International Institute for Applied Systems Analysis
IIASA
www.iiasa.ac.at

The Use of Collateral Data in Credibility Theory: A Hierarchical Model

Jewell, W.S.

IIASA Research Memorandum
June 1975

RM-75-24

THE USE OF COLLATERAL DATA IN CREDIBILITY THEORY: A HIERARCHICAL MODEL

William S. Jewell

June 1975

The Use of Collateral Data in Credibility Theory: A Hierarchical Model

William S. Jewell*

Abstract

In classical credibility theory, a linearized Bayesian forecast of the fair premium for an individual risk contract is made using prior estimates of the collective fair premium and individual experience data. However, collateral data from other contracts in the same portfolio is not used, in spite of intuitive feelings that this data would contain additional evidence about the quality of the risk collective from which the portfolio was drawn. By using a hierarchical model, one makes the individual risk parameters exchangeable, in the sense of de Finetti, and a modified credibility formula is obtained which uses the collateral data in an intuitively satisfying manner. The homogeneous formula of Bühlmann and Straub is obtained as a limiting case when the hyperprior distribution becomes "diffuse".

0. Introduction

In the usual collective model of risk theory [1], the random variables generated by individual risks are assumed to be independent, once the individual risk parameters are known. However, a priori, only collective (portfolio) statistics are available, taken from a distribution which is mixed over a prior distribution of the parameter. We assume that unlimited statistics are available for the collective as a whole, and a limited amount of experience (sample) data for individual risks drawn at random from the collective.

---

*University of California, Berkeley, and International Institute for Applied Systems Analysis, Laxenburg, Austria

In classical credibility theory, we make a linearized Bayesian forecast of the next observation of a particular individual risk, using his experience data and the statistics from the collective; the resulting formula, which has been known in various forms for over fifty years, requires only the individual sample mean, and the first and second moments from the collective.

If one attempts to use collateral data from other risks in a credibility forecast of a certain individual risk, it turns out that this cohort data has zero weight, and is discarded in favor of the assumed-known collective statistics. This is essentially because the various individual risk parameters are assumed to be independent and representative samples from the prior distribution.

This result is disturbing to many analysts, who feel that data from other risks in the portfolio contains valuable collateral information about the collective. In several of their models, Bühlmann and Straub [3,4] argue that, since the (mixed) moments of the collective must be estimated anyway, a credibility forecast should be only in terms of cohort data. They achieve a partial result of this kind by using a proportional function of all experience data; this forces the use of cohort data into an estimate of the collective mean, but the second moment components are still required. In [12], the author describes a model in which the individual risk parameters were correlated through an "externalities" model; the resulting formula uses both cohort sample data and the first and second collective moments. In [18], Taylor describes a model in which the "manual premium" (collective mean) is itself a random variable, and also obtains a formula in which collateral data is used. Finally, we should mention that similar arguments are advanced about the use of cohort data in the otherwise unrelated "empirical Bayes" models [14, 16].

In this paper, we attempt a reconciliation of these approaches, based upon the ideas of hierarchical models [13,14, 15] and model identification [17,19]. Although we obtain results similar to those already described in [12], the justification is completely different, and, we believe, provides a more natural explication of the situations in which collateral data should be used.

1. The Basic Model

In the basic model of the collective, we imagine that individual risk contracts are characterized by a risk parameter, θ, which is drawn from a known prior density, $p(\theta)$. A cohort, or portfolio, of such contracts consists of a finite population $[\theta_1, \theta_2, \ldots, \theta_r]$, whose members are drawn independently from the same density.

Then, given θ_i, we suppose that we have likelihood densities, $p_i(x_{it}|\theta_i)$,[1] which govern the generation of n_i independent

[1]We adopt the usual convention that all densities are indicated by p(.), the arguments indicating the appropriate random variable(s). The random variables, themselves, are indicated where necessary by a tilde. Finally, to avoid complicated (continued)

and identical realizations of the risk random variable, $\tilde{x}_{it}(t = 1,2,\ldots,n_i)$. In other words, from the total portfolio, we have r <u>individual experience data records</u>, $\underline{x}_i = [x_{i1},x_{i2},\ldots,x_{in_i}]$, which, together, we refer to as the <u>total experience</u>, X. Note that each process is stationary over time, but that we (temporarily) permit the individual risks to have different distributions. In particular, we need to define the first two conditional moments:

$$m_i(\theta_i) = \mathscr{E}\{\tilde{x}_{it}|\theta_i\} \quad ; \qquad v_i(\theta_i) = \mathscr{V}\{\tilde{x}_{it}|\theta_i\} \ . \qquad (1.1)$$

