 - z_0)m + z_0 Y_0 .
 \end{aligned}$$

Here, the collateral data is organized into  $r+1$  linear sufficient statistics:

$$(1.7) \quad y_i = (\sum x_{it} / n_i) ; \quad Y_0 = (\sum z_i y_i / \sum z_j) ;$$

and  $r+1$  credibility factors:

$$(1.8) \quad z_i = \frac{n_i}{n_i + (f/g)} ; \quad z_0 = \frac{\sum z_i}{\sum z_j + (g/h)} ;$$

for each risk ( $i=1,2,\dots,r$ ) and for the portfolio as a whole .  $f_0(\mathcal{D})$  turns out to be the predictive mean at the portfolio level,  $\xi\{\tilde{\varphi} | \mathcal{D}\}$ :

additional interpretations will be given below in Section 9.

(1.6) is also the credibility formula (best linear predictive mean in the least-squares sense) for hierarchical models with arbitrary  $p(x|\theta)$ ,  $p(\theta|\varphi)$ , and  $p(\varphi)$  (Jewell [1975b]).

Thus, it would seem that the normal-normal-normal hierarchical model would be satisfactory for most situations in which collateral data is available, since, as we shall see below, the full-distributional predictive results are also easily obtained. Recently, however, the author and Hans Buhlmann have been examining credibility approximations for the second moments of  $(\tilde{w}_1|\mathcal{D})$ , in which various second-order statistics from the portfolio are used to augment a second-moment forecast using individual data (Jewell & Schnieper [1985], Jewell [1987], Jewell & Buhlmann [1987]). The least-squares analysis, while messy, is straightforward. However, (1.5) turns out *not* to be a useful test case for these approximate formulae, since all of the posterior second moments are *homoscedastic* in the data, by which we mean that they depend only upon the  $r+1$  sampling design parameters  $(n_1, r)$ , and not upon the actual data values. For instance, for all data  $\mathcal{D}$ :

$$(1.9) \quad \mathcal{V}\{\tilde{\mu}_1|\mathcal{D}\} = g(1 - z_1) \quad ; \quad \mathcal{V}\{\tilde{\varphi}|\mathcal{D}\} = h(1 - z_0) \quad ;$$

and similarly for the predictive variance of the observables (see Section 4).

This is, we believe, a serious limitation of the normal-normal-normal model, since we would expect in a more general setting to be able to learn about any unknown variances from sufficient portfolio data. In what follows, we present our attempts to generalize (1.5) in order to obtain

heteroscedastic forecast formulae, while retaining the simplicity of the credible mean forecast (1.6). The resulting model turns out to be a generalization of the classical unknown mean and precision normal model for the non-hierarchical case. The formulation also clarifies the numerical integration that would be necessary to extend the generalization further.

## 2. A Generalized Hierarchical Model

To obtain heteroscedastical results from the normal hierarchical model, we must permit the variances at each level to be random quantities. Specifically, for risk  $i$  at time  $t$ , we assume:

$$\begin{aligned}
 (\tilde{x}_{it} | \mu_i, \varphi, \omega, \gamma, \eta) &\sim (\tilde{x}_{it} | \mu_i, \omega) \sim No(\mu_i; \omega^{-1}) ; & (i=1, 2, \dots, r) \\
 (\tilde{\mu}_i | \varphi, \gamma, \eta) &\sim (\tilde{\mu}_i | \varphi, \gamma) \sim No(\varphi; \gamma^{-1}) ; & (t=1, 2, \dots) \\
 (\tilde{\varphi} | \eta) &\sim \tilde{\varphi} \sim No(m; \eta^{-1}) .
 \end{aligned}
 \tag{2.1}$$

where  $\tilde{\omega}$ ,  $\tilde{\gamma}$ , and  $\tilde{\eta}$  are new random precisions, corresponding to  $f^{-1}$ ,  $g^{-1}$ , and  $h^{-1}$ , respectively. As before, the variable length observed data,  $\mathcal{D} = \{x_{it}\}$ , is to be used to estimate all of the unknown parameters, now denoted by  $\tilde{\Theta} = (\tilde{\mu}, \tilde{\varphi}; \tilde{\omega}, \tilde{\gamma}, \tilde{\eta})$ , plus make predictions of future  $\tilde{x}_{it}$ .  $\tilde{\mu}$  and  $\tilde{\varphi}$  are the conditional mean parameters, as before. We let  $\tilde{\Omega} = (\tilde{\omega}, \tilde{\gamma}, \tilde{\eta})$  denote the group of unknown precision parameters; we assume that they are

statistically independent of the  $r+1$  mean parameters. Temporarily, we assume that the precisions are governed by some known prior joint density,  $p(\Omega)$ .

Using the statistics  $y_i$  in (1.7), and setting  $y_{ii} = (\sum x_{it}^2/n_i)$ , we find the joint density of the data and the mean parameters, conditional on the precisions, as:

$$(2.2) \quad p(\mathcal{D}, \underline{\mu}, \varphi | \Omega) = \left(\frac{\omega}{2\pi}\right)^{\sum n_i/2} \left(\frac{\gamma}{2\pi}\right)^{r/2} \left(\frac{\eta}{2\pi}\right)^{1/2} \times$$

$$\exp \left\{ -\omega \sum_i [n_i y_{ii} - 2\mu_i n_i y_i + n_i \mu_i^2] / 2 \right\} \times$$

$$\exp \left\{ -\gamma \sum_i [\mu_i - \varphi]^2 / 2 - \eta [\varphi - m]^2 / 2 \right\}.$$

We then extract the conditional densities of the mean parameters in the usual tedious way.

