RIA-77-U1082

A ⎮ EDIBILITY THEORY

by

WILLIAM S. JEWELL

19970807 025

DTIC QUALITY INSPECTED 3

OPERATIONS
RESEARCH
CENTER

UNIVERSITY OF CALIFORNIA · BERKELEY

A SURVEY OF CREDIBILITY THEORY

Operations Research Center Research Report No. 76-31

William S. Jewell

October 1976

U. S. Army Research Office - Research Triangle Park

DAAG29-77-G-0040

Department of Industrial Engineering
and Operations Research
University of California, Berkeley

SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER ORC 76-31·	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE *(and Subtitle)* A SURVEY OF CREDIBILITY THEORY		5. TYPE OF REPORT & PERIOD COVERED Research Report
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) William S. Jewell		8. CONTRACT OR GRANT NUMBER(s) DAAG29-77-G-0040
9. PERFORMING ORGANIZATION NAME AND ADDRESS Operations Research Center University of California Berkeley, California 94720		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS P-14240-M
11. CONTROLLING OFFICE NAME AND ADDRESS U. S. Army Research Office P.O. Box 12211 Research Triangle Park, North Carolina 27709		12. REPORT DATE October 1976
		13. NUMBER OF PAGES 39
14. MONITORING AGENCY NAME & ADDRESS*(if different from Controlling Office)*		15. SECURITY CLASS. *(of this report)* Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT *(of this Report)*

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT *(of the abstract entered in Block 20, if different from Report)*

18. SUPPLEMENTARY NOTES

19. KEY WORDS *(Continue on reverse side if necessary and identify by block number)*

Forecasting
Bayesian Prediction
Estimation
Regression
Risk Models

20. ABSTRACT *(Continue on reverse side if necessary and identify by block number)*

(SEE ABSTRACT)

DD $_{1\ JAN\ 73}^{FORM}$ 1473   EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-LF-014-6601

ABSTRACT

Credibility theory is the name given by American actuaries to
linear estimation formulae developed to experience-rate insurance
premiums. These formulae can be viewed as linear Bayesian Forecasts
of a conditional mean, exact under certain conditions, and best least-
squares approximations otherwise. This paper surveys the recent
theoretical developments in the actuarial literature, relates these
results to other linear estimation methods, and describes a variety
of special models and applications.

A SURVEY OF CREDIBILITY THEORY

by

William S. Jewell

1. INTRODUCTION

Credibility theory is the name given by American actuaries to heuristic
linear estimation formulae developed in the 1920's for insurance rate-making
problems. These results and their recent extensions are not only useful in
practice, but have interesting relationships with other estimating and
forecasting methods, such as the classical formula for the combination of
observations due to Gauss, maximum-likelihood estimators, Bayesian esti-
mation, and linear filter theory. Credibility forecasts can be viewed as
linear Bayesian forecasts of a conditional mean, exact under certain con-
ditions, and best least-squares approximations otherwise.

Many new theoretical results and special models have appeared in the
actuarial literature; credibility theory was the theme of a recent actu-
arial research conference [38]. In this paper, we shall survey these
results, and relate them to linear estimation results from other fields.
Other noninsurance application of credibility will also be described.

2. THE BASIC CREDIBILITY ·FORMULA

In the original insurance *experience rating* problem which give rise to credibility, we consider a *collective* of similar but somewhat heterogeneous insurance contracts which are grouped together to "spread the risk." It is assumed that detailed prior statistics are available from this pool; in particular, the *manual* or *collective fair premium*, $m = E\{\tilde{x}\}$, is the average value of the risk random variable of interest, $\tilde{x}$, such as number of accidents per year, total dollar claims per unit exposure, etc.

Now suppose a new insurance contract of unknown risk characteristics is underwritten, and assigned to this pool. At the beginning, the *individual fair premium* charged would be just the collective premium m ; however, as n years' *individual experience data* $[x_1, x_2, \ldots, x_n]$ is obtained on this risk, it seems reasonable that the *individual sample mean*, $\bar{x} = \sum x_t/n$, would tend to reflect more nearly the risk characteristics of the individual, except for the large variability in $\bar{x}$ with small n .

Using heuristic reasoning on the pooling of data (and considering only the number of claims per year), the early actuarial literature argued for an *experience-rated fair premium* for next year's risk, $\tilde{x}_{n+1}$, of the form

$$(2.1) \qquad E\{\tilde{x}_{n+1} \mid x_1, x_2, \ldots, x_n\} \approx (1 - Z)m + Z\bar{x} ,$$

with

$$(2.2) \qquad Z = \frac{n}{n + N} .$$

Z was called the *credibility factor*; it mixes the manual premium, m , and the experience premium, $\bar{x}$, with increasing "credibility" attached to the latter as n increases. The time constant N was essentially

determined by trial and error for different types of insurance. This credibility formula was successfully used in American casualty insurance rate-making for more than 50 years, with innumerable variation and elaboration. A survey, with references, may be found in Longley-Cook [41]; see also [22] and [23].

3. THE BAYESIAN APPROACH

The modern development of credibility theory begins with the resurgence of interest in Bayesian ideas in the 1950's, and with the works of Bailey [2], [3] and Mayerson [42], who showed that the experience-rating problem could be formulated as finding a Bayesian conditional mean, as already implied by the notation in (2.1).

Let each member of the risk collective be characterized by a (scalar- or vector-valued) *risk parameter* $\tilde{\theta}$; the heterogeneity of the collective is then described by a *prior density* $u(\theta)$, from which each risk draws an independent sample. Given θ , the distribution of an individual's risk variable for one year, $\tilde{x} = x$, is given by a *likelihood density*, $p(x \mid \theta)$; on an individual basis, the fair premium is

$$(3.1) \qquad m(\theta) = E\{\tilde{x} \mid \theta\} = \int x p(x \mid \theta) dx \; ,$$

and the individual variance is

$$(3.2) \qquad v(\theta) = V\{\tilde{x} \mid \theta\} = \int (x - m(\theta))^2 p(x \mid \theta) dx \; .$$

The pooled statistics from the collective of risks, however, have a *mixed collective density* $p(x) = Ep(x \mid \tilde{\theta})$, and this implies the collective fair premium is:

$$(3.3) \qquad m = E\{\tilde{x}\} = Em(\tilde{\theta}) \; ,$$

with total *collective variance*

$$(3.4) \qquad v = V\{\tilde{x}\} = E + D \; ; \; E = Ev(\tilde{\theta}) \; ; \; D = Vm(\tilde{\theta}) \; .$$

Using standard Bayesian arguments, the exact experience-rated fair

premium is:

$$E\{\tilde{x}_{n+1} \mid x_1, x_2, \ldots, x_n\} = E\{m(\tilde{\theta}) \mid x_1, x_2, \ldots, x_n\}$$

(3.5)
$$= \int m(\theta) \left[\frac{\prod\limits_{t=1}^{n} p(x_t \mid \theta) u(\theta)}{\prod\limits_{t=1}^{n} p(x_t \mid \alpha) u(\alpha) d\alpha} \right] d\theta ,$$

where we have assumed that each successive year's experience is independent, for a given (constant) θ . The term in square brackets is the posterior-to-data density of $\tilde{\theta}$ for this risk.

