CREDIBILITY USING SEMIPARAMETRIC MODELS

BY VIRGINIA R. YOUNG

School of Business
University of Wisconsin-Madison

ABSTRACT

To use Bayesian analysis to model insurance losses, one usually chooses a parametric conditional loss distribution for each risk and a parametric prior distribution to describe how the conditional distributions vary across the risks. A criticism of this method is that the prior distribution can be difficult to choose and the resulting model may not represent the loss data very well. In this paper, we apply techniques from nonparametric density estimation to estimate the prior. We use the estimated model to calculate the predictive mean of future claims given past claims. We illustrate our method with simulated data from a mixture of a lognormal conditional over a lognormal prior and find that the estimated predictive mean is more accurate than the linear Bühlmann credibility estimator, even when we use a conditional that is not lognormal.

KEYWORDS

Kernel density estimation, claim estimation, Bayesian estimation.

1. INTRODUCTION

In a portfolio of insurance policyholders (also called risks), risks are heterogeneous; that is, the insurance losses of different risks follow different loss distributions. The premium an insurer charges a given risk depends on the information available concerning the loss distribution of that risk. If the insurer knew the exact loss distribution of a risk, then the appropriate net premium to charge would be the expectation of that loss distribution. On the other hand, if the insurer has no information about a specific policyholder, then the net premium is the expectation over the entire portfolio of policyholders. For the situation between these two extremes, suppose the insurer has prior claim data for the risk, then the net premium is the conditional expectation of future claims given the prior claims.

To use Bayesian analysis to model insurance losses, one usually chooses a parametric conditional loss distribution for each risk and a parametric prior distribution to describe how the conditional distributions vary across the risks. A criticism of this method is that the prior distribution can be difficult to choose and the resulting model may not represent the loss data very well. One method of circumventing this problem is to apply empirical Bayesian analysis in which one uses the data to estimate the parameters of the model (Klugman, 1992).

In this paper, we use a semiparametric mixture model to represent the insurance losses of a portfolio of risks: We choose a flexible parametric conditional loss distribution for each risk with unknown conditional mean that varies across the risks. This conditional distribution may depend on parameters other than the mean, and we use the data to estimate those parameters. Then, we apply techniques from nonparametric density estimation to estimate the distribution of the conditional means.

In Section 2, we describe a mixture model for insurance claims and estimate the prior density using kernel density estimation. In Section 3, we calculate the credibility estimator assuming squared-error loss and also give the projection of that estimator onto the space of linear functions. Finally, in Section 4, we apply our methodology to simulated data from a mixture of a lognormal conditional over a lognormal prior. We show that our method can lead to good credibility formulas, as measured by the mean squared error of the claim predictor, even when we use a gamma conditional instead of a lognormal conditional.

2. SEMIPARAMETRIC MIXTURE MODEL

2.1. Notation and Assumptions

Assume that the underlying claim of risk i per unit of exposure is a conditional random variable $Y|\theta_i$, $i = 1, 2, \ldots, r$, with probability density function $f(y|\theta_i)$. For each of the r risks, we observe the average claims per unit of exposur $\boldsymbol{x}_i = (x_{i1}, x_{i2}, \ldots, x_{in_i})$ with an associated exposure vector $\boldsymbol{w}_i = (w_{i1}, w_{i2}, \ldots, w_{in_i})$, $i = 1, 2, \ldots, r$. Thus, the observed average claim x_{ij} is the arithmetic average of w_{ij} claims, each of which is an independent realization of the conditional random variable $Y|\theta_i$. For example, if a risk is a group policyholder, then x_{ij} may be the average claim per insured member of the group in the j^{th} policy period and w_{ij} is the number of members in the group during the j^{th} policy period. For the data from Hachemeister (1975), a risk is the collection of insureds in a particular state covered by bodily injury automobile insurance, x_{ij} represents the average claim severity during period j, and w_{ij} is the corresponding number of claims.