At the individual risk level, we find that the conditional joint density of the mean parameters,  $p(\underline{\mu} | \varphi, \Omega, \mathcal{D})$ , is a product of independent normal densities, with means equal to:

$$(2.3) \quad \mathcal{E}\{\tilde{\mu}_i | \varphi, \Omega, \mathcal{D}\} = f_i(\varphi, \Omega, \mathcal{D}) = [1 - z_i(\omega, \gamma)] \varphi + z_i(\omega, \gamma) y_i \quad .$$

and variances:

$$(2.4) \quad \mathcal{V}\{\tilde{\mu}_i | \varphi, \Omega, \mathcal{D}\} = (\gamma + \omega n_i)^{-1} = \frac{1}{\gamma} [1 - z_i(\omega, \gamma)].$$

for  $i=1, 2, \dots, r$ . In parallel with the simpler model, we have defined (conditional) individual credibilities:

$$(2.5) \quad z_i(\varphi, \Omega) = z_i(\omega, \tau) = \left( \frac{\omega n_i}{\tau + \omega n_i} \right) .$$

and (conditional) individual credibility forecasts  $f_i(\varphi, \Omega, \mathcal{D})$ , for all risks. Note that neither the  $z_i$  nor the variances depend upon the data  $\mathcal{D}$  nor the highest-level precision parameter  $\eta$ . Thus, we are back to the homoscedastic case, as expected.

Removing  $\mu$  from (2.2), and again completing the square, we find that the conditional density of the cohort (portfolio) mean,  $p(\varphi | \Omega, \mathcal{D})$ , is also normal, with a (conditional) cohort credibility forecast:

$$(2.6) \quad \mathcal{E}\{\tilde{\varphi} | \Omega, \mathcal{D}\} = f_0(\Omega, \mathcal{D}) = [1 - z_0(\Omega)] m + z_0(\Omega) Y_0(\Omega),$$

where, as before, we define a (conditional) cohort credibility factor:

$$(2.7) \quad z_0(\Omega) = \frac{\tau \sum z_i(\omega, \tau)}{\eta + \tau \sum z_j(\omega, \tau)} .$$

and the (conditional) credibility-weighted average observation:

$$(2.8) \quad Y_0(\Omega) = Y_0(\omega, \tau) = \frac{\sum z_i(\omega, \tau) y_i}{\sum z_j(\omega, \tau)} .$$

(We suppress the obvious dependence of  $Y_0$  upon  $\mathcal{D}$ .) Note that  $z_0$  does not depend upon the data, and depends upon  $\omega$  only through the  $(z_i)$ . The variance of the conditional cohort mean parameter is:

$$(2.9) \quad \mathcal{V}\{\tilde{\varphi} | \Omega, \mathcal{D}\} = \mathcal{V}\{\tilde{\varphi} | \Omega\} = (\eta + \tau \sum z_i(\omega, \tau))^{-1} = \frac{1}{\eta} [1 - z_0(\Omega)].$$

which is also homoscedastic, and depends upon  $\omega$  only through the  $(z_i)$ .

The remainder of (2.2) can now be matched with the still-general prior  $p(\Omega)$  to give the conditional posterior joint density of the precision parameters as:

$$(2.10) \quad p(\Omega|\mathcal{D}) \propto p(\Omega) \omega^{\sum n_i/2} \left\{ \prod_i [1-z_i(\omega, \gamma)]^{1/2} \right\} \left\{ [1-z_0(\Omega)]^{1/2} \right\} \times \\ \exp\{-A(\Omega, \mathcal{D})/2\} \quad .$$

where the function  $A$  is:

$$(2.11a) \quad A(\Omega, \mathcal{D}) = \omega \sum_i [n_i y_{ii} - n_i z_i(\omega, \gamma) y_i^2] + \eta \{ m^2 - [1-z_0(\Omega)]^{-1} f_0^2(\Omega, \mathcal{D}) \}.$$

By expansion, and the use of the new definitions:

$$(2.12) \quad N = \sum n_i \quad ; \quad \bar{z}(\omega, \gamma) = N^{-1} \sum n_i z_i(\omega, \gamma) \quad ;$$

$$Z(\omega, \gamma) = \sum z_i(\omega, \gamma) \quad ; \quad Y_{00} = N^{-1} \sum n_i y_{ii} \quad ;$$

the exponent can be put into several equivalent forms, such as:

$$(2.11b) \quad A(\Omega, \mathcal{D}) = \omega [N Y_{00} - \sum n_i z_i(\omega, \gamma) y_i^2] + \eta z_0(\Omega) [m - Y_0(\omega, \gamma)]^2 \\ - \gamma Z(\omega, \gamma) Y_0^2(\omega, \gamma) \quad .$$

or

$$(2.11c) \quad A(\Omega, \mathcal{D}) = N\omega \left\{ [Y_{00} - Y_0^2(\omega, \tau)] + N^{-1} \sum_i z_i(\omega, \tau) [Y_0^2(\omega, \tau) - y_i^2] \right. \\ \left. + [1 - \bar{z}(\omega, \tau)][1 - z_0(\Omega)][m - Y_0(\omega, \tau)]^2 \right\}.$$

We apologize for this heavy notation, but it is important for the special models to follow so that we know explicitly where the various unknown precisions occur. In any case, it should be obvious from (2.10) and any of the (2.11) formulas that there is no "magic" prior  $p(\Omega)$  that will give a tractable, closed-form posterior density,  $p(\Omega|\mathcal{D})$  !