Prior to the data, $p(\theta)$ is the same prior density for any arbitrary risk drawn from the collective; thus, a priori, we have the following average moments for risks of the i^{th} and j^{th} types:

$$m_i = \mathscr{E}\{\tilde{x}_{it}\} = \mathscr{E}\{m_i(\tilde{\theta}_i)\} \quad ; \qquad (1.2)$$

$$E_{ij} = \mathscr{E}\mathscr{C}\{\tilde{x}_{it};\tilde{x}_{ju}|\tilde{\theta}_i;\tilde{\theta}_j\} = \begin{cases} 0 & (i \neq j) \\ \mathscr{E}\{v_i(\tilde{\theta}_i)\} & (i = j) \end{cases} \qquad (1.3)$$

$$D_{ij} = \mathscr{C}\{m_i(\tilde{\theta}_i);\ m_j(\tilde{\theta}_j)\} = \begin{cases} 0 & (i \neq j) \\ \mathscr{V}\{m_i(\tilde{\theta}_i)\} & (i = j) \end{cases} \qquad (1.4)$$

---

1 (cont'd) subscripts, we define the multiple conditional expectation:

$$\mathscr{E}\mathscr{E}\mathscr{E}\{f(\tilde{a},\tilde{b},\tilde{c},)|\tilde{b}|\tilde{c}\}$$

as being the expectation of $f(a,b,c)$ using measure $p(a|b,c)$, followed by the expectation using measure $p(b|c)$, followed by the expectation using $p(c)$. Any of these arguments may be multiple, and other operators, such as variance, $\mathscr{V}$, and covariance, $\mathscr{C}$, may be used.

Note in particular that there are no covariances between risks i and $j \neq i$ for two reasons:

(i) assumed independence between $\tilde{x}_{it}$ and $\tilde{x}_{ju}$, given θ_i and θ_j;

(ii) assumed independence between $\tilde{\theta}_i$ and $\tilde{\theta}_j$.

The total prior-to-data covariance between individual risks is then:

$$\mathscr{C}\{\tilde{x}_{it};\tilde{x}_{ju}\} = \begin{cases} E_{ii} + D_{ii} & (i = j, t = u) \\ D_{ii} & (i = j, t \neq u) \\ 0 & (i \neq j) \end{cases} \quad . \tag{1.5}$$

The basic problem of credibility theory is to forecast the next observation, $\tilde{x}_{s,n_s+1}$, of a selected risk, s, given the total data from all risks, $X = [x_i|(i = 1,2,\ldots,r)]$, and using the linear function:

$$f_s(X) = a_o + \sum_{i=1}^{r} \sum_{t=1}^{n_i} a_{it}\, x_{it} \;, \tag{1.6}$$

in which the coefficients $(a_o;a_{it})$ are chosen so as to approximate the conditional mean $\mathscr{E}\{\tilde{x}_{s,n_s+1}|X\}$ in the least-squares sense, over all prior possible data records, $p(X)$.

The appropriate least-squares formulae have been presented elsewhere (see, e.g., [7,12]). It turns out, for the basic model described above, that:

(i) $a_{it} = a_i$ $(i = 1,2,\ldots,r)(t = 1,2,\ldots,n_i)$ because of the stationarity assumption;

(ii) $a_i = 0$ $(i \neq 0,s)$ because $D_{sj} = 0$ $(j \neq s)$, that is, $\tilde{\theta}_j$ and $\tilde{\theta}_s$ are independent.

Defining the i^{th} credibility factor, Z_i, and time constant, N_i, as:

$$Z_i = n_i/(n_i + N_i) \quad ; \qquad N_i = E_{ii}/D_{ii} \quad ; \tag{1.7}$$

and the i^{th} experience sample mean, $\bar{x}_i$, as:

$$\bar{x}_i = \frac{1}{n_i} \sum_{t=1}^{n_i} x_{it} \quad , \tag{1.8}$$

we obtain the final credibility forecast as:

$$f_s(X) = (1 - Z_s)m_s + Z_s\bar{x}_s + O(x_{i\neq s,t}) \quad . \tag{1.9}$$

Various interesting interpretations of this classical result are possible [7,8,12], and it is known that (1.9) is, in fact, the exact Bayesian conditional mean for a large and important class of prior and likelihood densities [9,10].