Assume that the parameter θ is the conditional mean, $E[Y|\theta] = \theta$. There may be other parameters that characterize the conditional distribution, such as the shape parameter α for the gamma density. However, in this paper, we assume that parameters, other than the conditional mean, are fixed across the risks. The loss distribution of a given risk is, therefore, characterized by its conditional mean, although that mean is generally unknown. Denote the probability density function of θ by $\pi(\theta)$, also called the *structure function* (Bühlmann, 1970). The structure function characterizes how the conditional mean θ varies from risk to risk. We argue that assuming θ to be continuous is reasonable because in the

Bayesian paradigm, our uncertainty about θ for any particular risk would be represented by a continuous random variable. Also, if r is large, then the variable θ can be well approximated by a continuous random variable. Even if r is not large, the collection of r risks may be a sample from a larger population of risks whose distribution can be approximated by a continuous distribution. Assume that the experience of different risks is independent.

Note that our model is a special case of the one given by Bühlmann and Straub (1970). Because X_{ij} is the random variable of an average of w_{ij} iid claims $Y_1, Y_2, ..., Y_{w_{ij}}$, given θ_j, we have that $E[X_{ij}|\theta_i] = E[Y|\theta_i] = \theta_i$ is independent of the period j. It also follows that

$$Cov[X_{ij}, X_{ik}|\theta_i] = \begin{cases} \frac{Var[Y|\theta_i]}{w_{ij}} & if \quad j = k, \\ 0, & if \quad j \neq k, \end{cases}$$

as in the Bühlmann-Straub model. In the literature, $E[Y|\theta_i]$ is called the *hypothetical mean* and $Var[Y|\theta_i]$ the *process variance*. Note that we assume the observations for a risk arise as arithmetic averages of an underlying claim random variable $Y|\theta$, while Bühlmann and Straub (1970) do not assume this in their more general model.

The goal of credibility theory is to predict the future claim y (or an average of future claims) of a risk, given that the risk's claim experience is $\boldsymbol{x}$ and exposure $\boldsymbol{w}$. In this paper, we restrict our attention to credibility formulas that are functions of a single statistic because they are easier to estimate and to use. We choose the sample mean as our statistic, $\bar{x}_i = \frac{\sum_{j=1}^{n_i} w_{ij} x_{ij}}{\sum_{j=1}^{n_i} w_{ij}}$ because the claim experience $\boldsymbol{x}$ is a vector of averages. However, we do not restrict a claim estimator to be linear.

To pick a parametric conditional distribution for $Y|\theta$, we use the following criteria:

- $E[Y|\theta] = \theta$
- The sample mean is a sufficient statistic for θ.
- The functional form of $f(y|\theta)$ is closed under averaging. That is, if $\bar{X}$ is an average of w claims that follow the distribution given by $f(y|\theta)$, then the density of $\bar{X}$ has the same functional form as $f(y|\theta)$.

Three such families of densities are commonly used in actuarial science to model insurance losses—(1) the normal, with mean θ and fixed variance σ^2, (2) the gamma, with mean $\theta = \frac{\alpha}{\beta}$ and fixed shape parameter α, and (3) the inverse gaussian, with mean θ and fixed $\lambda = \frac{\theta^3}{Var[X|\theta]}$. Indeed, $Y|\theta \sim N(\theta, \sigma^2)$ implies that if $\bar{X}$ is an average of w iid claims $Y_1, Y_2, ..., Y_w$, given θ, then $\bar{X}|\theta \sim N\left(\theta, \sigma^2/_w\right)$. Similarly, if $Y|\theta \sim G(\theta, \alpha)$, then $\bar{X}|\theta \sim G(\theta, w\alpha)$, and the probability density function of $Y|\theta$ is

$$f(y|\theta) = \frac{\alpha^\alpha}{\Gamma(\alpha)\theta^\alpha} y^{\alpha-1} e^{-\frac{\alpha y}{\theta}}, \quad y > 0.$$

Finally, if $Y|\theta \sim InvG(\theta, \lambda)$, then $\overline{X}|\theta \sim InvG(\theta, w\lambda)$ and the probability density function of $Y|\theta$ is

$$f(y|\theta) = \sqrt{\frac{\lambda}{2\pi y^3}} \exp\left[-\frac{\lambda}{2}\frac{(y-\theta)^2}{y\theta^2}\right], \quad y > 0.$$

We use the family of gamma conditional distributions in an example in Section 4. In practice, one might use the normal conditional if the conditional variance is assumed constant across the risks. One might use the gamma conditional if the conditional coefficient of variation is assumed constant across risks or the inverse gaussian conditional if one wanted to use a loss distribution with a long tail. Note that for these three families, the predictive mean is a function of the sample mean for any prior distribution π. See Young (1997) for examples of credibility estimators that are functions of a one-dimensional sufficient statistic, not necessarily the sample mean.