### 3. Partially Unconditional Parameter Posteriors and Forecasts

Before progressing to various special cases, we first find several useful "marginal" densities, conditional only upon  $\Omega$  and  $\mathcal{D}$ .

Noting that the exponent of (2.2) is quadratic in  $\varphi$ , and that  $p(\underline{\mu}|\varphi, \Omega, \mathcal{D})$  is normal, we deduce that  $p(\underline{\mu}|\Omega, \mathcal{D})$  must be *multinormal*, with moments defined as, say:

$$(3.1) \quad \underline{\epsilon}\{\underline{\mu}|\Omega, \mathcal{D}\} = \underline{f}(\Omega, \mathcal{D}) \quad ; \quad \mathcal{V}\{\underline{\mu}|\Omega, \mathcal{D}\} = \Sigma(\Omega, \mathcal{D}) .$$

Unconditioning (2.3)(2.4) using (2.6)(2.9), we find, for the components of the mean vector:

$$(3.2) \quad [\underline{f}(\Omega, \mathcal{D})]_i = f_i(\Omega, \mathcal{D}) = [1 - z_i(\omega, \tau)] f_0(\Omega, \mathcal{D}) + z_i(\omega, \tau) y_i \quad ;$$

and, for the components of the covariance matrix:

$$(3.3) \quad [\Sigma(\Omega, \mathcal{D})]_{i,j} = \sigma_{ij}(\Omega, \mathcal{D}) =$$

$$\left\{ \begin{array}{ll} \gamma^{-1}[1-z_i(\omega, \gamma)] + \eta^{-1}[1-z_i(\omega, \gamma)]^2[1-z_0(\Omega)] & (i=j) \\ \eta^{-1}[1-z_i(\omega, \gamma)][1-z_j(\omega, \gamma)][1-z_0(\Omega)] & (i \neq j) \end{array} \right\} .$$

with  $i, j = 1, 2, \dots, r$ . (Henceforth, we shall omit obvious ranges on indices.) Again, we see that even these partially unconditioned covariances are still homoscedastic.

It is also important to be able to find the predictive densities for future values of the individual risk variables. Let  $\tilde{w}_{it} = \tilde{x}_{it}$  for any  $t > n_i$ . Since  $E\{\tilde{w}_{it} | \Theta\} = \mu_i$  and  $V\{\tilde{w}_{it} | \Theta\} = \omega^{-1}$ , for all  $i$  and all future  $t$ , it follows from the above that  $p(\underline{w} | \Omega, \mathcal{D})$  is also multinormal, with means:

$$(3.4) \quad E\{\tilde{w}_{it} | \Omega, \mathcal{D}\} = f_i(\Omega, \mathcal{D}) .$$

and covariances:

$$(3.5) \quad E\{\tilde{w}_{it}; \tilde{w}_{ju} | \Omega, \mathcal{D}\} = \left\{ \begin{array}{ll} \omega^{-1} + \sigma_{ii}(\Omega, \mathcal{D}) & (i=j) \quad (t=u) \\ \sigma_{ii}(\Omega, \mathcal{D}) & (i=j) \quad (t \neq u) \\ \sigma_{ij}(\Omega, \mathcal{D}) & (i \neq j) \end{array} \right\} .$$

for all  $t > n_i$  and  $u > n_j$ .

When considering the entire cohort, it may be useful to predict the total cohort risk sum,  $\sum_i \tilde{w}_{it}$ , or, more usually, the cohort-average future risk,  $\tilde{w}_0 = r^{-1} \sum_i \tilde{w}_{it}$  (for any future time  $t > \max(n_i)$ ). Defining a cohort-average credibility factor,

$$(3.6) \quad z_c(\Omega) = 1 - [1 - z_0(\Omega)][1 - r^{-1} \sum_i z_i(\omega, \tau)] = [1 + \eta(r\tau)^{-1}] z_0(\Omega) \quad .$$

we combine (3.2) and (2.6) to find:

$$(3.7) \quad E\{\tilde{w}_0 | \Omega, \mathcal{D}\} = f_c(\Omega, \mathcal{D}) = [1 - z_c(\Omega)] m + z_c(\Omega) Y_0(\omega, \tau) \quad .$$

and:

$$(3.8) \quad v\{\tilde{w}_0 | \Omega, \mathcal{D}\} = (r\omega)^{-1} + r^{-2} \sum \sum \sigma_{ij}(\Omega, \mathcal{D}) \quad .$$

Note that forecasting  $\tilde{w}_0$  is not the same as estimating  $\tilde{\varphi}$ , since the former is from a finite, possibly biased, sample of individual risks that is fixed once and for all. The reader may easily find the covariance between successive future values of  $\tilde{w}_0$ .

