A MIXTURE MODEL FOR PAYMENTS AND PAYMENT NUMBERS IN CLAIMS RESERVING

BY

PATRIZIA GIGANTE, LIVIANA PICECH AND LUCIANO SIGALOTTI

ABSTRACT

We consider a Tweedie's compound Poisson regression model with fixed and random effects, to describe the payment numbers and the incremental payments, jointly, in claims reserving. The parameter estimates are obtained within the framework of hierarchical generalized linear models, by applying the h-likelihood approach. Regression structures are allowed for the means and also for the dispersions. Predictions and prediction errors of the claims reserves are evaluated. Through the parameters of the distributions of the random effects, some external information (e.g. a development pattern of industry wide-data) can be incorporated into the model. A numerical example shows the impact of external data on the reserve and prediction error evaluations.

KEYWORDS

Claims reserving, Tweedie's compound Poisson model, conditional mean square error of prediction, hierarchical generalized linear models, h-likelihood.

1. INTRODUCTION

The evaluation of claims reserves, which are set aside to meet the liabilities for outstanding claims, is one of the main actuarial tasks in non-life insurance. In recent years, also because of the new solvency regulations, insurance companies are interested in providing best estimates of the outstanding claims and also in quantifying the uncertainty of such estimates. For this purpose, appropriate stochastic claims reserving models are needed (England and Verrall, 2002; Wüthrich and Merz, 2008).

In classical reserving methods, the prediction of future payments usually relies on run-off triangles of payments. However, it is quite common for the insurers to have available more information on the claims development process, e.g. the numbers of reported claims, the numbers of payments, the incurred losses. Therefore, it emerges the need of introducing stochastic models designed to incorporate such additional data in order to get more accurate models for the reserve evaluation.

In this paper, we assume that the available data consist of the two run-off triangles of the payment numbers and of the standardized incremental payments. Models for standardized claim payments, with respect to some exposure measure, are often used in the literature (see e.g. Taylor, 2000).

The reserving problem with payment numbers and standardized incremental payments is addressed in Wüthrich (2003) and Boucher and Davidov (2011) in the context of regression models and under the assumption that the standardized incremental payments are Tweedie's compound Poisson distributed. In Wüthrich (2003), the model is estimated within the framework of Generalized Linear Models (GLMs) with constant dispersion parameter. In Boucher and Davidov (2011), the model introduced by Smyth and Jørgensen (2002) for tarification is applied to claims reserving. Besides of the means, also the dispersion parameters are modeled through a regression structure in order to enhance the variance estimation. In fact, in some situations, models with constant dispersion could be inappropriate. The parameter estimates are obtained by two interconnected GLMs in the framework of Double Generalized Linear Models (DGLMs, see e.g. Smyth, 1989; Smyth and Verbyla, 1999; Nelder and Lee, 1991).

In recent papers (see Verrall *et al.*, 2010; Miranda-Martinez *et al.*, 2011; Miranda-Martinez, Nielsen and Verrall, 2012; Miranda-Martinez, Nielsen and Wüthrich, 2012), the Double Chain Ladder (CL) approach is used to evaluate the claims reserve through a micro-level description of the claim development process, based on the numbers of reported claims and the numbers of payments, in addition to the payments.

In the above papers, the regression parameters are fixed effects related to observable covariates, typically the origin and the development years. In the literature, models with random effects related to unobservable risk parameters are also considered. Such models are studied by following the Bayesian or the credibility approaches (e.g. Mack, 2000; de Alba, 2002; Ntzoufras and Dellaportas, 2002; Verrall, 2004; Verrall and England, 2005; England and Verrall, 2006; England *et al.*, 2012; Bühlmann and Moriconi, 2015; Taylor, 2015). Within regression models, the techniques of GLMs are combined with those of credibility theory (e.g. Nelder and Verrall, 1997; Ohlsson and Johansson, 2006; Ohlsson, 2008) or the Generalized Linear Mixed Models are used (Antonio *et al.*, 2006; Antonio and Beirlant, 2007). To estimate GLMs with random effects, Lee and Nelder (1996), Lee and Nelder (2001), Lee *et al.* (2006) suggest the hierarchical or *h*-likelihood approach. Such models are called Hierarchical Generalized Linear Models (HGLMs). The use of HGLMs in loss reserving has been considered in Gigante *et al.* (2013a) and Gigante *et al.* (2013b).

In this paper, we assume for the couples payment numbers and standardized payments a mixture model with risk parameters related to the origin or the development year or both of them. The risk parameters are introduced in the regression structure through random effects, hence the model belongs to the class of mixed models. It extends the fixed effect model in Smyth and Jørgensen (2002) and Boucher and Davidov (2011) to a model with random effects for the means of the payments, whereas the dispersion parameters follow a regression structure with fixed effects. The payment numbers and the incremental payments conditioned to the risk parameters are independent and the conditional incremental payments are compound Poisson distributed.

Because of the risk parameters, the couples number of payments and incremental payments in the run-off table are not independent. In this way, we can take account of dependencies due to unobservable effects related to the origin years and to the development years, such as correlation patterns among payments of a given origin year, or residual heterogeneity.

Through the parameters of the distributions of the random effects, the model allows us to incorporate some external information. This can be useful, in particular, when the observed development figures within a given run-off triangle are scarce or when they fluctuate, so that it can be important to rely on industry-wide development patterns or on development patterns of other similar lines of business. The need of taking account of external data (e.g. initial estimates of the ultimate claim amount, the development pattern of industry wide-data, market information or expert opinion) when assessing the claims reserve is a well-known problem. Some recent references on these aspects are Gisler and Wüthrich (2008), Saluz *et al.* (2011), Saluz (2015).

To estimate the model parameters, we follow the h-likelihood approach. In this way, we get estimates of the fixed and random effects, and also estimates of the variance–covariance matrix of the estimators, that can be used to evaluate prediction errors.

Some numerical results are supplied for illustrative purposes.

The rest of the paper is structured as follows. In Section 2, we introduce the model assumptions and we describe the estimation procedure based on the h-likelihood. By appropriately assessing the distributions of the risk parameters, the model h-likelihood coincides with that of an HGLM. In Section 3, regression structures for the dispersion parameters of the payments and of the risk parameters are added to the model. To estimate the model, by following a suggestion in Lee *et al.* (2006), we approximate the h-likelihood components related to the risk parameters by the respective extended quasi-likelihoods. The estimation procedure is reduced to fitting four interconnected GLMs, iteratively. In Section 4, we discuss the prediction problem in claims reserving and give approximate formulae, that can be easily calculated, in order to evaluate the prediction uncertainty. In Section 5, some numerical results are provided by using the data in Wüthrich (2003) and in Taylor (2000), and by assuming that external information on the claims development pattern is available. The results are compared with those obtained by DGLMs.

2. Model assumptions

We assume that the data of a portfolio consist of the two run-off triangles of the payment numbers and of the incremental payments standardized with respect to some exposure measure. Let n_{ij} be the number of payments, p_{ij} the incremental payments, ω_{ij} the exposure measure and $y_{ij} = p_{ij}/\omega_{ij}$ the standardized payments, $i, j = 0, \ldots, t$, $i + j \leq t$, where i denotes the origin year (accident year, underwriting year, ...), j the development year and t the most recent origin year, assumed to be equal to the latest development year (we do not consider tail factors).

We note that the data could have a more general shape i.e. $i = 0, \ldots, t_1$, $j = 0, \ldots, t_2$, with $t_1 \neq t_2$, as in the examples discussed in the last section of the paper. However, in the model description, we assume the usual triangular shape, to simplify the notations.

In connection with the above data set, we introduce the random process $\{(N_{ij}, Y_{ij}), i, j = 0, \ldots, t\}$, where $Y_{ij} = P_{ij}/\omega_{ij}$ and the exposure measure ω_{ij} is assumed to be known for any $i, j = 0, \ldots, t$.

