Mathematical Statistics
Stockholm University

Credibility estimators in multiplicative models

Esbjörn Ohlsson

Research Report 2006:3

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 **2006:3**,
http://www.math.su.se/matstat

Credibility estimators in multiplicative models

Esbjörn Ohlsson*

March 24, 2006

Abstract

Rating of non-life insurance contracts commonly employs multiplicative models that are estimated by Generalized Linear Models (GLMs); another useful tool for rate making is credibility models. The main goal of this paper is to demonstrate how these can be combined to solve important practical problems, in particular the car classification problem in motor insurance. This is achieved by reformulating the credibility models as multiplicative random effects models, where the GLM estimates are used as "a priori differences". Our models have various elements in common with, in cronological order, Sundt (1987), Dannenburg, Kaas & Goovaerts (1996), Nelder & Verall (1997) and Bühlmann & Gisler (2005, Chapter 4.13). The methods are illustrated on data from a Swedish insurance company.

Keywords

Credibility theory, Hierarchical credibility, Generalized linear models, GLMs, Car model classification, Motor insurance, Multi-level factor.

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

1 Introduction

Credibility estimators are useful in the process of rating non-life insurance contracts, but in our experience, credibility is not used as much in practice as it deserves. Non-life actuaries commonly employ multiplicative models, where a number of rating factors are estimated by GLMs (Generalized Linear Models). Most texts on credibility theory tend to ignore this fact, giving little or no guidance on how to use credibility estimators in this multiplicative GLM enviroment. To bridge the gap between GLMs and credibility, we suggest here that credibility models are stated as multiplicative random effects models. Credibility and GLMs can then easily be combined to mixed models (GLMMs) with the familiar multiplicative structure for the mean (log-link).

Another fact that “hampers the acceptance of credibility techniques by practicioners” was noted by Dannenburg, Kaas & Goovaerts (1996): that “credibility is currently taught in a needlessly complicated way” and that credibility is “set in a Bayesian framework”. They suggest replacing the Bayesian setting, with an abstract risk parameter Θ, by variance component models (*additive* random effects models). We find this a very clever and useful reformulation of credibility theory, but we prefer using *multiplicative* models, since these are common practice in GLM rating.

The idea of combining credibility and GLMs is not new: it was introduced by Nelder & Verall (1997), using the HGLMs (Hierarchical Generalized Linear Models) of Lee & Nelder (1996); however, their approach uses the concept of hierarchical likelihood which, presumably, is not familiar to most actuaries. We prefer here to stay with the distribution-free and simple BLP (best linear predictor) approach that is standard in credibility theory.

One might say that the present paper combines an idea by Dannenburg et al.—deriving traditional credibility estimators by using random effects models—with an idea by Nelder & Verall: combining GLMs and credibility.

In the basic Bühlmann-Straub case, even though our *model* is stated different, our *results* have a connection to those of Bühlmann & Gisler (2005, Proposition 2.11 and Section 4.13) on using *a priori* information in credibility—as will be explained in Section 3. In the hierarchical case, there are similarities between our results and Sundt (1987). What is new here, in relation to these two texts, is the connection to GLM and the implication this has on the estimators.

In summary our goal is to

- present new theory on the use of a priori information from GLMs in credibility, in particular in hierarchical credibility;
- demonstrate to the practicing actuary how credibility can be be a useful tool within the standard multiplicative GLM framework;
- show how credibility can be taught to actuarial students in a simple way, within the framework of multiplicative models and GLMs.

Section 2 presents our reformulation of the standard Bühlmann-Straub model. Section 3 discusses the use of GLM estimates as a priori information in credibility. The results are extended to the hierarchical case in Section 4. Finally, Section 5 presents an application to so called *car model classification* in private motor insurance, on data from Länsförsäkringar Alliance, Sweden.

2 The Bühlmann-Straub model

In non-life insurance rating, one works with some *key ratio* Y; this may be the *risk premium*—the ratio $Y = X/w$ of observed claim cost X to exposure w measured in policy years. With GLMs it is customary to carry out separate analyses with the key ratios *claim frequency* and average *claim severity*, respectively. Our exposition below covers all these cases, but for simplicity, we will mostly mention the risk premium only.

In the Bühlmann-Straub model, we only have one rating factor, dividing the collective into J groups. We observe Y_{jt}, where j is the group and repeated observations are indexed by

t. Let μ be the overall mean for the collective and let u_j be the price relativity for group j; $j = 1, 2, \ldots J$; measuring the departure of the group mean from μ. With sufficient data, the u_j's can be estimated by standard GLM techniques. Credibility theory, on the other hand, consideres the case when we do not have enough observations in all groups for the u_j's to be estimated with accuracy, but still want to make the best use of the information we have. Credibility estimators can be derived under the assumption that price relativities are random effects U_j. The basic multiplicative model is then

$$E(Y_{jt}|U_j) = \mu U_j \,,$$

where $E(U_j) = 1$, in order to avoid redundancy in the parameters. As we shall see, this reduces the number of parameters to estimate from $J+1$ in the GLM case, to only 3 (μ plus two variance components).

Even though price relativities like U_j are standard in GLM rating, it will be more convenient in our derivations to work with the variable $V_j \doteq \mu U_j$. By assumption on U_j, $E(V_j) = \mu$, and

$$E(Y_{jt}|V_j) = V_j \,. \tag{2.1}$$

In rating with GLMs, it is usually assumed that the data follow a *Tweedie model.* The most important part of this assumption is that the variance is proportional to the p:th power of the expectation. In our case we must condition on the random effect; the Tweedie variance is then

$$\mathrm{Var}\,(Y_{jt}|V_j) = \frac{\phi\, V_j^p}{w_{jt}} \,, \tag{2.2}$$

where ϕ is the GLM dispersion parameter. For $p = 1$ we get the (overdispersed) Poisson distribution used for the claim frequency; with $p = 2$ we get the gamma distribution case, often used for claim severity; for $1 < p < 2$ we have a compound Poisson distribution, suitable for the risk premium, cf. Jørgensen & Paes de Souza (1994). For more imformation on Tweedie models, see Jørgensen (1997). In general, it is a very natural assumption in premium rating that the standard deviation of Y_{jt} is proportional to some power of the mean (and inversely proportional to the exposure weight w).

Assuming that the V_j are identically distributed, we may introduce $\sigma^2 \doteq \phi E[V_j^p]$, independent of j. From (2.2) we then have

$$E\left[\operatorname{Var}\left(Y_{jt}|V_j\right)\right] = \frac{\sigma^2}{w_{jt}}, \tag{2.3}$$

This is really the only fact about the variance of Y_{jt} that we need here. We now collect all our assumptions for the reformulation of the Bühlmann-Straub model, some of which have already appeared. Without further notice, any random variable X in this text will be assumed to have finite second moment, $E(X^2) < \infty$.

Assumption 2.1 *(a) The groups are independent, i.e.* (Y_{jt}, V_j) *and* $(Y_{j't'}, V_{j'})$ *are independent as soon as* $j \neq j'$.

(b) The V_j; $j = 1, 2, \ldots, J$; *are identically distributed with* $E[V_j] = \mu > 0$.