#### 4. The Constant Variance Model

As our first use of the above formulae, we note the obvious fact that setting:

$$(4.1) \quad \omega^{-1} = f \quad ; \quad \tau^{-1} = g \quad ; \quad \eta^{-1} = h \quad .$$

will give the full-distribution results for the simple hierarchical model described in Section 1. We get the simplification  $Y_o = y_1$ , and the various credibility factors become simply:

$$(4.2) \quad z_i = \frac{n_i}{n_o + n_i} \quad ; \quad z_o = \frac{\sum z_j}{r_o + \sum z_j} \quad ; \quad \bar{z} = r^{-1} \sum z_i \quad ; \quad z_c = \left[1 + \frac{r_o}{r}\right] z_o \quad .$$

where  $n_o = f/g$  and  $r_o = g/h$ .  $p(\underline{\mu}|\mathcal{D})$ ,  $p(\varphi|\mathcal{D})$ ,  $p(\underline{w}|\mathcal{D})$ , and  $p(w_o|\mathcal{D})$  are multinormal, with means given by (2.3)(2.6) and (3.7), respectively.

Interpretation of these results will be given below in Section 9.

As mentioned previously, all of the variances and covariances of these rv's are homoscedastic, in the sense that they depend upon the sampling design parameters  $(n_i, r)$ , but not upon the actual data values in  $\mathcal{D}$ . For instance, (3.3) simplifies to:

$$(4.3) \quad \mathcal{C}\{\tilde{\mu}_i; \tilde{\mu}_j | \mathcal{D}\} = \sigma_{ij}(\mathcal{D}) = \left\{ \begin{array}{ll} g(1-z_i) + h(1-z_i)^2(1-z_o) & (i=j) \\ h(1-z_i)(1-z_j)(1-z_o) & (i \neq j) \end{array} \right\} .$$

### 5. A Special Non-Hierarchical Model

For our second special case, let us assume that knowledge about  $\tilde{\varphi}$  is "tight", by letting  $\tilde{\eta} \rightarrow \infty$ . This makes  $\tilde{\varphi} \rightarrow m$ , almost surely, and effectively removes the hierarchical nature of the model. Thus,

$\tilde{\mu}_i \sim N_0(m; \tau^{-1})$ , and the unknown parameters reduce to  $\Theta = (\mu; \omega, \tau)$ . Note also that, in the limit,  $z_0 \rightarrow 0$ , but that  $(\eta z_0) \rightarrow \tau \Sigma z_i(\omega, \tau)$ .

Furthermore, the  $z_i(\omega, \tau)$  still depend upon both unknown precisions, so that the conditional joint density (2.10) is still rather complicated.

However, by examining the individual credibility factors (2.5), we see that they, in fact, depend only upon the ratio  $\tau/\omega$  of the remaining two unknown precisions! Therefore, if we additionally assume that  $\tilde{\omega}$ , say, has a prior density  $p(\omega)$ , and that the ratio is fixed at some positive value  $n_0$ :

$$(5.1) \quad \tilde{\tau} = n_0 \tilde{\omega} .$$

we obtain a great simplification, in that the  $z_i = n_i/(n_0 + n_i)$  are now independent of the precision parameters, as are the forecasts  $f_i(\mathcal{D})$  in (3.2) (with  $f_0 = m$ ), and the sufficient statistic  $Y_0$  in (2.6) and (3.7). Thus,  $p(\underline{\mu} | \omega, \mathcal{D})$  is multinormal with means  $f_i(\mathcal{D})$  and simplified covariances:

$$(5.2) \quad \mathcal{V}\{\tilde{\mu}_i; \tilde{\mu}_j | \omega, \mathcal{D}\} = \sigma_{ij}(\omega, \mathcal{D}) = \begin{cases} (\omega n_0)^{-1} (1 - z_i) & (i=j) \\ 0 & (i \neq j) \end{cases} .$$

In other words, the  $(\tilde{\mu}_i | \omega, \mathcal{D})$  are conditionally independent. Similar results hold for the moments of the  $(\tilde{w}_{it})$  and for  $\tilde{w}_0$ , where now  $z_c = \tau^{-1} \Sigma z_i$ .

More importantly, (2.10) now simplifies to:

$$(5.3) \quad p(\omega | \mathcal{D}) = p(\omega) \omega^{N/2} \exp(-A(\omega, \mathcal{D})/2) .$$

with the exponent:

$$(5.4) \quad \Lambda(\omega, \mathcal{D}) = \omega N \left\{ [Y_{00} - Y_0^2] + (1-\bar{z})(1-z_0)(m-Y_0)^2 + N^{-1} \sum_i z_i (Y_0^2 - y_i^2) \right\} \\ = \omega N B(\mathcal{D}) \text{ , say.}$$

now a linear function of  $\omega$ ! Thus, unconditioning w.r.t.  $\omega$  can be carried out exactly, if we additionally assume that the natural conjugate gamma prior is used, that is:

$$(5.5) \quad p(\omega) = \frac{\beta^\alpha \omega^{\alpha-1} e^{-\beta\omega}}{\Gamma(\alpha)} \text{ .}$$

which we write as  $\tilde{\omega} \sim \text{Ga}(\alpha, \beta)$ .

