Credibility Using a Loss Function from Spline Theory: Parametric Models with a One-Dimensional Sufficient Statistic

Virginia R. Young*

Abstract

Current formulas in credibility theory often estimate expected claims as a function of the sample mean of the experience claims of a policyholder. An actuary may wish to estimate future claims as a function of some statistic other than the sample arithmetic mean of claims, such as the sample geometric mean. This can be suggested to the actuary through the exercise of regressing claims on the geometric mean of prior claims. It can also be suggested through a particular probabilistic model of claims, such as a model that assumes a lognormal conditional distribution. In the first case, the actuary may lean towards using a linear function of the geometric mean, depending on the results of the data analysis. On the other hand, through a probabilistic model, the actuary may want to use the most accurate estimator of future claims, as measured by squared-error loss. However, this estimator might not be linear.

In this paper, I provide a method for balancing the conflicting goals of linearity and accuracy. The credibility estimator proposed minimizes the expectation of a linear combination of a squared-error term and a second-derivative term. The squared-error term measures the accuracy of the estimator, while the second-derivative term constrains the estimator to be close to linear. I consider only those families of distributions with a one-dimensional sufficient statistic and estimators that are functions of that sufficient statistic or of the sample mean. Claim estimators are evaluated by comparing their conditional mean squared errors. In general, functions of the sufficient statistics prove to be better credibility estimators than functions of the sample mean.

1. Introduction

Current formulas in credibility theory often calculate the net premium as a weighted sum of the average experience of the policyholder and the average experience of the entire collection of policyholders. Because these formulas are linear, they are easy to use. Another advantage of linear formulas is that the claim estimate changes a fixed amount per change in claim experience; if an insurer uses such a formula, then the policyholder can readily calculate the change in premium. On the other hand, Venter (1990) points out that in some cases, the loss of accuracy makes a linear formula undesirable, such as in the mixture of a lognormal conditional over a lognormal prior.

An actuary may wish to estimate future claims as a function of some statistic other than the sample arithmetic mean of claims, such as the sample geometric mean. This can be suggested to the actuary through the exercise of regressing claims on the geometric mean of prior claims. That is, the geometric mean may prove to be a better explanatory variable of future claims than the arithmetic mean. It can also be suggested through a particular probabilistic model of claims, such as a model that assumes a lognormal conditional distribution. In the first case, the actuary may lean towards using a linear function of the geometric mean, depending on the results of data analysis. On the other hand, through a probabilistic model, the actuary may want to use the most accurate estimator of future claims, as measured by

*Virginia R. Young, F.S.A., Ph.D., is Assistant Professor of Business, School of Business, 975 University Avenue, University of Wisconsin–Madison, Madison, Wisconsin 53706.

squared-error loss. However, this estimator might not be linear.

In an earlier paper (Young 1997), I applied decision theory to develop a credibility formula that minimizes a linear combination of a squared-error term and a second-derivative term. The squared-error term measures the accuracy of the estimator, while the second-derivative term constrains the estimator to be close to linear. The sometimes conflicting goals of accuracy and linearity can be balanced by changing a single parameter in this loss function.[1] An actuary can choose the value of the parameter to target a specified curvature in the resulting claim estimator. This curvature may reflect the confidence that the actuary has in the model, as suggested by data analysis, versus his or her belief in the validity of a given probabilistic model.

The resulting credibility estimator is difficult to calculate analytically in all but a few simple cases. In this paper, I develop tools by which practitioners can use this credibility formula more easily in parametric models. In ongoing research, I examine how to apply the method using semiparametric models.

Earlier work in credibility theory is reviewed in Section 2. Section 3 describes the method for calculating credibility estimates of expected insurance claims. Sections 4 and 5 consider two parametric models of claims, for which there is a one-dimensional sufficient statistic. In the first example, the mixture of a lognormal conditional over a lognormal prior, a sufficient statistic is the geometric mean of prior claims. In the second example, the mixture of an inverse gamma conditional over a gamma prior, a sufficient statistic is the harmonic mean. In both examples, functions of the sufficient statistics prove to be better credibility estimators, in general, than functions of the sample mean. These flexible probabilistic models can be used to develop claim estimators that are functions of the geometric mean or the harmonic mean, respectively.

2. BACKGROUND

2.1 Notation and Assumptions

Assume that the total claims of a risk, or policyholder, in the i-th policy period, is a random variable $X_i|(\Theta=\theta)$, or more simply, $X_i|\theta$, $i=1, \ldots, n+1$. In some instances, $X_i|\theta$ may represent the random claim size of the i-th claim. For a fixed value $\Theta=\theta$, assume that the random variables $X_i|\theta$, $i=1, \ldots, n+1$, are independent and identically distributed according to a probability density function (pdf) $f(x|\theta)$, $x\in[a, b]$, a possibly infinite interval. Also, assume that the marginal density of X is non-zero on $[a, b]$, except possibly at $x=a$ or b, and that all necessary derivatives and expectations exist.

Assume that the value θ is fixed for a given risk, although it is generally unknown. Denote the pdf of Θ by $\pi(\theta)$, also called the *structure function* (Bühlmann 1970). Bayesians interpret $\pi(\theta)$ as the pdf of the prior distribution of Θ—the distribution that represents one's uncertainty about the parameter Θ before observing claim data for a given policyholder. This work holds equally well for discrete Θ, but for simplicity of notation, Θ is assumed to be continuous.

The primary goal of credibility theory is to calculate a net premium for period $n+1$ of a policyholder, given that the policyholder's claim experience in the first n periods is $\mathbf{X}=\mathbf{x}=<x_1, x_2, \ldots, x_n>\in[a, b]^n$. We consider credibility formulas that are not necessarily linear, denoted by $d(\mathbf{x})$, in which d is a real-valued function on $[a, b]^n$. We use d throughout this paper to denote a generic claim estimator that is a function of the entire claim experience $\mathbf{x}$ or of a statistic of $\mathbf{x}$, such as the sample arithmetic mean, $\bar{x}$, or the geometric mean,

$$\sqrt[n]{\prod_{i=1}^{n} x_i}.$$

(If $X_i|\theta$ represents the i-th claim, then the credibility formula estimates the $n+1^{\text{st}}$ claim amount given the n claim amounts $x_1, x_2, \ldots, x_n$.)

2.2 Some Earlier Work in Credibility

To estimate the future claims of a risk, $X_{n+1}|\theta$, with unknown θ, Bühlmann (1967, 1970) minimizes the expectation of the squared-error loss function:

$$(d(\mathbf{x}) - \mathrm{E}[X_{n+1}|\theta])^2.$$

(Also see Mayerson (1964a) for an earlier paper that considers Bayesian credibility for some specific claim models and Mayerson (1964b) for applications of credibility in property insurance ratemaking.) The resulting premium $d(\mathbf{x})$ is the expectation of $X_{n+1}|\mathbf{x}$, called the *predictive mean*; thus,

[1]Our loss function is similar to one used in spline theory, although in a different context, Cox (1988), Cox and O'Sullivan (1990), and Wahba (1990).

$$d(\mathbf{x}) = \mathrm{E}[X_{n+1}|\mathbf{x}] = \int\int \mathrm{E}[X_{n+1}|\theta]\ \pi(\theta|\mathbf{x})\ d\theta.$$

For simplicity, write $\mu(\mathbf{x})=\mathrm{E}[X_{n+1}|\mathbf{x}]$ to represent the predictive mean. Throughout, we denote the predictive mean by μ, whether as a function of $\mathbf{x}$ itself or of a statistic of $\mathbf{x}$.

By restricting the form of the renewal premium $d(\mathbf{x})$ to be an affine combination of the claim experience, $x_1, \ldots, x_n$, and by using the same squared-error loss function, Bühlmann (1967, 1970) obtains the following linear credibility formula

$$d(\mathbf{x}) = (1 - Z)\ \mathrm{E}[X] + Z\bar{x},$$

in which $\mathrm{E}[X]=E_\Theta \mathrm{E}[X|\theta]$ is the overall mean; $Z=n/(n+k)$; $k=\mathrm{E}_\Theta \mathrm{Var}[X|\theta]/\mathrm{Var}_\Theta \mathrm{E}[X|\theta]$; and $\bar{x}$ is the sample mean. The numerator of k is called the *expected process variance*; the denominator, the *variance of the hypothetical means*. This linear credibility formula can be viewed as a projection of the predictive mean onto the space of linear functions. This idea is discussed further in Section 3.

In certain cases, the predictive mean $\mu(\mathbf{x})$ is linear and, thus, equals the linear credibility formula. Jewell (1974a, 1974b) verifies conditions under which this *exact credibility* occurs: Under certain regularity conditions, exact credibility occurs for probability distributions from the linear exponential family when the natural conjugate prior is used. Also, see Willmot (1994) and Herzog (1996) for introductions to credibility theory.

Venter (1990) demonstrates that linear credibility formulas may lose predictive ability when the predictive mean is not linear. He analyzes the case for which the conditional distribution is lognormal and the distribution of Θ is also lognormal. He calculates errors obtained when the linear Bühlmann credibility formula is used in place of the predictive mean.

Venter also notes that the predictive mean in the mixture of a lognormal conditional over a lognormal prior is the exponentiation of an affine combination of the natural logarithms of the data. He, therefore, proposes that the linear least-squares credibility be applied to the logarithms of the data and the resulting expression exponentiated. This credibility formula is called the *loglinear* credibility estimator. He examines the loglinear estimator in the case for which the conditional distribution is inverse gamma and the distribution of Θ is gamma. We also compute the loglinear credibility estimator for the inverse gamma-gamma example (Section 5) and find that it performs better than the function of the sample mean.