For this process, we consider a mixture model depending on a vector of risk parameters connected with the origin and the development years. Let $(\boldsymbol{U}, \boldsymbol{V}) = (U_0, \ldots, U_t, V_0, \ldots, V_t)$ the vector of the risk parameters, where U_i is related to the origin year i and V_j to the development year j.

We assume that, conditionally on (U_i, V_j), the incremental payments $P_{ij} = \sum_{h=1}^{N_{ij}} Z_{ij}^{(h)}$ are compound Poisson distributed and that the size of each claim is gamma distributed with shape parameter α. From Jørgensen (1987, 1997) and Jørgensen and de Souza (1994), it is known that conditionally on (U_i, V_j), the distribution of $Y_{ij} = P_{ij}/\omega_{ij}$ belongs to an Exponential Dispersion Family (EDF) with $\text{var}[Y_{ij}|(U_i, V_j) = (u_i, v_j)] = \phi_{ij}\mu_{ij}^p/\omega_{ij}$, where $\mu_{ij} = E[Y_{ij}|(U_i, V_j) = (u_i, v_j)]$, $p = (\alpha + 2)/(\alpha + 1)$ and ϕ_{ij} is the dispersion parameter. Hence, $Y_{ij}|(U_i, V_j) = (u_i, v_j)$ follows a Tweedie's model, with $1 < p < 2$ implied by the positivity of α. Moreover, the joint density of $(N_{ij}, Y_{ij})|(U_i, V_j) = (u_i, v_j)$ can be parameterized in terms of μ_{ij}, ϕ_{ij} and p, that characterize the mean and the variance of the standardized incremental payments $Y_{ij}|(U_i, V_j) = (u_i, v_j)$. In fact, we have

$$f_{(N_{ij}, Y_{ij})|(U_i, V_j)=(u_i,v_j)}(n, y; \theta_{ij}, \phi_{ij}, p) = a(n, y) \exp\left\{\frac{\omega_{ij}}{\phi_{ij}}(y\theta_{ij} - b_p(\theta_{ij}))\right\}, \quad (2.1)$$

where

$$a(n, y) = \begin{cases} 1 & n = 0, y = 0 \\ \frac{1}{n!\Gamma(\alpha n)y}\left[\frac{y^{\alpha}(\omega_{ij}/\phi_{ij})^{(\alpha+1)}}{(p-1)^{\alpha}(2-p)}\right]^n & n > 0, y > 0, \end{cases}$$

with $1 < p < 2$, $\alpha = (2 - p)/(p - 1)$, $\theta_{ij} = \frac{\mu_{ij}^{1-p}}{1-p}$ and $b_p(\theta) = \frac{1}{2-p}[(1 - p)\theta]^{\frac{2-p}{1-p}}$. In terms of μ_{ij}, the density is

$$f_{(N_{ij}, Y_{ij})|(U_i, V_j)=(u_i,v_j)}(n, y; \mu_{ij}, \phi_{ij}, p) = a(n, y) \exp\left\{\frac{\omega_{ij}}{\phi_{ij}}\left(y\frac{\mu_{ij}^{1-p}}{1-p} - \frac{\mu_{ij}^{2-p}}{2-p}\right)\right\}. \quad (2.2)$$

For the means of the conditional standardized payments, we assume a regression structure as in GLMs, $\mu_{ij} = g^{-1}(\eta_{ij})$, where g is a link function and η_{ij} is a linear predictor that depends on the values of observable covariates $\boldsymbol{x}_{ij}$, but, differently from GLMs, also on random effects. The observable covariates that can be numerical or categorical are chosen among the origin year, the development year and the payment year. The random effects are additive terms obtained by transforming the risk parameters by means of strictly monotone and regular functions.

Now, we state the whole model.

a1. *Independence assumptions*
The components of the risk parameter $(\boldsymbol{U}, \boldsymbol{V}) = (U_0, \ldots, U_t, V_0, \ldots, V_t)$ are independent.
Conditionally on $(\boldsymbol{U}, \boldsymbol{V})$, the couples (N_{ij}, Y_{ij}), $i, j = 0, \ldots, t$, are independent.
With respect to the risk parameters $(\boldsymbol{U}, \boldsymbol{V})$, the distribution of $(N_{ij}, Y_{ij})|(\boldsymbol{U}, \boldsymbol{V}) = (\boldsymbol{u}, \boldsymbol{v})$ only depends on $U_i = u_i$ and $V_j = v_j$; i.e. $\left[(N_{ij}, Y_{ij})|(\boldsymbol{U}, \boldsymbol{V}) = (\boldsymbol{u}, \boldsymbol{v})\right] \overset{d}{=} \left[(N_{ij}, Y_{ij})|(U_i, V_j) = (u_i, v_j)\right]$.

a2. *Distributional assumptions for the responses conditional on the risk parameters*
The distribution of $(N_{ij}, Y_{ij})|(U_i, V_j) = (u_i, v_j)$ has density (2.1) or, equivalently, (2.2), for a given parameter p.

a3. *Structural assumptions for the expected standardized payments*
The expectations of the conditional standardized payments are given by

$$E[Y_{ij}|(U_i, V_j) = (u_i, v_j)] = \mu_{ij} = g^{-1}(\boldsymbol{x}_{ij}^T \boldsymbol{\beta} + \boldsymbol{z}_{ij}^T \boldsymbol{w}),$$

where $\boldsymbol{x}_{ij}$ and $\boldsymbol{z}_{ij}$ are vectors of covariates; $\boldsymbol{\beta}$ are the regression parameters, called *fixed effects*; $\boldsymbol{w} = (\boldsymbol{w}_U, \boldsymbol{w}_V) = (w_{U,0}, \ldots, w_{U,t}, w_{V,0}, \ldots, w_{V,t})$ are the *random effects*, with $w_{U,i} = g_U(u_i)$ and $w_{V,j} = g_V(v_j)$. The functions g, g_U and g_V are strictly monotone with first- and second-order continuous derivatives.
The vector $\boldsymbol{z}_{ij}$ is such that $\boldsymbol{z}_{ij}^T \boldsymbol{w} = w_{U,i} + w_{V,j}$ or $\boldsymbol{z}_{ij}^T \boldsymbol{w} = w_{U,i}$ or $\boldsymbol{z}_{ij}^T \boldsymbol{w} = w_{V,j}$. In the following, we assume that $\boldsymbol{z}_{ij}^T \boldsymbol{w} = w_{U,i} + w_{V,j}$; the other two cases can be treated by simple adjustments.

a4. *Distributional assumptions for the risk parameters*
Let $W_{U,i} = g_U(U_i)$ and $W_{V,j} = g_V(V_j)$. We assume that the densities of $W_{U,i}$ and $W_{V,j}$ are

$$f_{W_{U,i}}(w) = \exp\left\{\frac{1}{\lambda_{U,i}}\left(\psi_{U,i}\theta_U - b_U(\theta_U)\right)\right\} c_U(\psi_{U,i}, \lambda_{U,i}),$$

$$f_{W_{V,j}}(w) = \exp\left\{\frac{1}{\lambda_{V,j}}\left(\psi_{V,j}\theta_V - b_V(\theta_V)\right)\right\} c_V(\psi_{V,j}, \lambda_{V,j}),$$

where b_U, b_V are cumulant functions of EDFs, $\theta_U = b_U'^{-1}(g_U^{-1}(w))$, $\theta_V = b_V'^{-1}(g_V^{-1}(w))$, $\psi_{U,i}$, $\lambda_{U,i}$, $\psi_{V,j}$, $\lambda_{V,j}$ are parameters and $c_U(\psi_{U,i}, \lambda_{U,i})$, $c_V(\psi_{V,j}, \lambda_{V,j})$ are normalizing functions.