It then follows that the posterior-to-data density is closed under sampling, and  $(\tilde{\omega}|\mathcal{D}) \sim \text{Ga}(\alpha', \beta')$ , with updated parameters:

$$(5.6) \quad \alpha' = \alpha + N/2 \quad ; \quad \beta' = \beta + NB(\mathcal{D})/2 \text{ .}$$

In this way, the completely unconditional densities  $p(\underline{u}|\mathcal{D})$ ,  $p(\underline{w}|\mathcal{D})$ , and  $p(\underline{w}_0|\mathcal{D})$  can be obtained as Student-t densities, and variances computed with the help of  $E\{\tilde{\omega}^{-1}|\mathcal{D}\} = \beta'/(\alpha' - 1)$ .

We say that  $\underline{u}$  has the  $r$ -dimensional multivariate Student-t density with  $\nu$  degrees of freedom, mean vector  $\underline{m}$ , and precision matrix  $\Psi$ , written  $\underline{u} \sim \text{St}_r(\nu; \underline{m}; \Psi)$ , if:

$$(5.7) \quad p(\underline{u}) = \frac{\Gamma((\nu + r)/2)}{(\pi\nu)^{r/2} \Gamma(\nu/2)} \left(\frac{\nu}{\nu-2}\right) |\Psi| [1 + (\nu-2)^{-1} (\underline{u}-\underline{m})' \Psi (\underline{u}-\underline{m})]^{-(\nu+r)/2} \text{ ,}$$

with, of course,  $E\{\tilde{\underline{u}}\} = \underline{m}$ , and  $\mathcal{V}\{\tilde{\underline{u}}\} = \Psi^{-1}$ . ( $r=1$  is the ordinary Student-t density.) The proof that  $p(\underline{\mu}|\mathcal{D})$  has this form follows from the fact that  $\omega$  does not enter into the  $f_i(\mathcal{D})$ ; it also follows that the  $(\tilde{\mu}_i|\mathcal{D})$  are still independent but with density  $St_1(\nu; m_i; \psi_{ii})$ , where:

$$(5.8) \quad E\{\tilde{\mu}_i|\mathcal{D}\} = m_i = f_i(\mathcal{D}) \quad .$$

$$(5.9) \quad \mathcal{V}\{\tilde{\mu}_i|\mathcal{D}\} = \psi_{ii}^{-1} = (n_0 + n_i)^{-1} E\{\tilde{\omega}^{-1}|\mathcal{D}\} \quad .$$

A separate analysis shows there are  $\nu = 2\alpha' = 2\alpha + N$  degrees of freedom.

Similarly,  $(\tilde{w}_{it}|\mathcal{D}) \sim St_1\left[(2\alpha + N) ; f_i(\mathcal{D}) ; [1 + (n_0 + n_i)^{-1}] E\{\tilde{\omega}^{-1}|\mathcal{D}\}\right]$ ,

for all  $t > n_i$ .

To see these results in a more familiar context, consider the case of one risk, in which:

$$(5.10) \quad Y_0 = y_1 \quad ; \quad Y_{00} = y_{11} \quad ; \quad Z = \bar{z} = z_1 \quad ; \quad N = n_1 \quad .$$

and the exponent factor in (5.4) becomes simply:

$$(5.11) \quad B(\mathcal{D}) = (y_{11} - y_1^2) + (1 - z_1)(m - y_1)^2 \quad .$$

Then, the posterior-to-data variance of any future value of  $\tilde{x}_{1t}$  can be predicted from the credibility formula:

$$(5.12) \quad \mathcal{V}\{\tilde{w}_{1t}|\mathcal{D}\} = \left[ \frac{n_0 + n_1 + 1}{2\alpha - 2 + n_1} \right] \left\{ \left[ \frac{2\alpha - 2}{n_0 + 1} \right] (1 - z_1)\nu_0 + z_1(1 - z_1)(m - y_1)^2 + z_1(y_{11} - y_1^2) \right\} .$$

where  $v_0 = \gamma\{\tilde{w}_{1t}\} = (1+n_0^{-1}) \xi\{\tilde{\omega}^{-1}\}$ . If we take the "natural" choice of  $\alpha = (n_0+3)/2$  (see remarks in Jewell & Schnieper [1985]), we obtain the well-known natural-conjugate results for the one-dimensional normal with unknown mean and variance. In other words, the results of this section can be viewed as the generalization of this simple individual risk model when collateral data is available.

#### 6. A General Hierarchical Model with All Variances Linked

We are now ready to tackle a general heteroscedastical hierarchical model. The success of the non-hierarchical model of the last section was due primarily to assumption (5.1), which removes the dependency of the  $z_i$  upon any random quantity. This suggests that an appropriate assumption in the hierarchical case would be to link all three precisions together, by assuming:

$$(6.1) \quad \tilde{\gamma} = n_0 \tilde{\omega} ; \tilde{\eta} = r_0 \tilde{\gamma} = r_0 n_0 \tilde{\omega} .$$

for appropriate positive  $n_0$  and  $r_0$ . This is equivalent to saying that we have very tight knowledge about how the total variance is split up among the three levels of the model, but the value of the total variance is unknown. This assumption not only simplifies the various credibility factors:

$$(6.2) \quad z_i = n_i / (n_0 + n_i) ; z_0 = Z / (r_0 + Z) ; \bar{z} = \sum n_i z_i / N ; z_c = (1 + \frac{r_0}{r}) z_0 .$$

but also makes the  $(f_i)$ ,  $f_0$ ,  $Y_0$ , and  $f_c$  independent of  $\tilde{\omega}$ ! Thus, the unconditioning on  $\omega$  required in the general formulae of Sections 2 and 3 reduces to simple expectations over  $(\tilde{\omega}|\mathcal{D})$ , with the possibility of using the natural conjugate Gamma prior of (5.5)! For completeness, we now display all of the final results, using notation and arguments developed previously.