(c) For any j, *conditional on* V_j, *the* Y_{jt}*'s are mutually independent, with mean given by (2.1) and variance satisfying (2.3).*

By part (b) we may write

$$\tau^2 \doteq \operatorname{Var}(V_j)\,. \tag{2.4}$$

Note that

$$\operatorname{Var}\left(Y_{jt}\right) = \operatorname{Var}[E(Y_{jt}|V_j)] + E[\operatorname{Var}(Y_{jt}|V_j)] = \tau^2 + \frac{\sigma^2}{w_{jt}}. \tag{2.5}$$

indicating that we actually have a *variance component model.*

With enough data, we would treat the rating factor as a fixed effect and an obvious estimator of V_j would be the weighted mean

$$\overline{Y}_{j\cdot} = \frac{\sum_t w_{jt} Y_{jt}}{\sum_t w_{jt}}\,. \tag{2.6}$$

If, on the other hand, there where no data at all, a reasonable risk premium would be just μ. A credibility estimator is a compromise between these two extremes, viz.

$$\widehat{V}_j = z_j \overline{Y}_{j\cdot} + (1 - z_j) \cdot \mu, \tag{2.7}$$

where the *credibility factor* z_j satisfies $0 \leq z_j \leq 1$. To be more precise, one defines the *credibility estimator* as that linear function of the observations $h(\boldsymbol{Y})$ which minimizes the mean square error of prediction (MSEP)

$$E\left[(h(\boldsymbol{Y}) - V_j)^2\right]. \tag{2.8}$$

Even though our setting is a bit different, the solution to this minimization problem is, of course, nothing but the famous estimator by Bühlmann & Straub (1970).

Theorem 2.1 (Bühlmann-Straub) *Under Assumption 2.1, the credibility estimator of V_j is given by (2.7) with*

$$z_j = \frac{w_{j\cdot}}{w_{j\cdot} + \sigma^2/\tau^2}. \tag{2.9}$$

For the sake of completeness, a proof is given in the appendix. For a proof in the standard setting see, e.g., Bühlmann & Gisler (2005, Theorem 4.2).

Note. Some authors would call $\widehat{V}_j$ a predictor, since V_j is a random variable and not a parameter; in this context it is the BLP – best linear predictor. The credibility tradition is, however, to call it an estimator, and we adhere to this terminology. □

Finally, a credibility estimator of the random effect U_j is, of course, given by $\widehat{U}_j = z_j\overline{Y}_{j\cdot}/\mu + (1 - z_j)$. It remains to estimate the parameters, of which μ might be given *a priori*, e.g. by a tariff; else it could be estimated by a weighted mean of all observations. Unbiased estimators of the variance parameters σ^2 and τ^2 can be found in Bühlmann & Gisler (2005, Section 4.8).

2.1 Comparison with other notation

In a traditional text, such as Bühlmann & Gisler (2005), there is no random effect, but rather an abstract (random) risk parameter Θ_j. The task is then to estimate $\mu(\Theta_j) = E[Y_{jt}|\Theta_j]$. The correspondence to our notation is $\mu(\Theta_j) = V_j$. Since no inference is made on the parameter Θ_j itself, there is no loss in generality in our approach and the estimator is the same.

Our model is close to the one in Dannenburg et al. (1996), with the random effects ANOVA model

$$Y_{jt} = \mu + \Xi_j + \Xi_{jt}\,,$$

where Ξ_j are independent and identically distributed with $E[\Xi_j] = 0$ and $\mathrm{Var}[\Xi_j] = \tau^2$, and Ξ_{jt} are independent and identically distributed with $E[\Xi_{jt}] = 0$ and $\mathrm{Var}[\Xi_{jt}] = \sigma^2/w_{jt}$. The resulting estimator in their Theorem 2.2.2 is the same. This has the merit of being a familiar ANOVA model, while our approach was designed for the use in another familiar setting: multiplicative models and GLMs, which is the topic of the next section.

3 Credibility estimators in multiplicative models

In practice, we often have a large number of rating factors, especially in the private lines. In, e.g., motor insurance these might include the age and gender of the insured person, the age of the car, the annual mileage, etc. Some of these rating factors are categorical with just a few classes, e.g. gender, while others are continuous or ordinal and can be grouped into new variables, like age class or mileage class.

There are also categorical variables with a large number of levels that can not be grouped *a priori* by risk homogeneity. One example is *car model*, which in Sweden has about 2 500 levels. Even though cars may be grouped by technical variables such as weight or engine power, there is typically a lot of residual differences left within such groups, see Section 5. Since data are usually too sparce for the vast majority of models there is a need for credibility estimators. Furthermore, running a GLM with the 2 500 car models cross-classified by all the other rating factors may be impractical. A rating factor such as this will be called *a multi-level factor* (MLF).

Other examples of MLFs are geographical region (defined by e.g. zip codes), and the customer herself (experience rating and bonus/malus systems).

Our key ratio is now denoted Y_{ijt}, where the i refers to a cell in the tariff given by the *ordinary* *rating factors*, i.e. the non-MLFs. This tariff cell is the cross-tabulation of R different rating factors. If γ_j^i denotes the price relativity for factor number j for insurances in cell i, the multiplicative model is,

$$E(Y_{ijt}|U_j) = \mu\gamma_1^i\gamma_2^i\cdots\gamma_R^i U_j \,,$$

generalizing (2.1). Note that μ is now the so called *base premium*, the rating in the *base cell*, where $\gamma_r^i = 1$; $r = 1, \ldots, R$. By initially disregarding U_j, we can estimate μ and the γ_j^i's by standard GLM methods. Given such estimates, we now look for a credibility estimator of U_j. For simplicity in notation we introduce

$$\gamma_i = \gamma_1^i\gamma_2^i\cdots\gamma_R^i \,,$$

and again $V_j = \mu U_j$, so that the multiplicative model can be written

$$E(Y_{ijt}|V_j) = \gamma_i V_j \,. \tag{3.1}$$

As discussed in connection to (2.2), the standard GLM models for $Y_{jt}|V_j$ are Tweedie models, and so we have

$$\mathrm{Var}(Y_{ijt}|V_j) = \frac{\phi\,(\gamma_i V_j)^p}{w_{ijt}} \,.$$

Again let $\sigma^2 \doteq \phi E[V_j^p]$, so that

$$E[\mathrm{Var}(Y_{ijt}|V_j)] = \frac{\gamma_i^p\sigma^2}{w_{ijt}} \,. \tag{3.2}$$

We have the following generalization of Assumption 2.1.

Assumption 3.1 *(a) The groups are independent, i.e.* (Y_{ijt}, V_j) *and* $(Y_{i'j't'}, V_{j'})$ *are independent as soon as* $j \neq j'$.

(b) The V_j; $j = 1, 2, \ldots, J$; *are identically distributed with* $E[V_j] = \mu > 0$ *and again we write* $\tau^2 \doteq \mathrm{Var}[V_j]$.

(c) For any j*, conditional on* V_j*, the* Y_{ijt}*'s are mutually independent, with mean given by (3.1) and variance satisfying (3.2).*

Note. The γ_i act as *a priori* information in the sense of Bühlmann & Gisler (2005, Section 4.13); however, in their setting, all insurances in group j have the same a priori information. This may well be the case in some applications, e.g. in experience rating of large businesses based on their claims history, where j is a single business. In many cases, though, insurances in group j may occur among several a priori tariff cells i. In motor insurance, e.g., drivers of different age and sex drive the same car model j, and so the observations for that car are spread over many cells in the a priori tariff. In cases like this, there is a need for our slightly more general setting.