2. Objections and Previous Results

Two practical objections to the result (1.9) seem to be raised in the literature. The first is that three prior-to-data moments, m_s, E_{ss}, and D_{ss}, must be estimated from the collective for each risk which is forecast. Even in the more usual, identical-risk case, where $m_i = m$, $E_{ii} = E$, and $D_{ii} = D$, for all samples $i = 1,2,\ldots,r$, (1.9) provides no assistance in estimating the common moments. This concern is related to the second objection, namely, that there ought to be some use for the cohort data, $\{x_{i\neq s,t}\}$, since it is precisely from this data that one would attempt to form estimates of the first and second moments in actual practice. This <u>collateral data</u> ought, then, to be used either to form initial estimates of m, E, and D, or, in the case in which one had vague prior estimates of them, to somehow revise them as more portfolio-wide data becomes available. Notice that we are not talking about any problems of non-stationarity, such as inflation, or shifts in the risk environment, but just the vague notion that our collective might, in some way, be different from the initially-assumed statistics.

Bühlmann and Straub [3] were the first to point out that one can force all the data in X to be used by setting a_o in (1.6) equal to zero, and constraining the remaining coefficients to give a forecast which is unbiased, as in (1.9). For the simple model of the last section, in which the $\tilde{x}_{it}$ are not identically distributed, we obtain:

$$f_s(X) = (1 - Z_s) \left\{ m_s \frac{\sum_{i=1}^{r} \left(\frac{m_i}{D_{ii}}\right) Z_i \bar{x}_i}{\sum_{j=1}^{r} \left(\frac{m_j^2}{D_{jj}}\right) Z_j} \right\} + Z_s \bar{x}_s \quad . \qquad (2.1)$$

The term in braces, which used all the sample data, even that of risk s, is a substitute for m_s in (1.9); however, there is no simplification as far as collective moments to be estimated are concerned, since all the m_i, E_{ii}, and D_{ii} are used.

But in the important case where all risks are assumed to be identically distributed, for the same value of θ, (2.1) simplifies to:

$$f_s(X) = (1 - Z_s)\left\{\frac{\sum_{i=1}^{r} Z_i\bar{x}_i}{\sum_{j=1}^{r} Z_j}\right\} + Z_s\bar{x}_s \quad , \tag{2.2}$$

and now the forecast depends upon $Z_i = n_i/(n_i + N)$, with $N = E/D$ as a ratio between variance components which must be estimated from the collective. Of course, the forecast (2.2) must give a higher value to the mean-square error which was used to find (1.9).

If all data records are of the same length, $n_i = n$ and $Z_i = Z = n/(n + N)$, $(i = 1,2,\ldots,r)$, the surrogate for m_s in the braces in (2.2) becomes simply:

$$\sum_{i=1}^{r} \bar{x}_i/r = \sum_{i=1}^{r}\sum_{t=1}^{n} x_{it}/rn \quad , \tag{2.3}$$

the grand sample mean of all cohort data!

In some work on "related risk" models [12], the author assumed a situation in which the risk parameters $\underline{\tilde{\theta}} = [\tilde{\theta}_1,\tilde{\theta}_2,\ldots,\tilde{\theta}_r]$ are statistically <u>dependent</u>, with known joint prior. The only effect of this assumption is to introduce non-zero terms into the last line of (1.4), viz.:

$$D_{ij} = \mathscr{C}\{m_i(\tilde{\theta}_i);\ m_j(\tilde{\theta}_j)\} \tag{2.4}$$

for all i,j. If the underlying risk likelihoods are different, then a multidimensional credibility model [7,11] must be used with an $r \times r$ system of equations solved to find a matrix of credibility factors. However, in the important special case where the risks are identically distributed, given $\underline{\theta}$, $p(\underline{\theta})$ consists of exchangeable random variables, and there are only four collective moments, m, E, and, say, D_{11} and D_{12} for the cases in which $i = j$ and $i \neq j$, respectively, in (2.4). One may easily show that, with this correlation between risk parameters added, (1.9) becomes:

$$f_s(X) = (1 - Z_s) \left\{ \frac{(D_{11} - D_{12})m + D_{12} \sum_{i=1}^{r} Z_i \bar{x}_i}{(D_{11} - D_{12}) + \sum_{j=1}^{r} Z_j} \right\} + Z_s \bar{x}_s \ , \tag{2.5}$$

where the credibility factors now require a modified correlation time constant, N_{12}:

$$Z_i = n_i/(n_i + N_{12}) \quad ; \quad N_{12} = E/(D_{11} - D_{12}) \ . \tag{2.6}$$