The above assumptions define a mixture model with mixing distribution the distribution of $\boldsymbol{W} = (W_{U,0}, \ldots, W_{U,t}, W_{V,0}, \ldots, W_{V,t})$.

If, in particular, g_U is the canonical link of b_U, that is $g_U = b_U'^{-1}$, then we have $\theta_U = w$ and the distribution of $W_{U,i} = b_U'^{-1}(U_i)$ belongs to the conjugate family of the EDF with cumulant b_U. Its density is

$$f_{W_{U,i}}(w) = \exp\left\{\frac{1}{\lambda_{U,i}}\left(\psi_{U,i}w - b_U(w)\right)\right\} c_U(\psi_{U,i}, \lambda_{U,i}). \tag{2.3}$$

In this case, under suitable hypotheses, the so-called hyperparameters $\boldsymbol{\psi_U} = (\psi_{U,0}, \ldots \psi_{U,t})$, $\boldsymbol{\lambda_U} = (\lambda_{U,0}, \ldots \lambda_{U,t})$ are related to the moments of the risk parameter $\boldsymbol{U}$. Indeed, if $f_{W_{U,i}}$ disappears on the boundary of the canonical parameter space of the EDF with cumulant b_U, for any $\psi_{U,i}$, then $\psi_{U,i} = E(U_i)$ (see e.g. Jewell, 1974; Bühlmann and Gisler, 2005). If, in addition to the above hypothesis, the first derivative of $f_{W_{U,i}}$ disappears on the boundary of the canonical parameter space, then $\lambda_{U,i} = \mathrm{var}(U_i)/E(U_i^p)$ (see Ohlsson and Johansson, 2006). Similar considerations apply to $W_{V,j}$.

In the following, we assume that the parameters $\boldsymbol{\psi_U}$, $\boldsymbol{\psi_V}$ are given, so that the distribution of $\boldsymbol{W}$ only depends on the *dispersion parameters* $\boldsymbol{\lambda_U}$, $\boldsymbol{\lambda_V}$. We remark that the values of the parameters $\boldsymbol{\psi_U}$, $\boldsymbol{\psi_V}$ can be used to incorporate external information into the model (see the examples in Section 5).

If $b_p = b_U = b_V$ and $g = g_U = g_V = b_p'^{-1}$, then the distributions of both $W_{U,i} = b_p'^{-1}(U_i)$ and $W_{V,j} = b_p'^{-1}(V_j)$ are conjugate of the distribution of $Y_{ij}|(U_i, V_j) = (u_i, v_j)$. We have

$$f_{W_{U,i}}(w) = \exp\left\{\frac{1}{\lambda_{U,i}}\left(\psi_{U,i}w - b_p(w)\right)\right\} c(\psi_{U,i}, \lambda_{U,i}), \tag{2.4}$$

$$f_{W_{V,j}}(w) = \exp\left\{\frac{1}{\lambda_{V,j}}\left(\psi_{V,j}w - b_p(w)\right)\right\} c(\psi_{V,j}, \lambda_{V,j}). \tag{2.5}$$

We remark that the model hypotheses for the process $\{(N_{ij}, Y_{ij}), i, j = 0, \ldots, t\}$, conditional on $(\boldsymbol{U}, \boldsymbol{V})$, are the same as in Smyth and Jørgensen (2002). Moreover, if the random effects are only those related to the origin years, that is $\boldsymbol{z}_{ij}^T \boldsymbol{w} = w_{U,i}$, the process $\{U_i, Y_{ij}, i, j = 0, \ldots, t\}$ falls within the models dealt with in Gigante *et al.* (2013a).

In order to estimate the parameters in models with fixed and random effects, Lee and Nelder (1996), Lee and Nelder (2001), Lee *et al.* (2006) introduced the *hierarchical log-likelihood* or *h-loglihood*. In our problem, the *h*-loglihood is the joint log-density evaluated at the data $\boldsymbol{n} = (n_{ij}, i+j \le t)$, $\boldsymbol{y} = (y_{ij}, i+j \le t)$,

$$
\begin{aligned}
h &= \log f_{(\boldsymbol{N},\boldsymbol{Y},\boldsymbol{W})} = l_{(\boldsymbol{N},\boldsymbol{Y})|\boldsymbol{W}=\boldsymbol{w}} + l_{\boldsymbol{W}} \\
&= \sum_{i,j:i+j\le t} \Bigg\{ n_{ij} \log \frac{(\omega_{ij}/\phi_{ij})^{\alpha+1} y_{ij}^{\alpha}}{(p-1)^{\alpha}(2-p)} - \log\left(n_{ij}!\Gamma(\alpha n_{ij}) y_{ij}\right) \\
&\quad + \frac{\omega_{ij}}{\phi_{ij}} \left[y_{ij}\theta_{ij} - b_p(\theta_{ij}) \right] \Bigg\} \\
&\quad + \sum_{i=0}^{t} \left\{ \frac{1}{\lambda_{U,i}} \left[\psi_{U,i}\theta_{U,i} - b_U(\theta_{U,i}) \right] + \log\ c_U(\psi_{U,i}, \lambda_{U,i}) \right\} \\
&\quad + \sum_{j=0}^{t} \left\{ \frac{1}{\lambda_{V,j}} \left[\psi_{V,j}\theta_{V,j} - b_V(\theta_{V,j}) \right] + \log\ c_V(\psi_{V,j}, \lambda_{V,j}) \right\},
\end{aligned}
\tag{2.6}
$$

where $f_{(\boldsymbol{N},\boldsymbol{Y},\boldsymbol{W})}$ denotes the joint density of $(\boldsymbol{N}, \boldsymbol{Y}, \boldsymbol{W})$, $l_{(\boldsymbol{N},\boldsymbol{Y})|\boldsymbol{W}=\boldsymbol{w}}$ the log-likelihood of $(\boldsymbol{N}, \boldsymbol{Y})|\boldsymbol{W} = \boldsymbol{w}$, which is equal to the log-likelihood of $(\boldsymbol{N}, \boldsymbol{Y})|(\boldsymbol{U}, \boldsymbol{V}) = (\boldsymbol{u}, \boldsymbol{v})$, and $l_{\boldsymbol{W}}$ is the logarithm of the density of $\boldsymbol{W}$ and

$$
\begin{aligned}
\theta_{ij} &= b_p'^{-1}(g^{-1}(\boldsymbol{x}_{ij}^T\boldsymbol{\beta} + w_{U,i} + w_{V,j}), \\
\theta_{U,i} &= b_U'^{-1}(g_U^{-1}(w_{U,i})), \\
\theta_{V,j} &= b_V'^{-1}(g_V^{-1}(w_{V,j})).
\end{aligned}
$$

Hence, if in addition to p, $\boldsymbol{\psi}_{\boldsymbol{U}}$, $\boldsymbol{\psi}_V$ and $\boldsymbol{\omega} = (\omega_{ij}, i, j = 0, \dots, t)$, also the dispersion parameters $\boldsymbol{\phi} = (\phi_{ij}, i, j = 0, \dots, t)$, related to the standardized payments, and $\boldsymbol{\lambda}_U$, $\boldsymbol{\lambda}_V$, related to the risk parameters, are known, ignoring irrelevant constant terms, we get

$$
\begin{aligned}
h(\boldsymbol{\beta}, \boldsymbol{w}; \boldsymbol{\phi}, \boldsymbol{\lambda}_{\boldsymbol{U}}, \boldsymbol{\lambda}_V; \boldsymbol{n}, \boldsymbol{y}, \boldsymbol{\psi}_{\boldsymbol{U}}, \boldsymbol{\psi}_V, \boldsymbol{\omega}) &= \sum_{i,j:i+j\le t} \frac{\omega_{ij}}{\phi_{ij}} \left[y_{ij}\theta_{ij} - b_p(\theta_{ij}) \right] \\
&\quad + \sum_{i=0}^{t} \frac{1}{\lambda_{U,i}} \left[\psi_{U,i}\theta_{U,i} - b_U(\theta_{U,i}) \right] + \sum_{j=0}^{t} \frac{1}{\lambda_{V,j}} \left[\psi_{V,j}\theta_{V,j} - b_V(\theta_{V,j}) \right].
\end{aligned}
\tag{2.7}
$$