Bailey and Mayerson showed that the exact result (3.5) could be rearranged into the credibility form (2.1) for the special prior-likelihood combinations: Beta-Binominal, Gamma-Poisson, Gamma-Exponential, and Normal-Normal (known variances). m was calculated by (3.3), and N in (2.2) was a function of the *hyperparameters* of the prior, $u(\theta)$.

6

4. LEAST-SQUARES

The next step in the development of credibility was through least-squares theory. Suppose we have a vector-valued random variable $\underset{\sim}{\mathbf{x}}$ from whose observations $\underline{x}$ we are trying to predict a scalar random variable $\tilde{y}$ through a forecast function $f(\underline{x})$. Assuming we know the joint distribution $P(y,\underline{x}) = Pr\{\tilde{y} \leq y; \underset{\sim}{\tilde{x}} \leq \underline{x}\}$, the classical means of evaluating any f is the mean-square error norm;

$$(4.1) \qquad I = \int (y - f(\underline{x}))^2 dP(y,\underline{x}) \ .$$

It is known that the integrable function f^o which minimizes (4.1) at value I^o is the conditional mean:

$$(4.2) \qquad f^o(\underline{x}) = E\{\tilde{y} \mid \underset{\sim}{\tilde{x}} = \underline{x}\} \ .$$

In Bayesian terminology, the conditional mean minimizes quadratic Bayes' risk.

In many cases the exact conditional calculation (3.5) is too difficult and an approximate forecast function is sought. Since completion of the square shows that

$$(4.3) \qquad \begin{aligned} I &= I^o + \int (f^o(\underline{x}) - f(\underline{x}))^2 dP(\underline{x}) \ , \\ I^o &= EV\{\tilde{y} \mid \underset{\sim}{\tilde{x}}\} = V\{\tilde{y}\} - Vf^o(\underset{\sim}{\tilde{x}}) \ , \end{aligned}$$

then any approximate forecast f can be evaluated in terms of a least-squares fit to the conditional mean over the observation space.

A typical choice of an approximate forecast is a *linear function*

$$(4.4) \qquad f(\underline{x}) = a_o + \sum a_j x_j \ ,$$

where the parameters are adjusted to minimize (4.1) or (4.3). It is well

known that the optimal value $\underline{a}^*$ of the vector $\underline{a} = [a_j \mid j \neq 0]$ is given by the "normal" system of equations

$$(4.5) \qquad\qquad C\underline{a}^* = \underline{b} ,$$

with a_o^* selected so as to make the optimal linear forecast f^* unbiased, e.g.,

$$(4.6) \qquad Ef^*(\underline{\tilde{x}}) = E\{\tilde{y}\} ; \quad a_o^* = E\{\tilde{y}\} - \underline{a}^{*\prime}\underline{m} ;$$

and the covariance matrix C and the vectors $\underline{b}$ and $\underline{m}$ are:[†]

$$(4.7) \qquad C = V\{\underline{\tilde{x}}\} ; \quad \underline{b} = C\{\underline{\tilde{x}};\tilde{y}\} ; \quad \underline{m} = E\{\underline{\tilde{x}}\} .$$

The prior variance of the optimal linear forecast is:

$$(4.8) \qquad Vf^*(\underline{\tilde{x}}) = \underline{a}^{*\prime}C\underline{a}^* = \underline{a}^{*\prime}\underline{b} = \underline{b}'C^{-1}\underline{b}o = C\{f^*(\underline{\tilde{x}});\tilde{y}\}$$

giving minimal approximation mean-square error:

$$(4.9) \qquad I^* - I^o = Vf^o(\underline{x}) - \underline{b}'C^{-1}\underline{b} ,$$

which is smaller, the closer the conditional mean $E\{\tilde{y} \mid \underline{x}\}$ is to a linear form. In this sense, the optimal linear $f^*(\underline{x})$ is a best least-squares linearized Bayesian approximation.

In 1967, Bühlmann [4], [5] showed the important result that, for the collective model of Section 3, the optimal linear estimator for the experience-rated fair premium is exactly the credibility form (2.1), provided

---

[†]For any two random vectors or scalars $\underline{\tilde{\mu}}$ and $\underline{\tilde{\nu}}$, we define the (possibly nonsquare and unsymmetric) covariance matrix: $C\{\underline{\tilde{\mu}};\underline{\nu}\} = E\{\underline{\tilde{\mu}}\underline{\tilde{\nu}}'\} - E\{\underline{\tilde{\mu}}\}E\{\underline{\tilde{\nu}}'\}$, and call $C\{\underline{\tilde{\mu}};\underline{\tilde{\mu}}\} = V\{\underline{\tilde{\mu}}\}$, the usual covariance matrix of $\underline{\tilde{\mu}}$ on itself.

8

that the time constant in the credibility factor is chosen as the ratio of the components of the collective variance (3.4), i.e.,

$$(4.10) \qquad\qquad N = Ev(\tilde{\theta})/Vm(\tilde{\theta}) = E/D \ .$$

This shows that the basic credibility formula is robust, and has mean-square error (estimation error variance)

$$(4.11) \qquad\qquad I^* = E + (1 - Z)D \ ,$$

which shows clearly how increasing experience data improves the estimate.

The sample mean alone, $\bar{x}$, is a poorer estimate because it has $I = (1 + n^{-1})E$, which is always larger than I^* . However, if the *prior variation*, $D = Vm(\tilde{\theta})$, is very large compared to $E = Ev(\tilde{\theta})$ (a "diffuse prior") then N is very small, and f^* and $\bar{x}$ are practically the same.

Notice the important result

$$(4.12) \qquad\qquad C\{\tilde{y} - f^*(\underset{\sim}{\tilde{x}}); f^*(\underset{\sim}{\tilde{x}})\} = 0 \ ;$$

that is, any error remaining in the optimal forecast is uncorrelated with the predictor.

5. EXACT CREDIBILITY

In 1973, the author showed that the class of likelihood-prior families for which the credibility approximation (2.1), (2.2), (4.10) was *exactly* the Bayesian conditional mean could be extended [25], [27].

Consider the Koopman-Pitman-Darmis *exponential family* of likelihoods in which the sample mean is the only sufficient statistic and *natural parametrization* is chosen, i.e.,

$$(5.1) \qquad p(x \mid \theta) = \frac{a(x)e^{-\theta x}}{c(\theta)} \qquad (x \in X)$$

for continuous or discrete measure in a given range X, determined by the nonvanishing of $a(x)$. $c(\theta)$ is a normalizing factor to make

$$\int_X p(x \mid \theta)dx = 1 .$$

The *natural conjugate prior* corresponding to the likelihood (5.1) is

$$(5.2) \qquad u(\theta) = \frac{[c(\theta)]^{-n_o} e^{-\theta x_o}}{d(n_o, x_o)} , \qquad (\theta \in \Theta)$$

defined over a *natural parameter space*, Θ, for which (5.1) is a density, i.e., for all values of θ for which $c(\theta)$ is finite. Restrictions on the hyperparameters (n_o, x_o) may be necessary to make (5.2) a density as well, i.e., to make the normalization $d(n_o, x_o)$ finite. We shall henceforth assume $n_o > 0$.