In the Bayesian spirit, for a given loss function $L = L(y, d(\overline{x}))$ of the future claim y and the claim predictor d, we propose that the credibility estimator d be the function that minimizes the expected loss

$$E[L(y, d(\overline{x}))],$$

in which we take the expectation with respect to the joint density of the sample mean and future claim. In our mixture model, this joint density is $\int f(y|\theta)\, f(\overline{x}|\theta)\, \pi(\theta)\, d\theta$ Therefore, we require an estimate of the density $\pi(\theta)$.

2.2. Kernel Density Estimation

We use kernel density estimation (Silverman, 1986) to estimate the probability density $\pi(\theta)$. A *kernel* K acts as a weight function and satisfies the condition

$$\int_{-\infty}^{\infty} K(t)dt = 1.$$

If we were to observe directly the conditional means $\theta_1, \theta_2, \ldots, \theta_r$, then the kernel density estimate of $\pi(\theta)$ with kernel K would be given by

$$\frac{1}{r}\sum_{i=1}^{r}\frac{1}{h_i}K\left(\frac{\theta-\theta_i}{h_i}\right), \tag{2.1}$$

in which h_i is a positive parameter called the *window width*, or *bandwidth*. Assume that the kernel is symmetric; therefore, the expectation of θ is the sample mean.

Because we observe only data x_i and w_i and not the true conditional means θ_i, we rely on the law of large numbers and use the sample mean $\overline{x}_i$ to estimate θ_i consistently, $i = 1, 2, \ldots, r$, (Serfling, 1980). In the expression in (2.1), one may wish to weight the terms in the sum according to the relative number of claims for the i^{th} risk so that the expectation of θ is the sample mean $\bar{\bar{x}} = \frac{\sum_{i=1}^{r}\sum_{j=1}^{n_i} w_{ij}x_{ij}}{\sum_{i=1}^{r}\sum_{j=1}^{n_i} w_{ij}} = \frac{\sum_{i=1}^{r} w_i\bar{x}_i}{\sum_{i=1}^{r} w_i}$ in which $w_i = \sum_{j=1}^{n_i} w_{ij}$. We, therefore, propose the following kernel density estimator for $\pi(\theta)$

$$\hat{\pi}(\theta) = \sum_{i=1}^{r} \frac{w_i}{w_{tot}}\frac{1}{h_i}K\left(\frac{\theta - \bar{x}_i}{h_i}\right), \tag{2.2}$$

in which $w_{tot} = \sum_{i=1}^{r} w_i = \sum_{i=1}^{r}\sum_{j=1}^{n_i} w_{ij}$. See the Appendix for a discussion of the asymptotic mean square consistency of $\hat{\pi}(\theta)$.

Two commonly used symmetric kernels are (1) the Gaussian kernel, G,

$$G(t) = \frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}, \quad -\infty < t < \infty,$$

and (2) the Epanechnikov kernel, Epa,

$$Epa(t) = \begin{cases} \frac{3}{4}\frac{\left(1 - \frac{t^2}{5}\right)}{\sqrt{5}}, & -\sqrt{5} < t < \sqrt{5}, \\ 0, & else. \end{cases} \tag{2.3}$$

In our example in Section 4, we use the Epanechnikov kernel because its domain is bounded, and we can, therefore, easily restrict the support of $\hat{\pi}(\theta)$ to lie in the positive real numbers.

Remark: The Epanechnikov is optimal with respect to mean integrated square error (Silverman, 1986). The efficiency of the Gaussian kernel with respect to the optimal Epanechnikov kernel is roughly 95% (Silverman, 1986), so one does not lose much efficiency by using the Gaussian kernel. Silverman, therefore, suggests that one choose the kernel according to auxiliary requirements, such as ease of computing. □

There are many techniques for choosing the window width h_i; see, for example, Silverman (1986, Section 3.4) and Jones, Marron, and Sheather (1996). In our example in Section 4, we use a (modified) fixed window width selected by reference to a standard distribution (Silverman, 1986, Section 3.4.2). The window width h that minimizes the mean integrated squared error is given by

$$h = \left\{\int t^2\, K(t)\, dt\right\}^{-2/5}\left\{\int K(t)^2\, dt\right\}^{1/5}\left\{\int \pi''(\theta)\, d\theta\right\}^{-1/5} r^{-1/5}. \tag{2.4}$$