The h-loglihood (2.7) formally coincides with that of an HGLM with responses the standardized payments and with random effects the vector $\boldsymbol{w}$. Remark that it is the same as if the n_{ij} were not observed. It can be viewed as the log-likelihood of an *augmented* GLM for the data $\boldsymbol{y}$ and *pseudo-data* $\boldsymbol{\psi}_{\boldsymbol{U}}$, $\boldsymbol{\psi}_V$, with weights ω_{ij}/ϕ_{ij}, $i+j \le t$, $1/\lambda_{U,i}$, $i = 0, \dots, t$, $1/\lambda_{V,j}$, $j = 0, \dots, t$, respectively, and dispersion parameter 1. Notice that we use this terminology even if, in order to interpret (2.7) as the log-likelihood of a GLM, we should have $b_p = b_U = b_V$ and $g = g_U = g_V$. In this case, if g is the canonical link, the HGLM is called *conjugate*.

The augmented GLM has the following structure, denoting by $\overline{Y}_{ij}$, $i+j \leq t$, $\Psi_{U,i}$, $i = 0, \ldots, t$, $\Psi_{V,j}$, $j = 0, \ldots, t$, the response variables and by V, V_U, V_V the variance functions of the EDFs with cumulants b_p, b_U, b_V, respectively, the expected values and the variances of the response variables are

$$
\begin{array}{lll}
E(\overline{Y}_{ij}) = \mu_{ij}, & \operatorname{var}(\overline{Y}_{ij}) = \frac{\phi_{ij}}{\omega_{ij}} V(\mu_{ij}), & i+j \leq t, \\
E(\Psi_{U,i}) = u_i, & \operatorname{var}(\Psi_{U,i}) = \lambda_{U,i} V_U(u_i), & i = 0, \ldots, t, \\
E(\Psi_{V,j}) = v_j, & \operatorname{var}(\Psi_{V,j}) = \lambda_{V,j} V_V(v_j), & j = 0, \ldots, t.
\end{array}
$$

The linear predictors are

$$
\begin{array}{ll}
\eta_{ij} = g(\mu_{ij}) = \boldsymbol{x}_{ij}^T \boldsymbol{\beta} + w_{U,i} + w_{V,j}, & i+j \leq t, \\
\eta_{U,i} = g_U(u_i) = w_{U,i}, & i = 0, \ldots, t, \\
\eta_{V,j} = g_V(v_j) = w_{V,j}, & j = 0, \ldots, t.
\end{array}
$$

The design matrix of the model is

$$
\boldsymbol{T} = \begin{bmatrix} \boldsymbol{X} & \boldsymbol{Z}_U & \boldsymbol{Z}_V \\ \boldsymbol{0} & \boldsymbol{I}_{t+1} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{I}_{t+1} \end{bmatrix}, \tag{2.8}
$$

where $\boldsymbol{X} = [\boldsymbol{x}_{ij}^T]$ denotes the design matrix for the fixed effects, $\boldsymbol{I}_{t+1}$ is the identity matrix of order $t+1$, $\boldsymbol{Z}_U$ is the design matrix for the random effects $\boldsymbol{w}_U$, i.e. a $((t+1)(t+2)/2) \times (t+1)$ block indicator matrix, whose element in column k ($k = 0, \ldots, t$) corresponding to the observation y_{ij} is 1 if $k = i$, and $\boldsymbol{Z}_V$ is the design matrix for the random effects $\boldsymbol{w}_V$, i.e. a $((t+1)(t+2)/2) \times (t+1)$ block matrix where the block corresponding to the payments of the origin year i is the identity matrix of order $t - i + 1$ followed by null columns.

The maximum h-loglihood estimates of the fixed and random effects $\boldsymbol{\delta} = (\boldsymbol{\beta}^T, \boldsymbol{w}^T)^T$ are the solutions of the system

$$
\begin{cases} \partial h / \partial \boldsymbol{\beta} = \boldsymbol{0} \\ \partial h / \partial \boldsymbol{w} = \boldsymbol{0}, \end{cases}
$$

that can be solved by the Iterative Weighted Least Squares (IWLS) algorithm. The updating step of the algorithm is

$$
\boldsymbol{T}^T \boldsymbol{W}_a \boldsymbol{T} \boldsymbol{\delta} = \boldsymbol{T}^T \boldsymbol{W}_a \boldsymbol{z}_a, \tag{2.9}
$$

where $\boldsymbol{T}$ is the matrix in (2.8), $\boldsymbol{W}_a$ is a diagonal block matrix where the three non-null blocks are

$$\operatorname{diag}\left[\frac{\omega_{ij}}{\phi_{ij} g'(\mu_{ij})^2 V(\mu_{ij})}\right], \quad \operatorname{diag}\left[\frac{1}{\lambda_{U,i} g_U'(u_i)^2 V_U(u_i)}\right],$$
$$\operatorname{diag}\left[\frac{1}{\lambda_{V,j} g_V'(v_j)^2 V_V(v_j)}\right].$$

The dependent variables are defined by $\boldsymbol{z}_a = (\boldsymbol{z}^T, \boldsymbol{z}_U^T, \boldsymbol{z}_V^T)^T$, where the components of $\boldsymbol{z}$, $\boldsymbol{z}_U$, $\boldsymbol{z}_V$ are, respectively,

$$\begin{aligned}
z_{ij} &= \eta_{ij} + g'(\mu_{ij})(y_{ij} - \mu_{ij}), && i + j \leq t,\\
z_{U,i} &= w_{U,i} + g_U'(u_i)(\psi_{U,i} - u_i), && i = 0, \ldots, t,\\
z_{V,j} &= w_{V,j} + g_V'(v_j)(\psi_{V,j} - v_j), && j = 0, \ldots, t.
\end{aligned}$$

We remark that the inverse $\mathcal{I}(\hat{\boldsymbol{\delta}})^{-1}$ of the Fisher information matrix of the augmented GLM, evaluated at the estimate $\hat{\boldsymbol{\delta}}$, is an estimate of the variance–covariance matrix

$$\operatorname{var}\begin{bmatrix} \tilde{\boldsymbol{\beta}} \\ \tilde{\boldsymbol{w}} - \boldsymbol{W} \end{bmatrix},$$

where $\tilde{\boldsymbol{\beta}}$, $\tilde{\boldsymbol{w}}$ are the estimators of the fixed and random effects. Note that the estimator $\tilde{\boldsymbol{w}}$ of the parameter $\boldsymbol{w}$ is a predictor of the unobservable random vector $\boldsymbol{W}$.

In this way, by estimating the augmented GLM, we get estimates of the model parameters and of the standard errors of their estimators.

