Mathematical Statistics
Stockholm University

Credibility theory and GLM revisited

Esbjörn Ohlsson
Björn Johansson

Research Report 2003:15

ISSN 0282-9150

Postal address:
Mathematical Statistics
Dept. of Mathematics
Stockholm University
SE-106 91 Stockholm
Sweden

Internet:
http://www.math.su.se/matstat

Mathematical Statistics
Stockholm University
Research Report **2003:15**,
http://www.math.su.se/matstat

Credibility theory and GLM revisited

Esbjörn Ohlsson * Björn Johansson†

October 2003

Abstract

In non-life insurance, credibility estimates are appropriate for rating factors with many levels, where some have insufficient data. Traditional credibility theory models such *multi-level factors* as random effects, but does not treat the situation where there are also ordinary rating factors (fixed effects). Nelder and Verrall (1997) suggested using generalized linear models (GLMs) with both fixed and random effects and showed that in the absence of fixed effects, their result reduces to the classical credibility estimate. In this paper we further explore the relation between credibility and GLMs and derive MSE-based predictors of the random effects, by giving an extension of Jewell's theorem to the setting with both types of rating factors (fixed and random). We also derive sums-of-squares variance estimators in the spirit of traditional credibility theory and present an application to car model rating.

Keywords

Premium rating, credibility theory, Jewell's theorem, generalized linear models, Tweedie models, multi-level factor.

*Mathematical Statistics, Stockholm University, and Länsförsäkringar insurance group. E-mail: esbjorn.ohlsson@math.su.se

†Länsförsäkringar insurance group. E-mail: bjorn.johansson@lansforsakringar.se

Contents

1 Introduction

The most common rating technique in non-life insurance is to estimate the price relativities of a number of rating factors in a multiplicative model. Usually, these rating factors are either categorical with a few levels (e.g. sex) or continuous (e.g. age). In the latter case it is customary to discretize by forming groups of adjacent values. However, a problem arises when you have categorical rating factors with many levels without an inherent ordering, such as car model or geographical region. The problem is that there is no natural way of forming groups with sufficient data, as you do with ordered variables like age or annual mileage. In this paper we use the term *multi-level factors* for rating factors with many levels where there is little data for some levels, while there might be much data for others — for instance there is a number of very common car models but also some very uncommon ones. Another example of a multi-level factor is company — the common situation in experience rating in the commercial lines where the rating of different companies is based to a certain extent on the individual claims experience.

Rating of a multi-level factor is a standard context for employing credibility theory, where the premium for a certain level takes into consideration the amount of data for that level. However, in most cases you have ordinary rating factors, like age and sex, alongside with the multi-level factors, but traditional credibility theory only treats the analysis of multi-level factors by themselves.

In recent years, it has become standard practice in non-life insurance to use generalized linear models (GLMs) for the analysis of rating factors. Nelder and Verrall (1997) show that credibility-like properties can be achieved by introducing multi-level factors as *random effects* in GLMs (although they do not use the term multi-level factor). For estimation in these random effects models, they use the theory of Lee and Nelder

(1996), where *hierarchical likelihood* is applied to develop *hierarchical generalized linear models*, HGLM. Nelder and Verrall show that for a single multi-level factor in the absence of other rating factors, the traditional credibility estimator of Bühlmann is recovered.

In this paper we explore further the connection between credibility theory and random effects in GLMs. In traditional credibility theory random effects are estimated by means of minimum mean square error (MSE) predictors rather than the likelihood-based approach used in HGLMs. We study the possibility of finding MSE-based predictors for the GLMs most used in actuarial practice: multiplicative models with variance function of the power type. We suggest that understanding the random effects models from the actuarial point of view of credibility may enhance the use of these important models by actuaries.

Our main result is an extension of the famous theorem by Jewell (1974) — in it's generalized, weighted form given by Kaas et al. (1997) — to the situation where one has other rating factors alongside with the multi-level factor. The resulting predictor of the random "risk parameter" associated with the multi-level factor specializes to the classical Bühlmann-Straub estimator in the case with no other rating factors than just a single multi-level factor. We will further see that for the cases studied here, the unbiased predictors are close to the HGLM predictors, and in the Poisson case they actually coincide.

Besides the dispersion parameter ϕ of standard GLMs, the random effects model introduces an additional variance-related parameter α that has to be estimated. Lee and Nelder suggests a likelihood-based approach, whereas in credibility theory, unbiased estimators of variance parameters based on sums of squares are standard, see e.g. Goovaerts and Hoogstad (1987). We show that for the random effects GLMs we consider, one can find unbiased estimators of variance parameters that generalizes the standard sums of squares estimators. This leads to an alternative estimator of α, closely related to the estimators of variance parameters in classical credibility.

Here is an outline of the paper. In Section 2 we first review the important *Tweedie models* which include the most common models used in actuarial practice. We then prove our extension of Jewell's theorem. In Section 3 we derive estimators of variance parameters. In Section 4, finally, we present an application of our method to car insurance.

2 Extension of Jewell's theorem

In non-life insurance pricing, one studies the effect of rating factors on some key ratio Y_i, typically the risk premium, claims frequency or average claim amount. Nowadays many companies use generalized linear models (GLMs) in this process. By far, the most commonly used models are GLMs with a variance function of the form $v(\mu) = \mu^p$ for some p. We start out by repeating some basic facts about these so called Tweedie models, before deriving our extension of Jewell's theorem.