Furthermore, Bühlmann & Gisler use the weight μ_i, where we allow μ_i^p for some $p \geq 0$. In fact, they motivate their choice of weight partly by looking at the Poisson case—corresponding to $p = 1$ here. Since $p = 2$ is standard for analysing claim severity, and $1 < p < 2$ can be used for risk premium, our generalization has a bearing in practice. □

Next we transform the observations so that we can bring back this situation to the classical Bühlmann-Straub model in Section 2.

$$\widetilde{Y}_{ijt} = \frac{Y_{ijt}}{\gamma_i} \qquad \widetilde{w}_{ijt} = w_{ijt}\gamma_i^{2-p} \,. \tag{3.3}$$

Note that

$$E(\widetilde{Y}_{ijt}|V_j) = V_j \,, \tag{3.4}$$

and

$$E[\mathrm{Var}(\widetilde{Y}_{ijt}|V_j)] = \frac{\sigma^2}{\widetilde{w}_{ijt}} \,. \tag{3.5}$$

By this and Assumption 3.1, the entire Assumption 2.1 is fulfilled for $\widetilde{Y}_{ijt}$ with the weights $\widetilde{w}_{ijt}$ and we get the following result from (2.6), (2.7) and (2.9).

Corollary 3.1 *Under Assumption 3.1, the credibility estimator of V_j is*

$$\widehat{V}_j = z_j \overline{\overline{\widetilde{Y}}}_{\cdot j \cdot} + (1 - z_j)\mu \,, \tag{3.6}$$

where

$$z_j = \frac{\widetilde{w}_{\cdot j \cdot}}{\widetilde{w}_{\cdot j \cdot} + \sigma^2/\tau^2} \,, \tag{3.7}$$

and

$$\overline{\overline{Y}}_{\cdot j \cdot} = \frac{\sum_{i,t} \widetilde{w}_{ijt} \widetilde{Y}_{ijt}}{\widetilde{w}_{\cdot j \cdot}} = \frac{\sum_{i,t} \widetilde{w}_{ijt} Y_{ijt}/\gamma_i}{\sum_{i,t} \widetilde{w}_{ijt}} . \tag{3.8}$$

with $\widetilde{Y}_{ijt}$ *and* $\widetilde{w}_{ijt}$ *defined in (3.3).*

Consequently,

$$\widehat{U}_j = z_j \frac{\overline{\overline{Y}}_{\cdot j \cdot}}{\mu} + (1 - z_j), \tag{3.9}$$

and the rating for insurances with MLF level j in a priori tariff cell i is $\mu \gamma_i \widehat{U}_j$.

Note. When Y_{ijt} is *claim frequency*, with w_{ijt} as the number of insurance years, we use the Poisson distribution ($p = 1$) in a GLM, so that $\widetilde{w}_{ijt} = w_{ijt}\gamma_i$, a quantity that is called "normalized insurance years" in Campbell (1986). Then

$$\frac{\overline{\overline{Y}}_{\cdot j \cdot}}{\mu} = \frac{\sum_{i,t} w_{ijt} Y_{ijt}}{\sum_{i,t} w_{ijt} \mu \gamma_i} ,$$

i.e. the number of claims in group j divided by the expected number of claims in the same group: a very natural estimator of U_j.

For the case when Y_{ijt} is *claim severity* and w_{ijt} is the number of claims, the standard approach is to use a GLM with $p = 2$, corresponding to a gamma distribution. Here

$$\frac{\overline{\overline{Y}}_{\cdot j \cdot}}{\mu} = \frac{\sum_{i,t} w_{ijt} Y_{ijt}/(\mu \gamma_i)}{\sum_{i,t} w_{ijt}} ,$$

i.e. a weighted average of the observed relative deviance of Y_{ijt} from its expectation $\mu\gamma_i$, again a simple and natural estimator of U_j. □

Throughout this derivation, μ and γ_i were supposed to be known a priori. Recall that γ_i is the product of the price relativities $\gamma_1^i, \ldots, \gamma_R^i$ in cell i; these can be estimated by GLM, treating $\widehat{U}_j$ as a known *offset* variable. So we must know $\widehat{U}_j$ to estimate μ and γ_i and vice versa. The solution to this dilemma is to use an iterative procedure as described in the next section.

Note. Ohlsson & Johansson (2006) derived the estimator in Corollary 3.1 as a so called *exact credibility* estimator, viz. by assuming that the observations are drawn from a Tweedie

GLM and the random effect follows the natural conjugate prior distribution. The present section can be seen as a non-parametric counterpart of that result. □

3.1 Iteration between GLM and credibility

The above considerations suggest the following algorithm for simultaneous rating of ordinary factors $\gamma_1, \gamma_2, \ldots, \gamma_R$ by GLMs and a multi-level factor U_j by credibility, in the multiplicative model $\mu\gamma_1^i\gamma_2^i \cdots \gamma_R^i U_j$.

(0) Initially, let $\widehat{U}_j = 1$ for all j.

(1) Estimate the parameters for the ordinary rating factors by a Tweedie GLM (e.g. Poisson or Gamma) with log-link, using $\log(\widehat{U}_j)$ as an *offset*-variable. This yields $\hat{\mu}$ and $\hat{\gamma}_1^i, \ldots, \hat{\gamma}_R^i$.

(2) Compute $\hat{\sigma}^2$ and $\hat{\tau^2}$ as described in Section 3.2 below, using $\hat{\mu}$ and $\hat{\gamma}_1^i, \ldots, \hat{\gamma}_R^i$ from Step 1.

(3) Use (3.9) to compute $\widehat{U}_j$, using the estimates from Step 1 and 2.

(4) Return to Step 1 with the new $\widehat{U}_j$ from Step 3.

Repeat Step 1–4 until convergence. In our experience this is a matter of something like 3–5 iterations, if there are no extreme outliers in the data.

Note. Suppose that in reality we had enough data to use U_j as an ordinary rating factor in our GLM; then we would have a large exposure weight $w_{\cdot j \cdot}$, resulting in high credibility, with z_j close to 1. Then $\widehat{U}_j$ would be approximately equal to $\overline{\overline{Y}}_{\cdot j \cdot}/\mu$; it is not hard to show that the resulting equation $\widehat{U}_j = \overline{\overline{Y}}_{\cdot j \cdot}/\mu$, defines the maximum likelihood estimating equation for U_j in the corresponding GLM; hence the estimate would be the same that would be obtained with standard GLMs. This is a nice property, with the practical implication that we do not have to worry about whether a few U_j's should better be altered from random to fixed effects—the estimates will stay the same anyway. □

3.2 Estimation of variance parameters

It remains to estimate the variance parameters σ^2 and τ^2. Unbiased estimators can be obtained from the corresponding estimators in Bühlmann & Gisler (2005, Section 4.8); note, though, that they assume that n_k—the number of observations for group j—is equal for all groups. The straight-forward extension of σ^2 to the case with different n_k can be found in Remark 2.3.4 of Dannenburg et al. (1996), while the formula for τ^2 remains unchanged. Let

$$\hat{\sigma}_j^2 = \frac{1}{n_j - 1} \sum_{i,t} \widetilde{w}_{ijt}(\widetilde{Y}_{ijt} - \overline{\overline{\widetilde{Y}}}_{\cdot j \cdot})^2 \,. \tag{3.10}$$

Then

$$\hat{\sigma}^2 = \frac{\sum_j (n_j - 1)\hat{\sigma}_j^2}{\sum_j (n_j - 1)} \,, \tag{3.11}$$

and

$$\hat{\tau}^2 = \frac{\sum_j \widetilde{w}_{\cdot j \cdot}(\overline{\overline{\widetilde{Y}}}_{\cdot j \cdot} - \overline{\overline{\widetilde{Y}}}_{\cdots})^2 - (J-1)\hat{\sigma}^2}{\widetilde{w}_{\cdot\cdot} - \sum_i \widetilde{w}_{\cdot j \cdot}^2 / \widetilde{w}_{\cdots}} \,. \tag{3.12}$$

where $\overline{\overline{\widetilde{Y}}}_{\cdots}$ is the $\widetilde{w}_{\cdot j \cdot}$-weighted average of the $\overline{\overline{\widetilde{Y}}}_{\cdot j \cdot}$'s. Note that in our case these estimators are strictly unbiased only if γ_i is known, while in practice we plug in a GLM estimate here.

