EQUITY AND EXACT CREDIBILITY

BY

S. DAVID PROMISLOW AND VIRGINIA R. YOUNG

ABSTRACT

We consider an alternative to the usual credibility premium that arises from squared-error loss, namely, a so-called *equitable* credibility premium (Promislow and Young, 1999). We derive formulas for the credibility weight in certain cases and give sufficient conditions for exact credibility.

1. INTRODUCTION

When setting premiums for insurance, inequities will necessarily arise when, due to imperfect information, some policyholders are charged more than they should be and others less. By building on the previous work of Promislow (1987, 1991), we deal with the problem of choosing premiums to minimize this inequity (Promislow and Young, 1999). Much of our work parallels classical credibility theory, but in place of the traditional squared-error loss functions, we use the family of entropy loss functions. This is a familiar family that has frequently appeared in the economics literature for the purpose of measuring income inequality. We obtain formulas for the optimal premiums, and in certain cases, we obtain explicit formulas for the best affine approximation to the optimal premiums. A natural question, then, is to ask how good the affine approximations are. A basic result of the classical squared-error approach is that they are often exact. This occurs (given certain regularity conditions) when probability distributions are chosen from the linear exponential family with conjugate priors (Jewell, 1974a,b). The purpose of this note is to investigate conditions of exactness for a particular case of an entropy loss function.

In Section 2, we set our notation and assumptions and briefly review previous work in credibility theory, including some of our work in Promislow and Young (1999). We consider a specific case of our equitable credibility estimator. In Section 3, we study the case for which the equitable credibility premium is constrained to be an affine function of the claim data.

For the special case investigated in this paper, we have an explicit expression for the credibility weight and determine a sufficient condition for exact credibility.

2. BACKGROUND

Assume that the total claims of a policyholder, or risk, in the ith policy period, is a random variable X_i whose distribution depends on $\theta, i = 1, 2, ...$, in which θ varies across policyholders and may be vector-valued. Assume that the X_i are independent (conditional on θ) and identically distributed. The value of θ completely determines the claim distribution of the policyholder. Assume that the value θ is fixed for a given risk, although it is generally unknown and unobservable. Denote the probability (density) function of θ by $\pi(\theta)$, also called the *structure function* (Bühlmann, 1970).

One goal of credibility theory is to calculate a premium for period $n + 1$ of a policyholder, given that the policyholder's claim experience in the first n periods is $X_n = x_n = (x_1, x_2, ..., x_n) \in (\mathbf{R}^+)^n$, or more generally given any information, such as a demographic data. Consider general credibility estimators, denoted by Y, in which Y is a real-valued function on the information given, such as $(\mathbf{R}^+)^n$, if the information is prior claim data. We use a capital letter to denote the credibility premium Y to emphasize that it is a random variable. If we constrain Y to be a linear function of the claim data x, then we write L for Y.

If one knew the value of θ that determines the claim distribution of a policyholder, then $E(X_{n+1}|\theta)$ would be the most equitable premium for period $n + 1$, or more simply $E(X|\theta)$. Let $\mu(\theta)$ denote $E(X|\theta)$; also, let μ denote EX. The inequity of any other premium is measured relative to this most equitable premium. A general procedure is to select an appropriate loss (or unfairness) function U and then to choose $Y(x_n)$ to minimize

$$EU[(Y(x_n), \mu(\theta))].$$

In Bühlmann's classical theory (1967, 1970), the loss function U is taken to be the traditional squared error. That is,

$$U[(Y(x_n), \mu(\theta))] = (Y(x_n) - \mu(\theta))^2. \tag{2.1}$$

The resulting credibility premium is the posterior expected value of the conditional mean

$$Y(x_n) = \int E(X_{n+1}|\theta)\pi(\theta|x_n)d\theta. \tag{2.2}$$

By restricting the form of the credibility premium L to be a linear combination of prior claims, and by assuming that the claims are conditionally independent and identically distributed, one deduces the credibility estimator

$$L(x_n) = (1 - Z)\mu + Z\bar{x}, \tag{2.3}$$

in which $\mu = EX = E[\mu(\theta)]$ is the overall, or grand, mean; $\bar{x}$ is the sample mean,

$$Z = \frac{n}{n + k}, \tag{2.4}$$

in which

$$k = \frac{E[\text{Var}(X|\theta)]}{\text{Var}[\mu(\theta)]} \tag{2.5}$$

is the ratio of the expected process variance to the variance of the hypothetical means.