To approximate this optimal window width h, ones assumes that $\pi(\theta)$ is say, normal, with mean 0 and standard deviation σ. In that case, the term $\int \pi''(\theta)\, d\theta$ equals $\frac{3}{8}\pi^{-1/2}\sigma^{-5}$. We modify the window width h at each point $\bar{x}_i$ to ensure that the density has support on the nonnegative real numbers. Specifically, we set h_i equal to h, if $h < \frac{\bar{x}_i}{\sqrt{5}}$ otherwise, we set h_i equal to $\frac{\bar{x}_i}{\sqrt{5}}$.

3. Credibility using Squared-Error Loss

In this section, we use squared-error loss to determine a credibility estimator, as is used in greatest accuracy credibility theory, (Willmot, 1994) or (Herzog, 1996). The squared-error loss function has the form

$$L(y,\ d(\bar{x})) = (y - d(\bar{x}))^2.$$

It is straightforward to show that the minimizer of the expected loss is the predictive mean (Bühlmann, 1967), which in this case is the posterior mean of θ given the sample mean $\bar{x}$ which we estimate by

$$\hat{\mu}(\bar{x}) = \int E[Y|\theta]\hat{\pi}(\theta|\bar{x})d\theta = \hat{E}[\theta|\bar{x}].$$

For a general kernel K and bandwidths h_i, this estimated posterior mean of θ can be written

$$\hat{E}[\theta|\bar{x}] = \frac{\int \theta\, f(\bar{x}|\theta)\hat{\pi}(\theta)d\theta}{\int f(\bar{x}|\theta)\hat{\pi}(\theta)d\theta}$$

$$= \frac{\sum_{i=1}^{r} \frac{w_i}{h_i} \int \theta\, f(\bar{x}|\theta) K\left(\frac{\theta - \bar{x}_i}{h_i}\right) d\theta}{\sum_{i=1}^{r} \frac{w_i}{h_i} \int f(\bar{x}|\theta) K\left(\frac{\theta - \bar{x}_i}{h_i}\right) d\theta} \tag{3.1}$$

Recall that $\bar{x}$ is an average of w iid claims, each of which follows the density $f(y|\theta)$, as in Section 2.1. If we constrain the estimator d to be linear, then it is well-known that the least-squares linear estimator of $E[Y|\bar{x}] = E[\theta|\bar{x}]$ is

$$d(\bar{x}) = (1 - Z)E[Y] + Z\bar{x}, \tag{3.2}$$

in which $Z = \dfrac{w}{w + k}$ with $k = \dfrac{\hat{E}Var[Y|\theta]}{V\hat{a}r[\theta]}$ (Bühlmann, 1967). Using our estimate for the prior density (2.2), we obtain $\hat{E}[Y] = \hat{E}[\theta] = \bar{\bar{x}}$, as noted in Section 2.2. In the case of the normal conditional; $k = \dfrac{\sigma^2}{\hat{E}[\theta^2] - \bar{\bar{x}}^2}$ in the case of the gamma conditional, $k = \dfrac{\hat{E}[\theta^2]}{\alpha\left(\hat{E}[\theta^2] - \bar{\bar{x}}^2\right)}$; and in the case of the inverse gaussian conditional,

$$k = \frac{\hat{E}[\theta^3]}{\lambda\left(\hat{E}[\theta^2] - \bar{\bar{x}}^2\right)}.$$

To end this section, we show that as w approaches ∞, $\hat{\mu}(\overline{x})$ approaches the true expected value θ_0, for the given risk. Because $\overline{X}|\theta$ has mean θ and variance $\frac{Var(Y|\theta)}{w}$ under certain regularity conditions, (DeGroot, 1970) and (Walker, 1969), the density $f(\overline{x}|\theta)$ approaches the delta function with its mass concentrated at the point $\overline{x} = \theta_0$ Then,

$$\lim_{w\to\infty} \hat{\mu}(\overline{x}) = \lim_{w\to\infty} \frac{\int \theta f(\overline{x}|\theta)\hat{\pi}(\theta)d\theta}{\int f(\overline{x}|\theta)\hat{\pi}(\theta)d\theta} = \frac{\theta_0\hat{\pi}(\theta_0)}{\hat{\pi}(\theta_0)} = \theta_0, \; w.p. 1.$$

Thus, as an actuary gets more claim information for a given policyholder (w gets large), the estimated expected claim approaches the true expected claim with probability 1.