4 Hierarchical credibility in multiplicative models

In the example with 2 500 *car models*, data indicate that cars of the same brand (Renault, Volvo, etc.) have some risk characteristics in common, even if they represent different models (Renault Megane, Renault Laguna, etc.); see Figure 2 in Section 5.1 below. We like to include *brand* as a rating factor in our models; this is of particular importance when a new model of a well-known brand is introduced on the market, with no insurance data available. For some brands (say Volvo) we have lots of data, while for others (say Subaru) we have very few insurances in our data files; hence *brand*, like *car model*, is an MLF, for which we want to use credibility estimation.

Since the car models are hierarchically ordered under the brands, this is a case for using *hierarchical credibility*, as suggested already by Sundt (1987). As in the previous section, we like to take into account a priori information from the standard rating factors (age, gender, mileage, etc.) estimated by GLM and generalize our estimators from Section 3 to the hierarchical case.

Note. Another hierarchical MLF in motor insurance is the geographical zone, where we may have three or more levels, such as zip codes within counties, within states. For the sake of simplicity, we only consider two-level hierarchical models here, though. □

Let Y_{ijkt} be the observed key ratio, with denominator (weight) w_{ijkt}, in a priori tariff cell i, for *sector* j (e.g. car brand, county) and *group* k (e.g. car model, zip code area) within sector j. Note that while k is hierarchical under j, a particular i can be combined with any j, k and is hence *not* part of the hierarchical ordering. The full multiplicative model is now, with U_j as the random effect for sector and U_{jk} as the random effect for group within sector

$$E(Y_{ijkt}|U_j, U_{jk}) = \mu\gamma_1^i\gamma_2^i \cdots \gamma_R^i U_j U_{jk} \,.$$

Here we assume $E[U_j] = 1$ and $E(U_{jk}|U_j) = 1$. As before, we use an abbreviated notation for the standrad rating factors, $\gamma_i = \gamma_1^i\gamma_2^i \cdots \gamma_R^i$. We need credibility estimators of both U_j and U_{jk}, but again find it easier to work with $V_j = \mu U_j$, and also with $V_{jk} = \mu U_j U_{jk} = V_j U_{jk}$, so that

$$E(Y_{ijkt}|V_j, V_{jk}) = \gamma_i V_{jk} \,. \tag{4.1}$$

If, given the random effects, Y follows a Tweedie GLM model, then

$$\mathrm{Var}(Y_{ijkt}|V_j, V_{jk}) = \frac{\phi(\gamma_i V_{jk})^p}{w_{ijkt}},$$

where ϕ is the dispersion parameter. From this we get

$$E[\mathrm{Var}(Y_{ijkt}|V_j, V_{jk})] = \frac{\gamma_i^p \sigma^2}{w_{ijkt}}, \tag{4.2}$$

where $\sigma^2 = \phi E[V_{jk}^p]$. We now generalize Assumption 3.1 to the hierarchical case.

Assumption 4.1 *(a) The sectors are independent, i.e.* (Y_{ijkt}, V_j, V_{jk}) *and* $(Y_{i'j'k't'}, V_{k'}, V_{j'k'})$ *are independent as soon as* $j \neq j'$.

(b) For every j*, conditional on the sector effect* V_j*, the groups are independent, i.e.* (Y_{ijkt}, V_{jk}) *and* $(Y_{i'jk't'}, V_{jk'})$ *are conditionally independent as soon as* $k \neq k'$.

(c) All the pairs (V_j, V_{jk}); $j = 1, 2, \ldots, J; k = 1, 2, \ldots, K_j$; *are identically distributed, with*

$$E[V_j] = \mu > 0 \quad \textit{and} \quad E(V_{jk}|V_j) = V_j \, .$$

We use the notation

$$\tau^2 \doteq \mathrm{Var}[V_j] \quad \textit{and} \quad \nu^2 \doteq E[\mathrm{Var}(V_{jk}|V_j)] \, .$$

(d) For any (j, k)*, conditional on* (V_j, V_{jk})*, the* Y_{ijkt} *are independent, with mean given by (4.1) and with variance satisfying (4.2).*

We first look for a credibility estimator of V_j, and in the same vein as (3.3) and (3.8) we introduce

$$\widetilde{Y}_{ijkt} = \frac{Y_{ijkt}}{\gamma_i} \, , \qquad \widetilde{w}_{ijkt} = w_{ijkt}\mu_i^{2-p} \, , \quad \text{and} \quad \overline{\widetilde{Y}}_{\cdot jk\cdot} = \frac{\sum_{i,t} \widetilde{w}_{ijkt}\widetilde{Y}_{ijkt}}{\sum_{i,t} \widetilde{w}_{ijkt}} \, . \tag{4.3}$$

By (4.1) and Assumption 4.1 (c)

$$E(\overline{\widetilde{Y}}_{\cdot jk\cdot}|V_j, V_{jk}) = V_{jk} \quad \text{and} \quad E(\overline{\widetilde{Y}}_{\cdot jk\cdot}|V_j) = V_j \, . \tag{4.4}$$

As for the variance we find, by the rule of calculating variance by conditioning and (4.2)

$$\begin{aligned} \mathrm{Var}(\overline{\widetilde{Y}}_{\cdot jk\cdot}|V_j) &= E[\mathrm{Var}(\overline{\widetilde{Y}}_{\cdot jk\cdot}|V_j, V_{jk})|V_j] + \mathrm{Var}[E(\overline{\widetilde{Y}}_{\cdot jk\cdot}|V_j, V_{jk})|V_j] \\ \Rightarrow \quad E[\mathrm{Var}(\overline{\widetilde{Y}}_{\cdot jk\cdot}|V_j)] &= E[\mathrm{Var}(\overline{\widetilde{Y}}_{\cdot jk\cdot}|V_j, V_{jk})] + E[\mathrm{Var}(V_{jk}|V_j)] \\ &= \frac{\sum_{i,t} \widetilde{w}_{ijkt}^2 E[\mathrm{Var}(Y_{ijkt}|V_j, V_{jk})]/\gamma_i^2}{\widetilde{w}_{\cdot jk\cdot}^2} + \nu^2 = \frac{\sigma^2}{\widetilde{w}_{\cdot jk\cdot}} + \nu^2 \end{aligned}$$

If we define

$$z_{jk} = \frac{\widetilde{w}_{\cdot jk\cdot}}{\widetilde{w}_{\cdot jk\cdot} + \sigma^2/\nu^2} \, , \tag{4.5}$$

this can be rewritten as

$$E[\mathrm{Var}(\overline{\widetilde{Y}}_{\cdot jk\cdot}|V_j)] = \frac{\nu^2}{z_{jk}} \, .$$

Note also that by Assumption 4.1(b), the $\overline{\overline{Y}}_{\cdot jk\cdot}$ are independent for different k, given V_j. Now all parts of Assumption 2.1 are fulfilled with Y_{jt}, w_{jt} and σ^2 replaced by, respectively, $\overline{\overline{Y}}_{\cdot jk\cdot}$, z_{jk} and ν^2, and we get the following result from Theorem 2.1.