The advantage of a natural conjugate pair is that the family is *closed under sampling*, that is, the density of $\tilde{\theta}$ *posterior to the data*

is of the same form as (5.2), with hyperparameter updating:

$$n_o \leftarrow n_o + n$$

(5.3)

$$x_o \leftarrow x_o + \sum_{t=1}^{n} x_t \ .$$

Since $m(\theta) = -c'(\theta)/c(\theta)$ for this family, integration of (5.2) by parts gives

(5.4)
$$Em(\tilde{\theta}) = \frac{u(\theta) \mid \Theta}{n_o} + \frac{x_o}{n_o} \ .$$

$c(\theta)$ is analytic in the interior of Θ , and, in most cases of interest, vanishes at the boundary as well, making the first term on the RHS of (5.4) zero. The precise regularity conditions under which this happens are complex, and are covered in [27].

Assuming these conditions are satisfied, (5.3) then implies

(5.5)
$$E\{m(\tilde{\theta}) \mid x_1, x_2, \ \cdots, \ x_n\} = \frac{x_o + \sum x_t}{n_o + n} \ .$$

The final steps, which are similar, show that $x_o/n_o = m$, and n_o is the time constant N (4.10) of Bühlmann, thus proving credibility is exact for (5.1), (5.2).

Additional examples beyond those of Bailey and Mayerson are given in [25], and the extension to credibility mixing of more general statistics for arbitrary exponential families is also demonstrated.

6. MULTIDIMENSIONAL CREDIBILITY

The previous results are easily extended to prediction of a multivariate conditional mean [23], [26]. Let $\tilde{\underline{x}}$ be a p-dimensional random variable, depending upon a risk parameter $\tilde{\theta}$ through a likelihood density $p(\underline{x} \mid \theta)$; a prior density on θ is assumed known. For $t = 1, 2, \ldots, n$, we observe n independent realizations, $\tilde{\underline{x}}_t = \underline{x}_t$, of this random variable for a fixed θ . The problem is to make a (vector) forecast $\underline{f}(X)$ of the next observation, $E\{\tilde{\underline{x}}_{-n+1} \mid X\}$, where X is the $p \times n$ matrix of data $\{\underline{x}_t \mid t = 1, 2, \ldots, n\}$.

From the likelihood and the prior, we calculate the vector means:

$$(6.1) \qquad \underline{m}(\theta) = E\{\tilde{\underline{x}} \mid \theta\} \; ; \; \underline{m} = E\underline{m}(\tilde{\theta}) \; ;$$

and then the two $p \times p$ covariance matrices;

$$(6.2) \qquad E = EV\{\tilde{\underline{x}} \mid \tilde{\theta}\} \; ,$$

and

$$(6.3) \qquad D = V\{\underline{m}(\tilde{\theta})\} \; .$$

The total covariance matrix of any $\tilde{\underline{x}}_t$ on itself is $E + D$, but the covariance matrix between any $\tilde{\underline{x}}_t$ and $\tilde{\underline{x}}_u$ $(t \neq u)$ is just D .

Assuming that the forecast of each component of $\tilde{\underline{x}}_{-n+1}$ is linear in *all* the data X , then the use of least-squares theory gives after some algebra the multivariate credibility formula:

$$(6.4) \qquad E\{\tilde{\underline{x}}_{-n+1} \mid X\} \approx \underline{f}^*(X) = (I - Z)\underline{m} + Z\bar{\underline{x}} \; ,$$

where I is the $p \times p$ identity matrix, and $\bar{\underline{x}}$ is the vector of sample means,

$$\bar{\underline{x}} = \sum \underline{x}_t/n \ .$$

The $p \times p$ *credibility matrix* Z satisfies formulae analogous to the one-dimensional result (2.2):

(6.5) $\qquad Z = n(N + nI)^{-1} \ ; \ (I - Z) = \frac{1}{n} \ (ZN) = \frac{1}{n} \ (NZ) \ ;$

and the $p \times p$ *matrix of time constants*, N , is analogous to (4.10):

(6.6) $\qquad\qquad\qquad N = ED^{-1} \ .$

If the eigenvalues of N are $\{\nu_i\}$, then those of Z are $\{n/(n + \nu_i)\}$; one can show that in the nondegenerate case $\lim_{n \to \infty} Z = I$. In other words, the initial forecast (no data) is the prior mean $\underline{m}$; successive forecasts utilize linear mixtures of *all* sample means in varying proportions; but ultimately, each component of the risk is estimated only through its own sample mean, as $n \to \infty$. Specific examples are given in [23]. The $p \times p$ "preposterior" estimation error covariance matrix for the optimal vector forecast in (6.4) is:

(6.7) $\qquad\qquad V\{\tilde{\underline{x}}_{n+1} - \underline{f}^*(X)\} = E + (I - Z)D \ .$

This is similar to (4.11), but of course only the diagonal elements of the LHS of (6.7) were minimized in selecting the optimal coefficients.

There are also exact multi-dimensional results [26], corresponding to Section 5, for the *linear multivariate exponential family likelihood*:

(6.8) $\qquad p(\underline{x} \mid \underline{\theta}) = \dfrac{a(\underline{x})\exp\{-\underline{\theta}'\underline{x}\}}{c(\underline{\theta})} \ . \qquad\qquad (\underline{x} \ \varepsilon \ X)$

$\underline{\theta}$ is now a p-dimensional vector in the complete parameter space Θ for which the normalization, $c(\underline{\theta})$, is finite. This family has the vector sample mean, $\tilde{\underline{x}}$, as a sufficient statistic, and has been investigated somewhat; however, their natural conjugate priors have been little studied.

The simplest natural conjugate prior for (6.8), analogous to (5.2), is

$$(6.9) \qquad u(\underline{\theta}) \propto [c(\underline{\theta})]^{-n_o} \exp[-\underline{\theta}'\underline{x}_o] \qquad (\underline{\theta} \in \Theta)$$

where the scalar n_o and the vector $\underline{x}_o$ are hyperparameters; we assume always $u(\underline{\theta})$ vanishes at the boundary of Θ. Embarrassingly, in this case, although (6.4) is exactly the conditional mean, Z degenerates to a diagonal matrix because $N = n_o I$, and the forecasts for each component of $\tilde{\underline{x}}$ are independent!

To remedy this, the author develops in [26] an "enriched" version of (6.9), for likelihoods (6.8) in which $a(\underline{x})$ will factor into a product of p independent components when $\underline{x}$ is subject to a linear transformation. In this case, enough additional hyperparameters can be introduced to make N and Z non-diagonal, and all components of the sample mean are used in prediction of any one future value.

An important extension permits quadratic terms in the exponent of (6.8), and leads to the well-known *multinormal likelihood with unknown mean and unknown precision*. The usual mean-precision prior is a *normal-Wishart* distribution, due to Ando and Kaufman [1]; its "thinness" is well-known in the literature, and is similar to the degeneracy described above. Through the use of linear transformation, it is possible to extend the Ando-Kaufman prior, again giving a full-dimensional credibility formula (6.4) for the conditional mean [26]. One also finds the following interesting credibility formula for the conditional covariance matrix of this enriched multinormal:

$$V\{\tilde{\underline{x}}_{n+1} \mid X\} = (I - Z)V\{\tilde{\underline{x}}\} + Z\left[\frac{1}{n}\sum_{t=1}^{n}(\underline{x}_t - \bar{\underline{x}})(\underline{x}_t - \bar{\underline{x}})'\right]$$

(6.10)

$$+ (I - Z)(\underline{m} - \bar{\underline{x}})(\underline{m} - \bar{\underline{x}})'Z' \ .$$

Notice how the sample covariance ultimately dominates.