As in (2.2), the expression in braces in (2.5) is an estimate for the mean m_s, which can be seen to be different from m, because of the non-representative way in which the cohort of r risks may have been selected. As the correlation between the parameters vanishes, $D_{12} \to 0$, $D_{11} \to D$, and (2.5) reduces to the usual formula (1.9), with all the collateral data being thrown away.

Although this model is satisfactory from the mathematical point of view of explaining when cohort data would be used in a linear forecast, it does not show why there could be correlation in the collective, why the risk parameters should be exchangeable random variables, and under what conditions this correlation would be weak or strong. For this purpose, we need to extend the traditional model of the collective into a hierarchical model.

3. A Hierarchical Model

In our expanded model, the concepts of individual risk random variables, risk parameters, and a cohort of risks chosen from a collective are retained, but we imagine that our collective, the one under study, is not necessarily representative of other possible collectives which are drawn from some larger universe of collectives.

Formally, this means that there is a collective selection hyperparameter, $\tilde{\varphi}$, which describes how possible collectives may vary from one another, when chosen from some hyperprior density $p(\varphi)$. Once φ is chosen and the collective characteristics are defined, then the risk parameters $[\theta_i]$ are chosen for each of the r members of our cohort, independently, and identically distributed from a prior density $p(\theta|\varphi)$. Finally, the n_i experience samples for each individual risk i are drawn independently from a likelihood, $p_i(x_{it}|\theta_i,\varphi)$. Notice that the risk parameters and the individual risks are now independent only if φ is given; from the prior-to-selection-of-collective point of view, there is apparent correlation between cohort results because of the mixing on φ.

This somewhat abstract model has a very practical interpretation. Imagine an insurance company in which the individual risk is an individual insurance contract, and the collective is just a portfolio of similar coverages within our company. It is well recognized that portfolios vary from company to company, depending upon sales strategy, available customers, local risk conditions, etc.; our portfolio may be better or worse, than, say, the nationwide average. The universe of collectives, then, corresponds to the union of all possible risk contracts of this type in the nation, for which we may assume adequate statistics are available. Thus, in a hierarchical model, we hope to use nationwide statistics, together with all the data from our portfolio, not only to predict next year's fair premium for individual risks, but also to draw inferences about what kind of a portfolio we have.

For the development of a least-squares forecast, we start with the individual risk moments of $p(x_{it}|\theta_i,\varphi)$:

$$m_i(\theta_i,\varphi) = \mathscr{E}\{\tilde{x}_{it}|\theta_i,\varphi\} \quad ; \quad v_i(\theta_i,\varphi) = \mathscr{V}\{\tilde{x}_{it}|\theta_i,\varphi\} \quad , \tag{3.1}$$

and, from the usual conditional arguments, form the universal-average mean of the i^{th} type:

$$M_i = \mathscr{E}\{\tilde{x}_{it}\} = \mathscr{E}\mathscr{E}\{m_i(\tilde{\theta}_i,\tilde{\varphi})|\tilde{\theta}_i|\tilde{\varphi}\} \quad . \tag{3.2}$$

The universal covariances, using the conditional independence properties described above, are:

$$\mathscr{C}\{\tilde{x}_{it};\tilde{x}_{ju}\} = \begin{cases} F_{ii} + G_{ii} + H_{ii} & (i = j, t = u) \\ G_{ii} + H_{ii} & (i = j, t \neq u) \\ H_{ij} & (i \neq j) \quad , \end{cases} \tag{3.3}$$

where

$$F_{ii} = \mathscr{E}\mathscr{E}\{v_i(\tilde{\theta}_i,\tilde{\varphi})|\tilde{\theta}_i|\tilde{\varphi}\} \quad , \tag{3.4}$$

$$G_{ii} = \mathscr{E}\mathscr{V}\{m_i(\tilde{\theta}_i,\tilde{\varphi})|\tilde{\theta}_i|\tilde{\varphi}\} \quad , \tag{3.5}$$

and

$$H_{ij} = \mathscr{C}\left\{\mathscr{E}\{m_i(\tilde{\theta}_i,\tilde{\varphi})\,|\,\tilde{\varphi}\} \quad ; \mathscr{E}\{m_j(\tilde{\theta}_j,\tilde{\varphi})\,|\,\tilde{\varphi}\}\right\} \quad . \tag{3.6}$$