3. METHODOLOGY FOR CALCULATING CREDIBILITY ESTIMATORS

3.1 Loss Function from Spline Theory

I developed (Young 1997) a credibility formula that allows an actuary to balance the (sometimes) conflicting goals of accuracy (as measured by a squared-error term) and of linearity (as measured by a second-derivative term). We find the credibility estimator $d:[a, b]\rightarrow\mathbf{R}$ that minimizes the expected loss:

$$\mathrm{E}[(d(\bar{x}) - \mu(\bar{x}))^2] + h\mathrm{E}[(d''(\bar{x}))^2], \qquad (3.1)$$

in which $\mu(\bar{x})$ is the predictive mean of X_{n+1} given $\bar{X}=\bar{x}$, and we take the expectation with respect to the marginal density of $\bar{X}$. If we let h approach 0, then the optimal d is the predictive mean. At the other extreme, if we let h approach ∞, then the optimal d converges to the linear Bühlmann credibility estimator.

3.2 Parametric Models with a One-Dimensional Sufficient Statistic

In an earlier paper (Young 1997), I suggested that the possible credibility functions d be restricted to be functions of the statistic $\bar{x}$ because the resulting credibility estimator directly generalizes the one of Bühlmann. Also, the minimization problem becomes untractable if d is a function of several variables. Finally, for the case in which $X|\theta$ follows a distribution from a linear exponential family, the sample mean is a sufficient statistic for θ; therefore, the predictive mean is a function of the sample mean for any prior on Θ.

It is, therefore, natural to ask under what conditions there is a one-dimensional sufficient statistic for θ. Recall that we assume that the $X_i|\theta$ are iid with pdf $f(x|\theta)$ on $[a, b]$. If a and b are independent of θ and if there exists a one-dimensional sufficient statistic for the joint density of $X_1, \ldots, X_n$, given θ, then $X|\theta$ follows a distribution from a one-parameter exponential family, (Lehmann 1991, p. 44).

In this case, $X|\theta$ has pdf given by

$$f(x|\theta) = \exp[\eta(\theta)T(x) - B(\theta)] \cdot h(x),$$

in which η and B are real-valued functions of θ, and $T(x)$ is a real-valued statistic. Given claim data $\mathbf{X}=\mathbf{x}=\langle x_1, x_2, \ldots, x_n\rangle\in[a, b]^n$, a sufficient statistic for θ is $T = \Sigma_{i=1}^n T(x_i)$, or some monotone transformation of T. For the special case of the linear exponential family, $T(x)=x$, and the sample mean is, therefore, a sufficient statistic. In general, for a one-parameter exponential family, the pdf of the predictive distribution of $X_{n+1}|\mathbf{x}$ is a function of the sufficient statistic T, from which it follows that the predictive mean is also a function of T, $\mu(\mathbf{x})=\mu(t)$. Indeed,

$$\mu(\mathbf{x}) = \mathrm{E}[X_{n+1}|\mathbf{x}] = \int \mathrm{E}[X_{n+1}|\theta]\pi(\theta|\mathbf{x})d\theta,$$

and

$$\pi(\theta|\mathbf{x}) \propto \pi(\theta) \prod_{i=1}^{n} f(x_i|\theta) \propto \pi(\theta) \exp[\eta(\theta)t - nB(\theta)].$$

As noted above, we can use any monotone transformation of T as a sufficient statistic. For example, if $X|\theta$ follows the conditional lognormal distribution, then a sufficient statistic for θ is

$$\sum_{i=1}^{n} \ln(x_i) = \ln\left(\prod_{i=1}^{n} x_i\right),$$

but the geometric mean,

$$\left(\prod_{i=1}^{n} x_i\right)^{1/n},$$

is also a sufficient statistic and possibly more appealing to use in predicting future claims because of its parallel with the arithmetic mean.

Remark

To see that it is necessary to require $[a, b]$ to be independent of θ for the conditional distribution to be from an exponential family, consider $X|\theta \sim U(0, \theta)$, whose support $(0, \theta)$ depends on θ. A sufficient statistic for θ is $\max(x_1, \ldots, x_n)$, a one-dimensional statistic, but the uniform distribution on $(0, \theta)$ is not from an exponential family. □

In the parametric examples that follow, we find d: $[a, b]\rightarrow\mathbf{R}$ to minimize the expected loss:

$$\mathrm{E}[(d(t) - \mu(t))^2] + h\mathrm{E}[(d''(t))^2], \qquad (3.2)$$

in which $T=t$ lies in the interval $[a, b]$.

In Sections 4 and 5, we approximate solutions to the minimization problems (3.1) and (3.2) for the following mixtures of $X|\theta$ over Θ. (See the indicated section for the specific parametrization.) We choose to use the natural conjugate prior for each likelihood, because in most cases the natural conjugate prior is flexible enough to allow for a wide range of prior beliefs and because the calculations are usually more tractable (Berger 1985).

- Lognormal-lognormal, Section 4: This is the first case that Venter (1990) considers when he calculates the linearization error of the Bühlmann credibility estimator. This model can be used if the statistic of interest is the geometric mean because it is a sufficient statistic for the unknown parameter θ.
- Inverse gamma-gamma, Section 5: Venter (1990) considers this case when he computes the linearization error of the linear Bühlmann credibility estimator and of the loglinear credibility estimator. This model can be used if the statistic of interest is the harmonic mean because the harmonic mean is a sufficient statistic.

The next subsection describes how to minimize (3.1) and (3.2) approximately. Following that, we define the conditional mean squared error, which we use to measure accuracy of the various credibility estimators.

3.3 Approximating *d* with a Polynomial

Let U denote a sufficient statistic T or the sample mean $\bar{X}$. Then, we can combine the minimization problems (3.1) and (3.2) into the single problem of finding d:$[a, b]\rightarrow\mathbf{R}$ to minimize

$$\mathrm{E}[(d(u) - \mu(u))^2] + h\mathrm{E}[(d''(u))^2], \qquad (3.3)$$

in which $U=u$ takes values in the (possibly infinite) interval $[a, b]$. Practically, we restrict the interval $[a, b]$ to be finite by considering an interval that contains nearly all the support of the distribution of U.

The solution to (3.3) is difficult to calculate analytically in all but a few special cases (Young 1997); therefore, we approximate the optimal d by applying finite element methods from numerical analysis (Prenter 1975). We employ the Rayleigh-Ritz method, in which one approximates d by projecting the predictive mean onto a finite-dimensional subspace of a given space of functions, using an appropriate norm. (See the Appendix for more details.) This method generalizes the one used in linear least-squares credibility, in which one projects the predictive mean onto the two-dimensional subspace of linear functions. Similar geometric ideas applied to credibility theory appear in the work of Gerber and Jones (1975), De Vylder (1976a, 1976b), Taylor (1977), and De Vylder and Sundt (1982).

For the finite-dimensional space of functions on which to project the predictive mean, we choose the space of m-th degree polynomials on a finite interval

$[a, b]$, for m a fixed integer, $m=1, 2, 3, \ldots$. By the Weierstrass approximation theorem, we can approximate any continuous function on a closed, finite interval as closely as we want by a polynomial with sufficiently high degree m (Hewitt and Stromberg 1965). Consider the basis for the m-th degree polynomials on $[a, b]$ given by the Bernstein-Bezier (BB) polynomials (de Boor, 1978):

$$B_{im}(t) = \frac{m!}{i!(m-i)!}\frac{(b-t)^i\,(t-a)^{m-i}}{(b-a)^m}, \quad i = 0, 1, \ldots, m.$$

Write the approximation $\hat{d}$ of the minimizer d as a linear combination of m-th degree BB polynomials:

$$\hat{d} = \sum_{i=0}^{m} c_i\, B_{im}. \tag{3.4}$$

Then the system of equations in (A.2) becomes

$$\sum_{j=0}^{m}\left\{\int_a^b [B_{jm}\,(u)\,B_{im}\,(u) + h\,B''_{jm}\,(u)\,B''_{im}\,(u)]\,f\,(u)\,du\right\} c_j = \int_a^b [\mu(u)B_{im}\,(u)]\,f\,(u),\,du, \tag{3.5}$$

for each $i=0, 1, \ldots, m$. This system defines the matrix equation

$$A\mathbf{c} = \mathbf{v},$$

in which $A=(a_{ij})$, with $a_{ij}=\mathrm{E}[B_{jm}\;B_{im}+hB''_{jm}B''_{im}]$, and $v_i=\mathrm{E}[\mu B_{im}]$.

3.4 Conditional Mean Squared Error

Consider the following mean squared error, conditional on θ, in which the error is the difference between the claim in period $n+1$, x_{n+1}, and the estimate $g(u)$:

$$\int_X\int_U (x_{n+1} - g(u))^2\,f(u|\theta)\,f(x_{n+1}|\theta)\,du\,dx_{n+1} = \int_X (x_{n+1} - \mathrm{E}(X_{n+1}|\theta))^2\,f(x_{n+1}|\theta)\,dx_{n+1} + \int_U (\mathrm{E}(X_{n+1}|\theta) - g(u))^2\,f(u|\theta)\,du.$$

The first term in this sum is independent of the estimator g and of the statistic u, so in the following sections, we compare values of the conditional mean squared error

$$MSE(\theta) = \int_a^b [\mathrm{E}(X_{n+1}|\theta) - g(u)]^2\,f(u|\theta)\,du, \tag{3.6}$$

for various estimators g and statistics u, at six percentiles of Θ.