2.1 Tweedie models

The rating factors divide the portfolio into *tariff cells*, and the key ratio Y_i is computed over the policies in cell i. In GLMs, Y_i is assumed to have a frequency function of the form

$$f_{Y_i}(y_i; \theta_i, \phi) = \exp\left\{ \frac{y_i\theta_i - b(\theta_i)}{\phi/w_i} + c(y_i, \phi, w_i) \right\} \tag{2.1}$$

where ϕ is the *dispersion parameter* and w_i is the known exposure weight (the denominator of the key ratio). The function $b(\theta)$ is twice differentiable with a unique inverse for the first derivative $b'(\theta)$. With $\phi = 1$ and all $w_i = 1$, (2.1) would be the exponential family with canonical parameter considered by Jewell (1974). Jørgensen (1997) uses the name *reproductive exponential dispersion models* for (2.1).

From standard GLM theory we know that $\mu_i \doteq E(Y_i) = b'(\theta_i)$. If we further denote the inverse of b' by $h(\mu)$, we can express the variance as

$$\text{Var}(Y_i) = \phi\, b''(\theta_i)/w_i = \phi\, v(\mu_i)/w_i \tag{2.2}$$

where $v(\mu) = b''(h(\mu))$ is called the *variance function*. We will only consider a subclass of the reproductive exponential dispersion models here, viz. the ones that have

$$v(\mu) = \mu^p \tag{2.3}$$

In the terminology of Jørgensen, these are called *Tweedie models*. In the rest of this section we recapitulate some of their theory – for proofs, see Jørgensen (1997). The Tweedie models are defined only for p outside the interval $0 < p < 1$. Renshaw (1994) concludes that models with $p \leq 0$ "are of no practical concequence" in non-life insurance rating — one reason being that they have support on the whole real line, while our key ratios are non-negative. We thus restrict ourselves to the class with $p \geq 1$. Our calculations below carry through for $p = 0$ (Gaussian distribution), but that model is not appropriate for the non-negative key ratios with multiplicative rating factors that we consider here and is hence excluded.

Some special cases of Tweedie models with $p \geq 1$ are:

- $p = 1$: (Weighted) Poisson distribution.
- $1 < p < 2$: Compound Poisson distribution with gamma distributed summands.
- $p = 2$: Gamma distribution.
- $p = 3$: Inverse Gaussian distribution.

The case $1 < p < 2$ is applicable to risk premiums Y_i with a Poisson distributed number of claims and gamma distributed claim sizes. Here $p = (2 + \gamma)/(1 + \gamma)$, where γ is the shape parameter of the gamma distribution, see Jørgensen & Souza (1994), and hence $\gamma^{-1/2}$ is the coefficient of variation of that distribution.

As shown by Jørgensen (1997, Theorem 4.1), the Tweedie models are the only reproductive exponential distribution models that are closed under scale transformations. Since this is a desirable property of any economic model (a change of currency should not alter the distributional family), and since the distributions that are standard in actuarial applications of GLMs are included in the class — see the list above — the restriction to the Tweedie class with $p \geq 1$ is not serious.

The Tweedie models are continuous, except for $p = 1$ (Poisson) and $1 < p < 2$ (compound Poisson), the latter being continuous except for a positive probability at zero (Jørgensen p. 129).

From the relation between $v(\mu)$ and $b''(\theta)$, Jørgensen derives the functional form of the $b(\theta)$ corresponding to a variance function as in (2.3). The result is (for $p \geq 1$)

$$b(\theta) = \begin{cases} e^{\theta} & \text{for } p = 1; \\ -\log(-\theta) & \text{for } p = 2; \\ -\frac{1}{p-2}\left[-(p-1)\theta\right]^{(p-2)/(p-1)} & \text{for } 1 < p < 2 \text{ and } p > 2. \end{cases} \tag{2.4}$$

The canonical (maximal) parameter space M is

$$M = \begin{cases} -\infty < \theta < \infty & \text{for } p = 1; \\ -\infty < \theta < 0 & \text{for } 1 < p \leq 2; \\ -\infty < \theta \leq 0 & \text{for } p > 2. \end{cases} \tag{2.5}$$

We will also need expressions for the derivative $b'(\theta)$,

$$b'(\theta) = \begin{cases} e^{\theta} & \text{for } p = 1; \\ \left[-(p-1)\theta\right]^{-1/(p-1)} & \text{for } p > 1. \end{cases} \tag{2.6}$$

and its inverse $h(\mu)$,

$$h(\mu) = \begin{cases} \log(\mu) & p = 1; \\ -\frac{1}{p-1}\,\mu^{-(p-1)} & p > 1. \end{cases} \tag{2.7}$$

In GLM theory, $h(\mu)$ is called the canonical link function. Note, however, that we do not assume the use of canonical link, but instead use a log-link (multiplicative model) throughout. Multiplicative models are standard in insurance practice and usually a very reasonable choice.

2.2 Random effects in Tweedie models

Suppose we have a number of ordinary rating factors, dividing the portfolio into I tariff cells — by an ordinary rating factor we mean one that is not a multi-level factor. Suppose furthermore that we have a multi-level factor with K levels. Let w_{ik} denote the exposure weight (number of policy years for claim frequencies, number of claims for average claim amounts, etc.) in the ith tariff cell with respect to the ordinary rating factors and for the kth level of the multi-level factor. Let Y_{ik} denote the corresponding observed key ratio, considered as a random variable.

We shall assume the effect of the multi-level factor to be multiplicative. For level k of the multi-level factor, this effect is considered to be the outcome of a random variable U_k. We then have

$$E(Y_{ik}|U_k = u_k) = \mu_i u_k \tag{2.8}$$

Typically, the effects of the ordinary rating factors are also multiplicative, but this does not matter for what follows. Since the systematic effects are captured by μ_i, we can let the U_k:s be purely random, so that we have $E(U_k) = 1$, and hence $E(Y_{ik}) = \mu_i$.