7. CLASSICAL ESTIMATORS

We have emphasized the Bayesian role of credibility, either as an approximation to a conditional mean, or as an exact result for certain priors and likelihoods. However, there are other interpretations of credibility which show its relationship to classical least-squares estimators.

Suppose we rewrite (6.4) for the case in which only *one* p-dimensional observation $\tilde{\underline{x}} = \underline{x}_1$ is made:

$$(7.1) \qquad E\{\tilde{\underline{x}}_2 \mid \underline{x}_1\} \approx (I - Z_1)\underline{m} + Z_1\underline{x}_1 \, ,$$

with

$$(7.2) \qquad Z_1 = (I + ED^{-1})^{-1} \, .$$

Rewrite this as:

$$(7.3) \qquad E\{\tilde{\underline{x}}_2 \mid \underline{x}_1\} - \underline{m} \approx Z_1(\underline{x}_1 - \underline{m}) \, ,$$

and note that Z_1 can be written as:

$$(7.4) \qquad Z_1 = D(E + D)^{-1} = C\{\tilde{\underline{x}}_2 \; ; \; \tilde{\underline{x}}_1\}[V\{\tilde{\underline{x}}_1\}]^{-1} \, .$$

In this form, we recognize (7.3) as a well-known exact result, the *regression* of $\tilde{\underline{x}}_2$ on $\tilde{\underline{x}}_1$ for a joint multinormal distribution of $\{\tilde{\underline{x}}_1 \; ; \; \tilde{\underline{x}}_2\}$.

For the second interpretation, write in *linear model form*:

(7.5)
$$\underset{\sim}{x}_t = \underset{\sim}{m}(\tilde{\theta}) + \underset{\sim}{\tilde{u}}_t \qquad (t = 0,1,2, \ldots)$$

where $\underset{\sim}{\tilde{u}}_t$ is an appropriate *error variable*, independent of $\tilde{\theta}$ and other errors, and

(7.6)
$$E\{\underset{\sim}{\tilde{u}}_t\} = 0 \qquad (t = 0,1,2, \ldots) .$$

For $t = 1$, we observe $\underset{\sim}{\tilde{x}}_1 = \underset{\sim}{x}_1$, and this is an estimator of $E\{\underset{\sim}{\tilde{x}}_2 \mid \underset{\sim}{x}_1\} = E\{\underset{\sim}{m}(\tilde{\theta}) \mid \underset{\sim}{x}_1\}$ with error covariance matrix

(7.7)
$$V\{\underset{\sim}{\tilde{u}}_1\} = EV\{\underset{\sim}{\tilde{u}}_1 \mid \tilde{\theta}\} = E .$$

Now, a prior estimate can be thought of as an initial observation at $t = 0$, so that the initial credibility estimate $\underset{\sim}{\tilde{x}}_0 = \underset{\sim}{m}$ is also an unbiased estimate of $m(\tilde{\theta})$ before the other observations begin. Since $\underset{\sim}{m}$ is a constant, (7.5) shows that the error covariance of this estimate is:

(7.8)
$$V\{\underset{\sim}{\tilde{u}}_0\} = V\underset{\sim}{m}(\tilde{\theta}) = D .$$

By elementary manipulations:

(7.9)
$$E\{\underset{\sim}{m}(\tilde{\theta}) \mid \underset{\sim}{x}_0 = \underset{\sim}{m}, \underset{\sim}{x}_1\} \approx (E^{-1} + D^{-1})^{-1}\left[D^{-1}\underset{\sim}{m} + E^{-1}\underset{\sim}{x}_1\right] ,$$

and we recognize that the "two" observations $\underset{\sim}{x}_0$ and $\underset{\sim}{x}_1$ are combined by weighting with their respective *precisions*, D^{-1} and E^{-1}. This ancient formula for the combination of observations is due to Gauss, and is known to be exact for $\underset{\sim}{\tilde{u}}_1$ and $\underset{\sim}{\tilde{u}}_2$ independent and (multi-) normally distributed. The preposterior estimation error

precision is the inverse of (6.7) with $Z = Z_1$,

(7.10)
$$V^{-1}\{\tilde{\underline{x}}_2 - \underline{f}^*(\tilde{\underline{x}}_1)\} = D^{-1} + E^{-1} ,$$

the sum of the two observation precisions. Similar remarks apply to the Bayesian regression model of Section 10; see (10.10) and [34].

Notice that credibility formulae are mixtures of a prior mean and a classical maximum likelihood estimator. This is true for a large class of identifiable linear models (10.7), and in the exact case follows easily from the definition of exponential families.

Finally, we note that many articles in the statistical literature develop similar "wide-sense conditional expectations" and "parameter shrinkage" formulae, [14], [40], [17], [18], [20], [48], [53]; they are called "pseudo-Bayes estimators" in filter theory [47], and are no doubt being rediscovered in other fields as well. However, the intimate relationship between the Bayesian and the classical appraoch is not well appreciated, and there is widespread belief that these results are exact only for multinormal distributions, which is not true.

8.   OTHER CREDIBILITY MODELS

We now consider some of the many extensions of the credibility
model developed to deal with specific estimation problems in insurance.
These models all use least-square approximations, usually exploiting
the special structure to avoid numerical inversion of the covariance
matrix in (4.5).   See also the references in [13] and [38].

8.1   Other Functions

The first idea is that least-squares theory also applies to
functions of random variables, with appropriate modifications to (4.7).
Suppose in the basic model of (2.1), (2.2), (4.10), we replace $\tilde{x}$
by $I(u - \tilde{x})$ throughout; here u is some fixed value in the range
of $\tilde{x}$, and I is the unit-step function, unity for nonnegative
arguments, zero otherwise.   Since $\sum I(u - x_t)$ counts the number
of samples not greater than u , and $EI(u - \tilde{x}) = P(u)$, we get a
credible conditional distribution forecast [24]:

$$(8.1) \qquad P(u \mid \underline{x}) \approx (1 - Z)P(u) + Z\left[\sum_{t=1}^{n} I(u - x_t)/n\right] ,$$

as a mixture of the prior collective probability, $P(u)$, and the
experienced sample distribution. $Z = n/(n + N)$, as before, but the
time constant depends upon u :

$$(8.2) \qquad N = \frac{P(u)[1 - P(u)]}{VP(u \mid \tilde{\theta})} - 1 .$$

(8.1) is exact only in the simple case of Beta/Bernoulli prior/likelihood.
If $P(u)$ is continuous, it gives a mixed function of u which has

smaller mean-squared error for every u than using the sample dis-
tribution alone.

Clearly, the same idea also applies to estimating moments of $\tilde{x}$,
the difficulty being that higher collective moments corresponding to
(3.3), (3.4) must be known. In [5], Bühlmann develops a credibility
formula for the conditional variance $V\{\tilde{x}_{n+1} \mid x_1, x_2, \ldots, x_n\}$
based upon separation into a "variance" part and a "fluctuation"
part, and using several approximations; see also (6.10).

Estimating fractiles or order statistics by credibility seems
difficult; however Bühlmann [9] shows that one can estimate the
mass between any two ordered data points of given rank.