In certain cases, the predictive mean (2.2) is an affine function of the sample mean and, thus, equals the linear credibility estimator (2.3). Jewell (1974a,b) verifies conditions under which this *exact credibility* occurs: Under certain regularity conditions, exact credibility occurs for probability distributions from the linear exponential family when one uses the conjugate prior.

Promislow (1987, 1991) and Promislow and Young (1999) argue that squared error is inappropriate for measuring unfairness, and they justify using the entropy family in its place. Squared error is a function of the *absolute difference* between the charged premium and the true premium, while unfairness should depend on the *relative difference* between these two quantities. For example, we consider an individual who should be charged 1 unit but is actually charged 10 units to be treated more unfairly than an individual who should be charged 1001 units but is actually charged 1010 units.

Promislow (1987, 1991) and Promislow and Young (1999) show that appropriate loss functions to meet this objective are of the form

$$U(Y, \mu(\theta)) = \mu(\theta)g(r),$$

in which r denotes the ratio $Y/\mu(\theta)$. That is, loss is expressed as a function of the relative difference, weighted by the true premiums. It is shown, moreover, that the function g should be convex and satisfy $g(1) = 0$. In this paper, we will deal with the case in which $g(r) = r^2 - 1$, which leads to the loss function

$$U_2(Y, \mu(\theta)) = \frac{Y^2}{\mu(\theta)} - \mu(\theta) \tag{2.6}$$

In place of g, we could take the function h given by $h(r) = (r - 1)^2$. Since h differs additively from g by a multiple of $(r - 1)$, it is not difficult to see that there will be no effect on the result when we compute expectations. It is of interest to note that the classical squared error loss can be expressed in a similar form but at the cost of distorting the weights. We can write (2.1) in the form

$$U(Y, \mu(\theta)) = \mu(\theta)^2 h(r),$$

in which $h(r) = (r - 1)^2$. The weights now are the squares of the true premiums, which give much higher weight than before to the high cost situations. Also, we can also compare this with squared percentage error, where the loss function is

$$U(Y, \mu(\theta)) = h(r).$$

In this case, the weights are distorted by being independent of the true cost.

Note that the loss function in (2.6) equals $U_c(Y, \mu(\theta))$ from Promislow and Young (1999) in the special case for which $c = 2$. We will restrict our attention to this case for the remainder of this paper. In place of formula (2.2), one now gets an optimal premium of

$$Y_2(x_n) = \frac{\mu}{E\left(\left(E_{\theta|x_n}\left[\mu(\theta)^{-1}\right]\right)^{-1}\right)} \left(E_{\theta|x_n}\left[\mu(\theta)^{-1}\right]\right)^{-1}. \tag{2.7}$$

There is a convenient analogue of formulas (2.3) through (2.5). Indeed, (2.3) holds with Z replaced by z_2 given as follows:

$$z_2 = \frac{n}{n + \dfrac{J}{\mu W(\mu(\theta))}}, \tag{2.8}$$

in which $J = E\left[\dfrac{X^2}{\mu(\theta)}\right] - \mu = E\left[\dfrac{\mathrm{Var}(X|\theta)}{\mu(\theta)}\right]$, and $W(A) = E(A)\, E(A^{-1}) - 1$,

for any positive random variable A (Promislow and Young, 1999). Note that z_2 approaches 1 as n goes to infinity.

The expression in (2.8) is similar to the formula for Z, given by (2.4) and (2.5), with J replacing the expected process variance $E[\mathrm{Var}(X|\theta)]$ as a measure of the variability in X given a value of θ, and with $\mu W(\mu(\theta))$ replacing the variance of the hypothetical means $\mathrm{Var}(\mu(\theta))$ as a measure of the heterogeneity of the risks. See Promislow and Young (1999) for further discussion of the "variance" measures J and W and for the derivation of z_2.

3. CONDITIONS FOR EXACTNESS

For the linear exponential family and conjugate priors, we derive a sufficient condition for exact credibility. By exact credibility, we mean that the equitable credibility estimator $Y_2(x_n)$ given by (2.7) is an affine function of the sample mean and, therefore, equals the credibility estimator $L_2(x_n) = (1 - z_2)\mu + z_2\bar{x}$, with z_2 given by (2.8).