Conditionally on $U_k = u_k$ we assume that Y_{ik} follows a Tweedie model with expectation $\mu_i u_k$. Now, the frequency function in (2.1) is defined in terms of the canonic parameter θ, rather than the expectation μ, and in our case this parameter becomes $\theta'_{ik} = h(\mu_i u_k)$. We make the corresponding transformation of the random effect and introduce the random variable $\Theta_k = h(U_k)$, which corresponds to the *risk parameter* in Jewell (1974) and other sources on standard credibility theory, taking values $\theta_k = h(u_k)$.

Note that by (2.7)

$$\theta'_{ik} = h(\mu_i u_k) = \begin{cases} \log(\mu_i) + h(u_k) & p = 1; \\ \mu_i^{-(p-1)} h(u_k) & p > 1. \end{cases} \tag{2.9}$$

and then by (2.4),

$$\begin{aligned} b(\theta'_{ik}) &= b(h(\mu_i u_k)) \\ &= \begin{cases} \log(\mu_i) + b(\theta_k) & p = 2; \\ \mu_i^{2-p} b(\theta_k) & 1 \leq p < 2 \text{ and } p > 2. \end{cases} \end{aligned} \tag{2.10}$$

Now for all the p we consider, we can write

$$\begin{aligned} f_{Y_{ik}|\Theta_k}(y_{ik}|\theta_k) &= \exp\left\{\frac{y_{ik}\theta'_{ik} - b(\theta'_{ik})}{\phi/w_{ik}} + c_1\right\} \\ &= \exp\left\{\frac{w_{ik}}{\phi}\left[\frac{y_{ik}}{\mu_i^{p-1}}\,\theta_k - \frac{1}{\mu_i^{p-2}}\,b(\theta_k)\right] + c_2\right\} \end{aligned} \quad (2.11)$$

where c_1 and c_2 are constants that does not depend on θ_k, and into c_2 we have incorporated the terms $\log(\mu_i)$ appearing in (2.9) and (2.10).

Conditional on $\Theta_k = \theta_k$, or equivalently on $U_k = u_k$, the Y_{ik}:s are assumed independent. We can then perform a standard GLM analysis of the ordinary rating factors, using $\log(u_k)$ as an *offset* variable. Now, the U_k:s are of course non-observable and must be predicted. We will follow Jewell (1974) and assume that the density function of $\Theta = h(U)$ is the *natural conjugate prior* to the family in (2.1), which is given by

$$f_\Theta(\theta) = \frac{1}{c(\delta, \alpha)} \exp\left\{\frac{\theta\delta - b(\theta)}{1/\alpha}\right\} \quad (2.12)$$

for $\theta \epsilon M$ (the canonical parameter space of (2.1)), where $c(\delta, \alpha)$ is just a normalizing constant. For all $p \geq 1$, this is a proper distribution if $\alpha > 0$ and $\delta > 0$. The same is true for $p = 2$ if $-1 < \alpha < 0$ and $\delta < 0$; and for $p > 2$ if $\alpha < 0$ and $\delta < 0$.

Lemma 2.1 *Let $U = b'(\Theta)$, where Θ follows the distribution in (2.12) and let $\inf M$ and $\sup M$ denote the lower and upper bound of the interval M in (2.5).*

(a) Suppose that $f_\Theta(\inf M) = f_\Theta(\sup M) = 0$. Then

$$\delta = E(U)$$

(b) Suppose that $f'_\Theta(\inf M) = f'_\Theta(\sup M) = 0$. Then

$$\alpha = \frac{E\left[U^p\right]}{\mathrm{Var}(U)} \quad (2.13)$$

Proof. We have

$$\begin{aligned} f'_\Theta(\theta) &= \alpha(\delta - b'(\theta)) f_\Theta(\theta) \\ f''_\Theta(\theta) &= \alpha^2(\delta - b'(\theta))^2 f_\Theta(\theta) - \alpha b''(\theta) f_\Theta(\theta) \end{aligned}$$

Upon integrating these equations we get, under the assumptions of the limiting behavior of $f_\Theta(\theta)$ and $f'_\Theta(\theta)$, respectively,

$$\begin{aligned} 0 &= \alpha \int_M (\delta - b'(\theta)) f_\Theta(\theta)\, d\theta = \alpha(\delta - E[b'(\Theta)]) \\ 0 &= \alpha^2 \int_M (b'(\theta) - \delta)^2 f_\Theta(\theta)\, d\theta - \alpha \int_M b''(\theta) f_\Theta(\theta)\, d\theta \\ &= \alpha^2 \mathrm{Var}(b'(\Theta)) - \alpha E[b''(\Theta)] \end{aligned}$$

Now the fact that $u = b'(\theta)$ and that $b''(\theta) = b''(h(u)) = v(u) = u^p$ completes the proof. □

We next investigate to what extent the assumptions of the lemma are satisfied for our Tweedie models with $p \geq 1$.

Lemma 2.2 *(a) The assumptions of Lemma 2.1(a) are satisfied for* $1 \leq p < 2$ *and for* $p = 2$ *provided that* $\alpha > 0$*. They are not valid for* $p > 2$*.*

(b)The assumptions of Lemma 2.1(b) are satisfied for $1 \leq p < 2$*. For* $p = 2$ *they are satisfied if* $\alpha > 1$*, but not for* $\alpha \leq 1$*. For* $p > 2$ *they are invalid.*

The proof of this lemma is straightforward. For $p = 2$ one may add that when $\alpha \leq 1$ the variance $\mathrm{Var}(U)$ does not exist and when $\alpha \leq 0$ not even the expectation $E(U)$ exists.