Theorem 4.1 *Under Assumption 4.1, the credibility estimator of V_j is*

$$\widehat{V}_j = q_j \overline{\overline{Y}}^z_{\cdot j\cdot\cdot} + (1 - q_j)\mu \,, \tag{4.6}$$

where

$$\overline{\overline{Y}}^z_{\cdot j\cdot\cdot} = \frac{\sum_k z_{jk} \overline{\overline{Y}}_{\cdot jk\cdot}}{\sum_k z_{jk}} \qquad \text{and} \qquad q_j = \frac{z_{j\cdot}}{z_{j\cdot} + \nu^2/\tau^2} \,. \tag{4.7}$$

Note that the mean $\overline{\overline{Y}}^z_{\cdot j\cdot\cdot}$ is weighted by z instead of w. If the group variance ν^2 is small then $z_{jk} \approx \nu^2 \widetilde{w}_{\cdot jk}/\sigma^2$ and so

$$q_j \approx \frac{\widetilde{w}_{\cdot j\cdot}}{\widetilde{w}_{\cdot j\cdot} + \sigma^2/\tau^2} \,,$$

so that q_j, as it should, is the credibility factor in a model with no groups within the sectors, cf. (3.7). If, on the other hand, $\tau^2 \approx 0$, then $q_j \approx 0$ and so $\widehat{V}_j \approx \mu$, indicating that the sector level may be omitted, as it should when not contributing to the model variance.

We now turn to the estimation of V_{jk}. Conditional on V_j, the situation in a sector is very much as if we had a Bühlmann-Straub model with a priori differences γ_i and mean $\widehat{V}_j$. For the moment, this may serve as a motivation for the following theorem; the proof is postponed to the appendix.

Theorem 4.2 *The credibility estimator of V_{jk} is*

$$\widehat{V}_{jk} = z_{jk} \overline{\overline{Y}}_{\cdot jk\cdot} + (1 - z_{jk})\widehat{V}_j \,, \tag{4.8}$$

where $\overline{\overline{Y}}_{\cdot jk\cdot}$, z_{jk} and $\widehat{V}_j$ are given in (4.3), (4.5) and (4.6), respectively.

By the simple relation between the U's and V's we finally note that the random effects in the multiplicative model $\mu\gamma^i_1\gamma^i_2 \cdots \gamma^i_R U_j U_{jk}$ can be estimated as follows,

$$\widehat{U}_j = q_j \frac{\overline{\overline{Y}}^z_{\cdot j\cdot\cdot}}{\mu} + (1 - q_j) \,,$$

$$\widehat{U}_{jk} = z_{jk}\frac{\widetilde{\overline{Y}}_{\cdot jk\cdot}}{\widehat{V}_j} + (1 - z_{jk})\,.$$

4.1 Estimation of variance parameters

It remains to estimate the parameters σ^2, τ^2 and ν^2. Traditionally, iterative so called pseudo-estimators are suggested in the literature for the hierarchical case, but recently direct, unbiased-type estimators where derived by Gisler & Bühlmann (2005, Section 6.6) and Ohlsson (2005), independently. These can easily be extended to the case with a priori differences with the following result, where T_{jk} is the number of observations i and t for the group (j,k), K_j is the number of groups in sector j and J is the number of sectors.

$$\sigma^2 = \frac{1}{\sum_j \sum_k (T_{jk} - 1)} \sum_j \sum_k \sum_i \sum_t \widetilde{w}_{ijkt}(\widetilde{Y}_{ijkt} - \widetilde{\overline{Y}}_{\cdot jk\cdot})^2\,; \tag{4.9}$$

$$\hat{\nu}^2 = \frac{\sum_j \sum_k \widetilde{w}_{\cdot jk\cdot}(\widetilde{\overline{Y}}_{\cdot jk\cdot} - \widetilde{\overline{Y}}_{\cdot j\cdot\cdot})^2 - \hat{\sigma}^2 \sum_j (K_j - 1)}{\widetilde{w}_{\cdot\cdot\cdot\cdot} - \sum_j (\sum_k \widetilde{w}^2_{\cdot jk\cdot})/\widetilde{w}_{\cdot j\cdot\cdot}}\,; \tag{4.10}$$

$$\hat{\tau}^2 = \frac{\sum_j z_{j\cdot}(\widetilde{\overline{Y}}^z_{\cdot j\cdot\cdot} - \widetilde{\overline{Y}}^z_{\cdot\cdot\cdot\cdot})^2 - \hat{\nu}^2 (J-1)}{z_{\cdot\cdot} - \sum_j z^2_{j\cdot}/z_{\cdot\cdot}}\,. \tag{4.11}$$

where $\widetilde{\overline{Y}}_{\cdot jk\cdot}$ is given by (4.3), $\widetilde{\overline{Y}}^z_{\cdot j\cdot\cdot}$ by (4.7), and

$$\widetilde{\overline{Y}}_{\cdot j\cdot\cdot} = \frac{\sum_k \widetilde{w}_{jk}\widetilde{\overline{Y}}_{\cdot jk\cdot}}{\sum_k \widetilde{w}_{jk}} \quad \text{and} \quad \widetilde{\overline{Y}}^z_{\cdot\cdot\cdot\cdot} = \frac{\sum_j z_{j\cdot}\widetilde{\overline{Y}}^z_{\cdot j\cdot\cdot}}{\sum_j z_{j\cdot}}\,.$$

5 Application to car model classificiation

As an illustration of our methods, we present some results on *car model classification* in car hull insurance, using data from the Swedish insurance group Länsförsäkringar Alliance. The *car model* (such as Renault Scénic 1.8, Volvo V70 2.4) is an important rating factor in motor insurance. As already mentioned in Section 3, car model is a multi-level factor (MLF): in Sweden there are roughly 2 500 car models in a preliminary classification, and there is not enough data to get useful estimates of price relativities for all these models. The conclusion is that car model is a case for credibility estimation.

We also have a large number of ordinary rating factors such as sex, age, annual mileage etc., and a specific car model j may appear in any a priori tariff cell i constructed from the ordinary factors. The presence of both ordinary rating factors and an MLF makes the algorithm of Section 3.1 appropriate for car model classification.

As is standard in GLM rating, we make a separate analysis of *claim frequency* using a Poisson distribution (or rather a variance function μ^p with $p = 1$) and *average claim severity* using a gamma distribution (or rather $p = 2$). Here we present results only from the claim frequency part of the study.

The idea to use credibility theory in car model classification has been discussed earlier by Campbell (1986) and Sundt (1987). The credibility models of these authors resemble ours; however, these texts were written before GLM became a standard rating tool and so their analysis is on *risk premium*—not claim frequency and severity—with weights corresponding to the case $p = 1$ in (3.3), and they do not iterate as we suggest in Section 3.1.