In [11], de Vylder has studied the optimal *form* of the predictand
to be used in the semi-linear form:

(8.3)
$$f(\underline{x}) = a_0 + a_1 \sum_{t=1}^{n} g(x_t) ,$$

for arbitrary likelihood and prior. The resulting integral equation
uses the conditional density $p(x_2 \mid x_1)$ to find the optimal g ,
and seems most useful for discrete x . [12]

8.2 Compound Models

A basic concept in casualty insurance is that the total dollar
claims in a given exposure period is related to both *frequency* and
severity of a claim, once it occurs; this leads to a risk random
variable which is the random sum of other elementary random variables.
The major contributions are [7], [21], [23], [43], [44].

8.3 Bonus Hunger

It is a well-known fact that experience rating schemes induce a compensating behavior in the insured individual. For example, small claims will not be reported in order to keep future dividend payments down; insurance companies often encourage this, even though it makes estimation of the true risk more difficult. [45] examines the effect of this "bonus hunger" on credibility plans.

8.4 IBNR Models

Many insurance claims take a long time to "develop," that is, a claim in year t will incur "losses" in year $t, t+1, t+2, \ldots$; the total dollar claim is "incurred but not (fully) reported." Thus, at any epoch in time, one has an *IBNR triangle* of partially developed claims which can be used to estimate the final totals; correlations in observations in both the cohort and calender time dimensions are likely. These problems have been approached by Straub and Kamreiter using least-squares techniques [39], [49].

8.5 Time-Dependent Models

The analysis of time-dependencies is of great interest. In the general nonstationary case where $n + 1$ p-dimensional densities, $p_t(\underline{x}_t \mid \theta)$, $(t = 1, 2, \ldots, n+1)$, are available, the optimal linear predictor requires inversion of an $np \times np$ covariance matrix, which is hardly satisfactory. However, if the time-dependency is of *separable-mean type* [28], that is, for the means:

$$(8.4) \qquad m_{it}(\theta) = E\{\tilde{x}_{it} \mid \theta\} = k_{it} \cdot r_i(\theta) \ , \quad \begin{pmatrix} i = 1,2, \ \ldots, \ p \\ t = 1,2, \ \ldots, \ n+1 \end{pmatrix}$$

(with arbitrary time dependence of the covariance), then one can show that only matrices of order $p \times p$ need be inverted.

Another modeling approach is to consider that the risk parameter is itself changing over time even though the likelihood is stationary for a given θ . Thus, by specifying some joint prior $u(\theta_1, \theta_2, \ \ldots, \ \theta_t, \ \ldots)$ for the risk parameter in successive time periods, the *evolutionary mechanism* of the risk process is completely determined [28]. A special case of interest is the one-dimensional random shock model of Gerber and Jones [15], [16], [29] in which the evolutionary mechanism provides a sequence of mutually independent scale and location shifts $\{\tilde{k}_t \ , \ \tilde{s}_t\}$ to the location parameters of the risk variables, so that:

$$(8.5) \qquad E\{\tilde{x}_t \mid \theta_t, k_t, s_t\} = m_t(\theta_t) = k_t m_{t-1}(\theta_{t-1}) + s_t \ , \quad (t = 1,2, \ \ldots)$$

where $\tilde{\theta}_t$ and $\{\tilde{k}_u, \tilde{s}_u \mid u = t, t+1, \ \ldots\}$ are mutually independent. Forecasts for successive time periods follow a simple recursive calculation scheme; in many cases where the moments of $\tilde{k}_t \ , \ \tilde{s}_t$ are stable, the credibility weights are ultimately of geometric form.

In [29], the author explores in detail the "good" forms of C and $\underline{b}$ in (4.5) which lead to forecasts in either closed or recursive form.

8.6   Conditional Distributions

An obvious criticism of the multi-dimensional formula (6.4) is that in estimating the future value of a selected component, say

$\tilde{x}_{s,n+1}$, *all* of the remaining data is used in a linear approximation. If one could easily calculate the sth *conditional density*

$p_s(x_{st} \mid \underline{x}_t^-;\theta)$, ($\underline{x}_t^-$ is $\underline{x}_t$ without x_{st}) then one could use just a one-dimensional linear approximation in terms of $[x_{s1}, x_{s2}, \ldots, x_{sn}]$; now, however, the mean and variance components of this conditional density are quite complex, and are time-varying in the sense that different value of x_{-t}^- must be substituted. A complete development is given in [28]. As might be expected, simplification occurs only if the conditional mean is of separable type.

An important special model of this type has been analyzed by Bühlmann and Straub [6], [10], in which $\tilde{x}_{1t}$ is the *claim rate* (total portfolio \$ claims per unit volume of business), and x_{2t} is the given volume of business in year t . By elementary assumptions:

$$(8.6) \qquad E\{\tilde{x}_{1t} \mid x_{2t};\theta\} = m_0(\theta) \ ; \ V\{\tilde{x}_{1t} \mid x_{2t};\theta\} = v_0(\theta)/x_{2t} \ ,$$

where $m_0(\theta)$ and $v_0(\theta)$ are moments associated with a single unit volume. The credibility forecast is bilinear in $x_{1t}x_{2t}$ (total \$ claims) and x_{2t} (volume) and uses the latter as *operational time* for credibility, rather than n .

8.7 Minimax Credibility

The use of other than quadratic error norms seems mathematically intractable in approximating unknown conditional means. An interesting model by Bühlmann and Marazzi [8] adapts the point of view that Nature (who picks m , D , and E) is playing a game against the actuary; the resulting strategies depend strongly on the assumptions about the regions of play open to Nature.

9. COLLATERAL DATA AND HIERARCHIAL MODELS

Suppose that the different dimensions of $\underset{\sim}{x}$ refer to p *different* individual risks, each with a different risk parameter, $\tilde{\theta}_i$, (i = 1,2, ..., p) , *independently* distributed according to the same prior density, $u(\theta)$. In this case, it is easy to see that E , D , and Z of Section 6 are diagonal, and in predicting $E\{\tilde{x}_{s,n+1} \mid X\}$ for the sth risk, the experience data $\{x_{it} \mid (i \neq s) \ (t = 1,2, ..., n)\}$ from the other members of the cohort is not used. In other words, the one-dimensional form (2.1), mixing m and the sth component of $\bar{x}$, is optimal.

This result is disturbing to many practitioners, who feel that data from other risks in the same portfolio contains valuable collateral information. Similar arguments are advanced about the use of cohort data in the otherwise unrelated "empirical Bayes" approach.

Bühlmann and Straub [6], [10], [50] approach this problem by using a *homogeneous* linear least-squares forecast, in which $a_o = 0$ in (4.4), and the $\{a_j \mid j \neq 0\}$ are selected so as to minimize (4.1), but constrained so that the forecast is still unbiased. In the simplest credibility model (2.1), the collective prior mean, m , is then replaced by the estimator:

$$(9.1) \qquad \hat{m}(X) = \frac{1}{np} \sum_{i=1}^{p} \sum_{t=1}^{n} x_{it} \ ,$$

the *grand sample mean* of *all* cohort data. This also eliminates the problem of estimating m , but not that of estimating N . It also gives a larger mean-square error than (4.11).

In [30], the author constructs a heirarchial model, in which the insurance company and all risks in its portfolio are characterized by an additional hyperparameter ϕ , with a hyperprior distribution over the universe of all such collectives. ϕ is the "quality" of this particular portfolio.