Lemma 2.1 is fundamental to the proof of our main result. Therefore, from now on, we restrict ourselves to the case $1 \leq p \leq 2$, for which we conclude that $\delta = E(U) = 1$ so that in effect, we have just one parameter in the conjugate distribution, $\alpha > 0$, which can be interpreted in terms of moments of U if for $p = 2$ we restrict the conjugate distribution to $\alpha > 1$. For $p > 2$, the fundamental reduction of parameters by setting $\delta = 1$ does not work and neither does the interpretation of α in terms of moments of U — it is notable that Lee and Nelder (1996, p. 626) also comment on difficulties in using the conjugate distribution in the inverse Gaussian case $p = 3$. Fortunately, $1 \leq p \leq 2$ contains the most commonly used distributions in insurance applications of GLMs, namely the Poisson, gamma and compound Poisson-gamma distributions. (Note that the assumptions of Lemma 2.1(a) also appear in the original theorem by Jewell (1974) and so his results are not valid for Tweedie models with $p > 2$.)

2.3 Main result

In his classical result, Jewell (1974) assumed an exponential family of distributions for Y, conditionally on the so called *risk parameter* (our Θ_k). Kaas et al (1997) generalized Jewell's theorem to the exponential dispersion models used in GLMs, including weights w. Before presenting our extension of these results, we make some basic assumptions that are more or less standard in credibility theory.

Assumption 1 *(a)* $\Theta_k;\ k = 1, 2, \ldots, K$ *are independent and identically distributed random variables.*

(b) For $k = 1, 2, \ldots, K$, *the pairs* (Y_{ik}, Θ_k) *are independent.*

(c) Conditioned on Θ_k *the random variables* $Y_{1k}, Y_{2k}, \ldots, Y_{I_k,k}$ *are independent.*

By (2.8) we have $E(Y_{ik}|U_k) = \mu_i U_k$, where μ_i is the mean given by the ordinary rating factors, which can be estimated by standard GLM methods once we have the u_k. Hence, in our case the search for credibility "estimators" amounts to finding a predictor of U_k, for every k.

In analogy with classical credibility theory, we look for functions g of our data vector $\boldsymbol{Y}$ that minimize

$$E\left[(U_k - g(\boldsymbol{Y}))^2\right] \tag{2.14}$$

It is well known that the solution to this minimization problem is $g(\boldsymbol{Y}) = E[U_k|\boldsymbol{Y}]$. By assumption 1(b) we can restrict the conditioning to $\boldsymbol{Y}_k = \{Y_{ik}; i = 1, 2, \ldots, I_k\}$ and our optimal predictor is $g(\boldsymbol{Y}) = E[U_k|\boldsymbol{Y}_k] = E[b'(\Theta_k)|\boldsymbol{Y}_k]$. An expression for this *posterior mean* is given in the following extension of Jewell's theorem, which is our main result.

Theorem 2.1 *Suppose that conditionally on U_k we have a Tweedie model for Y_{ik} with $1 \leq p \leq 2$ and that $\Theta_k = h(U_k)$ follows the natural conjugate distribution given by (2.12), where $\alpha > 0$ and $\delta > 0$. Let Assumption 1(b) and (c) be satisfied. Then the optimal predictor $E(U_k|\boldsymbol{Y}_k)$ of the random effect U_k can be written as*

$$\hat{u}_k = \frac{\sum_i w_{ik} y_{ik}/\mu_i^{p-1} + \phi\alpha}{\sum_i w_{ik}\mu_i^{2-p} + \phi\alpha} \tag{2.15}$$

Proof. To compute the posterior expectation we need the posterior distribution of Θ_k, which we get from Bayes theorem, plus (2.11) and (2.12).

$$\begin{aligned} f_{\Theta_k|\boldsymbol{Y}_k}(\theta|\boldsymbol{y}_k) &\propto f_{\Theta_k}(\theta) f_{\boldsymbol{Y}_k|\Theta_k}(\boldsymbol{y}_k|\theta) = f_{\Theta_k}(\theta) \prod_i f_{Y_{ik}|\Theta_k}(y_{ik}|\theta) \\ &\propto \exp\{\alpha(\theta - b(\theta))\} \prod_i \exp\left\{\frac{w_{ik}}{\phi}\left[\frac{y_{ik}}{\mu_i^{p-1}}\,\theta - \mu_i^{2-p}\, b(\theta)\right]\right\} \qquad (2.16) \\ &= \exp\left\{\theta\left(\alpha + \frac{1}{\phi}\sum_i w_{ik}\frac{y_{ik}}{\mu_i^{p-1}}\right) - b(\theta)\left(\alpha + \frac{1}{\phi}\sum_i w_{ik}\,\mu_i^{2-p}\right)\right\} \end{aligned}$$

Since we are using a conjugate prior, it is no surprise that the posterior distribution is a member of the same family, with new "updated" parameters

$$\alpha' = \alpha + \frac{1}{\phi}\sum_i w_{ik}\,\mu_i^{2-p} \qquad \delta' = \hat{u}_k \tag{2.17}$$

where $\hat{u}_k$ is given by (2.15). Finally, from Lemma 2.1(a) and 2.2(a) the expectation of U_k in the posterior distribution is just δ' and the proof is complete. □

We can rewrite (2.15) in the form of a classical credibility estimator. First, introduce the weighted average

$$\overline{u}_k = \frac{\sum_i (w_{ik}\mu_i^{2-p})\, y_{ik}/\mu_i}{\sum_i w_{ik}\mu_i^{2-p}} \tag{2.18}$$