### 6.1 Posterior Parameter Results

If  $p(\omega)$  is  $Ga(\alpha, \beta)$ , then from (2.11c) we see that  $p(\omega|\mathcal{D})$  is  $Ga(\alpha', \beta')$ , with updated parameters:

$$(6.3) \quad \begin{aligned} \alpha' &= \alpha + N/2 & ; & & \beta' &= \beta + NB(\mathcal{D})/2 & ; \\ B(\mathcal{D}) &= [Y_{00} - Y_0^2] + (1-\bar{z})(1-z_0)[m - Y_0]^2 + N^{-1} \sum n_i z_i [Y_0^2 - y_i^2] . \end{aligned}$$

similar to (5.4). The most important moment formula for our purposes is  $E\{\tilde{\omega}^{-1}|\mathcal{D}\} = \alpha' / (\beta' - 1)$ , which we note can be written:

$$(6.4) \quad E\{\tilde{\omega}^{-1}|\mathcal{D}\} = (1 - z_{**}) E\{\tilde{\omega}^{-1}\} + z_{**} B(\mathcal{D}) ,$$

where we have eliminated  $\beta$  in favor of the prior mean variance and have defined a variance credibility factor:

$$(6.5) \quad z_{**} = N / (N + 2(\alpha - 1)) .$$

From (2.6) and (2.9), we find the posterior-to-data moments of the portfolio mean to be:

$$(6.6) \quad \xi\{\tilde{\varphi}|\mathcal{D}\} = f_0(\mathcal{D}) = (1 - z_0) m + z_0 Y_0 \quad ;$$

$$(6.7) \quad \mathcal{V}\{\tilde{\varphi}|\mathcal{D}\} = \left[ \frac{1 - z_0}{r_0 n_0} \right] \xi\{\tilde{\omega}^{-1}|\mathcal{D}\} \quad .$$

In fact, because of the position of  $\omega$  in the simplified version of  $p(\varphi|\Omega, \mathcal{D})$ , we see that  $p(\varphi|\mathcal{D})$  must be a one-dimensional Student-t density, with  $2\alpha' = 2\alpha + N$  degrees of freedom, and the above mean and variance.

Progressing to the individual risk means, we use the results of Section 3 to find the individual mean forecasts as before:

$$(6.8) \quad \xi\{\tilde{\mu}_i|\mathcal{D}\} = f_i(\mathcal{D}) = (1 - z_i) f_0(\mathcal{D}) + z_i y_i \quad .$$

and the new covariance structure:

$$(6.9) \quad \xi\{\tilde{\mu}_i; \tilde{\mu}_j|\mathcal{D}\} = \sigma_{ij}(\mathcal{D}) = \left[ \frac{(1-z_i)}{n_0} \delta_{ij} + \frac{(1-z_0)(1-z_i)(1-z_j)}{r_0 n_0} \right] \xi\{\tilde{\omega}^{-1}|\mathcal{D}\} \quad .$$

where  $\delta_{ij}$  is the Kronecker delta-function, and the data values enter through the use of (6.3)-(6.5). Again, because of the normal-gamma distributional assumption, one can argue that  $p(\underline{\mu}|\mathcal{D})$  is a  $r$ -dimensional Student-t density, with  $2\alpha + N$  degrees of freedom.

## 6.2 Posterior Predictive Results

Passing to the forecasts of future values, (3.4) (3.5) give us simply:

$$(6.10) \quad \mathcal{E}\{\tilde{w}_{it} | D\} = f_i(D) = (1 - z_i) m + z_i y_i \quad .$$

$$(6.11) \quad \mathcal{V}\{\tilde{w}_{it} : \tilde{w}_{ju} | \mathcal{D}\} = \delta_{ij} \delta_{ju} \mathcal{E}\{\tilde{\omega}^{-1} | \mathcal{D}\} + \sigma_{ij}(\mathcal{D}) \quad .$$

From these, various multivariate Student-t densities can be generated "over future times" for a single risk, "over risks" at one future epoch, or for various mixtures of risks and epochs.

Finally, the average future risk for the entire cohort (3.7)(3.8) has moments:

$$(6.12) \quad \mathcal{E}\{\tilde{w}_o | \mathcal{D}\} = f_c(\mathcal{D}) = (1 - z_c) m + z_c Y_o \quad ;$$

$$(6.13) \quad \mathcal{V}\{\tilde{w}_o | \mathcal{D}\} = \left\{ \frac{1}{r} \left[ 1 + \frac{1}{n_o} \left( 1 - \frac{z}{r} \right) \right] + \left[ \frac{1-z_o}{r_o n_o} \right] \left[ 1 - \frac{z}{r} \right]^2 \right\} \mathcal{E}\{\tilde{\omega}^{-1} | \mathcal{D}\} \quad .$$

It follows also from arguments similar to those above that  $p(w_o | \mathcal{D})$  will be a Student-t density.