4. SIMULATED DATA FROM A LOGNORMAL-LOGNORMAL MIXTURE

The lognormal distribution is 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 assume that we are given individual claim data; that is, $w_{ij} = 1$, for all risks i and policy periods j, and $X = Y$. We model the lognormal-lognormal mixture as follows:

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

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

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

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

Given claim data for a specific policyholder, $\boldsymbol{X} = \boldsymbol{x} = < x_1, x_2, \ldots, x_n > \in [0, \infty]^n$, the posterior distribution of $\phi|\boldsymbol{x}$ is lognormal; $(\ln\phi)|\boldsymbol{x} \sim N(\ln\mu^*, \tau^{*^2})$, in which

$$\mu^* = \exp\left(\frac{\sigma^2 \ln\mu + \tau^2 t}{\sigma^2 + n\tau^2}\right),$$

$t = \sum_{i=1}^{n} \ln(x_i)$ and

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

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

$$\mu(\boldsymbol{x}) = E(X_{n+1}|\boldsymbol{x}) = \exp\left(\frac{\sigma^2 \ln \mu + \tau^2 t}{\sigma^2 + n\tau^2} + \frac{\sigma^2(\sigma^2 + (n+1)\tau^2)}{2(\sigma^2 + n\tau^2)}\right). \tag{4.1}$$

We performed 200 simulations of a lognormal-lognormal mixture of claims. We let $\sigma^2 = 0.25$, $\tau^2 = 0.50$, and $\mu = 2000e^{-0.25}$. The marginal expectation of X is 2267, and the marginal standard deviation is 2395. For each simulation run, we simulated claim data from this lognormal-lognormal mixture for $r = 100$ risks (values of ϕ). For each of the 100 risks, we simulated $n_i = w_i = 5$ claims. To estimate the distribution of the conditional means, we used kernel density estimation with the Epanechnikov kernel, as given by (2.3). Also, we used a fixed window width h, chosen by reference to a normal distribution with mean 0 and standard deviation σ. We estimated the standard deviation by the interquartile range of the sample means, R, divided by 1.34 (Silverman, 1986, Section 3.4). The bandwidth h was calculated by $h = (1)^{-2/5}(0.268)^{1/5}(0.212)^{-1/5}\frac{R}{1.34}100^{-1/5}$ $\approx 0.312R$ as in (2.4). We truncated this bandwidth h for a given risk if, by otherwise using it, the prior density would have a negative support. Specifically, if $h > \frac{\overline{x}_i}{\sqrt{5}}$ then we set the bandwidth h_i equal to $\frac{\overline{x}_i}{\sqrt{5}}$ to guarantee that the support of the estimated density of θ be contained in the nonnegative real numbers, as described in Section 2.2.

Instead of assuming that the conditional is lognormal, we assumed that the coefficient of variation is constant from risk to risk and, therefore, fit a gamma conditional to each risk. In each simulation run, we estimated the parameter α by the median of the following sample statistic $\frac{\overline{x}_i^2}{\frac{1}{5-1}\sum_{j=1}^{5}\left(x_{ij} - \overline{x}_i\right)^2}$. We used the estimated prior density along with the gamma conditional to estimate the marginal density of X.

We used the estimated mixture model to estimate the predictive mean of X_{n+1} given claim data $\boldsymbol{x}$. We also computed the Bühlmann credibility estimator, $lin(\boldsymbol{x})$, for which we estimated the expected process variance by

$$E\hat{P}V = \frac{1}{100(5-1)}\sum_{i=1}^{100}\sum_{j=1}^{5}\left(x_{ij} - \overline{x}_i\right)^2$$

and the variance of the hypothetical means by

$$V\hat{H}M = \frac{1}{100-1}\sum_{i=1}^{100}\left(\overline{x}_i - \overline{\overline{x}}\right)^2 - \frac{E\hat{P}V}{5},$$

(Willmot, 1994, Section 5.1).