We can regard $\overline{u}_k$ as an *experience factor*, indicating how one might adjust the expected values μ_i to take into account the experiences y_{ik}. Note that $\overline{u}_k$ is the limiting value of (2.15) as $\alpha \to 0$. We can now rewrite the optimal predictor as

$$\hat{u}_k = z_k\, \overline{u}_k + (1 - z_k) \cdot 1 \tag{2.19}$$

where the *credibility factor* z_k is defined by

$$z_k \doteq \frac{\sum_i w_{ik}\mu_i^{2-p}}{\sum_i w_{ik}\mu_i^{2-p} + \phi\alpha} \tag{2.20}$$

Thus the optimal predictor $\hat{u}_k$ is a credibility weighted adjustment factor to the rating by the ordinary factors, which for each tariff cell is given by it's corresponding μ_i. High credibility, which occurs with large exposures w_{ik} and/or small $\phi\alpha$, gives large weight to the experience factor $\overline{u}_k$, while low credibility gives more weight to the factor 1, i.e. 'no adjustment'.

We proceed to rewrite z_k on a more familiar form. The variance parameters a and s^2 in Bühlmann-Straub theory (see e.g. Goovaerts & Hoogstad, 1987, p 47) here become, by (2.8), (2.2) and (2.3),

$$\begin{aligned} a_i &\doteq \mathrm{Var}(E[Y_{ik}|\Theta_k]) = \mu_i^2 \mathrm{Var}(U_k) \\ s_i^2 &\doteq E[w_{ik}\mathrm{Var}(Y_{ik}|\Theta_k)] = \phi\mu_i^p E[U_k^p] \end{aligned} \tag{2.21}$$

Hence, by (2.13),

$$z_k = \frac{\sum_i w_{ik}a_i/s_i^2}{\sum_i w_{ik}a_i/s_i^2 + 1} \tag{2.22}$$

We find that we have high credibility when the random effect variances a_i are large compared to the variances s_i^2 between the observations *and/or* when the exposures w_{ik} are large. This is in direct analogy with classical credibility.

2.3.1 Comparison with classical credibility theory

We next specialize to the classical case without any ordinary rating factors, i.e., we assume for the moment that all μ_i are equal, $\mu = \mu_i$. Hereby, a and s^2 in (2.21) do not depend on i and we have $z_k = aw_{\cdot k}/(aw_{\cdot k} + s^2)$ as in Goovaerts & Hoogstad (1987, p 47). Classical credibility seeks predictors of $E(Y_{ik}|\Theta_k)$, which in our notation equals μU_k and so we multiply our predictor $\hat{u}_k$ by μ and arrive at $\hat{u}_k = z_k\,\overline{y}_k + (1-z_k)\cdot\mu$, where we have used that $\mu\overline{u}_k$ here equals $\overline{y}_k = \sum_i w_{ik}y_k/w_{\cdot k}$.

This is readily seen to be the same as the predictor in (11) of Kaas et al. (1997), if we identify our μ as x_0 there, and note that our $s^2/a = \phi\alpha/\mu^{2-p}$ corresponds to their w_0. Note that the predictor by Kaas et al. corresponds to the classical Bühlmann-Straub estimator. Hence, our result is really an *extension* of classical credibility to cases where ordinary rating factors are present. It is not a proper *generalization* though, since our result is restricted to the (albeit important) Tweedie family with $1 \leq p \leq 2$.

2.3.2 Special cases

Finally, it is of interest to specialize our predictors to the important special cases $p = 1$ (Poisson) and $p = 2$ (gamma), by looking at the corresponding experience factors.

$$p = 1 \quad \Rightarrow \quad \overline{u}_k = \frac{\sum_i w_{ik}y_{ik}}{\sum_i w_{ik}\mu_i} \tag{2.23}$$

$$p = 2 \quad \Rightarrow \quad \overline{u}_k = \frac{\sum_i w_{ik}\frac{y_{ik}}{\mu_i}}{\sum_i w_{ik}} \tag{2.24}$$

It can be noted that equation (2.23) corresponds to an estimating equation in the so called *method of marginal totals* and that (2.24) corresponds to the so called *direct method* (see, e.g., Kaas et al., 2001, pp. 179-181).

In these cases — and in general — equation (2.18) is easily verified to be the estimating equation for u_k if considered a *fixed* effect in a standard GLM analysis (remembering that we are using a log-link). This is appealing: in a case with with very high credibility our predictors are the same as the estimators resulting from treating the random effect as just another covariate in our GLM.

3 Estimation

Lee and Nelder (1996) suggest iteration between GLM parameter estimation for fixed effects, estimation of the dispersion parameters ϕ and α and prediction of random effects u_k. In our setting such iteration leads to the following algorithm for simultaneous rating of ordinary (fixed effect) factors and multi-level (random effect) factors.

(0) Initially, let $\hat{u}_k = 1$ for all k.

(1) Estimate the μ_i in the usual way, with all ordinary rating factors as explanatory variables in a GLM, using a log-link and having $\log(\hat{u})$ as *offset*-variable.

(2) Estimate $\phi\alpha$ using $\hat{\mu}_i$ from Step 1 (this is discussed in Section 3.1 below).

(3) Compute $\hat{u}_k$ for $k = 1, 2, \ldots, K$, using the estimates from Step 1 and 2.

(4) Return to Step 1 with the offset-variable $\log(\hat{u})$ from Step 3.

Repeat Step 1-4 until convergence.

Lee and Nelder (1996) suggest a general likelihood-based approach, utilizing the concept of h-likelihood. For the models we consider (Tweedie with $1 \leq p \leq 2$), which are the most common in insurance practice, one has the alternative of using the minimum MSE predictor $\hat{u}_k$ in Step 3. For $p = 1$, our predictor (2.15) is a w-weighted version of the estimator on p. 623 in Lee & Nelder (1996). In the case $p = 2$, the predictor in Lee & Nelder's (2.12) is slightly different from (2.15) — a term $+1$ appears in the denominator. In most insurance applications the difference between the predictors can be expected to be small. The case $1 < p < 2$ is not explicitly treated in the paper by Lee & Nelder.