Any of the final (co)variance formulae can be put into a credibility-like form. For example, the formula corresponding to (5.12) for the total variance of a future observation of risk #1 can be written as:

$$(6.14) \quad \mathcal{V}\{\tilde{w}_{1t} | \mathcal{D}\} = \left[ \frac{r_o n_o + (1-z_1)r_o + (1-z_o)(1-z_1)^2}{r_o n_o + r_o + 1} \right] \times$$

$$\left\{ (1-z_m) \mathcal{V}\{\tilde{w}_{1t}\} + z_m \left[ \frac{r_o n_o + r_o + 1}{r_o n_o} \right] \left[ (Y_{oo} - Y_o^2) + (1-z)(1-z_o)(m - Y_o^2) + \right. \right.$$

$$\left. \left. N^{-1} \sum_i z_i (Y_o^2 - y_i^2) \right] \right\} \quad .$$

### 7. Similar Data Lengths

Notation simplifies dramatically if all data record lengths are the same, i.e.,  $n_i = n$  ( $i=1,2,\dots,r$ ). Then:

$$(7.1) \quad N=nr ; z_i = z = \bar{z} = n/(n_0 + n) ; Z=rz ; z_0 = rz/(r_0 + rz) ; z_{**} = nr/(nr + 2(\alpha - 1)).$$

and the overall statistics  $Y_0$  and  $Y_{00}$  can be replaced by the simpler versions:

$$(7.2) \quad y_0 = \Sigma y_i / r = \Sigma \Sigma x_{it} / nr ; y_{00} = \Sigma y_{ii} / r = \Sigma \Sigma x_{it}^2 / nr .$$

Then (6.3) simplifies to:

$$(7.3) \quad B(\mathcal{D}) = y_{00} - (1-z) y_0^2 - z(\Sigma y_i^2 / r) + (1-z)(1-z_0)(m-y_0)^2 .$$

For approximations and asymptotic studies, it is useful to eliminate  $y_0^2$  and  $\Sigma y_i^2 / r$  in favor of the statistics:

$$(7.4) \quad y_{0;0} = \Sigma y_{i;i} / r ; y_{i;i} = \left[ \frac{2}{n(n-1)} \right] \Sigma_{t < u} x_{it} x_{iu} ;$$

$$y_{**} = \left[ \frac{2}{r(r-1)} \right] \Sigma_{i < j} y_i y_j .$$

from which one can show that:

$$(7.5) \quad y_0^2 = \frac{1}{r} \left( \frac{1}{r} \Sigma y_i^2 \right) + \left( \frac{r-1}{r} \right) y_{**} ; \frac{1}{r} \Sigma y_i^2 = \frac{1}{n} y_{00} + \left( \frac{n-1}{n} \right) y_{0;0} .$$

$B(\mathcal{D})$  and the variances above can be manipulated into various useful forms. For instance,  $B(\mathcal{D})$  can be expressed solely in terms of the differences  $(y_{00} - y_{0;0})$  and  $(y_{0;0} - y_{\ast\ast\ast})$ , the various credibility factors, and the design parameters,  $n$  and  $r$ .

### 8. Single Unknown Variance

To illustrate the difficulties caused by more general precision structures, we consider the model analyzed by Berger [1985, 4.6], in which it is assumed (in our notation) that  $\omega = f^{-1}$  and  $\eta = h^{-1}$  are known with certainty, so that the unknown parameters are  $(\underline{\mu}, \underline{\varphi}; \gamma)$ . For simplicity, we assume all data lengths are equal to  $n$  (Berger has  $n = 1$ ).

Thus there are only two conditional credibility factors to consider:

$$(8.1) \quad z_1(\gamma) = \frac{n}{f\gamma+n} \quad ; \quad z_0(\gamma) = \frac{(rnh)\gamma}{(rnh+f)\gamma+n} \quad ;$$

plus the individual statistics  $(y_i, y_{ii})$  and the simplified portfolio statistics  $(y_0, y_{00})$ . Other factors in (2.12) are simply:  $N = nr$ ,  $\bar{z}(\gamma) = z_1(\gamma)$ , and  $Z(\gamma) = rz_1(\gamma)$ .

The posterior precision density (2.10) now becomes a one-dimensional formula that can be "simplified" to reveal the structure on  $\gamma$ :

$$(8.2) \quad p(\gamma|\mathcal{D}) = p(\gamma) \left[ \frac{f\gamma}{f\gamma+n} \right]^{r/2} \left[ \frac{f\gamma+n}{(rnh+f)\gamma+n} \right]^{1/2} \exp(-C(\gamma)/2) \quad ;$$

$$(8.3) \quad C(\gamma) = (rnf^{-1}) \left[ \left[ \frac{n}{f\gamma+n} \right] \left[ (y_0^2 - \Sigma y_i^2/r) + \left[ \frac{f\gamma}{(rnh+f)\gamma+n} \right] (m-y_0)^2 \right] \right] .$$

It is clear that no analytical prior will lead to a tractable posterior density, so that numerical methods will have to be used in this case. However, examination of the results of Section 3 shows that the essential work in finding the various posterior and predictive first and second moments will be in taking expectations of various powers of the two credibility factors, plus finding  $E\{\gamma^{-1}|\mathcal{D}\}$ . The situation is similar if either  $\tilde{\omega}$  or  $\tilde{\eta}$  is an unknown variance, except that, in the latter case, only  $z_0$  will be modified by the data.