Assume each individual risk $(i = 1,2, \ldots, p)$ has first and second moments

$$(9.2) \qquad m(\theta_i,\phi) = E\{\tilde{x}_{it} \mid \theta_i,\phi\} \; ; \; v(\theta_i,\phi) = V\{\tilde{x}_{it} \mid \theta_i,\phi\} \; .$$

Prior information now consists of an *universal mean* over all collectives,

$$(9.3) \qquad\qquad\qquad M = EEm(\tilde{\theta},\tilde{\phi})$$

and *three* components of *universal variance*:

$$(9.4) \qquad F = EEv(\tilde{\theta},\tilde{\phi}) \; ; \; G = EVm(\tilde{\theta},\tilde{\phi}) \; ; \; H = VEm(\tilde{\theta},\tilde{\phi}) \; .$$

(In the above expressions, the inner operation is on $\tilde{\theta}$, with $\tilde{\phi}$ fixed, and the outer operation is on $\tilde{\phi}$.) F and G correspond to the two terms in (3.4), averaged over all possible collectives, while H is a new term, corresponding to *inter-portfolio variation*. Presumably, one could easily estimate M , F , G , and H from a nationwide bank of experience data from different insurance companies. In predicting $x_{s,n+1}$, we now have to "learn" simultaneously about both θ_s and ϕ , and all collateral data from the portfolio will be used since all risks have the same ϕ .

The optimal estimator then consists of three terms:

$$(9.5) \quad E\{\tilde{x}_{s,n+1} \mid X\} = (1 - Z)[(1 - Z_c)M + Z_c\hat{m}(X)] + Z \sum_{t=1}^{n} (x_{st}/n) \; .$$

The *individual credibility factor*, Z , is given by (2.2), but in place of (4.10), N = F/G .

The term in square brackets represents the best current estimate of the fair premium for "our" collective; it mixes the universal mean, M , with the grand sample mean (9.1), using a *collective credibility factor*

$$(9.6) \qquad Z_c = \frac{npH}{F + nG + npH} .$$

Note that $Z_c < 1$ as $n \to \infty$; that is, $\hat{m}(X)$ is *not* ultimately "fully credible" for the fair premium of "our" collective; this is because we have only a finite sample of $\{\tilde{\theta}_i\}$ for a fixed ϕ .

If $H \to 0$, (9.2) reduces to the usual credibility formula. Bühlmann and Straub's result can be seen as a limiting case in which $H \to \infty$; that is, we have a "diffuse (hyper-) prior" on ϕ , and inter-portfolio variations are very large.

10. BAYESIAN REGRESSION MODELS

A very general mathematical structure which contains all of the previous models is the *linear (regression) model*

$$(10.1) \qquad \tilde{\underline{y}} = H\tilde{\underline{\beta}} + \tilde{\underline{u}} \, ,$$

where $\tilde{\underline{y}}$ and $\tilde{\underline{u}}$ are $n \times 1$ random vectors of observable *output variables* and unobservable *error variables*, respectively, H is a known $n \times p$ *design matrix*, and $\tilde{\underline{\beta}}$ is a $p \times 1$ random vector of unknown *regression parameters*; we assume that a prior joint density of $(\tilde{\underline{\beta}}, \tilde{\underline{u}})$ is known. Given an observation $\tilde{\underline{y}} = y$, the problem is to draw posterior-to-data inferences about $\tilde{\underline{\beta}}$, or about future values of $\tilde{\underline{y}}$ for some (possibly) different design matrix; this is a problem in *Bayesian regression*. A complete Bayesian regression analysis is very difficult, usually requiring restrictive distributional assumptions or complicated algebraic manipulations (see, e.g. [54]).

However, the linearized approach of credibility theory can be very useful if the goal is to update only mean values of $\tilde{\underline{\beta}}$ or $\tilde{\underline{y}}$; preposterior error covariances can also be determined.

Let the prior knowledge of $\{\tilde{\underline{\beta}}, \tilde{\underline{u}}\}$ be summarized in the mean vectors:

$$(10.2) \qquad E\{\tilde{\underline{\beta}}\} = \underline{b} \; ; \; E\{\tilde{\underline{u}} \mid \tilde{\underline{\beta}}\} = 0 \quad \text{(for all } \tilde{\underline{\beta}}) \; ;$$

and the covariance matrices:

$$(10.3) \qquad V\{\tilde{\underline{\beta}}\} = \Delta \; ; \; EV\{\tilde{\underline{y}} \mid \tilde{\underline{\beta}}\} = V\{\tilde{\underline{u}}\} = E \; ;$$

of order $p \times p$ and $n \times n$, respectively. We define also alternate-dimension versions of the covariances:

$$(10.4) \qquad D = H\Delta H' \; ; \; \epsilon = (H'E^{-1}H)^{-1} \; ;$$

which are $n \times n$ and $p \times p$, respectively. Even if E is positive

definite (most applications have E diagonal, i.e., "homoscedastic errors,"

or "white noise"), ε may not exist in many models of interest because

H is not of rank $\max(p,n)$.

There are two versions of the credibility forecast of the mean para-

meter values $\underline{f}(\underline{y}) \approx E\{\tilde{\underline{\beta}} \mid \underline{y}\}$. In the first version:

$$(10.5) \qquad \underline{f}(\underline{y}) = (I - ZH)\underline{b} + Z\underline{y} ,$$

where I is the $p \times p$ unit matrix, and Z is a $p \times n$ *credibility matrix*

$$(10.6) \qquad Z = \Delta H'(E + D)^{-1} .$$

This clearly exists if, say, E is positive definite, and H contains

only nonnegative elements; an $n \times n$ inversion is required, even if E^{-1}

is known, hence this form is suitable for limited-observation experiments

where $n < p$.

In the second version:

$$(10.7) \qquad \underline{f}(\underline{y}) = (I - z)\underline{b} + z\hat{\underline{\beta}}(\underline{y}) ,$$

where $\hat{\beta}(y)$ is the classical (generalized) least-squares estimator of $\tilde{\underline{\beta}}$:

$$(10.8) \qquad \hat{\underline{\beta}}(\underline{y}) = \varepsilon H'E^{-1}\underline{y} = (H'E^{-1}H)^{-1}H'E^{-1}y ,$$

and z is a $p \times p$ credibility matrix:

$$(10.9) \qquad z = (I + \varepsilon\Delta^{-1})^{-1} = \Delta(I + \varepsilon^{-1}\Delta)^{-1}\varepsilon^{-1} .$$

This matrix is analogous to the usual multidimensional credibility

matrix (6.5) with "one" sample, and show clearly the mixing of the prior

mean and the classical estimator. Moreover, in many applications $n > p$, and (10.9) shows that only one $p \times p$ inversion is required to find $\underline{f}(\underline{y})$, if E^{-1} is known, greatly reducing the computational labor. On the other hand, to find $\underline{\hat{\beta}}(\underline{y})$ *explicitly*, we require that ε exist, which leads to the classic problem of "identifiability," and the requirement that rank $(H) = p$.

The preposterior covariance of the parameter estimation error can be shown to be:

$$(10.10) \quad V\{\underline{\tilde{\beta}} - \underline{f}(\underline{\tilde{y}})\} = (I - ZH)\Delta = (I - z)\Delta = (\Delta^{-1} + \varepsilon^{-1})^{-1} .$$

Cf. (7.10).

Hachemeister [19] and Taylor [51] were the first to give special versions of (10.7), (10.8) in credibility terminology. Many other results and interpretations can be found in [34].