We will now show that in Step 2 it is possible to use unbiased estimators based on sums of squares — analogous to the estimation procedures used in classical crediblity.

3.1 Unbiased estimation of variance parameters

The derivations in this section are restricted to Tweedie models with $1 \leq p < 2$ and $p = 2$ with $\alpha > 1$, where the assumptions of Lemma 2.1(b) are satisfied. The idea is to compute separate unbiased estimators of $\sigma^2 = \phi E[U^p]$ and $\sigma_U^2 = \text{Var}(U)$, whose ratio is $\phi\alpha$, cf. (2.13). Note that these quantities are closely related to the variance parameters s^2 and a in the classical Bühlmann-Straub model, cf. (2.21).

For estimation of $\phi E[U^p]$ note that conditionally on U_k, the variables $X_{ik} \doteq Y_{ik}/\mu_i$ are independent with common expectation U_k. From standard GLM theory we have (2.2), which in the present conditional form is

$$\text{Var}(Y_{ik}|U_k = u_k) = \frac{\phi}{w_{ik}}(\mu_i u_k)^p \qquad \text{Var}(X_{ik}|U_k = u_k) = \frac{\phi}{\tilde{w}_{ik}} u_k^p \quad (3.1)$$

with new weights $\tilde{w}_{ik} = w_{ik}\mu_i^{2-p}$. Note that $\overline{u}_k$ in (2.18) is the $\tilde{w}_{ik}$-weighted average of the X_{ik}'s. Now standard results on weighted variances — collected in Lemma A.1 of the Appendix for the sake of completeness — supplies us with the following conditionally unbiased estimator of $\sigma_k^2 = \phi u_k^p$,

$$\hat{\sigma}_k^2 = \frac{1}{I_k - 1} \sum_i w_{ik}\mu_i^{2-p} \left(\frac{Y_{ik}}{\mu_i} - \overline{u}_k \right)^2$$

where I_k are the number of tariff cells i where we have $w_{ik} > 0$. We conclude that unconditionally we have $E(\hat{\sigma}_k^2) = E[E(\hat{\sigma}_k^2|U_k)] = \phi E[U_k^p]$.

For each k we get a separate estimator and it is natural to weigh them together with weights $I_k - 1$ to the overall estimator

$$\hat{\sigma}^2 = \frac{\sum_k (I_k - 1)\hat{\sigma}_k^2}{\sum_k (I_k - 1)} \quad (3.2)$$

This is our estimator of $\phi E[U^p]$. Our starting point for estimation of σ_U^2 is $\overline{u}_k$ in (2.18). By (2.8) we have

$$E(\overline{U}_k | U_k = u_k) = u_k \qquad E(\overline{U}_k) = 1 \tag{3.3}$$

Now

$$\begin{aligned} E[(\overline{U}_k - 1)^2] = \mathrm{Var}[\overline{U}_k] &= \mathrm{Var}[E(\overline{U}_k|U_k)] + E[\mathrm{Var}(\overline{U}_k|U_k)] \\ &= \mathrm{Var}(U_k) + E[\mathrm{Var}(\overline{U}_k|U_k)] \end{aligned}$$

By the definition of $\overline{u}_k$ in (2.18) plus the fact that the Y_{ik}'s are conditionally independent according to Assumption 1(c), and by (3.1) we get

$$\mathrm{Var}(\tilde{w}_{\cdot k}\overline{U}_k|U_k) = \sum_i (\tilde{w}_{ik})^2 \mathrm{Var}\left(\frac{Y_{ik}}{\mu_i} \mid U_k\right) = \sum_i \tilde{w}_{ik}\phi U_k^p = \tilde{w}_{\cdot k}\phi U_k^p$$

Hence

$$E[\tilde{w}_{\cdot k}(\overline{U}_k - 1)^2] = \tilde{w}_{\cdot k}\mathrm{Var}(U_k) + \phi E[U_k^p]$$

By Assumption 1(a) we can drop the k in $\mathrm{Var}(U_k)$ and $E[U_k^p]$. Summing over the index k we get

$$E[\sum_k \tilde{w}_{\cdot k}(\overline{U}_k - 1)^2] = \tilde{w}_{\cdot\cdot}\mathrm{Var}(U) + K\phi E[U^p]$$

where K is the number of classes k. We conclude that the following is an unbiased estimator of $\sigma_U^2 = \mathrm{Var}(U_k)$.

$$\hat{\sigma}_U^2 = \frac{\sum_k \tilde{w}_{\cdot k}(\overline{u}_k - 1)^2 - K\hat{\sigma}^2}{\tilde{w}_{\cdot\cdot}} = \frac{\sum_k \sum_i w_{ik}\mu_i^{2-p}(\overline{u}_k - 1)^2 - K\hat{\sigma}^2}{\sum_k \sum_i w_{ik}\mu_i^{2-p}} \tag{3.4}$$

Note. Even though the GLM dispersion parameter ϕ is included in some of our equations, it does not enter in the calculations. Hence, $\hat{u}_k$ does not depend on the choice of estimator of ϕ. □

3.1.1 Comparison with classical credibility theory

We return to the special case without any ordinary rating factors, see Section sec:classical, and compare the above estimators with the ones used in classical credibility theory, as presented by Goovaerts & Hoogstad, 1987, p. 48. Our estimator of s^2 is

$$\hat{s}^2 = \mu^p \hat{\sigma}^2 = \frac{1}{\sum_k (I_k - 1)} \sum_k \sum_i w_{ik} \left(Y_{ik} - \overline{Y}_{\cdot k} \right)^2$$

In case all I_k are equal, this is exactly the classical estimator of s^2.