4. LOGNORMAL-LOGNORMAL

The lognormal distribution is often used by actuaries to model the distribution of claim severity. It is also used to model the distribution of total claims in some lines of insurance, such as health insurance. In this section, we compute claim estimators for a mixture of a lognormal conditional over a lognormal prior and compare them via their conditional mean squared errors. We model the lognormal-lognormal mixture as follows:

$$f(x|\theta) = \frac{1}{\sigma x\sqrt{2\pi}}\exp\left\{-\frac{1}{2\sigma^2}\left[\ln\left(\frac{x}{\theta}\right)\right]^2\right\},\ x > 0,$$

in which $\sigma>0$ is a known parameter, and

$$\pi(\theta) = \frac{1}{\tau\theta\sqrt{2\pi}}\exp\left\{-\frac{1}{2\tau^2}\left[\ln\left(\frac{\theta}{\mu}\right)\right]^2\right\},\ \theta > 0,$$

in which $\mu>0$ and $\tau>0$ are known parameters. That is, $(\ln X)|\theta\sim N(\ln\theta,\ \sigma^2)$, and $\ln\Theta\sim N(\ln\mu,\ \tau^2)$. The marginal distribution of X is also lognormal; $\ln X\sim N(\ln\mu,\ \sigma^2+\tau^2)$.

Given claim data for a specific policyholder, $\mathbf{X}=\mathbf{x}=\langle x_1, x_2, \ldots, x_n\rangle\in[0, \infty]^n$, a sufficient statistic for θ is the geometric mean,

$$t = \left(\prod_{i=1}^{n} x_i\right)^{1/n}.$$

The distribution of T given θ, is lognormal with

$$\ln T|\theta \sim N\left(\ln\theta, \frac{\sigma^2}{n}\right).$$

Therefore, the marginal distribution of T is also lognormal with

$$\ln T \sim N\left(\ln\mu, \frac{\sigma^2}{n} + \tau^2\right).$$

The posterior distribution of $\Theta|\mathbf{x}$ is lognormal because we use the conjugate prior; $(\ln\Theta)|\mathbf{x}\sim N(\ln\mu^*,\ \tau^{*2})$, in which

$$\mu^* = t^{(n\tau^2)/(\sigma^2+n\tau^2)}\;\mu^{(\sigma^2)/(\sigma^2+n\tau^2)},$$

and

$$\tau^{*2} = \frac{\sigma^2\,\tau^2}{\sigma^2 + n\tau^2}.$$

Thus, the predictive distribution of $X_{n+1}|\mathbf{x}$ is lognormal; $(\ln X_{n+1})|\mathbf{x} \sim N(\ln\mu^*, \sigma^2+\tau^{*2})$. It follows that the predictive mean is a function of t, as expected:

$$\begin{aligned}\mu(t) &= \mathrm{E}(X_{n+1}|\mathbf{x}) \\ &= t^{(n\tau^2)/(\sigma^2+n\tau^2)}\ \mu^{(\sigma^2)/(\sigma^2+n\tau^2)} \\ &\quad \exp\left(\frac{\sigma^2\ (\sigma^2 + (n+1)\ \tau^2)}{2(\sigma^2 + n\tau^2)}\right). \qquad (4.1)\end{aligned}$$

Note that the predictive mean approaches a multiple of t as n gets large. Also, note that the linear Bühlmann credibility estimator is given by

$$(1 - Z)\ \mu\ \exp\left(\frac{\sigma^2 + \tau^2}{2}\right) + Z\overline{x},$$

in which $Z = \dfrac{n}{n+k}$, and $k = \dfrac{e^{\tau^2}\ (e^{\sigma^2}-1)}{e^{\tau^2}-1}$.

Example 4.1 (Similar to Venter 1990)

Let the lognormal-lognormal parameters be $\sigma^2=4$, $\mu=1$, and $\tau^2=2$. In this case, $\mathrm{E}(X)=20.086$ on some scale, say \$000s. Suppose we have $n=3$ years of claim experience. The expectation of the process variance is $\mathrm{E}(\mathrm{Var}(X|\theta))=159{,}774$, and the variance of the hypothetical means is $\mathrm{Var}(\mathrm{E}(X|\theta))=2{,}578$. Therefore, the linear Bühlmann credibility weight is rather small, $Z=0.046$.

Let $\hat{d}$ be the function of the sufficient statistic $T=t$ that is the credibility estimator obtained from the Rayleigh-Ritz method by projecting the predictive mean onto the space of fifth-degree polynomials on $[e^{-5}, e^5]$, as given by (3.4) and (3.5). We chose the fifth-degree polynomials so that the overall mean squared error

$$\int_{e^{-5}}^{e^5} (\mu(t) - \hat{d}(t))^2\ f(t)\ dt$$

is small relative to

$$\int_{e^{-5}}^{e^5} \mu(t)^2\ f(t)\ dt$$

when $h=0$. The interval $[e^{-5}, e^5]$ contains nearly all the distribution of T according to numerical calculation on Mathcad Plus 6.0; indeed,

$$\int_{e^{-5}}^{e^5} f(t)\ dt = 0.994.$$

The linearization penalty h can be chosen so that the curvature of $\hat{d}$ is some fraction of that of the predictive mean or so that the curvature is no more than some absolute amount. The curvature of a function g can be measured by

$$\int (g''(t))^2\ f(t)\ dt;$$

in this case, the parameter h acts as a Lagrange multiplier. For this example, we used the maximum of the second-derivative squared $\max_{t\in[a,b]} (g''(t))^2$ as a measure of curvature and targeted the specific value of 0.005. When $h=0$, the maximum of the second-derivative squared of $\hat{d}$ is 0.085. To achieve a curvature of no more than 0.005, the value of h is 47,500. In general, the curvature can be chosen to reflect the desired degree of linearity versus accuracy; the targeted curvature can also reflect the confidence we have in the probabilistic model.

See Table 1 for a comparison of the conditional mean squared errors for estimators that are functions of the geometric mean t and the arithmetic mean $\overline{x}$. The credibility estimator that is a linear function of t is the limiting estimator as h approaches infinity in (3.2).

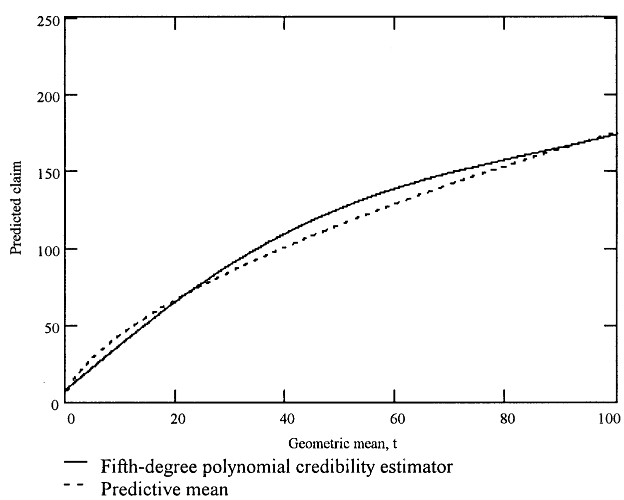
TABLE 1
LOGNORMAL-LOGNORMAL CONDITIONAL MEAN SQUARED ERRORS

	Percentile of θ					
	1	10	50	75	90	99
θ	0.037	0.163	1.000	2.596	6.125	26.84
$E(X\|\theta)$	0.275	1.206	7.389	19.18	45.26	198.3
Conditional Mean Squared Errors						
Functions of the sample arithmetic mean						
Sample mean	1.354	26.00	975.5	6,572	36,595	702,786
Bühlmann credibility	357.1	324.3	148.7	14.75	654.5	30,404
Functions of the sample geometric mean						
Predictive mean	5.125	26.12	164.4	397.3	908.9	14,057
Linear credibility	90.52	88.91	77.26	220.7	1,046	17,115
$\hat{d}$	54.12	57.30	113.9	365.2	1,097	14,262

See Figure 1 for a graph of the predictive mean and of the fifth-degree polynomial $\hat{d}$. From Table 1, we learn that no claim estimator is uniformly best, as measured by conditional mean squared error, over all the values of θ considered. In general, the estimators that are functions of the geometric mean have lower mean squared errors than those that are functions of the arithmetic mean. The mean squared errors of the fifth-degree polynomial $\hat{d}$ lie between those of the predictive mean and those of the credibility estimator that is a linear function t, for all but one value of θ. Of all these estimators, we recommend using the fifth-degree polynomial because its conditional mean squared errors are of the same order of magnitude as those of the predictive mean (except for small θ) and because its curvature is very low relative to that of the predictive mean, which has infinite curvature at zero (and curvature of 8.417×10^6 at $a=e^{-5}$).

FIGURE 1

FIFTH-DEGREE POLYNOMIAL CLAIM ESTIMATOR



5. INVERSE GAMMA-GAMMA

In this section, we compute claim estimators for a mixture of an inverse gamma conditional over a gamma prior and compare them via their conditional mean squared errors. The inverse gamma-gamma mixture is given as follows:

$$f(x|\theta) = \frac{\theta^r}{\Gamma(r)} x^{-r-1} e^{-\theta/x}, \; x > 0,$$

in which $r>0$ is a known parameter, and

$$\pi(\theta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \theta^{\alpha-1} e^{-\beta\theta}, \; \theta > 0,$$

in which $\alpha>0$ and $\beta>0$ are known parameters. That is, $(1/X)|\theta \sim G(r, \theta)$, and $\Theta \sim G(\alpha, \beta)$. The marginal distribution of X has pdf

$$f(x) = \frac{\Gamma(r+\alpha)}{\Gamma(r)\Gamma(\alpha)} \cdot \frac{\beta^\alpha x^{\alpha-1}}{(1+\beta x)^{r+\alpha}}, \; x > 0,$$

and expectation

$$EX = \frac{\alpha}{\beta(r-1)}, \; r > 1.$$

Given claim data for a specific policyholder, $\mathbf{X}=\mathbf{x}=\langle x_1, x_2, \ldots, x_n\rangle \in [0, \infty]^n$, a sufficient statistic for θ is the harmonic mean

$$t = n \Big/ \sum_{i=1}^{n} \frac{1}{x_i}.$$

The distribution of T given θ is inverse gamma with $(1/T)|\theta \sim G(nr, n\theta)$. Therefore, the marginal distribution of T is given by the pdf

$$f(t) = \frac{\Gamma(\alpha+nr)}{\Gamma(\alpha)\Gamma(nr)} \cdot \frac{\beta^\alpha (n/t)^{nr}}{t(\beta+n/t)^{\alpha+nr}}, \; t > 0.$$

The posterior distribution of $\Theta|\mathbf{x}$ is gamma because we use the conjugate prior; $\Theta|\mathbf{x} \sim G(\alpha+nr, \beta+n/t)$. Thus, the predictive distribution of $X_{n+1}|\mathbf{x}$ is similar to the marginal distribution of X with α replaced with $(\alpha+nr)$ and β replaced with $(\beta+n/t)$. It follows that the predictive mean is a function of t, as expected:

$$\mu(t) = \mathrm{E}(X_{n+1}|\mathbf{x}) = \frac{\alpha+nr}{(\beta+n/t)(r-1)}. \tag{5.1}$$

Note that the predictive mean approaches a multiple of t as n approaches infinity.