3. THE MODEL WITH STRUCTURED DISPERSION

In Section 2, we have assumed that the dispersion parameters $\boldsymbol{\phi}$, $\boldsymbol{\lambda}_U$ and $\boldsymbol{\lambda}_V$ are known. However, in practice, also these parameters should be estimated. As it has been pointed out by Smyth and Jørgensen (2002) and Boucher and Davidov (2011), sometimes models with constant dispersion could be inappropriate. Therefore, we assume a model that allows the μ_{ij} and also the dispersion parameters to vary depending on the values of the covariates. In addition to *(a1)–(a4)* in Section 2, here we assume

a5. *Structural assumptions for the dispersion parameters*
The dispersion parameters are given by

$$\begin{aligned}
\phi_{ij} &= g_\phi^{-1}(\boldsymbol{x}^{(\phi_{ij})T}\boldsymbol{\gamma}_\phi), && i + j \leq t,\\
\lambda_{U,i} &= g_{\lambda_U}^{-1}(\boldsymbol{x}^{(\lambda_{U,i})T}\boldsymbol{\gamma}_{\lambda_U}), && i = 0, \ldots, t,\\
\lambda_{V,j} &= g_{\lambda_V}^{-1}(\boldsymbol{x}^{(\lambda_{V,j})T}\boldsymbol{\gamma}_{\lambda_V}), && j = 0, \ldots, t,
\end{aligned}$$

where $\boldsymbol{x}^{(\phi_{ij})}$, $\boldsymbol{x}^{(\lambda_{U,i})}$, $\boldsymbol{x}^{(\lambda_{V,j})}$ are vectors of covariates and g_ϕ, g_{λ_U}, g_{λ_V} are link functions.

We note that in claims reserving, because of the scarcity of data, all the regression structures of the model must be chosen appropriately to avoid overparameterization.

When also the dispersion parameters have to be estimated, the normalizing functions in the h-loglihood (2.6) cannot be neglected anymore. However, an explicit form for such functions could not be available. Hence, to get a log-likelihood-based estimation of the dispersion parameters, some approximations are needed. By following the suggestion in Lee *et al.* (2006) for the *quasi-HGLMs*, we approximate the h-loglihood components concerning the random effects with the respective *Extended Quasi Likelihoods* (see Nelder and Pregibon, 1987). The approximation of h in (2.6), denoted by $\bar{h}$, is

$$
\begin{aligned}
&\bar{h}(\boldsymbol{\beta}, \boldsymbol{w}; \boldsymbol{\phi}, \boldsymbol{\lambda}_U, \boldsymbol{\lambda}_V; \boldsymbol{n}, \boldsymbol{y}, \boldsymbol{\psi}_U, \boldsymbol{\psi}_V, \boldsymbol{\omega}) \\
&\quad = \sum_{i,j:i+j\le t} \left\{ n_{ij} \log \frac{(\omega_{ij}/\phi_{ij})^{\alpha+1} y_{ij}^{\alpha}}{(p-1)^{\alpha}(2-p)} + \frac{\omega_{ij}}{\phi_{ij}} \left[y_{ij}\theta_{ij} - b_p(\theta_{ij}) \right] \right\} \\
&\qquad - \frac{1}{2} \sum_{i=0}^{t} \left\{ \frac{d_{U,i}}{\lambda_{U,i}} + \log \left[2\pi \lambda_{U,i} V_U(\psi_{U,i}) \right] \right\} \\
&\qquad - \frac{1}{2} \sum_{j=0}^{t} \left\{ \frac{d_{V,j}}{\lambda_{V,j}} + \log \left[2\pi \lambda_{V,j} V_V(\psi_{V,j}) \right] \right\},
\end{aligned}
\tag{3.1}
$$

where

$$
d_{U,i} = -2 \int_{\psi_{U,i}}^{u_i} \frac{\psi_{U,i} - s}{V_U(s)} ds, \qquad d_{V,j} = -2 \int_{\psi_{V,j}}^{v_j} \frac{\psi_{V,j} - s}{V_V(s)} ds,
$$

are the deviances of the GLM components related to U_i and V_j, respectively.

The parameter estimates are obtained by maximizing the function $\bar{h}$.

The derivatives of $\bar{h}$ with respect to the fixed and random effects $(\boldsymbol{\beta}, \boldsymbol{w})$ are equal to the derivatives of the h-loglihood (2.7). The derivatives of $\bar{h}$ with respect to $\boldsymbol{\gamma}_{\lambda_U}$ and $\boldsymbol{\gamma}_{\lambda_V}$ can be seen as the derivatives of the log-likelihoods of two GLMs with gamma distributed responses. If we properly define weights and responses, also the derivatives of $\bar{h}$ with respect to the regression parameters $\boldsymbol{\gamma}_\phi$ can be seen, formally, as the log-likelihood derivatives of a GLM. Therefore, the parameter estimates can be found through an algorithm in which four interconnected GLMs are fitted iteratively, according to the following steps (see Appendix A for the details).

Step 1 Given the dispersion parameters $\boldsymbol{\phi}, \boldsymbol{\lambda}_U, \boldsymbol{\lambda}_V$, the fixed and random effects $\boldsymbol{\delta} = (\boldsymbol{\beta}^T, \boldsymbol{w}^T)^T$ can be estimated by means of an augmented GLM. The IWLS equations are in (2.9).

Step 2 Given $\boldsymbol{\delta}, \boldsymbol{\lambda}_U, \boldsymbol{\lambda}_V$, the parameters $\boldsymbol{\gamma}_\phi$, and hence $\boldsymbol{\phi}$, can be estimated by a proper GLM with gamma distributed responses.

Step 3 Given $\boldsymbol{\delta}, \boldsymbol{\phi}$, the parameters $\boldsymbol{\gamma}_{\lambda_U}, \boldsymbol{\gamma}_{\lambda_V}$, and hence $\boldsymbol{\lambda}_U, \boldsymbol{\lambda}_V$, can be estimated by the two proper GLMs with gamma distributed responses.

At convergence, we can compute the standard errors of $(\tilde{\boldsymbol{\beta}}^T, (\tilde{\boldsymbol{w}} - \boldsymbol{W})^T)^T$ by means of the inverse $\mathcal{I}(\hat{\boldsymbol{\delta}})^{-1}$ of the Fisher information matrix of the augmented GLM in Step 1.

However, the estimates of the dispersion parameters provided by the above algorithm could under-estimate the variances. As recommended in Lee *et al.* (2006), in order to reduce the bias in estimating the dispersion parameters, we consider the REML estimation by using the adjusted profile loglihood

$$p_{\boldsymbol{\beta},\boldsymbol{w}}(\bar{h}) = \left\{ \bar{h} - \frac{1}{2} \log \left[det \left(\frac{\mathcal{I}(\boldsymbol{\delta})}{2\pi} \right) \right] \right\} \Big|_{\boldsymbol{\delta}=\hat{\boldsymbol{\delta}}(\boldsymbol{\gamma}_\phi, \boldsymbol{\gamma}_{\lambda_U}, \boldsymbol{\gamma}_{\lambda_V})}, \tag{3.2}$$

where $\hat{\boldsymbol{\delta}}(\boldsymbol{\gamma}_\phi, \boldsymbol{\gamma}_{\lambda_U}, \boldsymbol{\gamma}_{\lambda_V})$ solves $\partial \bar{h}/\partial \boldsymbol{\beta} = \mathbf{0}$, $\partial \bar{h}/\partial \boldsymbol{w} = \mathbf{0}$, for fixed $\boldsymbol{\gamma}_\phi, \boldsymbol{\gamma}_{\lambda_U}, \boldsymbol{\gamma}_{\lambda_V}$; hence, it is the estimate of the parameter vector $\boldsymbol{\delta}$ of the augmented GLM in Step 1. The matrix $\mathcal{I}(\hat{\boldsymbol{\delta}}(\boldsymbol{\gamma}_\phi, \boldsymbol{\gamma}_{\lambda_U}, \boldsymbol{\gamma}_{\lambda_V})) = \boldsymbol{T}^T \hat{\boldsymbol{W}}_a \boldsymbol{T}$ is the Fisher information matrix of the same GLM (see Appendix A for the details).

In the fitting algorithm, the Steps 2 and 3 above have to be adapted accordingly.