An application of the algorithm in Section 3.1 with all the ordinary rating factors and car model j gave the result shown by the non-filled bars (with the legend "No auxiliaries") in the histogram in Figure 1, where the bins are formed by rounding the $\widehat{U}_j$'s to the nearest first decimal number. We find that almost 30% of the car models received $\widehat{U}_j$'s in the span 0.95 to 1.05, but in most cases, the car model really makes a difference: the extremes differ by a factor of about $1.9/0.4 = 4.75$ from each other.

Another approach to car model classification is to introduce a number of ordinary rating factors that describe the car model technically, such as: the weight, the power of the engine, the type of car (saloon, estate car, convertible), etc. We call these variables *auxiliary classification factors* here, since they are not intended for separate use in the tariff, but only to help in the car model classification. Let us call this *the factor method* of car model classification; it has the advantage of not requiring large amounts of data for individual car models—a new model can be rated with the same accuracy as an old one.

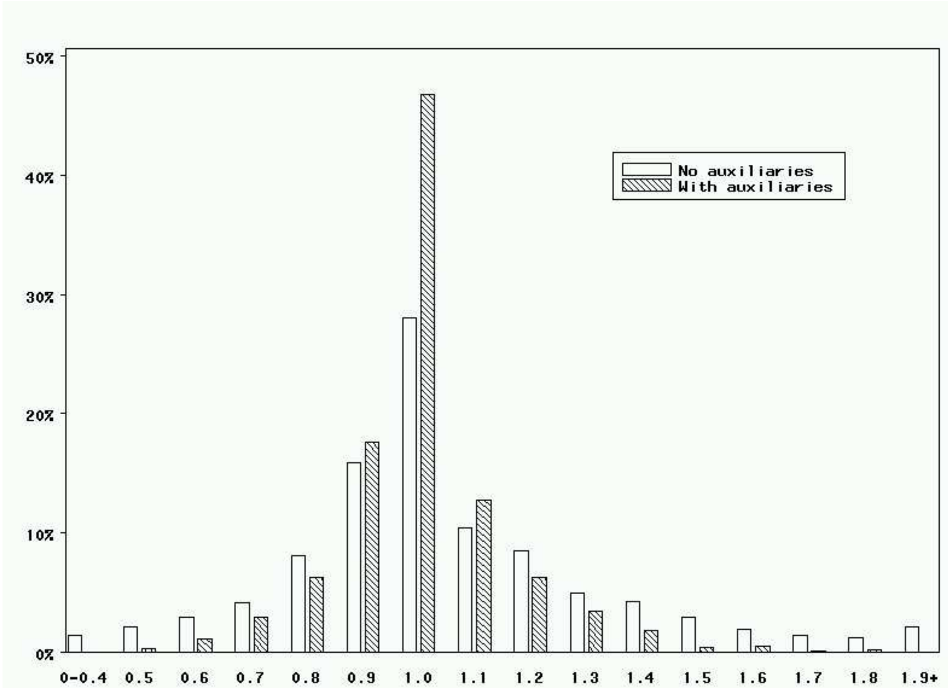

Figure 1: *Histogram of credibility estimators* $\widehat{U}_j$ *for car model, with and without auxiliary classification factors in the model.*

It is actually possible to have the best from both worlds, by combining the credibility method and the factor method as follows. Add the auxiliary classification factors to the ordinary factors in the GLM analysis and then again use the algorithm in Section 3.1. The resulting $\widehat{U}_j$'s are graphed in Figure 1, with the legend "With auxiliaries". Even though over 45% of the car models now get $\widehat{U}_j$'s close to 1, there is still a considerable variation among them; this shows that the factor method leaves a lot of residual variation, which can be estimated as a (residual) random effect U_j. A possible explanation is that differences in claim frequency is to a large extent a question of drivers behaviour and that different models attract different kinds of drivers, not only due to technical differences.

On the other hand, the introduction of the auxiliary factors reduce the variation of the $\widehat{U}_j$'s. This is an indication that we get a better tariff from inclusion of auxiliary factors in our credibility car model classification: a brand new or rare car model will get $\widehat{U}_j = 1$ because of lack of data, but the auxiliaries still provide some information on this car, thus reducing the non-explained variation in claim frequency. For models with large amounts of data, on the other hand, the introduction of auxiliary variables makes no difference.

This effect is shown in some detail in Table 5.1, for a sample of car models sorted by the exposure weights $w_{\cdot j\cdot}$ (number of policy years). As expected, with large $w_{\cdot j\cdot}$, the weighted averages of the observations $\overline{\overline{Y}}_{\cdot j\cdot}$ produce reliable estimates that are given high weight in the credibility formula, and the rating of car models is hardly affected by the introduction of auxiliaries, as seen by comparing the "No auxiliaries" $\widehat{U}_j$ to the "With auxiliaries" column c_j, where the product of the auxiliaries' price relativities is multiplied by the new $\widehat{U}_j$ from this round (note that the auxiliaries are tied to the car model j and not to the tariff cell i).

j	$w_{\cdot j\cdot}$	*No auxiliaries*			*With auxiliaries*			
		$\overline{\overline{Y}}_{\cdot j\cdot}$	$\widehat{U}_j$	z_j	$\overline{\overline{Y}}_{\cdot j\cdot}$	$\widehat{U}_j$	z_j	c_j
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 5.1: Selected car models j with number of policy years $w_{\cdot j\cdot}$, weighted averages $\overline{\overline{Y}}_{\cdot j\cdot}$, credibility estimators $\widehat{U}_j$ and credibility factors z_j; without and with auxiliary classification factors; the classification variable c_j is $\widehat{U}_j$ multiplied by the price relativities for the auxiliaries.

At the other end of the table, with data from very few policy years, credibility is low and the $\tilde{\overline{Y}}_{\cdot j\cdot}$'s are shaky. Here one has to rely, to a large extent, on the auxiliary factors. Without them, the classification would have been much closer to 1 for these cars.

Note. In applications we might have more than one MLF: in motor insurance, besides car model, we may use a geographical area defined by zip code or parish, say. In principle, the algorithm can be extended to the case with two or more independent MLFs in a rather straight-forward fashion; nevertheless, in practice we find it easier to rate one MLF at a time.□

5.1 The hierarchical case

As discussed at the beginning of Section 4, the *car brand* might contain valuable information that is not contained in the auxiliary (technical) data on the car. Now, car brand is itself an MLF: following Sundt (1987) we try an hierarchical credibility model for this situation. In our case the multiplicative model includes: first the ordinary rating factors, then the auxiliary factors based on technical properties of the car models, then the car brand stochastic effect, and then finally the car model stochastic effect, cf. Table 5.2.

Rating factors	Example	Notation
Ordinary rating factors	Age of driver	$\gamma_1^i, \gamma_2^i, \ldots, \gamma_R^i$
Auxiliary factors	Engine power	$\gamma_{R+1}^{jk}, \gamma_{R+2}^{jk}, \ldots, \gamma_{R+A}^{jk}$
Car brand	Volvo	U_j
Car model	Volvo V70 2.4	U_{jk}

Table 5.2: Variables in car model classification with hierarchical credibility and GLM.