Suppose $X|\theta$ is distributed according to a distribution from a linear exponential family. Specifically, the pf or pdf of $X|\theta$ is of the form

$$f(x|\theta) = \frac{p(x)e^{-x\theta}}{q(\theta)},$$

for $x \geq 0$ and for θ taking values in an interval (θ_0, θ_1), where $-\infty \leq \theta_0 < \theta_1 \leq \infty$. Note that q is the Laplace transform of p because $q(\theta) = \int_0^\infty p(x)e^{-x\theta}dx$. The conditional mean of $X|\theta$ is given by

$$\mu(\theta) = -\frac{q'(\theta)}{q(\theta)},$$

(Klugman et al., 1998). We concentrate on linear exponential families because if the sample mean is a sufficient statistic for θ and if the support of the pdf of the continuous random variable $X|\theta$ is independent of θ, then the distribution of $X|\theta$ comes from a linear exponential family (Lehmann, 1991, Theorem 5.4).

The natural conjugate prior of θ has the form

$$\pi(\theta) = \frac{\{q(\theta)\}^{-k}e^{-\mu k\theta}}{c(\mu, k)}, \quad \theta_0 < \theta < \theta_1,$$

for some μ and $k > 0$. The value $c(\mu, k)$ is a normalizing constant for given values of μ and k. Assume that $\pi(\theta_0) = \pi(\theta_1)$. It follows that $EX = \mu$, the posterior density of θ given x_n is of the same form as the prior with

$k^* = n + k$ and $\mu^* = \dfrac{\mu k + n\bar{x}}{k + n}$, and the predictive mean equals

$$E[X_{n+1}|x_n] = \mu^* = \frac{k}{n + k}\mu + \frac{n}{n + k}\bar{x},$$

in which $k = E[Var(X|\theta)]/Var[\mu(\theta)]$. Thus, we get exact credibility for the predictive mean.

To obtain exact credibility for the equitable estimator Y_2, assume that $\pi(\theta_0) = \pi(\theta_1)$ and that $\pi(\theta_0) v(\theta_0) = \pi(\theta_1) v(\theta_1)$, in which

$$v(\theta) = -\frac{q(\theta)}{q'(\theta)}$$

is the multiplicative inverse of the conditional mean. We next prove the following result for exact credibility.

Theorem 3.1 *Suppose that $\{f(X|\theta)\}$ is a linear exponential family and that the natural conjugate prior satisfies the regularity conditions on its boundary given above. If v satisfies the differential equation*

$$v'' = av'$$

for some constant a, then the equitable credibility estimator Y_2 is exact. Specifically,

$$Y_2(x) = (1 - z_2)\mu + z_2\bar{x}.$$

Moreover, $z_2 = \dfrac{n}{n + k - a/\mu}$.

Proof:

$$E[\mu(\theta)|x]^{-1} = \int_{\theta_0}^{\theta_1} v(\theta)\pi(\theta|\bar{x})d\theta = c(k^*, \mu^*)^{-1} \int_{\theta_0}^{\theta_1} v(\theta)q(\theta)^{-k^*} e^{-a\theta} e^{-(\mu^* k^* - a)\theta} d\theta.$$

Note that $\left[q^{-k*} ve^{-a\theta}\right]' = -k^* q^{-k^*-1} q' ve^{-a\theta} + q^{-k^*} v'e^{-a\theta} - aq^{-k^*} ve^{-a\theta}$. By using

the definition of v and the fact that $v' - av$ is a constant, we deduce that

$$\left[q^{-k^*} ve^{-a\theta}\right]' = Kq^{-k^*} ve^{-a\theta},$$

for some constant K.

We next integrate by parts and obtain

$$E[\mu(\theta)|x]^{-1} = \frac{K}{\mu^* k^* - a}.$$

Since $\mu^* k^* = \mu k + n\bar{x}$, we have that

$$Y_2(x) = \frac{\mu}{\mu(n + k) - a}(\mu k + n\bar{x} - a),$$

and the result follows.   □

Remarks:
(1) Note that z_2 will be equal to, greater than, or less than the corresponding Bühlmann credibility weight Z, according as a is zero, positive, or negative.
(2) The possibilities for v are limited. If $a = 0$, then $v(\theta) = c_1\theta + c_2$, for some constants c_1 and c_2. If $a \neq 0$, then $v(\theta) = c_1 e^{a\theta} + c_2$, for some constants c_1 and c_2. After the following examples, we determine the functions q and p that correspond to these forms of v.