Finally, we note that the estimation of the parameter p could be obtained by maximizing the adjusted profile loglihood with respect to the parameters p and $\boldsymbol{\phi}$.

4. Reserve prediction and prediction error

In claims reserving, we need to predict the outstanding claims and evaluate the quality of the prediction.

To tackle this issue, we restrict ourselves to considering the exposures dependent on the origin years only. Note that if the exposures also depend on the development years, forecast of the exposures in future cells are needed and an additional source of error arises.

Let

$$R_i = \sum_{i=t-i+1}^{t} P_{ij} = \sum_{i=t-i+1}^{t} \omega_i Y_{ij},$$

denote the outstanding claims of the origin year i, $i = 1, \ldots, t$, and

$$R = \sum_{i=1}^{t} R_i = \sum_{i,j:i+j>t} P_{ij} = \sum_{i,j:i+j>t} \omega_i Y_{ij}.$$

the total outstanding claims.

The conditional expectation of R, at time t, is

$$E(R|\mathcal{D}_t) = \sum_{i,j:i+j>t} \omega_i E(Y_{ij}|\mathcal{D}_t), \tag{4.1}$$

where $\mathcal{D}_t = \left\{(N_{ij}, Y_{ij}), i+j \leq t\right\}$.

By the tower property of the conditional expectation and the conditional independence of the Y_{ij}, given $(\boldsymbol{U}, \boldsymbol{V})$, we get

$$E(Y_{ij}|\mathcal{D}_t) = E\left[E(Y_{ij}|\mathcal{D}_t, \boldsymbol{U}, \boldsymbol{V})|\mathcal{D}_t\right] = E\left[g^{-1}(\boldsymbol{x}_{ij}^T\boldsymbol{\beta} + W_{U,i} + W_{V,j})|\mathcal{D}_t\right]. \tag{4.2}$$

Now, we assume that the parameter estimates $\hat{\boldsymbol{\delta}} = (\hat{\boldsymbol{\beta}}^T, \hat{\boldsymbol{w}}^T)^T$ and the corresponding estimators $\tilde{\boldsymbol{\delta}} = (\tilde{\boldsymbol{\beta}}^T, \tilde{\boldsymbol{w}}^T)^T$ provide estimates and estimators of the linear predictors $\boldsymbol{x}_{ij}^T\boldsymbol{\beta} + w_{U,i} + w_{V,j}$, also for $i + j > t$. Note that this does not allow the payment year to be considered as a categorical covariate.

As a predictor for Y_{ij}, we consider the following estimator of $E(Y_{ij}|\mathcal{D}_t)$:

$$\tilde{Y}_{ij} = g^{-1}(\boldsymbol{x}_{ij}^T\tilde{\boldsymbol{\beta}} + \tilde{w}_{U,i} + \tilde{w}_{V,j}).$$

The predictor of the total outstanding claims is

$$\tilde{R} = \sum_{i,j:i+j>t} \omega_i g^{-1}(\boldsymbol{x}_{ij}^T\tilde{\boldsymbol{\beta}} + \tilde{w}_{U,i} + \tilde{w}_{V,j}), \tag{4.3}$$

and the predicted value or the *total claims reserve estimate* is

$$\hat{R} = \sum_{i,j:i+j>t} \omega_i g^{-1}(\boldsymbol{x}_{ij}^T\hat{\boldsymbol{\beta}} + \hat{w}_{U,i} + \hat{w}_{V,j}). \tag{4.4}$$

As a measure of the prediction uncertainty, we use the *conditional mean square error of prediction* which is defined by

$$\mathrm{MSEP}_{R|\mathcal{D}_t}(\tilde{R}) = E\left[(R - \tilde{R})^2|\mathcal{D}_t\right]. \tag{4.5}$$

It takes account of the fluctuations of the outstanding claims around the predictor $\tilde{R}$. Since the predictor $\tilde{R}$ is $\mathcal{D}_t$-measurable, we get

$$\begin{aligned}\mathrm{MSEP}_{R|\mathcal{D}_t}(\tilde{R}) &= \mathrm{var}(R|\mathcal{D}_t) + E\left[\left(E(R|\mathcal{D}_t) - \tilde{R}\right)^2 |\mathcal{D}_t\right] \\ &= \mathrm{var}(R|\mathcal{D}_t) + \left(E(R|\mathcal{D}_t) - \tilde{R}\right)^2.\end{aligned} \tag{4.6}$$

By the tower property of the conditional expectation and the conditional independence of the Y_{ij}, given $(\boldsymbol{U}, \boldsymbol{V})$, the conditional MSEP (4.6) can be written as a sum of three terms as follows:

$$\mathrm{MSEP}_{R|\mathcal{D}_t}(\tilde{R}) = E[\mathrm{var}(R|\boldsymbol{U}, \boldsymbol{V})|\mathcal{D}_t] + \mathrm{var}[E(R|\boldsymbol{U}, \boldsymbol{V})|\mathcal{D}_t] + \left(E(R|\mathcal{D}_t) - \tilde{R}\right)^2. \tag{4.7}$$

In order to estimate the conditional MSEP, we approximate the three terms as in Gigante *et al.* (2013a) and we get the following estimates:

$$\begin{aligned}
\widehat{E}[\mathrm{var}(R|\boldsymbol{U}, \boldsymbol{V})|\mathcal{D}_t] &= \sum_{i,j:i+j>t} \frac{\hat{\phi}_{ij}}{\omega_i} V\left(g^{-1}(\boldsymbol{x}_{ij}^T\hat{\boldsymbol{\beta}} + \hat{w}_{U,i} + \hat{w}_{V,j})\right), \\
\widehat{\mathrm{var}}[E(R|\boldsymbol{U}, \boldsymbol{V})|\mathcal{D}_t] &= \left\{\boldsymbol{J}_r(\boldsymbol{w})\boldsymbol{H}_{22}^{-1}\boldsymbol{J}_r(\boldsymbol{w})^T\right\}\big|_{\hat{\boldsymbol{\delta}}}, \\
\left(\widehat{E(R|\mathcal{D}_t) - \tilde{R}}\right)^2 &= \left\{\boldsymbol{J}_f(\boldsymbol{\beta})\boldsymbol{G}^{-1}\boldsymbol{J}_f(\boldsymbol{\beta})^T\right\}\big|_{\hat{\boldsymbol{\delta}}},
\end{aligned} \tag{4.8}$$

where $\boldsymbol{J}_r$ and $\boldsymbol{J}_f$ denote the Jacobian matrices of the functions

$$\begin{aligned}
r(\boldsymbol{w}) &= \sum_{i,j:i+j>t} \omega_i g^{-1}(\boldsymbol{x}_{ij}^T\boldsymbol{\beta} + w_{U,i} + w_{V,i}), \\
f(\boldsymbol{\beta}) &= \sum_{i,j:i+j>t} \omega_i g^{-1}(\boldsymbol{x}_{ij}^T\boldsymbol{\beta} + \tilde{w}_{U,i}(\boldsymbol{\beta}) + \tilde{w}_{V,j}(\boldsymbol{\beta})),
\end{aligned}$$

with $\tilde{\boldsymbol{w}}(\boldsymbol{\beta})$ denoting the maximum h-loglihood estimator of $\boldsymbol{w}$ obtained for given $\boldsymbol{\beta}$. The Jacobian matrix of $\tilde{\boldsymbol{w}}(\boldsymbol{\beta})$ is given by $-\boldsymbol{H}_{22}^{-1}\boldsymbol{H}_{12}^T$ (Lee and Nelder, 1996, Appendix C; Lee and Ha, 2010). The matrices $\boldsymbol{H}_{22}^{-1}$, $\boldsymbol{G}^{-1}$ and $\boldsymbol{H}_{12}^T$ are obtained from the Fisher information matrix of the augmented GLM and its inverse

$$\mathcal{I}(\boldsymbol{\delta}) = \begin{bmatrix} \boldsymbol{H}_{11} & \boldsymbol{H}_{12} \\ \boldsymbol{H}_{12}^T & \boldsymbol{H}_{22} \end{bmatrix}, \qquad \mathcal{I}(\boldsymbol{\delta})^{-1} = \begin{bmatrix} \boldsymbol{G}^{-1} & \boldsymbol{F} \\ \boldsymbol{F}^T & \boldsymbol{C} \end{bmatrix}, \tag{4.9}$$

where $\boldsymbol{H}_{11}$ denotes the block of the derivatives of the h-loglihood taken both with respect to $\boldsymbol{\beta}$, $\boldsymbol{H}_{22}$ denotes the block of the derivatives taken both with respect to $\boldsymbol{w}$ and $\boldsymbol{H}_{12}$ the block of the mixed derivatives. The estimates in (4.8) can be easily obtained by matrix calculus, once the parameter estimates and the Fisher information matrix at the estimates are available.