Several remarks are in order. From one point of view, what we have done is to introduce correlation between risk parameters of members of the same collective, for on comparing the above with (1.5) as modified by (2.4), we get the formal equivalences:

$$E_{ii} \equiv F_{ii} \quad ; \quad D_{ii} \equiv G_{ii} + H_{ii} \quad ; \quad D_{ij} \equiv H_{ij} \quad (i \neq j) \quad . \tag{3.7}$$

However, the interpretation is completely different, as we have seen.

The second observation is that is might seem worth while to decouple the $\tilde{x}_{it}$ from $\tilde{\varphi}$, and make the likelihood only dependent upon $\tilde{\theta}_i$; this might simplify some of the computations above, but does not diminish the number of individual prior-to-selection-of-collective moments needed.

However, in the important special case where the individual risk contracts are similar, giving identical likelihoods, given θ_i and φ, it can be seen that only four moments remain: M, F, G, and H. These may be interpreted in terms of our simpler model by noticing that it is as if the moments of Section 1 had a hidden dependence upon an unknown parameter φ. Calling those moments, then, $m(\varphi)$, $E(\varphi)$, and $D(\varphi)$, we see that the universal moments are equivalent to:

$$M = \mathscr{E}m(\tilde{\varphi}) \quad ; \mathscr{E}F = \mathscr{E}E(\varphi) \quad ; \quad G = \mathscr{E}D(\varphi) \quad ; \quad H = \mathscr{V}m(\tilde{\varphi}) \quad . \tag{3.8}$$

In other words, M, F, and G are universe-averaged versions of our previous m, E, and D. H, however, is new, and represents the variance of the fair premium over all possible collectives.

4. Universal Forecasts

Continuing with the important special case of identical risk distributions, it follows easily from least-squares theory and the above definitions that the optimal credibility forecast for the hierarchical model is:

$$f_s(X) = (1 - Z_s) \left\{ \frac{GM + H \sum_{i=1}^{r} Z_i \bar{x}_i}{G + H \sum_{j=1}^{r} Z_j} \right\} + Z_s \bar{x}_s \qquad (4.1)$$

where now a new universal time constant, N_U, appears in the credibility factors:

$$N_U = F/G \quad ; \quad Z_i = n_i/(n_i + N_U) \quad . \qquad (4.2)$$

Alternatively, we can get (4.1) from (2.5) and (3.7).

Following an idea of Taylor for his model [18], we note that (4.1) can be split into two parts:

$$f_s(X) = (1 - Z_s)\, \hat{M}(X) + Z_s \bar{x}_s \quad ; \qquad (4.3)$$

$$\hat{M}(X) = \frac{GM + H \sum_{i=1}^{r} Z_i \bar{x}_i}{G + H \sum_{j=1}^{r} Z_j} \quad . \qquad (4.4)$$

The second formula may be regarded as a revision of the "prior expected manual premium", M, using the experience data of all members of the cohort to obtain an "adjusted manual premium", $\hat{M}(X)$. This revised manual premium is then used in an ordinary credibility formula with the appropriate individual credibility factor, Z_s, for the forecast risk s.

The credibility revision of the universal mean (4.4) depends in a complicated manner upon the amount of data from each risk. However, if all data records are of the same length n, then $Z_i = Z = n/(n + N_U)$ for all i, and (4.4) can be rewritten:

$$M(X) = (1 - Z_C)M + Z_C\left(\frac{1}{r} \sum_{i=1}^{r} \bar{x}_i\right) , \tag{4.5}$$

where the collective credibility factor, Z_C, is:

$$Z_C = \frac{rnH}{F + nG + rnH} = \left[\frac{rH}{G + rH}\right]\left[\frac{n}{n + (F/G + rH))}\right] . \tag{4.6}$$

If rH is large compared to G, this function increases at first more rapidly than the common individual credibility factor Z, as n increases; however, Z_C has an asymptotic limit less than unity, so that (4.5) is not a credibility formula in the usual sense; that is, the grand sample mean is not ultimately "fully credible" for $m(\varphi)$.