The linear Bühlmann credibility estimator is given by

$$(1-Z)\frac{\alpha}{\beta(r-1)} + Z\bar{x},$$

in which $Z=n/(n+k)$, and $k=(\alpha+1)/(r-2)$. The log-linear credibility estimator (Venter 1990) is given by

$$\frac{\alpha\beta^{z-1}}{r-1}\left[\frac{\Gamma(r)\Gamma(\alpha)}{\Gamma(r-z/n)\Gamma(\alpha+z/n)}\right]^n \exp\left(\frac{z}{n}\sum_{i=1}^{n} \ln x_i\right),$$

in which $z=n/(n+\tilde{k})$, and $\tilde{k}=[\Psi'(r)]/[\Psi'(\alpha)]$. The term Ψ is the digamma function, or the derivative of the natural logarithm of the gamma function.

Example 5.1 (Similar to Venter 1990)

Let the inverse gamma-gamma parameters be $r=3$, $\alpha=10$, and $\beta=0.1$. In this case, $E(X)=50$ on some scale, say $000s. Suppose we have $n=3$ years of claim experience. The expectation of the process variance is $E(Var(X|\theta))=2{,}750$, and the variance of the hypothetical means is $Var(E(X|\theta))=250$. Therefore, the linear Bühlmann credibility weight is a bit larger than in Example 4.1, $Z=0.214$.

Let $\hat{d}$ be the function of the sufficient statistic $T=t$ that is the credibility estimator obtained from the Rayleigh-Ritz method by projecting the predictive mean onto the space of fifth-degree polynomials on $[e^2, e^5]$, as given by (3.4) and (3.5). We chose the fifth-degree polynomials so that the overall mean squared error

$$\int_{e^2}^{e^5} (\mu(t) - \hat{d}(t))^2 f(t)\, dt$$

is small relative to

$$\int_{e^2}^{e^5} (\mu(t))^2 f(t)\, dt$$

when $h=0$. The interval $[e^2, e^5]$ contains nearly all of the distribution of T according to numerical calculation on Mathcad Plus 6.0; indeed,

$$\int_{e^2}^{e^5} f(t)\, dt = 0.998.$$

As in Example 4.1, we used the maximum of the second-derivative squared $\max_{t\in[a,b]} (g''(t))^2$ as a measure of curvature and targeted the specific value of 0.0001. When $h=0$, the maximum of the second-derivative squared of $\hat{d}$ is 0.00463. To achieve a curvature of no more than 0.0001, the value of h is 110,000.

See Table 2 for a comparison of the conditional mean squared errors for the estimators that are functions of the arithmetic mean $\bar{x}$ and the harmonic mean t. The credibility estimator that is a linear function of t is the limiting estimator as h approaches infinity in (3.2). We also include the loglinear credibility estimator.

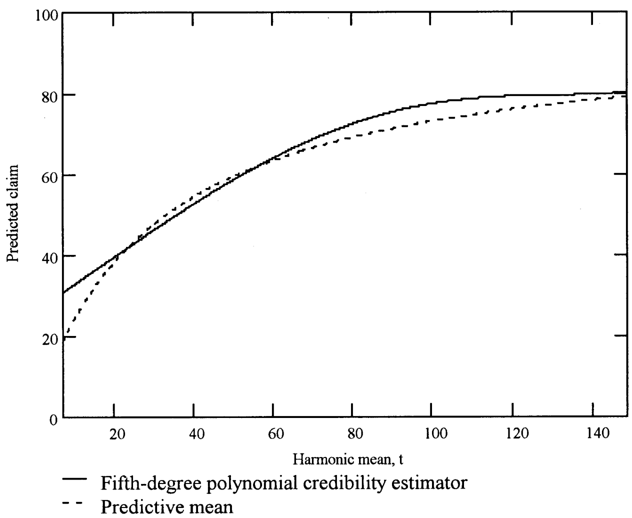
See Figure 2 for a graph of the predictive mean and of the fifth-degree polynomial $\hat{d}$. From Table 2, we learn that no claim estimator is uniformly best, as measured by conditional mean squared error, over all the values of θ considered, as occurred in Example 4.1. In general, the estimators that are functions of the harmonic mean have lower mean squared errors than those that are functions of the arithmetic mean. The mean squared errors of the fifth-degree polynomial $\hat{d}$ lie between those of the predictive mean and those of the credibility estimator that is a linear function t, for all but one value of θ. Again, the fifth-degree polynomial could be used because its conditional mean squared errors are of the same order of magnitude as those of the predictive mean, and its curvature is very low relative to that of the predictive mean, which has curvature 0.012 at $a=e^2$. Also, in this case, the loglinear credibility formula performs well.

TABLE 2
INVERSE GAMMA-GAMMA CONDITIONAL MEAN SQUARED ERRORS

	Percentile of θ					
	1	**10**	**50**	**75**	**90**	**99**
θ	41.30	62.21	96.69	119.1	142.1	187.8
$E(X\|\theta)$	20.65	31.11	48.34	59.57	71.03	93.92
Conditional Mean Squared Errors						
Functions of the sample arithmetic mean						
Sample mean	142.2	322.5	779.0	1,183	1,682	2,940
Bühlmann credibility	538.3	235.2	37.47	110.8	350.3	1,325
Functions of the sample harmonic mean						
Predictive mean	167.7	145.4	65.79	77.19	196.4	882.1
Linear credibility	321.6	153.2	53.52	109.1	263.3	870.6
$\hat{d}$	256.8	140.5	62.81	98.80	224.6	837.8
Function of the sample geometric mean						
Loglinear credibility	222.5	151.1	73.95	104.9	230.6	828.5

FIGURE 2

FIFTH-DEGREE POLYNOMIAL CLAIM ESTIMATOR



6. SUMMARY AND CONCLUSIONS

In this paper, we demonstrated how to approximate the minimizing solution of

$$E[(d(u) - \mu(u))^2] + hE[(d''(u))^2],$$

by using techniques from approximation theory. Specifically, we showed how to calculate the projection of d onto a finite-dimensional space of functions by using the norm defined by the loss function. We restricted our work to those families of distributions for which there exists a one-dimensional sufficient statistic, such as the geometric or harmonic mean, and considered estimators that are functions of that sufficient statistic or of the sample mean. We evaluated various claim estimators by comparing their conditional mean squared errors, evaluated at six percentiles of the parameter.

In practice, an actuary may wish to apply techniques from point estimation to fit parametric loss distributions (Hogg and Klugman 1984) and then apply the method described in this paper to compute a credibility estimator. Thereby, the actuary will trade accuracy for linearity—this trade-off can be especially desirable if the actuary has less than full confidence in the probabilistic claim model. In future research, we examine the problem of determining credibility estimators in a semiparametric setting.

ACKNOWLEDGMENTS

I thank the Committee for Knowledge Extension and Research of the Society of Actuaries and the Actuarial and Education Research Fund for financial support and guidance. I especially thank the Project Oversight Group (Hans Gerber and Gary Venter, led by Thomas Herzog and assisted by Warren Luckner of the SOA) for insightful and helpful comments. I also thank the Graduate School of the University of Wisconsin–Madison for financial support of this project.

REFERENCES

DE BOOR, C. 1978. *A Practical Guide to Splines*, New York: Springer-Verlag.

BERGER, J. O. 1985. *Statistical Decision Theory and Bayesian Analysis*, 2nd ed. New York: Springer-Verlag.

BÜHLMANN, H. 1967. "Experience Rating and Credibility," *ASTIN Bulletin* 4:199–207.

BÜHLMANN, H. 1970. *Mathematical Models in Risk Theory*, New York: Springer-Verlag.

COX, D. 1988. "Approximation of Method of Regularization Estimators," *Annals of Statistics*, 16:713–32.

COX, D., AND O'SULLIVAN, F. 1990. "Asymptotic Analysis of Penalized Likelihood and Related Estimators," *Annals of Statistics* 18:1676–95.

DE VYLDER, F. E. 1976a. "Geometrical Credibility," *Scandinavian Actuarial Journal*, 121–49.

DE VYLDER, F. E. 1976b. "Optimal Semilinear Credibility," *Journal of the Swiss Association of Actuaries* 27–40.

DE VYLDER, F. E. AND SUNDT, B. 1982. "Constrained Credibility Estimators in the Regression Model," *Scandinavian Actuarial Journal*, 23–37.

GERBER, H. U., AND JONES, D. A. 1975. "Credibility Formulas of the Updating Type," *Transactions of the Society of Actuaries* XXVII:31–52.

HERZOG, T. L. 1996. *Introduction to Credibility Theory*, 2nd ed. Abington, Conn.: Actex.

HEWITT, E., AND STROMBERG, K. 1965. *Real and Abstract Analysis*. New York: Springer-Verlag.

HOGG, R., AND KLUGMAN, S. 1984. *Loss Distributions*. New York: Wiley.

JEWELL, W. S. 1974a. "Credibility Is Exact Bayesian for Exponential Families," *ASTIN Bulletin* 8:77–90.