TABLE 4.1
DESCRIPTIVE STATISTICS OF h, MSE, MSEB, AND RATIO

Variable	Mean	Median	StDev	Q1	Q3
h	564.35	561.00	91.64	500.25	623.75
MSE	16,450	12,111	13,146	7,808	21,623
$MSEB$	74,559	69,595	37,539	44,466	94,878
$Ratio$	0.2984	0.1777	0.3239	0.0890	0.3819

For $n = w = 1$, we compared the estimated predictive mean, $\hat{\mu}(x)$ and the Bühlmann credibility estimator, $lin(x)$, with the true predictive mean, $\mu(x)$. To compare these credibility estimators numerically, for each of the 200 simulation runs, we calculated the mean squared errors up to the 95^{th} percentile of X, namely 6,500: $MSE = \int_0^{6500} (\hat{\mu}(x) - \mu(x))^2 f(x)dx$ and $MSEB = \int_0^{6500} (lin(x) - \mu(x))^2 f(x)dx$. See Table 4.1 for descriptive statistics of the bandwidth h; the mean squared errors, MSE and $MSEB$; and the ratio of MSE to $MSEB$, $Ratio$.

Thus, we see that up to the 95^{th} percentile, on average, our estimated predictive mean performs much better than the linear Bühlmann credibility estimator. See Figure 4.1 for a scatter plot of MSE versus h. Note the quadratic relationship between the two variables and that the minimum of MSE occurs near the average value of h, 564. We fit a quadratic to these observations by minimizing the sum of the absolute values of the errors and obtained the fitted model

$$\hat{MSE} = 196{,}603 - 691.36h + 0.6402h^2,$$

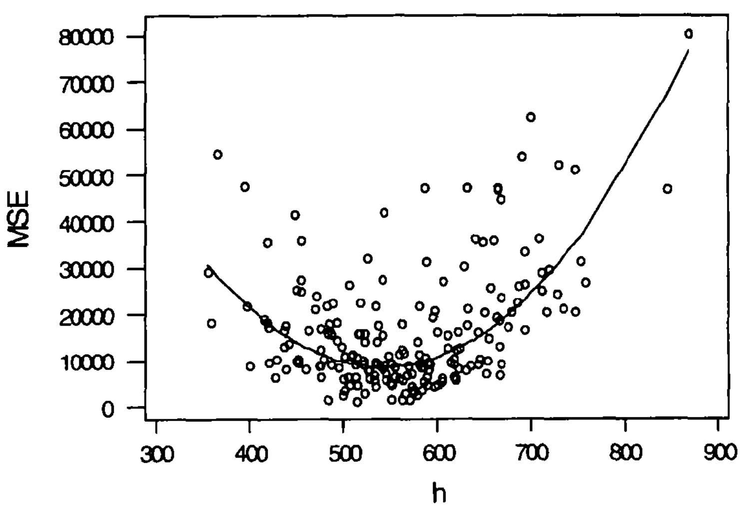

FIGURE 4.1: Scatter Plot of MSE versus h with Quadratic Superimposed.

with vertex at 542. See Figure 4.1 for a graph of this quadratic superimposed on a scatter plot of the observations.

We also computed some of the mean squared errors up to the 99^{th} percentile and found that the estimated predictive mean compared poorly relative to the Bühlmann credibility estimator. We conclude that our estimate of the prior density at larger conditional means may suffer. Silverman (1986) suggests a variable bandwidth approach for estimating densities with long tails which uses

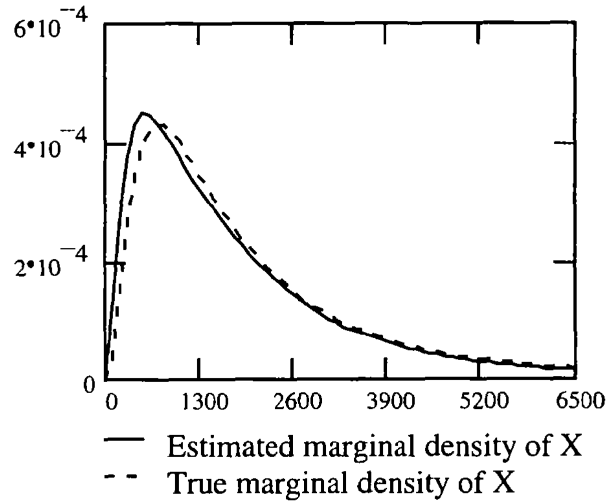

FIGURE 4.2: Estimated and True Marginal Densities of Claims.

larger bandwidths in the regions of lower density. We tried this method without increased accuracy in the upper percentiles of our claim estimator. We suspect that the poor fit at the higher percentiles may be due to our using a medium-tailed gamma conditional to model a heavy-tailed lognormal. We encourage the interested reader to investigate using an inverse gaussian instead of a gamma conditional to model the conditional claim distribution.