Turning to a our estimator is

$$\hat{a} = \mu^2 \hat{\sigma}_U^2 = \frac{\sum_k \sum_i w_{ik} (\overline{Y}_{\cdot k} - \mu)^2 - K \hat{s}^2}{w_{\cdot\cdot}}$$

while the classical estimator is

$$\frac{\sum_k \sum_i w_{ik} (\overline{Y}_{\cdot k} - \overline{Y}_{\cdot\cdot})^2 - (K-1)\hat{s}^2}{w_{\cdot\cdot} - \sum_k w_{\cdot k}^2 / w_{\cdot\cdot}}$$

with $\overline{Y}_{\cdot\cdot}$ denoting the grand w_{ik}-weighted average and $w_{\cdot\cdot}$ the corresponding sum of weights. These estimates are very similar — notice that the natural, GLM estimate of μ will be $\overline{Y}_{\cdot\cdot}$. Seemingly, the other differences are due to that the classical estimator of a does take into account the randomness of $\overline{Y}_{\cdot\cdot}$, while our analysis is conditional on the GLM estimation of ordinary rating factors — in this case of μ.

Note. In the special case above, the index i should point to a single observation rather than to a tariff cell. In general, our estimators of variance parameters are presented at the aggregated level of tariff cells, while they could alternatively be applied at the unaggregate level of single observations. The latter form has been avoided for simplicity in presentation. □

4 Application

As an illustration of the method discussed in this paper, we present some results from a study of car hull insurance, using data from the Swedish insurance group Länsförsäkringar. The multi-level factor in this case was car model, having roughly 2 500 levels. The analysis was made separately for claim frequency using a Poisson distribution ($p = 1$) and average claim amount using a gamma distribution ($p = 2$). In step 1 of the algorithm in Section 3 the price relativities of the ordinary rating factors where estimated by standard GLM software; in step 2 the estimators in (3.2) and (3.4) were used; the u-predictor in step 3 was (2.15). We will show some results from the claim frequency part of the study.

To achieve a fair rating of car models the idea is to describe the models using auxiliary rating factors like weight and weight/power ratio as far as possible; then the residual variation is taken care of by the u-predictors. (These factors are called auxiliary since, unlike other ordinary rating factors, like sex and age of the policyholder, they are introduced mainly as as an aid in the risk classification of car models.) The introduction of auxiliaries decreases the (residual) variation of the car models, with the effect that the $\overline{u}_k$-values become more concentrated around 1. This is illustrated by the bar chart in Figure 1.

The effect of using auxiliary factors is shown in more detail in Table 4.1, for a sample of car models sorted by the exposure weights $w_{\cdot k}$ (number of policy years). As expected, with large $w_{\cdot k}$, the experience factors $\overline{u}_k$ produce reliable estimates, and the rating of car models is hardly affected by the introduction of auxiliaries, as seen by comparing the "No auxiliaries" $\hat{u}_k$ to the "With auxiliaries" column $\hat{\mu}_k \hat{u}_k$, where $\hat{\mu}_k$ is the product of the rating factors for the auxiliaries.

At the other end of the table, with data from very few policy years, credibility is low and the experience factors $\overline{u}_k$ are shaky. Here one has

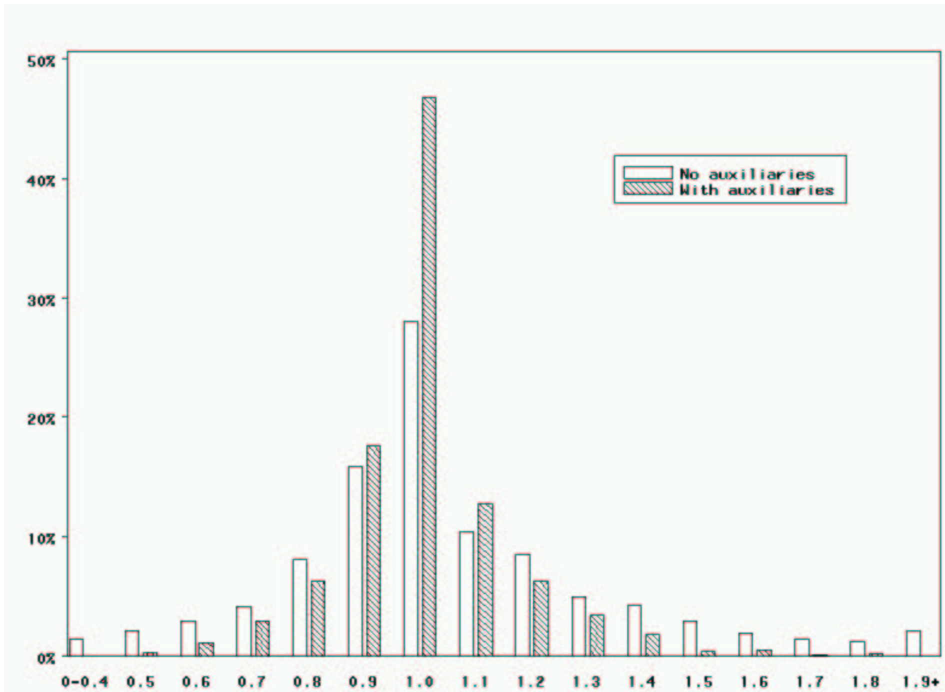

Figure 1: *Histogram of credibility predictors* $\hat{u}_k$ *with and without auxiliary rating factors.*

to rely, to a large extent, on the auxiliary car model rating factors.