The estimate of the conditional MSEP (4.6) takes account of the variability in the estimates of both the regression parameters $\boldsymbol{\beta}$ and the random effects $\boldsymbol{w}$. However, it does not allow for the variability in the dispersion parameter estimates, $\hat{\boldsymbol{\gamma}}_{\phi}$, $\hat{\boldsymbol{\gamma}}_{\lambda_U}$ and $\hat{\boldsymbol{\gamma}}_{\lambda_V}$. An insight into this aspect can be obtained from the standard errors estimated through the Fisher information matrices of the GLMs used to estimate such parameters: high standard errors with respect to the parameter estimates could indicate low accuracy and reliability of the estimates.

Finally, we note that the conditional MSEP has been estimated by taking first-order approximations in the three terms in (4.7) and plugging the parameter estimates into the formula. The estimator is biased, as arises in similar cases (see e.g. Booth and Hobert, 1998, Maiti *et al.*, 2014). A better estimator could be obtained by introducing correction terms or resorting to simulation. However, developing a correction term could be a difficult task. On the other hand, simulation requires repeated estimation of the model parameters on the basis of re-sampled data and this is often computationally demanding in mixture models. The advantage of the approximations in (4.8) is to have a simple formula to evaluate the quality of the predictions, that takes into account both the process and the estimation error. In Appendix B, we have reported the results of a simulation study in order to appreciate the effect of the approximations.

5. Numerical results

We illustrate the model by applying it to two data sets.

Example 1. The first data set concerns the claims development figures of a Swiss Motor Insurance portfolio, given in Wüthrich (2003). The same data have been used by Boucher and Davidov (2011) to estimate a Tweedie's compound Poisson model where, in addition to the means of the payments, also the dispersion parameters follow a regression structure, through a DGLM. Note that in our mixture model, conditionally on the risk parameters, we assume the same hypotheses, so that the results can be compared.

The observations consist of the number of payments and the incremental payments for nine accident years and eleven development years. The number of reported claims for accident year i, approximated by the number of claims reported in the first two development years of the same accident year, is assumed as exposure measure w_i.

We have pointed out in the introduction that in some situations, in particular, when the observed development figures within a given run-off triangle are scarce or fluctuate, the actuaries often rely on industry-wide development patterns or on development patterns of other similar lines of business. Similarly, when a business line of a portfolio consists of different business units, the actuary typically sets up reserves for each business unit, but he could also rely on the claims development figures of other business units (see Gisler and Wüthrich, 2008).

The model presented in this paper is suitable to be applied in such situations. To illustrate this aspect, we assume that, in addition to the specific claims data described above, we have some external information on the CL development factors, $f_0, \ldots, f_{t-1}$, of similar business units, obtained from industry-wide or company data. Such external development factors are reported in the first column in Table 1 and they are compared with the CL development factors $f_0^P, \ldots, f_{t-1}^P$, obtained from the payment data of the specific portfolio under consideration. As it is well known, the CL development factors can be related to the proportions, $r_0, \ldots, r_t$, of the expected ultimate claim amounts, $E(C_{it})$, settled in the different development years: $E(P_{ij}) = E(C_{it})r_j$. We note that the

TABLE 1

CL DEVELOPMENT FACTORS AND PROPORTIONS.

Development Year	External Data		Portfolio Data	
j	f_j	r_j	f_j^P	r_j^P
0	1.300	0.731211	1.327720	0.711932
1	1.020	0.219363	1.030096	0.233315
2	1.010	0.019011	1.010728	0.028449
3	1.010	0.009696	1.007601	0.010446
4	1.004	0.009793	1.002959	0.007481
5	1.003	0.003956	1.002012	0.002935
6	1.001	0.002979	1.001912	0.002001
7	1.001	0.000996	1.000766	0.001905
8	1.001	0.000997	1.000761	0.000765
9	1.001	0.000998	1.000012	0.000761
10		0.000999		0.000012

$r_0^P, \ldots, r_t^P$, evaluated for the specific portfolio show a lower proportion of the amount paid in the first development year and higher proportions settled in the following three development years, in particular in development year 2; the proportions are lower in all the following development years, except in development year 7.

For such data, we assume the mixture model with the following specifications.

In $(a2)$, the conditional distribution of $(N_{ij}, Y_{ij})|(U_i, V_j) = (u_i, v_j)$ has density (2.1), with $p = 1.7981$. For the sake of comparison, the value of the parameter p is the same as that of the Model IV in Boucher and Davidov (2011). Some comments on the effect of different choices for this parameter are reported at the end of the example.

In $(a3)$, the link function g and the functions g_U, g_V, that transform the risk parameters, are the logarithm. As fixed effect, we consider only a base level μ. Hence, we have

$$E[Y_{ij}|(U_i, V_j)] = \exp(\mu + \log(U_i) + \log(V_j)) = e^{\mu} U_i V_j.$$

We note that, the regression structure is as in Boucher and Davidov (2011), but, instead of the fixed effects related to the origin and development years, here we have random effects.

In $(a4)$, the distributions of $W_{U,i} = \log(U_i)$ and $W_{V,j} = \log(V_j)$ are conjugate of the Poisson EDF, that is

$$f_{W_{U,i}}(w) = \exp\left\{\frac{1}{\lambda_{U,i}}\left(\psi_{U,i} w - \exp(w)\right)\right\} c_U(\psi_{U,i}, \lambda_{U,i}), \qquad (5.1)$$

TABLE 2

MODEL ESTIMATES.

i	$\hat{u}_i$	j	$\hat{v}_j$	Mixture Coefficient	$\hat{\phi}_j$
0	0.984081	0	0.712510	0.970038	240
1	0.996163	1	0.232338	0.929956	402
2	1.016670	2	0.027861	0.937781	2,301
3	1.008071	3	0.010372	0.901602	6,374
4	1.004320	4	0.007821	0.852800	14,598
5	0.997471	5	0.003067	0.870612	23,842
6	1.003331	6	0.002211	0.785431	47,064
7	0.990424	7	0.001657	0.727510	62,122
8	0.999469	8	0.000816	0.779175	79,357
		9	0.000848	0.630315	113,219
		10	0.000499	0.506797	113,219

$$f_{W_{V,j}}(w) = \exp\left\{\frac{1}{\lambda_{V,j}}\left(\psi_{V,j}w - \exp(w)\right)\right\} c_V(\psi_{V,j}, \lambda_{V,j}). \tag{5.2}$$

Hence, U_i and V_j are gamma distributed.