Example 3.2 (Gamma-Gamma) Let $X|\theta \sim \text{Gamma}(\gamma, \theta)$ with conditional mean $\dfrac{\gamma}{\theta}$ and conditional variance $\dfrac{\gamma}{\theta^2}$, in which the shape parameter $\gamma > 0$ is known, and let $\theta \sim \text{Gamma}(\alpha, \beta)$. The differential equation of Theorem 3.1 holds with

$$a = 0, \quad z_2 = \frac{n}{n + \dfrac{\alpha - 1}{\gamma}} = Z, \quad Y_2(x) = (1 - z_2) \cdot \frac{\gamma \beta}{\alpha - 1} + z_2 \cdot \bar{x} = \text{E}[X_{n+1}|x_n]. \quad \square$$

Example 3.3 (Poisson-Gamma) Let $X|\lambda \sim \text{Poisson}(\lambda)$ with conditional mean λ, and let $\lambda \sim \text{Gamma}(\alpha, \beta)$. To put this in standard form, let $\theta = -\ln(\lambda)$. Then, $v(\theta) = e^\theta$, and the differential equation for v holds with $a = 1$.

The credibility weight z_2 equals $\dfrac{n}{n + \beta - \beta/\alpha}$ versus $Z = \dfrac{n}{n + \beta}$, and $Y_2(x) = (1 - z_2) \cdot \dfrac{\alpha}{\beta} + z_2 \cdot \bar{x}$. Note that $Z < z_2$, so that the equitable

premium Y_2 gives more weight to the policyholder's experience than in the Bühlmann credibility estimator. $\square$

Example 3.4 (Binomial-Beta) Let $X|p \sim \text{Binomial}(r, p)$ with conditional mean rp, and let $p \sim \text{Beta}(\alpha, \beta)$. To put this in standard form, let $\theta = -\ln(p/(1 - p))$. Then, $v(\theta) = (1 + e^\theta)/r$, and the differential equation for v holds with $a = 1$. The credibility weight z_2 equals $\dfrac{n}{n + \dfrac{(\alpha + \beta)(\alpha - 1)}{r\alpha}}$

versus $Z = \dfrac{n}{n + \dfrac{\alpha + \beta}{r}}$, and $Y_2(x) = (1 - z_2) \cdot \dfrac{r\alpha}{\alpha + \beta} + z_2 \cdot \bar{x}$. Note that

$Z < z_2$ because $a > 0$, as in Example 3.3. $\square$

Example 3.5 (Negative Binomial-Beta) Let $X|p \sim \text{Negative Binomial}(r, p)$ with probability function

$$f(x|p) = \binom{r + x - 1}{x} p^r (1 - p)^x, \quad x = 0, 1, 2, \ldots,$$

in which $r > 0$ is known, and let $p \sim \text{Beta}(\alpha, \beta)$. To put this in standard form, let $\theta = -\ln(1 - p)$. Then, $v(\theta) = (e^\theta - 1)/r$, and the differential equation for v holds with $a = 1$. The credibility weight z_2 equals

$$\frac{n}{n + \dfrac{(\alpha - 1)(\beta - 1)}{r\beta}} \text{ versus } Z = \frac{n}{n + \dfrac{\alpha - 1}{r}}, \text{ and } Y_2(x) = (1 - z_2) \cdot \frac{r\beta}{\alpha - 1} + z_2 \cdot \bar{x}.$$

Note that $Z < z_2$ because $a > 0$, as in Examples 3.3 and 3.4.   □

It is not always the case for the linear exponential family with conjugate prior, that the equitable premium Y_2 is an affine function of the sample mean $\bar{x}$, as we see in the next example.

Example 3.6 Let $X|\theta$ have pdf $f(x|\theta) = \dfrac{(x+1)\theta^2 e^{-x\theta}}{\theta+1}$, for $x > 0$, a member of the linear exponential family, and let θ have pdf proportional to $\dfrac{\theta^2 e^{-\theta}}{\theta+1}$, for $\theta > 0$, the natural conjugate prior for θ. After some tedious calculation, one finds that for $n = 1$,

$$Y_2(x) \propto c\left[\frac{3}{8b^4} - \frac{3}{4b^3} + \frac{7}{4b^2} - \frac{15}{2b} - e^{2b}\int_1^\infty \frac{e^{-2bu}}{u}du + 32e^{2b}\int_2^\infty \frac{e^{-2bu}}{u}du\right]^{-1},$$

in which $c = \left[\dfrac{1}{4b^3} - \dfrac{1}{2b^2} + \dfrac{3}{2b} - 4e^{2b}\displaystyle\int_1^\infty \frac{e^{-2bu}}{u}du + e^{2b}\int_1^\infty \frac{e^{-2bu}}{u^2}du\right]\mu$, and $b = \dfrac{x+1}{2}$. Via numerical calculation, one can verify that Y_2 is not linear in x.   □

Now, we return to the problem of determining which distributions of $X|\theta$ lead to $v'' = av'$, for some constant a. We consider the following cases:

(1) $a = 0 \Rightarrow v(\theta) = c_1\theta + c_2$, for some constants c_1 and c_2, not both 0.