Our conclusion is that GLMs that combine fixed and random effects are useful tools for the simultaneous analysis of ordinary and multi-level factors, with efficient handling of the problem of having varying amounts of data for the different levels of the latter. The credibility approach taken in this paper has, in our experience, been helpful in the understanding of the estimators and in the communication of the results to both actuaries and non-actuaries.

k	$w_{\cdot k}$	*No auxiliaries*			*With auxiliaries*			
		$\overline{u}_k$	$\hat{u}_k$	z_k	$\overline{u}_k$	$\hat{u}_k$	z_k	$\hat{\mu}_k\hat{u}_k$
1	41275	0.74	0.74	1.00	0.98	0.98	0.99	0.75
2	39626	0.58	0.58	1.00	0.89	0.89	0.99	0.59
3	39188	0.59	0.59	1.00	0.86	0.86	0.99	0.60
4	31240	0.82	0.82	1.00	0.93	0.93	0.99	0.82
5	28159	0.49	0.50	1.00	0.74	0.75	0.98	0.50
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
401	803	2.08	1.95	0.88	1.43	1.35	0.82	1.99
402	802	0.97	0.97	0.86	1.11	1.08	0.70	0.95
403	801	1.77	1.66	0.86	1.54	1.40	0.74	1.62
404	799	0.74	0.78	0.86	0.83	0.88	0.69	0.79
405	798	1.32	1.27	0.86	0.73	0.78	0.82	1.41
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
901	181	1.38	1.22	0.58	1.14	1.06	0.42	1.29
902	180	1.61	1.38	0.63	0.91	0.95	0.56	1.70
903	180	2.28	1.76	0.59	1.35	1.18	0.51	2.01
904	179	0.79	0.88	0.56	0.86	0.95	0.34	0.88
905	179	2.38	1.80	0.58	1.52	1.25	0.48	1.98
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
1801	7	2.39	1.07	0.05	2.05	1.03	0.03	1.22
1802	7	4.63	1.19	0.05	3.86	1.08	0.03	1.31
1803	7	0.00	0.96	0.04	0.00	0.99	0.01	0.55
1804	7	0.00	0.95	0.05	0.00	0.98	0.02	0.87
1805	7	0.00	0.94	0.06	0.00	0.98	0.02	0.58
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

Table 4.1: Selected car models k with number of policy years $w_{\cdot k}$, experience factors $\overline{u}_k$, credibility predictors $\hat{u}_k$ and credibility factors z_k; without and with auxiliary rating factors, the product of the latter being $\hat{\mu}_k$.

References

Goovaerts, M.J. and Hoogstad, W.J. (1987): *Credibility Theory.* Survey of Actuarial Studies No. 4, Nationale-Nederlanden N.V.

Jewell, W.S: (1974): *Credible means are exact Bayesian for exponential families.* ASTIN Bulletin, 8, 77-90.

Jørgensen, B. (1997): The Theory of Dispersion Models. Chapman and Hall, London.

Jørgensen, B. and Paes de Souza, M. C. (1994): *Fitting Tweedie's compound Poisson model to insurance claim data.* Scand. Actuarial J. 69-93.

Kaas, R., Dannenburg, D. and Goovaerts, M.J. (1997): *Exact credibility for weighted observations.* ASTIN Bulletin, 27, 287-295.

Kaas, R., Goovaerts, M.J., Dhaene, J. and Denuit, M. (2001): *Modern Actuarial Risk Theory.* Kluwer Academic Publishers, Dordrecht.

Lee, Y. and Nelder, J. A. (1996): *Hierarchical Generalized linear models.* J. R. Statist. Soc. B Vol. 58:4, 619-678.

Nelder, J. A. and Verrall, R. J. (1997): *Credibility Theory and Generalized linear models.* ASTIN Bulletin, Vol. 27:1, 71-82.

Renshaw, A. E. (1994): *Modelling the claims process in the precense of covariates.* ASTIN Bulletin, Vol. 24:2, 265-285.

Esbjörn Ohlsson
Mathematical Statistics
Department of Mathematics
Stockholm University
SE-106 91 Stockholm, Sweden
E-mail: esbjorn.ohlsson@math.su.se

A Appendix

Lemma A.1 *Let $X_1, X_2, \ldots, X_n$ be a sequence of uncorrelated random variables with common mean μ and variance inversely proportional to weights w_i, i.e.* $\mathrm{Var}(X_i) = \sigma^2/w_i$; $i = 1, 2, \ldots, n$. *With $w. = \sum_i w_i$ we let*

$$\overline{X} = \frac{1}{w.}\sum_i w_i X_i \qquad \textit{and} \qquad s^2 = \frac{1}{n-1}\sum_i w_i (X_i - \overline{X})^2$$

Then s^2 is an unbiased estimator of σ^2.

Proof.

$$\begin{aligned}\mathrm{Var}(X_i - \overline{X}) &= \mathrm{Var}\left(X_i(1 - \frac{w_i}{w.}) - \sum_{k \neq i} \frac{w_k}{w.} X_k\right) \\ &= (1 - \frac{w_i}{w.})^2 \, \frac{\sigma^2}{w_i} + \sum_{k \neq i} \frac{w_k}{w.^2} \, \sigma^2 = \frac{\sigma^2}{w_i}(1 - \frac{w_i}{w.})\end{aligned}$$

and we get

$$E\left[\sum_{i=1}^{n} w_i (X_i - \overline{X})^2\right] = \sum_{i=1}^{n} w_i \mathrm{Var}(X_i - \overline{X}) = \sigma^2 (n-1)$$

□