JEWELL, W. S. 1974b. "Regularity Conditions for Exact Credibility," *ASTIN Bulletin* 8:336–41.

LEHMANN, E. L. 1991. *Theory of Point Estimation*. Pacific Grove, Calif.: Wadsworth.

MAYERSON, A. L. 1964a. "A Bayesian View of Credibility," *Proceedings of the Casualty Actuarial Society* 51:85–104.

MAYERSON, A. L. 1964b. "The Uses of Credibility in Property Insurance Ratemaking," *Giornale dell'Instituto Italiano degli Attuari* 27.

PRENTER, P. M. 1975. *Splines and Variational Methods*. New York: Wiley.

TAYLOR, G. C. 1977. "Abstract Credibility," *Scandinavian Actuarial Journal*, 149–68.

VENTER, G. 1990. "Credibility," in *Foundations of Casualty Actuarial Science*. New York: Casualty Actuarial Society.

WAHBA, G. 1990. *Spline Models for Observational Data*. Philadelphia, Penn.: Society for Industrial and Applied Mathematics.
WILLMOT, G. E. 1994. *Introductory Credibility Theory*. Waterloo, Ontario: Institute of Insurance and Pension Research, University of Waterloo.
YOUNG, V. R. 1997. "Credibility Using a Loss Function from Spline Theory," *Scandinavian Actuarial Journal*, 160–85.

APPENDIX

MINIMIZING (3.2) USING THE FRAMEWORK OF AN INNER PRODUCT SPACE OF FUNCTIONS

Most of the material in this appendix is from Prenter (1975) and de Boor (1978); also see Young (1997) for additional background specific to this problem. Let $L_2[a, b]$ represent the space of real-valued functions on $[a, b]$ that are square-integrable with respect to the measure determined by $f(u)\,du$. Define an inner product on $L_2[a, b]$ by

$$\langle z_1, z_2\rangle_f = \int_a^b z_1(u)\, z_2(u)\, f(u)\, du = \mathrm{E}[z_1(u)\, z_2(u)].$$

In Young (1997), we show that the solution d to the minimization problem (3.3) lies in the space of functions F defined as follows:

Definition

Let F be the linear subspace of $L_2[a, b]$ of functions $\{z:[a, b] \to \mathbf{R}\}$ that satisfy the properties:

(a) $$\int_a^b [z(u)]^2 f(u)\, du < \infty.$$

(b) If z is in F, then

$$\lim_{x\to a,b} [f(u)\, z''(u)] = 0,$$

and

$$\lim_{x\to a,b} [f(u)\, z''(u)]' = 0.$$

(c) If z_1 and z_2 are in F, then

$$\lim_{x\to a,b} z_1'(u)\, [f(u)\, z_2''(u)] = 0,$$

and

$$\lim_{x\to a,b} z_1(u)\, [f(u)\, z_2''(u)]' = 0.$$

We also assume that if z is in F and if $\mu(u)$ is the predictive mean, then

$$\lim_{x\to a,b} \mu'(u)\, [f(u)\, z''(u)] = 0,$$

and

$$\lim_{x\to a,b} \mu(u)\, [f(u)\, z''(u)]' = 0.$$

□

Remark

Note that if we define the linear operator L_f on F by

$$L_f(z)(u) = \frac{h}{f(u)}\,(f(u)z''(u))'' + z(u),$$

then

$$\langle z_1, L_f\, z_2\rangle_f = \langle L_f\, z_1, z_2\rangle_f \geq 0,$$

for z_1, z_2 in F. Thus, $\langle z_1, z_2\rangle_L \equiv \langle L_f\, z_1, z_2\rangle_f$ defines a new inner product on the subspace F. □

Now, let $\{\phi_0, \phi_1, \ldots, \phi_m\}$ be a linearly independent subset of functions in F, and let F_m be the span of $\{\phi_0, \phi_1, \ldots, \phi_m\}$. The Rayleigh-Ritz method consists of finding the least-squares fit $d_m \in F_m$ to the minimizing solution $d \in F$ with respect to the inner product given by $\langle z_1, z_2\rangle_L = \langle L_f\, z_1, z_2\rangle_f$. That is, d_m solves the following minimization problem:

$$\min_{\tilde{d}\in F_m} < \tilde{d} - d, \tilde{d} - d >_L .$$

This problem is equivalent to finding $d_m \in F_m$ that minimizes

$$\mathrm{E}[(d(u) - \mu(u))^2] + h\mathrm{E}[(d''(u))^2]. \qquad \text{(A.1)}$$

For example, if F_m were the subspace of linear functions, then d_m would be the linear least-squares estimate of μ.

If we write

$$d_m = c_0\, \phi_0 + c_1\, \phi_1 + \ldots + c_m\, \phi_m,$$

in which the c_i, $i=0, 1, \ldots, m$, are real constants, then the c_i satisfy the following system of equations

$$\sum_{j=0}^{m} < \phi_j, \phi_i >_L \quad c_j = < \mu, \phi_i >_f,$$

for $i=0, 1, \ldots, m$. This system becomes

$$\sum_{j=0}^{m} \left\{ \int_a^b [\phi_j(u)\, \phi_i(u) + h\phi_j''(u)\, \phi_i''(u)]\, f(u)\, du \right\} c_j = \int_a^b [\mu(u)\phi_i(u)]\, f(u)\, du, \qquad \text{(A.2)}$$

for $i=0, 1, \ldots, m$. By noticing that

$$< \phi_j, \phi_i >_L = \int_a^b [\phi_j(u)\, \phi_i(u) + h\phi_j''(u)\, \phi_i''(u)]\, f(u)\, du$$

defines an inner product on a less restrictive space than F_m, we can expand F_m to include functions ϕ on $[a, b]$ for which

$$\int_a^b [\phi(u)^2 + h\phi''(u)^2]\, f(u)\, du < \infty.$$

DISCUSSIONS

F. ETIENNE DE VYLDER*

Dr. Young considers a very general problem in credibility theory, and she solves it numerically by a polynomial approximation under particular distribution assumptions. Hereafter I reconstruct the steps leading to this beautiful problem, and I indicate another solution.

1. A Triple Generalization of Bühlmann's Linear Credibility Problem

We consider the random vector

$$(\Theta, X_1, X_2, \ldots, X_n, X_{n+1}), \tag{i}$$

where $X_1, \ldots, X_{n+1}$ are square-integrable random variables conditionally iid for fixed Θ. The distribution of (i) is determined by the couple

$$(\text{distribution of } \Theta, \text{ distribution of } X_i/\Theta=\theta), \tag{ii}$$

where the conditional distribution depends on θ but not on $i=1, \ldots, n+1$. The random variables $X_1, \ldots, X_n$ are regarded as observable, but not X_{n+1}. Of course, the structure variable Θ is not observable.

Bühlmann approximates $E(X_{n+1}/\Theta)$ by $\alpha+\beta T$, where $T=(X_1+\ldots+X_n)/n$. Precisely, Bühlmann's problem is to minimize the expectation

$$E[E(X_{n+1}/\Theta) - (\alpha + \beta T)]^2 \tag{iii}$$

by the optimal choice of the real numbers α and β. The following is a triple generalization of this problem.

(a) Instead of the mean T, more general statistics $T=\varphi(X_1, \ldots, X_n)$ are considered.
(b) Instead of affine functions $\alpha+\beta T$, more general functions $d(T)$ of T are allowed.
(c) A compromise between proximity to $E(X_{n+1}/\Theta)$ and linearity is introduced.

The general problem is to minimize

$$E[E(X_{n+1}/\Theta) - d(T)]^2 + h\, E[d''(T)]^2 \tag{iv}$$

by the optimal choice of the function $d(\cdot)$ (with obvious properties implied by the statement of the problem). Here $h\geq 0$ is a fixed constant. The linearity impact is weak when h is small. It is strong when h is large.

2. A Related Credibility Problem

Instead of (iv), Dr. Young considers the minimization of

$$E[E(X_{n+1}/X_1, \ldots, X_n) - d(T)]^2 + h\, E[d''(T)]^2 \tag{v}$$

by the optimal choice of $d(\cdot)$. Hence, she replaces the conditional expectation $E(X_{n+1}/\Theta)$ by the predictable mean $E(X_{n+1}/X_1, \ldots, X_n)$. In the Bühlmann case, the problems defined by (iv) and (v) are equivalent; that is, they have the same solution. This equivalence remains if the predictive mean is replaced by X_{n+1} in (v). In the general case, the solutions of both problems are probably close to each other. Anyway, the problem defined by (v) is interesting enough, and it can be considered for itself.

Dr. Young's numerical tests based on the evaluation of the conditional mean squared error $MSE(\theta)$ seem to indicate that she indeed regards the solution of the problem defined by (v) as an approximation of the solution of the initial general problem. Both problems remain impossible, analytically as well as numerically, if no further distribution assumptions are introduced. We notice that the expectations involved in (iv) and (v) are multiple integrals depending on the distributions (ii).

3. Reduction to a One-Dimensional Problem

A decisive step performed by Dr. Young is the detection of triples

$$(\text{distribution of } \Theta, \text{ distribution of } X_i/\Theta, \text{ statistic } T) \tag{vi}$$

with the following properties.

*F. Etienne De Vylder, Ph.D., is a former Professor of Actuarial Science at a Belgian university. He can be reached at Broederlijke-Weversplein 32, 9000 Gent, Belgium.

(a) The predictive mean $E(X_{n+1}/X_1, \ldots, X_n)$ can be expressed as a deterministic function $\mu(T)$ of the statistic T.
(b) The function $\mu(\cdot)$ can be found analytically (or can be evaluated very simply).
(c) The density $f(\cdot)$ of T can be found analytically (or can be evaluated very simply).