This puzzling result can be explained by remembering that the risk parameters of the cohort $[\theta_i | i = 1,2,...,r]$, once picked, remain the same for all n. Therefore, if one estimates a fair premium for an arbitrary new member of the portfolio, say, with risk parameter θ_{r+1}, then there remains the possibility that the cohort sample is biased. Thus Z_C does not approach unity with increasing n, unless $rH >> G$, which means that a large enough portfolio contains a representative sample of risk parameters. This effect is not important in our estimate of $\tilde{x}_{s,n+1}$ because of the factor $(1 - Z_s)$ in (4.1).

If, on the other hand, we did wish to estimate the fair premium averaged over the current portfolio:

$$\mathscr{E}\left\{\frac{1}{r} \sum_{i=1}^{r} \tilde{x}_{i,n+1} | X\right\} ,$$

then one can show that (4.5) is still correct if a different credibility factor,

$$Z_o = (nG + rnH)/(F + nG + rnH) , \qquad (4.7)$$

is used; this does approach unity with increasing n.

5. Limiting Cases

The time constant $N_U = F/G$ is just the universe-average version of the classical Bühlmann time constant $N = E/D$, so that (4.3) is in a certain sense similar to (1.9). However, the factor $H = \mathscr{V}m(\tilde{\varphi})$ is completely new, and it is interesting to examine limiting cases.

If $H \to 0$, then we may say that all collectives are representative samples from the rather narrow universe of collectives in which there is little variance in fair premium. Thus, $M \to m$, $G \to D$, $N_U \to N$, and $Z_C \to 0$. No updating of the fair premium is necessary from the collateral data, and (4.3)-(4.4) reduce to the classical model (1.9).

On the other hand, if $H \to \infty$, this means that collectives are drastically different from one another, or in Bayesian language, we have a "diffuse prior" on $m(\tilde{\varphi})$. Then from (4.4) or (4.6), we see that, whenever there is cohort data, it is "fully credible" for $m(\tilde{\varphi})$, and (4.1) reduces to the Bühlmann-Straub proportional forecast (2.2)!

The same effect occurs in (4.6) as $r \to \infty$, but for a different reason: the grand sample mean of X is almost surely the correct mean, $m(\varphi)$, for our collective, and thus M is eliminated.

6. Approximation Error

The value of any forecast must be judged in terms of the mean-square error:

$$I = \mathscr{E}\left\{[\tilde{x}_{s,n_s+1} - f_s(\tilde{X})]^2\right\} \quad . \tag{6.1}$$

A certain portion of this error is due to individual fluctuation, and cannot be removed; the remainder is essentially an approximation error between the chosen forecast and the optimal Bayesian forecast, $\mathscr{E}\{\tilde{x}_{s,n_s+1}|X\}$. (See, e.g.,[12].) We now examine the mean-square error for several of the forecasts suggested previously.

The first and simplest possibility is to take the universal mean, $f_s(X) = M$, as an estimator. Then:

$$I_1 = F + G + H \quad , \tag{6.2}$$

that is, no component of variance is removed.

The second possibility, suggested by the surrogate for the collective mean in (2.2), is to take the credibility-weighted mean of all cohort data, $f_s(X) = \Sigma Z_i \bar{x}_i / \Sigma Z_j$, giving:

$$I_2 = F + G\left[1 + \frac{(1 - 2Z_s)}{\Sigma Z_j}\right] , \tag{6.3}$$

which removes the fluctuation component H, but may increase the second term for $Z_s < \frac{1}{2}$.