See Figure 4.2 for graphs of the estimated and true marginal densities of X for one of the simulations[1]. Of the graphs we plotted, Figure 4.2 is typical, in that the estimated marginal density of X is less skewed than the true density.

See Figure 4.3 for the corresponding graphs of the estimated and true predictive means. Notice how closely the estimated predictive mean follows the true predictive mean, compared with the linear Bühlmann estimator for claims less than 4000. Also note how the estimated predictive mean diverges upward for claims larger than 4000. This phenomenon occurred in all of the several graphs that we plotted and is due, we believe, to the fact that we used a gamma conditional to estimate a lognormal. It may also be due to computational errors

[1] In this run, h = 476, MSE = 12,076, and $MSEB$ = 84,571. Recall that $n = 1$ and that the claim amount 6,500 is the 95^{th} percentile of X.

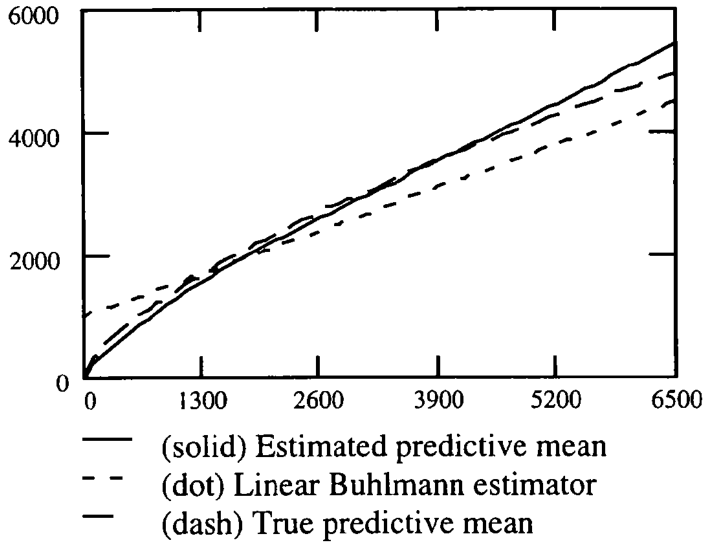

FIGURE 4.3: Credibility Estimators.

because there are only a few simulated claims in the right tail. One way to adjust the estimated predictive mean to eliminate this divergence is to extend it linearly beyond some large value of the sample mean. Another solution may be to use a conditional distribution with a longer tail, such as the inverse gaussian. Yet another solution may be to apply my method of blending the criteria of accuracy and linearity (Young, 1997).

5. SUMMARY AND CONCLUSIONS

The Bühlmann-Straub credibility method results in a linear estimator with a different slope (or credibility weight) for each risk. Therefore, to apply their method to a risk not used to construct the original model, one would be required to recalculate the model to obtain a linear estimator for the new risk. An advantage of our method is that it is applicable to risks outside the original data set, if one assumes that the average claims and corresponding exposures of the new risk come from the same parent (mixture) population as the data. Another advantage of our method is increased accuracy over a linear estimator, as demonstrated in the example in Section 4, even when we use an 'incorrect' conditional density.

One may wish to use the underlying mixture model and kernel density estimation in combination with other loss functions, such as a linear combination of a squared-error term and a second-derivative term to blend the goals of accuracy and linearity (Young, 1997). Also, it would be interesting if one were to extend the model to include a trend component, as in Hachemeister (1975), and apply kernel density estimation in the more general model.

ACKNOWLEDGMENTS

I thank the Committee for Knowledge Extension and Research of the Society of Actuaries (SOA) and the Actuarial and Education Research Fund for financial support. I especially thank my SOA Project Oversight Group (Hans Gerber and Gary Venter, led by Thomas Herzog and assisted by Warren Luckner of the SOA) for helpful guidance.