Then the problem becomes the minimization of

$$E[d(T) - \mu(T)]^2 + h\,E[d''(T)]^2 \qquad \text{(vii)}$$

by the optimal choice of $d(\cdot)$. Now only one-dimensional integrals are involved. This allows the research of a numerical solution.

Triples (vi) with properties 3a, 3b, and 3c result from the statistical theory of point estimation related to exponential families of distributions, priors, conjugate priors and sufficient statistics. Only one-parameter families can be considered because $d(T)$ is a function of a single statistic T.

4. The Numerical Optimization Problem

Let us assume that the density $f(\cdot)$ of T vanishes outside the interval $[a, b]$. Then (vii) equals

$$\int_{[a,b]} [d(t) - \mu(t)]^2 f(t)dt + h \int_{[a,b]} [d''(t)]^2 f(t)dt. \qquad \text{(viii)}$$

Neglecting some small amount of probability mass if necessary, it can be assumed that the interval $[a, b]$ is bounded in practice.

Dr. Young looks for the best function $d(\cdot)$ in the class of polynomials of the fifth degree on $[a, b]$, and she determines this polynomial in two examples. I here indicate another method for finding $d(\cdot)$ numerically. It is based on a discretization of the integrals of (viii). Let $\delta>0$ be the span of the discretization. Let $a=m\delta$, $b=(m+n)\delta$, where $m\geq 0$ and $n>0$ are integers. Let

$$d_i = d((m+i)\delta),\ \mu_i = \mu((m+i)\delta),$$
$$f_i = f((m+i)\delta)\ (i=0, 1, \ldots, n).$$

It is assumed that $f_0\geq 0$ and $f_i>0$ $(i=1, \ldots, n)$. We consider the following approximations. They are asymptotically exact as $\delta\downarrow 0$ (under obvious regularity assumptions on the involved functions).

$$\int_{[a,b]} [d(t)-\mu(t)]^2 f(t)dt \approx \sum_{0\leq i\leq n} (d_i-\mu_i)^2 f_i\delta,$$

$$d''((m+i)\delta) \approx \frac{d_{i-1} - 2d_i + d_{i+1}}{\delta^2}\ (i=1, \ldots, n-1),$$

$$\int_{[a,b]} [d''(t)]^2 f(t)dt \approx \sum_{1\leq i\leq n-1} \left[\frac{d_{i-1} - 2d_i + d_{i+1}}{\delta^2}\right]^2 f_i\,\delta.$$

Hence, the discrete version of the problem is the minimization of

$$Q = \sum_{0\leq i\leq n} (d_i-\mu_i)^2 f_i + \eta \sum_{1\leq i\leq n-1} (d_{i-1} - 2d_i + d_{i+1})^2 f_i$$

by the optimal choice of $d_0, d_1, \ldots, d_n$. Here $\eta=h/\delta^4$. Let us take $c_i=d_i-\mu_i$ $(i=0,1, \ldots, n)$ as unknown quantities. Then

$$Q = \sum_{0\leq i\leq n} c_i^2 f_i + \eta \sum_{1\leq i\leq n-1} (\gamma_{i-1} + c_{i-1} - 2c_i + c_{i+1})^2 f_i, \qquad \text{(ix)}$$

where

$$\gamma_{i-1} = \mu_{i-1} - 2\mu_i + \mu_{i+1}\ (i=1, \ldots, n-1).$$

The annulation of the partial derivative of Q with respect to c_j $(j=0,1, \ldots, n)$ furnishes the equations

$$c_0 f_0 + \eta(\gamma_0 + c_0 - 2c_1 + c_2)f_1 = 0, \qquad (x_0)$$

$$c_1 f_1 - 2\eta(\gamma_0 + c_0 - 2c_1 + c_2) f_1 + \eta(\gamma_1 + c_1 - 2c_2 + c_3)f_2 = 0, \qquad (x_1)$$

$$c_2 f_2 + \eta(\gamma_0 + c_0 - 2c_1 + c_2)f_1 - 2\eta(\gamma_1 + c_1 - 2c_2 + c_3)f_2 + \eta(\gamma_2 + c_2 - 2c_3 + c_4)f_3 = 0, \qquad (x_2)$$

$$c_3 f_3 + \eta(\gamma_1 + c_1 - 2c_2 + c_3)f_2 - 2\eta(\gamma_2 + c_2 - 2c_3 + c_4)f_3 + \eta(\gamma_3 + c_3 - 2c_4 + c_5)f_4 = 0, \qquad (x_3)$$

... ...

$$c_{n-2} f_{n-2} + \eta(\gamma_{n-4} + c_{n-4} - 2c_{n-3} + c_{n-2})f_{n-3} - 2\eta(\gamma_{n-3} + c_{n-3} - 2c_{n-2} + c_{n-1})f_{n-2} + \eta(\gamma_{n-2} + c_{n-2} - 2c_{n-1} + c_n)f_{n-1} = 0, \qquad (x_{n-2})$$

$$c_{n-1} f_{n-1} + \eta(\gamma_{n-3} + c_{n-3} - 2c_{n-2} + c_{n-1})f_{n-2} - 2\eta(\gamma_{n-2} + c_{n-2} - 2c_{n-1} + c_n)f_{n-1} = 0, \qquad (x_{n-1})$$

$$c_n f_n + \eta(\gamma_{n-2} + c_{n-2} - 2c_{n-1} + c_n)f_{n-1} = 0. \qquad (x_n)$$

By summation, $\Sigma\, c_i f_i=0$; that is, $\Sigma\, d_i f_i=\Sigma\, \mu_i f_i$. The continuous version of the latter relation (for $\delta\downarrow 0$) is $Ed(T)=E\mu(T)$. But $E\mu(T)=E(X_{n+1})$.

Hence $\mathrm{E}d(T)=\mathrm{E}(X_{n+1})$. This relation guarantees an equilibrium in the portfolio when the insureds pay the premium $d(T)$.

5. Numerical Solution of the Linear System

Let $\rho_i=f_i/f_{i+1}$ $(i=0, 1, \ldots, n-1)$. Let $c_0=\alpha$ and $c_1=\beta$. Then, by relations (x_0), (x_1), . . . , (x_{n-2}),

$$c_0 = \alpha, \qquad (\mathrm{xi}_0)$$

$$c_1 = \beta, \qquad (\mathrm{xi}_1)$$

$$c_2 = -c_0\rho_0/\eta - (\gamma_0 + c_0 - 2c_1), \qquad (\mathrm{xi}_2)$$

$$c_3 = -c_1\rho_1/\eta + 2(\gamma_0 + c_0 - 2c_1 + c_2)\rho_1 - (\gamma_1 + c_1 - 2c_2), \qquad (\mathrm{xi}_3)$$

$$c_{j+2} = -c_j\rho_j/\eta - (\gamma_{j-2} + c_{j-2} - 2c_{j-1} + c_j)\rho_{j-1}\rho_j + 2(\gamma_{j-1} + c_{j-1} - 2c_j + c_{j+1})\rho_j - (\gamma_j + c_j - 2c_{j+1})$$
$$(j=2, 3, \ldots, n-3, n-2) \qquad (\mathrm{xi}_{j+2})$$

and by (x_{n-1}) and (x_n),

$$c_{n-1}/\eta + (\gamma_{n-3} + c_{n-3} - 2c_{n-2} + c_{n-1})\rho_{n-2} - 2(\gamma_{n-2} + c_{n-2} - 2c_{n-1} + c_n) = 0, \qquad (\mathrm{xii}_1)$$

$$c_n/\eta + (\gamma_{n-2} + c_{n-2} - 2c_{n-1} + c_n)\rho_{n-1} = 0. \qquad (\mathrm{xii}_2)$$

By the relations (xi_0), (xi_1), . . . , (xi_n), the numbers c_0, c_1, . . . , c_n can successively be obtained as affine functions of α and β (but it is not necessary to perform these eliminations analytically). Then, by substitution in (xii_1) and (xii_2), the first member of these relations also is an affine function of α and β. Let these first members be $x\alpha+y\beta+z$ and $x'\alpha+y'\beta+z'$. The unknown coefficients x, y, z and x', y', z' can be found numerically. Indeed, for $\alpha=0$ and $\beta=0$, we obtain z and z'. For $\alpha=1$ and $\beta=0$, we obtain $x+z$ and $x'+z'$, hence x and x'. For $\alpha=0$ and $\beta=1$, we obtain $y+z$ and $y'+z'$, hence y and y'. Then it is enough to fix α and β in such a way that $x\alpha+y\beta+z=0$ and $x'\alpha+y'\beta+z'=0$, that is,

$$\alpha = \frac{yz' - zy'}{xy' - yx'}, \quad \beta = \frac{zx' - xz'}{xy' - yx'}.$$

As far as I know, this is a rather original procedure. It is obvious to what kind of linear systems it can be extended.