A third collective-wide possibility which has already been justified is the "adjusted manual premium", $\hat{M}(X)$, in (4.4), for which:

$$I_3 = F + G + H\left[\frac{G(1 - 2Z_s)}{G + H\Sigma Z_j}\right] . \tag{6.4}$$

Turning now to forecasts which use the data from the individual risk in a special way, we could use the Bühlmann-Straub homogenous formula (2.2), giving:

$$I_4 = F + G(1 - Z_s)\left[1 + \frac{(1 - Z_s)}{\Sigma Z_j}\right] . \tag{6.5}$$

Also of interest would be an individual forecast in which the cohort data is ignored, (1.9):

$$I_5 = F + G(1 - Z_s) + H(1 - Z_s)^2 . \tag{6.6}$$

Finally, we have the variance when the optimal universal forecast (4.1) is used:

$$I_6 = F + G(1 - Z_s) + H\left[\frac{G}{G + H\Sigma Z_j}\right](1 - Z_s)^2 . \tag{6.7}$$

Notice that none of the forecasts removes F; this is the irreducible variance component. Comparison of different forecasts depends in general upon the values of G, H, and the credibility factors; for example, one cannot say that I_2 is uniformly better than I_1.

The following relationships do hold, however, for all values of the coefficients:

$$I_6 < I_3 < I_2 \quad ;$$

$$I_6 < I_4 < I_2 \quad ;$$

$$I_6 < I_5 < I_1 \quad .$$

This effectively removes I_1 and I_2 from the second-rank contenders, after the optimal forecast I_6.

The Bühlmann-Straub formula, I_4, would seem to have special appeal because of the fact that H is removed completely. However, $I_6 < I_4$ always; and when $H \to \infty$, I_6 approaches a finite limit as well. Conversely, the classical individual credibility mean-square error, I_5, continues to increase as the universal prior becomes more diffuse, and this is the basic justification for including the cohort data.

7. Normal Hierarchical Family

A special case of interest is when all densities discussed in Section 3 are normal. If N(a,b) refers to the normal density with mean a and variance b, then by setting:

$$p(x_{it}|\theta_i,\varphi) = N(\theta_i,F) \quad ; \quad p(\theta_i|\varphi) = N(\varphi,G) \quad ; \quad p(\varphi) = N(M,H) \quad , \tag{7.1}$$

we find that the universal forecast (4.1) is exactly the Bayesian conditional mean $\mathscr{E}\{\tilde{x}_{s,n_s+1}|X\}$.

Further, the adjusted manual premium, $\hat{M}(X)$ (4.4), is $\mathscr{E}\{\tilde{\varphi}|X\}$. The joint distribution $p(\underline{\theta}|X)$, as well as $p(\phi|X)$, are both normal, and their precision matrices may be found by elementary calculations.

8. Related Work

A linear Bayesian model which is hierarchical in form has been given by Lindley and Smith [13,14,15]. In this model, $\underline{\tilde{x}}$, $\underline{\tilde{\theta}}$, and $\underline{\tilde{\varphi}}$ are random vectors for which $\mathscr{E}\{\underline{\tilde{x}}|\underline{\tilde{\theta}},\underline{\tilde{\varphi}}\} = A_1\underline{\tilde{\theta}}$, and $\mathscr{E}\{\underline{\tilde{\theta}}|\underline{\tilde{\varphi}}\} = A_2\underline{\tilde{\varphi}}$, A_1 and A_2 being matrices of appropriate dimension. The underlying distributions are all assumed to be multinormal, with $\mathscr{E}\{\underline{\tilde{\varphi}}\}$ and the covariances assumed to be known constants. When specialized to our model, results similar to Section 7 are obtained.

In [18], Taylor develops a credibility model in which the "manual premium", m, is revised according to "the average actual claim amount per unit risk in the entire collective in the year of experience". His assumptions are different from ours, in that m "has a prior distribution at the beginning of the year of experience", but "for fixed m, each $m(\theta_i)$ is fixed" (in our notation). I interpret this as saying, in effect, that there is a hidden parameter, φ, which is still left in $m = m(\varphi)$, after averaging over the θ_i. However, I have been unable to further relate the two models, and his formulae have the disadvantage that, as "the prior distribution on m" becomes degenerate, his forecast does not reduce to the usual credibility formula.