    (a) $c_1 = 0 \Rightarrow q(\theta) = c_3 e^{-\theta/c_2}$, for some constant c_3. Because p is the inverse Laplace transform of q, we have that $p(x)$ is a point mass at $x = 1/c_2$. It follows that $f(x|\theta) = 1$ if $x = 1/c_2$ and 0 otherwise.

    (b) $c_1 \neq 0 \Rightarrow q(\theta) = c_3(c_1\theta + c_2)^{-1/c_2}$, for some constant c_3. It follows that $p(x)$ is proportional to $x^{\gamma-1}e^{-\lambda x}$, in which $\gamma = 1/c_1$ and $\lambda = c_2/c_1$. Thus, $f(x|\theta)$ is proportional to $x^{\gamma-1}e^{-x(\lambda+\theta)}$, from which it follows that $X|\theta$ is distributed according to Gamma$(\gamma, \lambda+\theta)$. In Example 3.2, we saw a special instance of this case in which $\lambda = 0$.

(2) $a \neq 0 \Rightarrow v(\theta) = c_1 e^{a\theta} + c_2$, for some constants c_1 and c_2, not both 0.

    (a) $c_2 = 0 \Rightarrow q(\theta) = c_3\exp\left(\dfrac{1}{ac_1}e^{-a\theta}\right)$, for some constant c_3. It follows that $p(x)$ is proportional to

$$\delta(x) + \frac{1}{ac_1}\delta(x-a) + \frac{1}{2(ac_1)^2}\delta(x-2a) + \dots + \frac{1}{n!(ac_1)^n}\delta(x-na) + \dots,$$

in which δ is the Dirac delta function. Thus, $f(x|\theta) \propto \dfrac{e^{-x\theta}}{(ac_1)^{x/a}(x/a)!}$, for $x = 0, a, 2a, \dots$. If $a = c_1 = 1$, then we have that $X|\theta$ is distributed according to the Poisson distribution with conditional mean $e^{-\theta}$, as in Example 3.3.

(b) $c_2 \neq 0 \Rightarrow q(\theta) = c_3(c_1 + c_2 e^{-a\theta})^{1/c_2}$, for some constant c_3. It follows

that $p(x)$ is proportional to $\sum_{m=0}^{\infty} \binom{1/c_2}{m} (c_1/c_2)^{(1/c_2-m)} \delta(x - am)$,

under appropriate regularity conditions. For example, if $1/c_2 = r$, a positive integer, if $c_1 = c_2$, and if $a = 1$, then we have the binomial distribution, as in Example 3.4. If $1/c_2 = r$, a negative integer, if $c_1 = c_2$, and if $a = 1$, then we have the negative binomial distribution, as in Example 3.5.

We see that in some sense, Examples 3.2 through 3.5 cover the simplest of the interesting cases for the distribution of $X|\theta$ for which the conditions of Theorem 3.1 hold. Also, note from the above discussion that if $\mu(\theta) \geq 0$, then $a \geq 0$ from which it follows that $z_2 \geq Z$.

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.
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, 336-341.
KLUGMAN, S., PANJER, H., and WILLMOT, (1998). *Loss Models: From Data to Decisions*, Wiley, New York.
LEHMANN, E.L. (1991) *Theory of Point Estimation*, Wadsworth, Pacific Grove, California.
PROMISLOW, S.D. (1987) Measurement of equity, *Transactions of the Society of Actuaries* 39, 215-256.
PROMISLOW, S.D. (1991) An axiomatic characterization of some measures of unfairness, *Journal of Economic Theory* 53, 345-368.
PROMISLOW, S.D. and YOUNG, V.R. (1999) Equity and credibility, *Scandinavian Actuarial Journal*, to appear.
VENTER, G. (1990) *Credibility. Foundations of Casualty Actuarial Science*, Casualty Actuarial Society, New York.

PROF. V.R. YOUNG
Actuarial Science, School of Business
University of Wisconsin-Madison
975 University Avenue
Madison, Wisconsin 53706-1323
USA