Note that even though car brand is a random effect in the hierarchical credibility model, from a practical point of view it serves the same purpose as the auxiliaries: to improve the estimation of car model factors by extracting information that is common to several car models and hence may be estimated more accurately than the car models themselves. The full multiplicative model is then

$$E(Y_{ijkt}|U_j, U_{jk}) = \mu\, \gamma_1^i \cdots \gamma_R^i\, \gamma_{R+1}^{jk} \cdots, \gamma_{R+A}^{jk}\, U_j U_{jk}\,.$$

This model can be estimated using a straight-forward hierarchical credibility extension of the algorithm in Section 3.1, using the estimators from Section 4. Figure 2 displays $\widehat{U}_j$ for the 96 car brands; we can see that there are substantial differences between many of the 96 car brands, but not as large as between car models. We conclude that it is worthwhile to include the car brand level in our multiplicative model; this is especially important for the classification of car models on which we have none or few data.

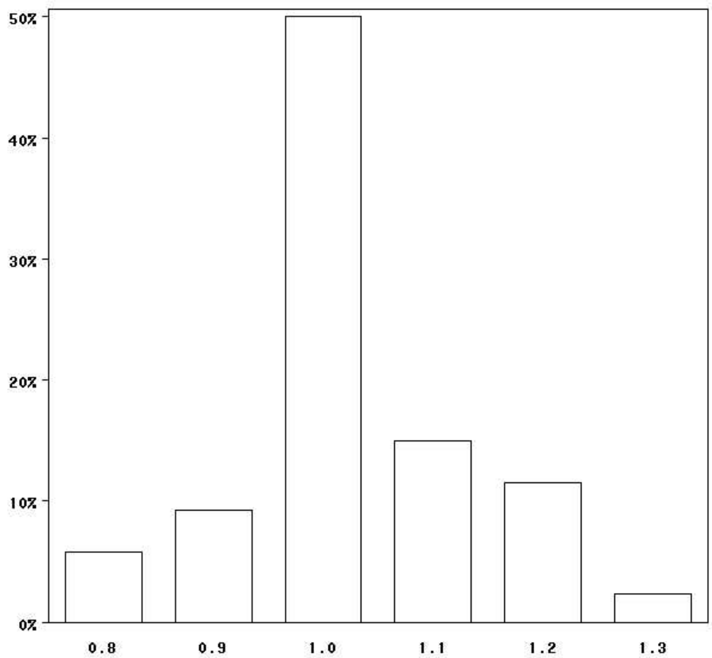

Figure 2: *Histogram of credibility estimators* $\widehat{U}_j$ *for car brands.*

After analysing claim frequency and claim severity by the above method, we multiply the auxiliary factors, the car brand factor $\widehat{U}_j$ and the car model factor $\widehat{U}_{jk}$, from each of these two analyses and get

$$c_{jk} = \gamma_{R+1}^{jk} \cdots , \gamma_{R+A}^{jk} \, \widehat{U}_j \widehat{U}_{jk} \, .$$

The c_{jk}'s for claim frequency and severity are then multiplied, giving a classification variable for the risk premium. The classification is finalised by calibrating this variable to maintain the total premium level, as is demonstrated in Sundt (1987, Section 3.4) and then forming classes from the calibrated variable.

Our overall conclusion is that the algorithm in Section 3.1, extended to the hierarchical case with the estimators from Section 4, is an efficient tool for car model classification and that the results are improved by the augmentation of the model with auxiliary technical variables as well as the car brand, estimated by hierarchical credibility in combination with GLMs.

6 Appendix

Here we give proofs of Theorems 2.1 and 4.2. We start by restating, without proof, a well-known lemma by Sundt (1980, Theorem 1(ii)). Recall that an estimator (predictor) $\hat{h}$ of the random variable V_j that is linear in the data $\boldsymbol{Y} = \{Y_{jt}\}$, is called the credibility estimator if it minimizes the MSEP $E[(h(\boldsymbol{Y}) - V_j)^2]$ among all linear estimators $h(\boldsymbol{Y})$.

Lemma 6.1 *A linear estimator $\hat{h}(\boldsymbol{Y})$ of V_j is the credibility estimator, if and only if,*

$$E(\hat{h}(\boldsymbol{Y})) = E(V_j)\,; \tag{6.1}$$

$$\mathrm{Cov}(\hat{h}(\boldsymbol{Y}), Y_{j't}) = \mathrm{Cov}(V_j, Y_{j't}); \quad \forall\, j', t\,. \tag{6.2}$$

Sundt also refers to a result by de Vylder, by which there always exists a unique credibility estimator; hence it is correct to speak of "*the* credibility estimator".

Below we will make repeated use of the following lemma, which is a simple extension of the standard rule for computing variances by conditioning, given here without proof.

Lemma 6.2 *Let X and Y be random variables, and Z a random vector, all with finite second moment.*

(a) Then

$$\mathrm{Cov}(X, Y) = E[\mathrm{Cov}(X, Y|Z)] + \mathrm{Cov}[E(X|Z), E(Y|Z)]\,.$$

(b) When X is a function of Z this specializes to

$$\mathrm{Cov}(X, Y) = \mathrm{Cov}[X, E(Y|Z)]\,.$$

Proof of Theorem 2.1. We shall prove that the equations of Lemma 6.1 are fulfilled for $\hat{h}(\boldsymbol{Y}) = \hat{V}_j$ in (2.7). Now since $E(V_j) = \mu$, and $E(Y_{jt}) = E[E(Y_{jt}|V_j)] = E(V_j) = \mu$, equation (6.1) is fulfilled.

Next note that because of the independence of groups, we have $\text{Cov}(V_j, Y_{j't}) = 0$ for $j' \neq j$, and since $\hat{V}_j$ only includes Y-values from group j, equation (6.2) is trivially fulfilled as soon as $j' \neq j$, and we only have to consider the case $j' = j$ in the following.

We use Lemma 6.2(b) to show that

$$\text{Cov}(V_j, Y_{jt}) = \text{Cov}[V_j, E(Y_t|V_j)] = \text{Var}(V_j) = \tau^2.$$

For $s \neq t$ we have by Assumption 2.1(c), this time using part (a) of Lemma 6.2,

$$\text{Cov}(Y_{js}, Y_{jt}) \quad = \quad E[\text{Cov}(Y_{js}, Y_{jt})|V_j] + \text{Cov}[E(Y_{js}|V_j), E(Y_{jt}|V_j)] = 0 + \text{Var}[V_j] = \tau^2.$$

For $s = t$, on the other hand, we have by (2.5),

$$\text{Cov}(Y_{jt}, Y_{jt}) = \text{Var}(Y_{jt}) = \tau^2 + \frac{\sigma^2}{w_{jt}}\,;$$

hence by (2.9)

$$\text{Cov}(\widehat{V}_j, Y_{jt}) = z_j \sum_s \frac{w_{js}}{w_{j\cdot}} \text{Cov}(Y_{js}, Y_{jt}) = z_j \left(\tau^2 + \frac{w_{jt}}{w_{j\cdot}} \frac{\sigma^2}{w_{jt}}\right) = \tau^2 z_j \left(\frac{w_{j\cdot} + \sigma^2/\tau^2}{w_{j\cdot}}\right) = \tau^2\,.$$

This finishes the proof of Theorem 2.1. □

Proof of Theorem 4.2. Again, we shall verify that the equations of Lemma 6.1 are fulfilled. In the present setting these equations are, if we divide both sides of the second equation by γ_i and recall that $\widetilde{Y}_{ijkt} = Y_{ijkt}/\gamma_i$,

$$E(\widehat{V}_{jk}) \quad = \quad E(V_{jk})\,, \tag{6.3}$$

$$\text{Cov}(\widehat{V}_{jk}, \widetilde{Y}_{ijk't}) \quad = \quad \text{Cov}(V_{jk}, \widetilde{Y}_{ijk't}); \quad \forall\, i, k', t\,. \tag{6.4}$$

Here we have excluded the $\widetilde{Y}_{ij'k't}$-variables with $j' \neq j$, since independence makes the corresponding covariances equal 0.