9. Conclusion

In conclusion, we mention that our hierarchical model implies that the joint distribution of the risk parameters at the level of the insurance company is:

$$p(\underline{\theta}) = \int \prod_{i=1}^{r} p(\theta_i|\varphi)\ p(\varphi)\ d\varphi \quad ,$$

which is equivalent to assuming that the risk parameters are exchangeable random variables. This powerful concept, due to de Finetti [5,6], is a natural modelling assumption for problems in which a random sample generates a finite population whose members are distinguishable only by their indices, as in our selection of a portfolio from an abstract collective. [14], Section 6, and [15] contain further discussions of the applicability of exchangeability. In a certain sense, what our model does is to use exchangeability to introduce correlation among the cohort θ_i, in the same way that a Bayesian prior introduces correlation among successive individual samples. In both cases, this prior correlation vanishes as the actual values of φ and $\underline{\theta}$ become identified.

G. Ferrara once asked how credibility experience rating could be used in a company where there are no prior statistics. By referring the prior estimation problem to a higher level of data collection, and by using all the experience data generated by the company's contracts as one learns about the actual portfolio quality, we believe that the model developed here goes a long way towards answering this question.

References

[1] Bühlmann, H. Mathematical Methods in Risk Theory, Springer-Verlag, Berlin, 1970.

[2] Bühlmann, H. "Experience Rating and Credibility", ASTIN Bulletin, 4, Part 3, (July, 1967), pp. 199-207.

[3] Bühlmann, H. and Straub, E. "Glaubwürdigkeit für Schadenssätze", Mitteilungen der Vereinigung Schweizerischer Versicherungsmathematiker, 70 (1970), pp. 111-113. Trans. C.E. Brooks, "Credibility for Loss Ratios," ARCH 1972.2.

[4] Bühlmann, H. "Credibility Procedures", Sixth Berkeley Symposium on Mathematical Statistics, (1971), pp. 515-525.

[5] De Finetti, B. Probability, Induction, and Statistics, J. Wiley, New York (1972), p. 266.

[6] De Finetti, B. Theory of Probability, Vol. II, J. Wiley, New York (to appear).

[7] Jewell, W.S. "Multi-Dimensional Credibility", ORC 73-7, Operations Research Center, University of California, Berkeley, (April, 1973). To appear, Journal of Risk and Insurance.

[8] Jewell, W.S. "The Credible Distribution", ORC 73-13, Operations Research Center, University of California, Berkeley, (August, 1973). ASTIN Bulletin, VII, Part 3 (March, 1974), pp. 237-269.

[9] Jewell, W.S. "Credible Means are Exact Bayesian for Simple Exponential Families", ORC 73-21, Operations Research Center, University of California, Berkeley, (October, 1973). ASTIN Bulletin, VIII, Part 3 (September, 1974), pp. 77-90.

[10] Jewell, W.S. "Regularity Conditions for Exact Credibility", ORC 74-22, Operations Research Center, University of California, Berkeley, (July, 1974). To appear, ASTIN Bulletin.

[11] Jewell, W.S. "Exact Multidimensional Credibility", ORC 74-14, Operations Research Center, University of California, Berkeley, (May, 1974). Mitteilungen der Vereinigung schweizerischer Versicherungsmathematiker, Band 74, Heft 2 (1974), pp. 193-214.

[12] Jewell, W.S. "Model Variations in Credibility Theory", ORC 74-25, Operations Research Center, University of California, Berkeley, (August, 1974). To appear Proceedings of Actuarial Research Conference on Credibility Theory, Berkeley, Sept., 1974. Academic Press, New York.

[13] Lindley, D.V. "Bayesian Least Squares", Bull. Inst. Internat. Statist., 43, 2, pp. 152-153.

[14] Lindley, D.V. Bayesian Statistics, A Review, Regional Conferences Series in Applied Mathematics, SIAM, Philadelphia, 83 pp.

[15] Lindley, D.V. and Smith, A.F.M. "Bayes Estimates for the Linear Model", Jour. Roy. Statist. Soc., B, 34, (1972), pp. 1-41.

[16] Maritz, J.S. Empirical Bayes Methods, Methuen, London (1970), 159 pp.

[17] Smallwood, R.C. "A Decision Analysis of Model Selection", IEEE Trans. on Systems Science and Cybernetics, SSC-4, 3 (Sept. 1969), pp. 333-342.

[18] Taylor, G.C. "Experience Rating with Credibility Adjustment of the Manual Premium", ASTIN Bulletin, 7, Part 3 (March, 1974).

[19] Wood, E.F. "A Bayesian Approach to Analyzing Uncertainty Among Stochastic Models", RR-74-16, International Institute for Applied Systems Analysis, Laxenburg, Austria (Sept., 1974).