### 9. Interpretation of Results

As previously mentioned, the various credibility results (the formulae with notation  $f(\mathcal{D})$ ) show that predictive means can be represented as linear convex combinations of a prior mean with a classical estimator for that mean. With special priors, this may even be true for an updated variance (6.4). As the number of data points increases, the credibility factor increases, so that we place more "credibility" in the classical estimator, until (in the simplest cases), the Bayes estimator is essentially the sampling-theory estimator. The rate at which this "learning" occurs depends upon a credibility "time constant" that describes how our prior uncertainty is split between observational variation and prior uncertainty in the mean parameter.

In the hierarchical model with known variances, (3.2) and (2.6) reveal that the individual mean forecasts,  $f_i(\mathcal{D})$ , are linear combinations between the classical estimators,  $y_i$ , and another credibility

forecast for the portfolio-level mean,  $\tilde{\varphi}$ .  $f_0(\mathcal{D})$  itself mixes a "universal" mean,  $m$ , with a portfolio-wide statistic,  $Y_0$ , according to a cohort learning curve,  $z_0$ , with a time constant that is the ratio of higher level variances. Note that the individual data  $(x_{it})$  from risk  $i$  all has the same unity weight over time in calculating the individual-level statistic,  $y_i$ , but that these then have differing weights,  $z_i$ , over risks in calculating the portfolio statistic,  $Y_0$ . Further, the  $z_i$  increase with increasing numbers of samples,  $n_i$ , but  $z_0$  increases only with increasing sum  $\sum z_j$  over the number of risks,  $r$ ! It can be shown that this appealing layer-by-layer compounding of credibility forecasts that depend only upon local sources of variation and lower-level credibility factors repeats itself in more general hierarchical models with many layers (Buhlmann & Jewell [1987]). An important practical remark is that, while the  $z_i$  approach unity with increasing  $n_i$ , this is not true at the portfolio level, since  $z_0$  remains less than unity for finite  $r$ . In other words, because our finite portfolio may represent a biased selection of possible  $(\tilde{\theta}_i)$ ,  $Y_0$  is not "fully credible" for  $\tilde{\varphi}$ , even with very large numbers of data points per risk.

In fact, it was examination of the credibility factors that revealed the essential simplification in the generalized hierarchical model (2.2). In order to keep the intuitive and appealing linear formulae for the predictive means, we must constrain the ratios of adjacent-level precisions so that the credibility time constants,  $n_0$  and  $r_0$ , will remain constant, so that only one precision is unknown! This has the very useful

byproduct of making the exponent  $A$  in (2.10)(2.11) linear in that precision, so that the  $\Gamma$  is a natural-conjugate prior density. The notation  $n_0$  and  $r_0$  suggests that the prior parameter densities behave as if prior information were equivalent to  $r_0$  risks each having  $n_0$  data samples with common value,  $m$ . Similar constraints are necessary in more complicated hierarchical models to keep the predictive means in credibility form; full details for the Lindley & Smith [1974] linear model will be in a forthcoming paper.

There are many other combinations of model and prior densities that have simple credibility forecasts using individual data. However, the author has been unable to extend these models to the hierarchical case and still retain the linear structure for a portfolio-level forecast using collateral data. I would be interested in hearing from others who have looked further into special structures for first- and second-moment forecasts in hierarchical models.

## REFERENCES

- Berger, J.O. (1985). *Statistical Decision Theory and Bayesian Analysis*. New York: Springer-Verlag 2nd ed.
- Bühlmann, H. (1967). Experience rating and credibility. *ASTIN Bulletin* 4, 199-207.
- Bühlmann, H. & Jewell, W.S. (1987). Hierarchical credibility revisited. *Bulletin Association of Swiss Actuaries* 1, 35-54.
- Jewell, W.S. (1974). Credible means are exact Bayesian for simple exponential families, *ASTIN Bulletin* 8, 77-90.
- Jewell, W.S. (1975a). Regularity conditions for exact credibility. *ASTIN Bulletin* 8, 336-341.
- Jewell, W.S. (1975b). The use of collateral data in credibility theory: a hierarchical model, *Giornale dell'Istituto Italiano degli Attuari* 38, 1-16.
- Jewell, W.S. & Schnieper, R. (1985). Credibility approximations for Bayesian prediction of second moments, *ASTIN Bulletin* 15, 103-121.
- Jewell, W.S. (1987). A general framework for second moment multi-dimensional credibility approximations, O.R. Center, U. California, Berkeley, in preparation.
- Jewell, W.S. & Bühlmann, H. (1987). Credibility approximations for second moments in hierarchical models, O.R. Center, U. California, Berkeley, in preparation.
- Lindley, D.V. & Smith, A.F.M. (1972). Bayes estimates for the linear model, *Journal Royal Statistical Society B*, 1-41.

END

10-87

DTIC