APPENDIX

ASYMPTOTIC MEAN SQUARE CONSISTENCY OF (2.2)

Let $\tilde{\pi}(\theta) = \sum_{i=1}^{r} \frac{w_i}{w_{tot}} \frac{1}{h_i} K\left(\frac{\theta - \theta_i}{h_i}\right)$ denote the kernel density estimator of π when we are given observations θ_i, $i = 1, 2, \ldots, r$. Consider the mean squared error of the density estimate $\hat{\pi}$ at a fixed value θ :

$$E\left[(\hat{\pi}(\theta) - \pi(\theta))^2\right] = E\left[(\hat{\pi}(\theta) - \tilde{\pi}(\theta))^2 + 2(\hat{\pi}(\theta) - \tilde{\pi}(\theta))(\tilde{\pi}(\theta) - \pi(\theta)) + (\tilde{\pi}(\theta) - \pi(\theta))^2\right]$$

$$= E\left[\sum_{i=1}^{r} \frac{w_i}{w_{tot}} \frac{1}{h_i} \left\{K\left(\frac{\theta - \bar{x}_i}{h_i}\right) - K\left(\frac{\theta - \theta_i}{h_i}\right)\right\}^2\right] + E\left[(\tilde{\pi}(\theta) - \pi(\theta))^2\right]$$

$$+2E\left[\sum_{i=1}^{r} \frac{w_i}{w_{tot}} \frac{1}{h_i} \left\{K\left(\frac{\theta - \bar{x}_i}{h_i}\right) - K\left(\frac{\theta - \theta_i}{h_i}\right)\right\} \cdot (\tilde{\pi}(\theta) - \pi(\theta))\right].$$

By the law of large numbers (Serfling, 1980), $\bar{x}_i$ approaches θ_i, with probability one, as w_i approaches infinity. Therefore, as w_i approaches infinity, the first term in the mean squared error goes to zero. By Silverman (1986) or Thompson and Tapia (1990), the second and third terms go to zero as r goes to infinity if $\lim_{r \to \infty} h_i = 0$ and $\lim_{r \to \infty} rh_i = \infty$.

REFERENCES

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

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

BÜHLMANN, H. and E. Straub (1970), Glaubwürdigkeit für Schadensätze, *Mitteilungen der Vereinigung Schweizerischer Versicherungs-Mathematiker*, 70: 111-133.

DE GROOT, M. H. (1970), *Optimal Statistical Decisions*, McGraw-Hill, New York.

HACHEMEISTER, C. A. (1975), Credibility for regression models with application to trend, 129-163, in *Credibility: Theory and Applications*, (ed. P. M. Kahn), 129-163, Academic Press, New York.

HERZOG, T. L. (1996), *Introduction to Credibility Theory*, second edition, ACTEX, Abington, Connecticut.

JONES, M. C., J. S. MARRON, and S. J. SHEATHER (1996), A brief survey of bandwidth selection for density estimation, *Journal of the American Statistical Association*, **91**: 401-407.

KLUGMAN, S. A. (1992), *Bayesian Statistics in Actuarial Science with Emphasis on Credibility*, Kluwer, Boston.

SERFLING, R. J. (1980), *Approximation Theorems of Mathematical Statistics*, Wiley, New York.

SILVERMAN, B. W. (1986), *Density Estimation for Statistics and Data Analysis*, Chapman & Hall, London.

THOMPSON, J. R. and R. A. TAPIA (1990), *Nonparametric Function Estimation, Modeling, and Simulation*, Society for Industrial and Applied Mathematics, Philadelphia.

WALKER, A. M. (1969), On the asymptotic behaviour of posterior distributions, *Journal of the Royal Statistical Society*, Series B, **31**: 80-88.

WILLMOT, G. E. (1994), *Introductory Credibility Theory*, Institute of Insurance and Pension Research, University of Waterloo, Waterloo, Ontario.

YOUNG, V. R. (1997), Credibility using a loss function from spline theory: Parametric models with a one-dimensional sufficient statistic, to appear, *North American Actuarial Journal*.

VIRGINIA R. YOUNG
School of Business
975 University Avenue
University of Wisconsin-Madison
Madison, WI USA 53706