As an illustration, this method is applied to Dr. Young's Example 5.1 (inverse gamma-gamma) with her numerical values $r=3$, $\alpha=10$, $\beta=0.1$, and $n=3$ (of course, here α, β and n are not those considered before). As integration interval $[a,b]$, we take [0, 200] instead of $[e^2,e^5]$ and consider different values of h (including the value 110,000 treated by Dr. Young). Numerical values $f(t)$, $\mu(t)$, and $c(t)=d(t)-\mu(t)$ are reproduced in Table 1. They have been evaluated with $\delta=1$. Smaller values of δ do not sensibly improve the results. It is safe to work with the greatest precision,

TABLE 1
NUMERICAL VALUES OF $f(t)$, $\mu(t)$, AND $c(t) = d(t) - \mu(t)$

			$c(t)$						
t	$f(t)$	$\mu(t)$	$h=10^3$	$h=10^4$	Young	$h=10^6$	$h=10^7$	$h=10^8$	$h=10^9$
0	0.000000	0.00	13.37	19.71	25.51	28.84	30.10	30.29	30.31
10	0.003133	23.75	2.29	5.28	8.66	10.82	11.67	11.80	11.82
20	0.023120	38.00	−0.23	0.15	1.29	2.30	2.74	2.81	2.82
30	0.027821	47.50	−0.20	−0.91	−1.46	−1.50	−1.44	−1.43	−1.43
40	0.019807	54.29	0.08	−0.41	−1.75	−2.62	−2.91	−2.96	−2.96
50	0.011637	59.34	0.13	0.33	−0.79	−2.12	−2.70	−2.79	−2.80
60	0.006425	63.33	0.09	0.76	0.66	−0.64	−1.37	−1.49	−1.50
70	0.003476	66.50	0.05	0.81	2.12	1.43	0.73	0.60	0.56
80	0.001890	69.09	0.03	0.63	3.26	3.80	3.36	3.26	3.24
90	0.001044	71.25	0.02	0.39	3.94	6.24	6.36	6.34	6.34
100	0.000589	73.08	0.02	0.19	4.12	8.58	9.36	9.75	9.76
110	0.000340	74.64	0.01	0.08	3.87	10.68	13.08	13.41	13.44
120	0.000200	76.00	0.01	0.04	3.30	12.43	16.63	17.27	17.34
130	0.000121	77.19	0.01	0.03	2.56	13.37	20.23	21.28	21.40
140	0.000074	78.24	0.01	0.04	1.77	14.66	23.84	25.42	25.49
150	0.000047	79.17	0.00	0.04	1.04	15.11	27.43	29.66	29.91
160	0.000030	80.00	0.00	0.03	0.45	15.16	31.00	33.99	34.32
170	0.000019	80.75	0.00	0.03	0.01	14.88	34.53	38.39	38.81
180	0.000013	81.43	0.00	0.02	−0.28	14.34	38.05	42.84	43.37
190	0.000009	82.05	0.00	0.03	−0.46	13.77	41.58	47.36	47.99
200	0.000006	82.61	0.01	0.07	−0.56	13.15	45.14	51.92	52.67

for instance, to use the *extended* real number data type in Pascal.

In the Young case, treated with $\delta=0.1$, I find that

$$c(t) > 0 \ (0 \le t \le 23.3)$$
$$c(t) < 0 \ (23.4 \le t \le 55.5)$$
$$c(t) > 0 \ (55.6 \le t \le 170.1)$$
$$c(t) < 0 \ (170.2 \le t \le 200)$$

These results are in reasonable but not complete agreement with Dr. Young's Figure 1. The slight differences can be explained only by her polynomial approximations. Let me recall that a test of correctness is the relation $\Sigma\ d_i f_i\ \delta \approx \mathrm{E}d(T)=\mathrm{E}(X_{n+1})$ (=50 in the Young case).

DONALD A. JONES*

As a longtime Bayesian actuary, I was pleased to see Equation (3.1) in Dr. Young's extremely well-written paper describing her skillful application of mathematics. That one equation contains the heart of two actuarial methods that had their roots in Bayesian theory before the birth of the neo-Bayesian statistics school. Credibility methods, which started in North America in the second decade of this century, were given a Bayesian foundation by Arthur Bailey in 1950, if not earlier by Keffer (1929). In 1919 Whittaker (1923) gave a Bayesian interpretation for his graduation method by starting with (3.1). I date the birth of the Bayesian statistics school near the publication of *The Foundations of Statistics* by Savage in 1954. Of course such historical dating is subjective (smile), but I think these are close to the seminal publications.

Again as a *longtime* Bayesian actuary, I appreciated the linear credibility estimators when the calculating tool on my desk was a four-function calculator without an exponential/logarithmic key. I was glad to read Bailey and others who put a Bayesian foundation under those linear estimators in some cases. However, I have considered us liberated by the technology explosion of the last quarter of this century. Now that Mathcad Plus is available, I no longer need to adopt assumptions to simplify my calculations to a point of linearity.

As I look at Tables 1 and 2, why would I use anything but the predictive mean? I am using mean squared error as the measure of the goodness of my estimators (via some loss function), and I know that the predictive mean will minimize the expected mean squared error. Why am I constraining toward linearity in my estimator? Why do I want to reduce curvature, which takes me away from my predictive mean?

I congratulate Dr. Young on the excellent mathematics and hope that we can encourage her to turn her talents to modeling the distributions involved.

REFERENCES

BAILEY, A. L. 1950. "Credibility Procedures," *Proceedings of the Casualty Actuarial Society* XXXVII:7–23.

KEFFER, R. 1929. "An Experience Rating Formula," *Transactions of the Actuarial Society of America* XXX:130–39.

SAVAGE, L. J. 1954. *The Foundations of Statistics*. New York: John Wiley & Sons.

WHITTAKER, E. T. 1923. "On a New Method of Graduation," *Proceedings of the Edinburgh Mathematical Society* XLI:63–75.

BJØRN SUNDT*

Dr. Young considers a situation where X_i denotes the total claims of an insurance policy incurred during the i-th year the policy is in force. It is assumed that the X_i's are conditionally independent and identically distributed given an unknown random risk parameter Θ that represents unknown risk characteristics of the policy. Let U denote the distribution of Θ (the structure distribution) and $F(\cdot|\theta)$ the conditional distribution of X_i given that $\Theta=\theta$.

After n years the available data are $\mathbf{X}_n=(X_1, \ldots, X_n)'$, and the insurance company wants to apply these data to set the premium for year $n+1$. In credibility theory one has traditionally applied minimization of the expected quadratic loss as the optimality criterion; that is, one wants to find a function $\dot{m}$ that minimizes $\mathrm{E}\ (\dot{m}(\mathbf{X}_n)-X_{n+1})^2$, which is equivalent to minimizing $\mathrm{E}(\dot{m}(\mathbf{X}_n)-m(\Theta))^2$, where $m(\Theta)=\mathrm{E}[X_{n+1}|\Theta]$. The optimal estimator is the Bayes estimator $\tilde{\tilde{\tilde{m}}}_{n+1}(\mathbf{X}_n) =\mathrm{E}[m(\Theta)|\mathbf{X}_n]$.

Unfortunately the Bayes estimator has some disadvantages. One has to specify the distributions U and F, and the function $\tilde{\tilde{\tilde{m}}}_{n+1}$ could have a complicated

*Donald A. Jones, F.S.A., Ph.D., is Director of the Actuarial Science Program at Oregon State University, 368 Kidder Hall, Corvallis, Oregon 97331-4605.

*Bjørn Sundt, Dr. Sci. Math., is a senior lecturer in the Department of Mathematics, University of Bergen, Johannes Bruns gate 12, N-5008 Bergen, Norway.

form. Intuitively one would want the premium to be increasing in the observed claims, but this is not always the case with the Bayes estimator.

Because of these deficiencies, one often restricts the class of permitted functions $\dot{m}$ to the class of functions that are linear in the observed claim amounts. The optimal estimator in this class is the credibility estimator

$$\tilde{m}_{n+1}(\mathbf{X}_n) = \frac{n}{n+\kappa}\overline{X}_n + \frac{\kappa}{n+\kappa}\mu$$

with

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i; \quad \kappa = \frac{\varphi}{\lambda}$$

$$\mu = \mathrm{E}m(\Theta); \quad \varphi = \mathrm{E}\,\mathrm{Var}\,[X_i|\Theta]; \quad \lambda = \mathrm{Var}\; m(\Theta).$$

The credibility estimator has an intuitive appeal and could easily be explained to nonactuarial insurance agents and customers. The premium is a weighted average of the mean claim amount in the portfolio and the claim experiences of the policy. It increases with the average claim amount of the policy; a policy with high claims should have a high premium. The weight given to the claim experience increases with the number of observed years; the more experience we have, the more weight (credibility) we can give it. This weight increases with λ; the greater heterogeneity between the policies, the less we can trust the portfolio mean when setting the premium for an individual policy. On the other hand, it decreases with φ; the more the claim amounts within one policy vary from year to year, the less we can trust the individual experience when setting the premiums.

As pointed out above, to evaluate the Bayes estimator, we need to fully specify the distributions U and F. However, we see that in the credibility estimator we need only the structure parameters μ, φ, and λ. In practice, one would usually not know the distributions U and F. However, if we assume that the policies within the portfolios are independent and have the same U and F, we can easily estimate the structure parameters from claim data from the policies within the portfolio; compare, for example, Sundt (1993).

The deduction of credibility estimators can be extended to more complicated models, for example, regression models, and as one restricts the class of permitted estimators to linear functions of the observations, the estimators will always depend on the distributions only through moments of first and second order; compare Sundt (1993). When saying *observations* in the present context, one can of course more generally consider specified functions of the observations, for example, letting the credibility estimator be a linear function of (powers up to a specified order of) a sufficient statistic.

In her paper, Dr. Young restricts the class of permitted estimators to functions linear in $T_n(\mathbf{X}_n)$, where T_n is a one-dimensional function. She considers two cases: $T_n(\mathbf{X}_n)=\overline{X}_n$ and the case in which $T_n(\mathbf{X}_n)$ is a sufficient statistic.

When $T_n(\mathbf{X}_n)$ is sufficient, one would normally have to assume a parametric class for the distribution F. In this case we have

$$\mathrm{E}[m(\Theta)|T_n(\mathbf{X}_n)] = \mathrm{E}[m(\Theta)|\mathbf{X}_n] = \approx{m}_{n+1}(\mathbf{X}_n);$$

that is, the Bayes estimator based on the sufficient statistic is equal to the Bayes estimator based on the claim amounts. However, the credibility estimator based on the sufficient statistic is equal to the credibility estimator based on the claim amounts if and only if the function T_n is linear.