However there are numerous nonBayesian versions of the above formulae in earlier statistical literature [46], [52]. Priority for these formulae probably belongs in the communications theory literature, where generalized least-squares and "pseudo-Bayes" estimators have been used in *linear (Wiener-Kalman-Bucy) filters* for many years, (see, e.g. [47], pp. 182-4); specialized jargon of this field has no doubt delayed recognition of the simularities between approaches. Also, filter theory emphasizes dynamic regression models, with recursive calculation of successive forecasts to reduce computational labor. An example of this approach for simple trends in regression parameters is given in [33].

11. OTHER APPLICATIONS

In closing, we describe several noninsurance applications of credibility theory.

11.1 Sampling Schemes

Classical statistics treats in great detail the analysis of variance for complex sampling schemes. However, in many real applications, one has also a prior estimate of the quantity being measured, and the experimental observations should be combined optimally with this prior knowledge. [37] explores this idea for various nested sampling schemes; as might be expected from Section 7, the prior mean value is combined with the classical estimators in a manner proportional to their relative precisions.

11.2 Material Accountability Systems

The rapid proliferation of nuclear material has induced development of statistical material accountability systems to monitor and "safeguard" the production, storage, and shipment processes. The basic tool is a material balance equation, which should balance out to zero if the *material unaccounted for* is also zero; however, there are very difficult instrumentation problems in measuring radioactive materials, and this balance can only be estimated statistically. [32] describes a simple batch material balance closure problem. More general dynamic multi-stage problems can be tackled using the formulae of Section 10.

11.3 Instrument Calibration and Measurement and Inverse Regression

An instrument can be characterized by the simple linear model

$$(11.1) \qquad \tilde{y} = \tilde{\beta}_0 + \tilde{\beta}_1 \tilde{x} + \tilde{u} \ .$$

In *calibration*, relatively precise *standard inputs* $x_1, x_2, \ldots$ are given to an uncalibrated instrument, resulting in *outputs* $y_1, y_2, \ldots$ which contain *observation errors* $u_1, u_2, \ldots$. We generally have some joint prior information about the *instrument parameters* $(\tilde{\beta}_0, \tilde{\beta}_1)$, and this is updated using the calibration data.

In *measurement*, we place a partially known quantity $\tilde{x}_0$ as input (with some prior on its value), and observe $\tilde{y}_0 = y_0$; the problem is then to make an *inverse regression* to estimate $\tilde{x}_0$.

Clearly, the general Bayesian formulation is to find $E\{\tilde{x}_0 \mid y_0; (x_1, y_1); (x_2, y_2); \ldots\}$. In [31], we show that the use of credibility breaks the estimation into two natural stages: (1) a credibility estimation of $(\tilde{\beta}_0, \tilde{\beta}_1)$ using the standards; (2) a linearized inverse regression using the new parameter estimates.

11.4 Network Flows

Many road traffic, telecommunication, and accounting processes can be modeled as *flows over networks*; the basic equation is Kirchoff's conservation law, in which the design matrix H of Section 10 is the node-arc incidence matrix. In one formulation, there is prior knowledge about arc flows and their prior precisions (usually highly correlated because they arise from origin/destination "path" flows); the problem is to make a few boundary or selected-arc flow measurements, and to infer the new arc flows. This problem is highly "unidentifiable" in the classic sense because the rank of H is one less than the number of nodes; however, the theory in (10.5) and (10.6) still applies [35].

11.5 Life Testing

The usual manner of inferring lifetime distributions in reliability studies is to place a large number of components, N, on a test stand, and run them for a fixed time T until $C = C(T) < < N$ of the components have failed. Assuming prior knowledge about the distribution parameters, the problem is to make a posterior-to-test inference using C *completed lifetimes* $[\tilde{x}_i = x_i \mid i = 1,2, \ldots, C]$ and $N - C$ *incomplete lifetimes* $[\tilde{x}_i > T \mid i = C+1, C+2, \ldots, N]$.

[36] examines the *proportional hazard* lifetime distribution:

$$(11.2) \qquad \Pr \{\tilde{x} > x \mid \theta\} = \exp \{-\theta Q(x)\} \, ,$$

where Q is a known *prototype failure function*, and θ has a Gamma prior; it is shown that the Bayesian estimate of $\tilde{\theta}^{-1}$ from this test is exactly a credibility mixture of the prior estimate of $\tilde{\theta}^{-1}$ together with a new maximum likelihood estimator:

$$(11.3) \qquad \frac{1}{C} \sum_{i=1}^{C} Q(x_i) + \left(\frac{N}{C} - 1\right) Q(T) \, ,$$

which generalizes the well-known *total-time-on-test* statistic.

REFERENCES

[1]   Ando, A. and G. Kaufman, "Bayesian Analysis of the Independent
         Multinormal Process-Neither Mean Nor Precision Known,"
         Journal of the American Statistical Association, Vol. 60,
         pp. 347-358, (1965).

[2]   Bailey, A. L., "A Generalized Theory of Credibility," PCAS,
         Vol. 32, pp. 13-20, (1945).

[3]   Bailey, A. L., "Credibility Procedures, LaPlace's Generalization
         of Bayes' Rule, and the Combination of Collateral Knowledge
         with Observed Data," PCAS, Vol. 37, pp. 7-23, (1950);
         Discussion, PCAS, Vol. 37, pp. 94-115, (1950).

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

[5]   Bühlmann, H., MATHEMATICAL METHODS IN RISK THEORY, Springer-Verlag,
         New York, (1970).

[6]   Bühlmann, H., "Credibility Procedures," Sixth Berkeley Symposium on
         Probability and Statistics, University of California Press,
         Berkeley, pp. 515-525, (1971).

[7]   Bühlmann, H., "A Comparison of Three Credibility Formulae Using
         Multidimensional Techniques," ASTIN-Bulletin, Vol. 7, Part 3,
         pp. 203-207, (March 1974).

[8]   Bühlmann, H., "Minimax Credibility," in [38].

[9]   Bühlmann, H., "The Mincing Machine Revisited," ASTIN-Bulletin,
         Vol. 8, Part 2, pp. 135-143, (September 1975).

[10] Bühlmann, H. and E. Straub, "Glaubwürdigkeit für Schadensätze,"
         Mitteilungen der Vereinigung Schweizerischer
         Versicherungsmathematiker, Vol. 70, pp. 111-133, (1970);
         translation by C. E. Brooks, "Credibility for Loss Ratios,"
         (unpublished).

[11] de Vylder, F., "Optimal Semilinear Credibility," Mitteilungen der
         Vereiningung Schweizerischer Versicherungsmathematiker,
         Vol. 76, pp. 27-40, (1976).

[12] de Vylder, F. and Y. Ballegeer, "A Numerical Illustration of Optimal
         Semilinear Credibility," University of Louvain, Belgium, (1975).

[13] de Pril, N., L. D'Hooge and M. J. Goovaerts, "A Bibliography on
         Credibility Theory and its Applications," University of Louvain,
         Belgium; to be published in Journal of Computation and Applied
         Mathematics.

[14] Ericson, W. A., "A Note on the Posterior Mean of a Population
      Mean," _Journal of the Royal Statistical Society_, (B),
      Vol. 31, No. 2, pp. 332-334, (1969).

[15] Gerber, H. V. and D. A. Jones, "Credibility Formulae with Geometric
      Weights," _Transactions of the Society of Actuaries_, Vol. 27,
      (1975).