Now by Assumption 4.1(c), $E(V_{jk}) = E[V_j] = \mu$. By (4.4) we have $E(\widetilde{\overline{Y}}_{\cdot jk\cdot}) = \mu$, which we use in (4.6) to find $E(\widehat{V}_j) = \mu$. By inserting this in (4.8), we obtain $E(\widehat{V}_{jk}) = \mu$, proving that (6.3) is fulfilled. We now turn to the covariances.

By (4.1), Assumption 4.1(b) and (c), Lemma 6.2 and (4.4),

$$\begin{aligned} & \mathrm{Cov}(V_{jk}, \widetilde{Y}_{ijk't}) = \mathrm{Cov}[V_{jk}, E(\widetilde{Y}_{ijk't}|V_j, V_{jk}, V_{jk'})] = \mathrm{Cov}(V_{jk}, V_{jk'}) \\ = \; & E[\mathrm{Cov}(V_{jk}, V_{jk'})|V_j] + \mathrm{Cov}[E(V_{jk}|V_j), E(V_{jk'}|V_j)] = \begin{cases} \nu^2 + \tau^2, & \text{if } k' = k; \\ \tau^2, & \text{if } k' \neq k. \end{cases} \end{aligned} \tag{6.5}$$

Next we turn to the the left-hand side of (6.4) where

$$\mathrm{Cov}(\widehat{V}_{jk}, \widetilde{Y}_{ijk't}) = z_{jk}\mathrm{Cov}(\widetilde{\overline{Y}}_{\cdot jk\cdot}, \widetilde{Y}_{ijk't}) + (1 - z_{jk})\mathrm{Cov}(\widehat{V}_j, \widetilde{Y}_{ijk't}) \,. \tag{6.6}$$

But since $\widehat{V}_j$ is a credibility estimator of V_j, we have by Lemma 6.1,

$$\mathrm{Cov}(\widehat{V}_j, \widetilde{Y}_{ijk't}) = \mathrm{Cov}(V_j, \widetilde{Y}_{ijk't}) = \mathrm{Cov}[V_j, E(\widetilde{Y}_{ijk't}|V_j)] = \mathrm{Cov}[V_j, V_j] = \tau^2 \,. \tag{6.7}$$

In the case $k' \neq k$, by Assumption 4.1(b),

$$\mathrm{Cov}(\widetilde{\overline{Y}}_{\cdot jk\cdot}, \widetilde{Y}_{ijk't}) = E[\mathrm{Cov}(\widetilde{\overline{Y}}_{\cdot jk\cdot}, \widetilde{Y}_{ijk't}|V_j)] + \mathrm{Cov}[E(\widetilde{\overline{Y}}_{\cdot jk\cdot}|V_j), E(\widetilde{Y}_{ijk't}|V_j)] = 0 + \mathrm{Cov}[V_j, V_j] = \tau^2 \,, \tag{6.8}$$

while in the case $k' = k$ we first note that

$$\begin{aligned} \mathrm{Cov}(\widetilde{\overline{Y}}_{\cdot jk\cdot}, \widetilde{Y}_{ijkt}|V_j) &= E[\mathrm{Cov}(\widetilde{\overline{Y}}_{\cdot jk\cdot}, \widetilde{Y}_{ijkt}|V_j, V_{jk})|V_j] + \mathrm{Cov}[E(\widetilde{\overline{Y}}_{\cdot jk\cdot}|V_j, V_{jk}), E(\widetilde{Y}_{ijkt}|V_j, V_{jk})|V_j] \\ &= \frac{\widetilde{w}_{ijkt}}{\widetilde{w}_{\cdot jk\cdot}} E[\mathrm{Var}(\widetilde{Y}_{ijkt}|V_j, V_{jk})|V_j] + \mathrm{Var}[V_{jk}|V_j] \,. \end{aligned}$$

We now proceed as in (6.8) and get, by (4.2), (4.3) and (4.5),

$$\begin{aligned} \mathrm{Cov}(\widetilde{\overline{Y}}_{\cdot jk\cdot}, \widetilde{Y}_{ijkt}) &= E[\mathrm{Cov}(\widetilde{\overline{Y}}_{\cdot jk\cdot}, \widetilde{Y}_{ijk't}|V_j, V_{jk})] + E[\mathrm{Cov}(V_{jk}, V_{jk}|V_j)] + \tau^2 \\ &= \frac{\widetilde{w}_{ijkt}}{\widetilde{w}_{\cdot jk\cdot}} E[\mathrm{Var}(\widetilde{Y}_{ijkt}|V_j, V_{jk})] + E[\mathrm{Var}(V_{jk}|V_j)] + \tau^2 \\ &= \frac{\sigma^2}{\widetilde{w}_{\cdot jk\cdot}} + \nu^2 + \tau^2 = \frac{\nu^2}{z_{jk}} + \tau^2 \,. \end{aligned}$$

By inserting this, (6.8) and (6.7) into (6.6) we conclude

$$\mathrm{Cov}(\widehat{V}_{jk}, \widetilde{Y}_{ijkt}) = \begin{cases} z_{jk}\left(\frac{\nu^2}{z_{jk}} + \tau^2\right) + (1 - z_{jk})\tau^2 = \nu^2 + \tau^2, & \text{if } k' = k\,; \\ z_{jk}\tau^2 + (1 - z_{jk})\tau^2 = \tau^2, & \text{if } k' \neq k\,. \end{cases}$$

But this is the right-hand side of (6.5); hence, (6.4) is fulfilled and the proof is complete. □

References

Bühlmann, H. & Gisler, A. (2005): *A Course in Credibility Theory and its Applications.* Springer Universitext.

Bühlmann, H. & Straub, E. (1970): *Glaubwürdigkeit für Schadensätze.* Bulletin of Swiss Association of Actuaries, 111-133.

Campbell, M. (1986): *An integrated system for estimating the risk premium of individual car models in motor insurance.* ASTIN Bulletin, Vol. 16:2, 165-183.

Dannenburg, D.R., Kaas, R. & Goovaerts, M.J. (1996): *Practical actuarial credibility models.* IAE, Amsterdam.

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

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

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.

Ohlsson, E. (2005): *Simplified estimation of structure parameters in hierarchical credibility.* The 36th ASTIN Colloquium, Zurich; http://www.astin200f.ch/.

Ohlsson, E. & Johansson, B. (2006): *Exact credibility and Tweedie models.* To appear in ASTIN Bulletin.

Sundt, B. (1980): *A Multi-level Hierarchical Credibility Regression Model.* Scandinavian Actuarial Journal 1:25-32.

Sundt, B. (1987): *Two Credibility Regression Approaches for the Classification of Passenger Cars in a Multiplicative Tariff.* ASTIN Bulletin, Vol. 17:1, 42-70.

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