In traditional credibility theory the desire for linearity has been expressed by restricting the class of permitted estimators to linear estimators. In her paper, Dr. Young takes a different approach. Because linearity is important for the decision-maker, a desire for linearity should also be expressed through the loss function. The author wants to find the function $\dot{m}$ that minimizes

$$\mathrm{E}[\dot{m}(T_n(\mathbf{X}_n)) - m(\Theta)]^2 + h\mathrm{E}[\dot{m}''(T_n(\mathbf{X}_n))]^2. \quad (1)$$

The idea is that when $\dot{m}$ is linear, then $\dot{m}''=0$, and thus the term $\mathrm{E}[\dot{m}''(T_n(\mathbf{X}_n))]^2$ can be interpreted as a measure of closeness to linearity. The constant h expresses the importance for the decision-maker of closeness to linearity as compared to accuracy of the estimator. Let $\approx{m}^{(h)}_{n+1}$ be the function that minimizes (1).

Earlier in this discussion I have given arguments for applying a linear estimator instead of a Bayes estimator, but does a desire for a linear estimator imply a desire for an estimator close to a linear estimator? I would say no.

One of the reasons for applying a linear estimator was simplicity. Linearity represents a true simplification, whereas closeness to linearity is no simplification; to be linear, or not to be linear—that is the question. The estimator $\approx{m}^{(h)}_{n+1}$ is likely to be more complicated than not only the credibility estimator, but also the Bayes estimator. Although it is more likely that $\approx{m}^{(h)}_{n+1}$ will be monotonic, the larger the value of h, monotonicity is not guaranteed as with the credibility estimator. The credibility estimator was easy to interpret and explain; that would usually not be the case with $\approx{m}^{(h)}_{n+1}$.

Another advantage of the credibility estimator was that it depended on only first- and second-order moments that could be estimated from portfolio data. To evaluate $\tilde{\tilde{m}}_{n+1}^{(h)}$, one would have to specify U and F.

Furthermore, even under parametric assumptions, it does not seem possible to obtain explicit expressions for $\tilde{\tilde{m}}_{n+1}^{(h)}$ analytically. The author suggests a numerical procedure. To apply that procedure, one has to assume that the support of F is restricted to a finite interval, and one also has to restrict the class of permitted estimators to the class of polynomials of a specified order. In her examples the author applies order five. But as far as I can see, there is no point in having an estimator in the shape of a polynomial of order five close to a linear estimator unless it is a linear estimator itself. Consequently one could just as well find the optimal estimator according to minimization of expected quadratic loss when the class of permitted estimators is restricted to polynomials of order five.

When restricting the class of permitted estimators to polynomials in $T_n(\mathbf{X}_n)$ of a specified order, the question is what order to choose. A lower order gives a simpler estimator, whereas a higher order implies a lower expected loss. It is therefore interesting to study how much the loss will increase when we restrict the class to polynomials of a lower order. Such problems have been discussed by Neuhaus (1985).

At the end of Section 1 I indicated the possibility of generalizing the deduction of credibility estimators to more complicated models. Considering the complications that occur already in the simple case, I believe that a generalization of the new approach to more complicated models, for example, Hachemeister's (1975) regression model, will be prohibitively complicated.

My opinion, based on the criticisms raised in this discussion, is that this new approach does not appear to be a fruitful development of credibility theory.

REFERENCES

HACHEMEISTER, C. A. 1975. "Credibility for Regression Models with Application to Trend," in *Credibility: Theory and Applications*, edited by P.M. Kahn. New York: Academic Press, pp. 129–63.

NEUHAUS, W. 1985. "Choice of Statistics in Linear Bayes Estimation," *Scandinavian Actuarial Journal*, 1–26.

SUNDT, B. 1993. *An introduction to Non-Life Insurance Mathematics*. 3rd ed. Karlsruhe: Verlag Versicherungswirtschaft e.V.

GREGORY C. TAYLOR*

The paper involves two very distinct ideas, as reflected in its title. These are:

- The use of sufficient statistics to maximize efficiency of a credibility estimator
- Compromising between estimation efficiency and smoothness in the choice of a credibility estimator.

In each case there are tracts of the literature that warrant discussion.

In relation to the first idea, the author makes passing reference to Taylor (1977) but does not mention the central result of that work. That result stated essentially that the most efficient credibility estimators are those based on sufficient statistics. This foretells the author's rather empirical conclusion in the Abstract that "functions of sufficient statistics prove to be better credibility estimators than functions of the sample mean."

To explore the second main idea of the paper, the author introduces (3.1), a composite loss function reflecting both estimation efficiency and smoothness. The title of the paper attributes this to spline theory, but many actuaries would be more familiar with it from Whittaker graduation.

This subject has recently been taken up in a short sequence of papers by Taylor (1992) and Verrall (1995, 1996). These papers provide a Bayesian interpretation of Whittaker graduation, thus placing it in the same setting as that from which credibility theory is derived.

One interesting result of this interpretation of Whittaker graduation is a precise meaning of the relativity constant. This is the Whittaker counterpart of the parameter h in the author's (3.1).

Taylor considers the situation in which $\{E[X_m|\theta], m=1, 2, \ldots, n\}$ (in Young's notation) has a prior distribution for which $\Delta^2 E[X_m|\theta] \sim N(0, \tau^2)$. Then the Whittaker graduation, essentially the same as Young's estimator based on (3.1), is an asymptotic (large sample) MLE of the Bayesian posterior distribution of $E[X_{n+1}|\theta]$ if h is chosen as σ^2/τ^2, where

$$\sigma^2 = V[X_m|\theta],$$

assumed independent of m.

In other words, there is a reasoned way of choosing Young's h; it is related to the inherent smoothness of the sequence $E[X_m|\theta]$ under consideration.

Verrall (1995) has extended Taylor's results. He shows that Taylor's Bayesian version of Whittaker

*Gregory C. Taylor, F.I.A., F.I.A.A., Ph.D., is with Tillinghast-Towers Perrin, Level 16, MLC Centre, 19-29 Martin Place, Sydney, NSW 2000, Australia.

graduation (of a set of mortality rates by age) is equivalent to application of a linear filter to the empirical rates and a particular order of differences of those rates. The sequence of differences is interpreted as a realization of a randomly varying parameter.

Translation of these results to Young's case means that her estimate $d(\bar{x})$ from (3.1) with $h=\sigma^2/\tau^2$ is the same as would be obtained by applying the Kalman filter to the state space vector (X_m, θ_m, β_m) where

$$X_m|\theta_m \sim N(\theta_m, \sigma^2)$$

$$\theta_m = \theta_{m-1} + \beta_m$$

$$\beta_m = \beta_{m-1} + \nu_m$$

$$\nu_m - N(0, \tau^2).$$

Subsequently, Verrall (1996) has shown how this filter (equivalently Young's credibility formula) may be implemented within a generalized linear model framework.

REFERENCES

TAYLOR, G. 1977. "Abstract Credibility," *Scandinavian Actuarial Journal*, 149–68.

TAYLOR, G. 1992. "A Bayesian Interpretation of Whittaker-Henderson Graduation," *Insurance Mathematics and Economics* 11:7–16.

VERRALL, R. J. 1995. "Whittaker Graduation and Parametric State Space Models," *City University Actuarial Research Paper No. 79*.

VERRALL, R. J. 1996. "A Unified Framework for Graduation," *City University Actuarial Research Paper No. 91*.

AUTHOR'S REPLY

VIRGINIA R. YOUNG

I thank the four discussants for their useful contribution to my work. I especially thank Dr. De Vylder for showing how to approximate the solutions to the boundary-value problems of my paper. His method yields more accurate results than the method I used, which is based on Bernstein-Bezier polynomials. I also thank Dr. Taylor for connecting his and Verrall's work in graduation to my loss function.

As a technical note, the optimality of functions of a sufficient statistic generally follows from Jensen's inequality as applied in the Rao-Blackwell theorem for a convex loss function, such as squared-error loss. Also, substituting the predictive mean for the conditional mean in squared-error loss results in the same minimizing solution d when d is a function of a sufficient statistic or the sample mean.

In response to the comments of Dr. Sundt and Dr. Jones: I introduce the loss function in Equation (3.2) for two reasons: An actuary may have evidence from two sources for choosing a credibility function. The first source is claim data itself, and the actuary may find that expected claims can be regressed as a linear function of a statistic, such as the sample, geometric, or harmonic mean. The second source is the actuary's subjective belief that the claims follow certain parametric loss distributions, such as the lognormal-lognormal or inverse gamma-gamma. To balance these sources of information, the actuary may wish to minimize the loss function I propose in Equation (3.2).

I also have a second motive for introducing such a loss function. I want to encourage actuaries to consider loss functions other than simple squared-error loss functions. In general, an actuary has goals other than accuracy, as embodied by squared-error loss. Those goals may include maximizing profit while maximizing the amount of business written. See Young (1996) in which I propose a loss function that balances these two conflicting goals. Also, an actuary may wish not to penalize too greatly those insureds who have had very poor experience. Minimizing the following loss function may aid in achieving this goal:

$$\int_a^b (\mu(t) - d(t))^2 f(t)dt + h\int_a^b (d'(t))^2\, dt,$$

in the notation of Equation (3.2). The squared-error term constrains the credibility formula d to be accurate, while the first-derivative penalty constrains d to be close to a constant and thus not penalize unlucky insureds.* Finally, Sundt (1992) proposes a loss function that blends the goals of accuracy and stability of premium from one period to the next. I encourage the reader to consider such alternative loss functions in developing credibility premiums.

REFERENCES

SUNDT, B. 1992. "On Greatest Accuracy Credibility with Limited Fluctuation," *Scandinavian Actuarial Journal*, 109–119.

YOUNG, V.R. 1996. "Credibility and Persistency," *ASTIN Bulletin* 26:53–69.

Additional discussions on this paper can be submitted until July 1, 1998. The author reserves the right to reply to any discussion. See the Submission Guidelines for Authors for detailed instructions on the submission of discussions.

*I am grateful to Dr. De Vylder for suggesting this loss function. Applying it in credibility theory is the subject of my current research with Dr. De Vylder.