[16] Gerber, H. V. and D. A. Jones, "Credibility Formulas of Updating
      Type," in [38].

[17] Goldstein, M., "Approximate Bayes Solution to Some Nonparametric
      Problems," _Annals of Statistics_, Vol. 3, No. 2, pp. 512-517,
      (1975).

[18] Goldstein, M., "Bayesian Nonparametric Estimates," _Annals of
      Statistics_, Vol. 3, No. 3, pp. 736-740, (1975).

[19] Hachemeister, C. A., "Credibility for Regression Models with
      Application to Trend," in [38].

[20] Hartigan, J. A., "Linear Bayesian Methods," _Journal of the Royal
      Statistical Society_, (B), Vol. 31, pp. 446-454, (1969).

[21] Hewitt, C. C., Jr., "Credibility for Severity," _PCAS_, Vol. 57, pp. 148-
      171, (1970); Discussion, _PCAS_, Vol. 58, pp. 25-31, (1971).

[22] Hickman, J. C., "Introduction and Historical Overview of Credibility,"
      in [38].

[23] Jewell, W. S., "Multi-Dimensional Credibility," ORC 73-7, (April 1973).

[24] Jewell, W. S., "The Credible Distribution," ORC 73-13, (August 1973);
      ASTIN-Bulletin, Vol. 7, Part 3, pp. 237-269, (March 1974).

[25] Jewell, W. S., "Credible Means are Exact Bayesian for Simple
      Exponential Families," ORC 73-21, (October 1973); _ASTIN-Bulletin_,
      Vol. 8, Part 1, pp. 77-90, (September 1974).

[26] Jewell, W. S., "Exact Multidimensional Credibility," ORC 74-14,
      (May 1974); _Mitteilungen der Vereinigung Schweizerischer
      Versicherungsmathematiker_, Vol. 74, pp. 193-214, (1974).

[27] Jewell, W. S., "Regularity Conditions for Exact Credibility,"
      ORC 74-22, (July 1974); _ASTIN-Bulletin_, Vol. 8, Part 3,
      pp. 336-341, (September 1975).

[28] Jewell, W. S., "Model Variations in Credibility Theory," ORC 74-25,
      (August 1974); also in [38].

[29] Jewell, W. S., "Two Classes of Covariance Matrices Giving Simple
      Linear Forecasts," RM-75-17, International Institute for Applied
      Systems Analysis, Laxenburg, Austria, (May 1975); _Scandinavian
      Actuarial Journal_, Vol. 1976, No. 1, pp. 15-29, (1976).

34

[30] Jewell, W. S., "The Use of Collateral Data in Credibility Theory:
         A Hierarchial Model," RM-75-24, International Institute of
         Applied Systems Analysis, Laxenburg, Austria, (June 1975);
         to appear in Giornale dell' Istituto Italiano degli Attuari.

[31] Jewell, W. S., "Bayesian Inverse Regression and Discrimination:
         An Application of Credibility Theory," (with R. Avenhaus),
         RM-75-27, International Institute for Applied Systems Analysis,
         Laxenburg, Austria, (June 1975).

[32] Jewell, W. S., "Application of Credibility Theory to Material
         Verification," (with R. Avenhaus), WP-75-89, International
         Institute for Applied Systems Analysis, Laxenburg, Austria,
         (July 1975).

[33] Jewell, W. S., "Credibility Regression with Simple Trends," ORC 75-16,
         (September 1975); to appear in ASTIN-Bulletin.

[34] Jewell, W. S., "Bayesian Regression and Credibility Theory," RM-75-63,
         International Institute for Applied Systems Analysis, Laxenburg,
         Austria, (November 1975).

[35] Jewell, W. S., "Linearized Bayesian Estimation of Network Flows,"
         ORC 76-2, (January 1976); IEEE Transactions: Circuits and
         Systems, Vol. 23, pp.   , (December 1976).

[36] Jewell, W. S., "Bayesian Life Testing Using the Total Q on Test,"
         ORC 76-3, (January 1976); The Theory and Applications of
         Reliability with Emphasis on Bayesian and Nonparametric Methods,
         Academic Press, New York.

[37] Jewell, W. S. and J. W. Kwiatkowski, "Estimates and Sampling Schemes
         for the Instrumentation of Accountability Systems," ORC 76-21,
         (July 1976).

[38] Kahn, P. M., (ed.), Credibility: Theory and Applications, Proceedings
         of Actuarial Research Conference on Credibility, Berkeley,
         September, 1974, Academic Press, New York, (1975).

[39] Kamseiter, H. and E. Straub, "On the Calculation of IBNR Reserves II,"
         Mitteilungen der Vereinigung Schweizerischer Versicherungs-
         mathematiker, Vol. 73, pp. 177-190, (1974).

[40] Lindley, D. V., "Bayesian Statistics: A Review," Regional Conference
         Series in Applied Mathematics, SIAM, Philadelphia, (1972).

[41] Longley-Cook, L. H., "An Introduction to Credibility Theory," PCAS,
         Vol. 49, pp. 194-221, (1962); available as a separate report
         from Casualty Actuarial Society, 200 East 42nd St., New York,
         N. Y. 10017.

[42] Mayerson, A. L., "A Bayesian View of Credibility," PCAS, Vol. 51, pp.
         85-104, (1964); Discussion, PCAS, Vol. 52, pp. 121-127, (1965).

[43] Mayerson, A. L., D. A. Jones and N. L. Bowers, Jr., "On the Credibility of the Pure Premium," PCAS, Vol. 55, pp. 175-185, (1968); Discussion, PCAS, Vol. 56, pp. 63-82, (1969).

[44] Miller, R. B. and J. C. Hickman, "Insurance Credibility Theory and Bayesian Extension," in [38].

[45] Norberg, R., "Credibility Premium Plans which Make Allowance for Bonus Hunger," Scandinavian Actuarial Journal, No. 2, pp. 73-86, (1976).

[46] Rao, C. R., LINEAR STATISTICAL INFERENCE AND ITS APPLICATIONS, Wiley, New York, (1965).

[47] Sage, A. P. and J. L. Melsa, ESTIMATION THEORY WITH APPLICATIONS TO COMMUNICATIONS AND CONTROL, McGraw-Hill, New York, (1971).

[48] Stone, M., "Robustness of Non-Ideal Decision Procedures," JASA, Vol. 58, No. 302, pp. 480-486, (June 1963).

[49] Straub, E., "On the Calculation of IBNR Reserves," IBNR - The Prize-Winning Papers in the Boleslaw Monie Fund Competition held in 1971. Nederlandse Reassurantie Groep, N. V., Amsterdam, (1972).

[50] Straub, E., "Credibility in Practice," in [38].

[51] Taylor, G. C., "Credibility for Time-Heterogeneous Loss Ratios," in [38].

[52] Theil, H., "On the Use of Incomplete Prior Information in Regression Analysis," Journal of the American Statistical Association, Vol. 58, pp. 401-414, (June 1963).

[53] Wedderburn, R. W. M., "Quasi-Likelihood Functions, Generalized Linear Models and the Gauss-Newton Method," Biometrika, Vol. 61, No. 3, pp. 439-447, (December 1974).

[54] Zellner, A., AN INTRODUCTION TO BAYESIAN INFERENCE IN ECONOMETRICS, Wiley, New York, (1971).