We assume $\psi_{U,i} = E(U_i) = 1$ and $\psi_{V,j} = E(V_j) = r_j$ in Table 1, so that the external information is incorporated into the model. It follows that the expected values of the unconditional standardized payments are $E(Y_{ij}) = \exp(\mu)r_j$ and the expected incremental payments are $E(P_{ij}) = \exp(\mu)\omega_i r_j$. Hence, according to the usual parameter interpretation, the expected ultimate claim amount of origin year i is assumed to be proportional to the exposure measure ω_i and the proportion of such amount paid in development year j is given by the external proportion r_j. Note that these are initial estimates, that do not take account of the specific payment data.

As in Boucher and Davidov (2011), we assume for the dispersion parameters related to the standardized payments the structure $\phi_{ij} = \phi_j = \exp(\gamma_{\phi,j})$, i.e. we suppose only a development year effect. Moreover, the parameters of the last two development years are assumed to be equal. The dispersion parameters related to the risk parameters are assumed constant with i and j, respectively, i.e. $\lambda_{U,i} = \lambda_U$, $\lambda_{V,j} = \lambda_V$.

Now, we come to the model estimate (for implementation, we have developed our own code in SAS).

The estimate of $\exp(\mu)$ is 254.62, that can be interpreted as the expected ultimate claim amount per unit of exposure.

In Table 2 are reported the estimates of the risk parameters $\hat{u}_i = \exp(\hat{w}_{U,i})$ and $\hat{v}_j = \exp(\hat{w}_{V,j})$. Note that such estimates can be interpreted as the expected risk parameters updated through the observations in the run-off triangles. As for $\hat{u}_i$, the estimate is close to the initial estimate $\psi_{U,i} = 1$. Hence, the origin

TABLE 3

RESERVE AND PREDICTION ERROR ESTIMATES.

	Current Model		Boucher–Davidov	
Origin Year	Reserve	Prediction Error	Reserve	Prediction Error
1	13,961	21,313	325	800
2	36,755	37,916	21,357	29,928
3	56,673	44,275	40,205	37,325
4	96,846	58,461	87,224	58,549
5	155,421	73,602	138,317	71,041
6	220,232	81,908	202,512	79,777
7	393,922	107,005	359,344	101,663
8	621,890	122,057	596,578	119,780
Total	1,595,700	268,149	1,445,862	258,926

year effect is well explained by the exposure measure ω_i. The estimate $\hat{v}_j$ can be interpreted as the updated proportion of the ultimate claim amount to be paid in development year j. It is interesting to compare such estimates with the initial estimates $\psi_{V,j} = r_j$. The updated proportions can be seen as mixtures of the external proportions r_j and the portfolio proportions r_j^P; the mixture coefficients are reported in Table 2. The weights assigned to the portfolio proportions are always greater than those assigned to the external proportions; they are particularly high in the earlier development years for which we have more data. However, the weights are not monotonically decreasing. Note in particular the weights for the development years 2 and 7, that show remarkably high differences between the external and the portfolio proportions.

The last column in Table 2 shows the estimates of the dispersion parameters ϕ_j; they are very close to the estimates obtained by Boucher and Davidov (2011). As expected, the dispersion parameters related to the standardized payments are increasing in the development year. Remark that such estimates are substantially different and they show the importance of modeling the dispersion. Actually, for these data, a constant dispersion model could be inappropriate.

As for the estimates of the dispersion parameters related to the risk parameters, we get $\hat{\lambda}_U = 0.000274$ and $\hat{\lambda}_V = 0.000781$. Note that, in the current model, we have $\lambda_U = \text{var}(U_i)/E(U_i)$ and $\lambda_V = \text{var}(V_j)/E(V_j)$. It follows that, whereas the estimates of the coefficients of variation of the risk parameters U_i, $(\lambda_U/E(U_i))^{1/2}$, are low (about 1.7%), the estimates of the coefficients of variation of the risk parameters V_j are rather high for $j \geq 3$ (greater than 25%). Therefore, a model with random effects, at least for the development years, seems suitable for such data. Moreover, the inclusion of random effects is useful, due to the possibility of incorporating external information into the model.

The estimates of the reserves and of the prediction errors, given by the square roots of the MSEPs (4.6), are reported in Table 3. We note that, as

TABLE 4

STANDARD ERRORS FOR THE REGRESSION PARAMETERS OF THE DISPERSION PARAMETERS.

Parameter	Estimate	Standard Error
$\gamma_{\phi,0}$	5.480954	0.003861
$\gamma_{\phi,1}$	5.996669	0.005596
$\gamma_{\phi,2}$	7.740931	0.016581
$\gamma_{\phi,3}$	8.759934	0.032124
$\gamma_{\phi,4}$	9.588662	0.053120
$\gamma_{\phi,5}$	10.079189	0.080083
$\gamma_{\phi,6}$	10.759273	0.127157
$\gamma_{\phi,7}$	11.036859	0.167146
$\gamma_{\phi,8}$	11.281706	0.233496
$\gamma_{\phi,9} = \gamma_{\phi,10}$	11.637080	0.282901
γ_{λ_U}	−8.203300	0.860619
γ_{λ_V}	−7.155162	0.504638

usual, there is considerable uncertainty in the reserve estimates in the earlier origin years and then the relative prediction errors decrease. The prediction error for the whole reserve as a percentage of the claims reserve is about 16.8%. We remark that the conditional MSEP estimate allows for the variability in the estimates of the regression parameter μ and the random effects $\boldsymbol{w} = (w_{U,0}, \ldots, w_{U,t}, w_{V,0}, \ldots, w_{V,t})$, but not in the dispersion parameter estimates. The uncertainty in such estimates can be provided by the standard errors of the respective parameter estimates reported in Table 4. The relative standard errors of the parameters $\gamma_{\phi,j}$ are rather low in particular for the earlier development years for which more data are available. The relative standard errors of the parameters γ_{λ_U} and γ_{λ_V} are higher, hence the estimates are less precise.

For a comparison, we have reported in Table 3 the reserves and the prediction errors estimated by the DGLM Model IV in Boucher and Davidov (2011). Our reserves are higher than those in the quoted paper, for all of the origin years. The total reserve is 10% higher. In particular, note the reserve for origin year 1. The high value of this reserve is mainly determined by the estimate of the risk parameter V_{10}, $\hat{v}_{10} = 0.000499$. In fact, such estimate takes account of the external estimate $r_{10} = 0.000999$ which is quite higher than the portfolio estimate $r_{10}^{P} = 0.000012$ (see Table 1). The total prediction error is higher in our model, whereas the prediction error as a percentage of the claims reserve (16.8%) is lower than in the DGLM (17.9%). Notice that we estimate two more parameters, but we use more data, i.e. the external data in addition to the run-off data.

The differences in the reserve estimates are mainly caused by the effect of the external estimates. To appreciate this aspect, we have evaluated the reserve and the prediction error estimates starting from other two different sets of external values for the proportions of the expected ultimate claim amount to be paid in

TABLE 5

EXTERNAL AND ESTIMATED PROPORTIONS.

	Worst case			Best case		
Development Year j	Link Ratio f_j	External Proportion r_j	Estimated Proportion $\hat{\upsilon}_j$	Link Ratio f_j	External Proportion r_j	Estimated Proportion $\hat{\upsilon}_j$
0	1.417151	0.647315	0.710065	1.297105	0.745633	0.713474
1	1.041802	0.270028	0.233982	1.022409	0.221531	0.233162
2	1.015575	0.038347	0.028617	1.005731	0.021673	0.028253
3	1.014566	0.014885	0.010555	1.002547	0.005667	0.010259
4	1.004352	0.014137	0.007824	1.001381	0.002533	0.007126
5	1.004539	0.004286	0.002992	1.000595	0.001377	0.002830
6	1.004060	0.004490	0.002194	1.000353	0.000594	0.001848
7	1.001534	0.004034	0.002160	1.000046	0.000353	0.001668
8	1.000936	0.001531	0.000830	1.000582	0.000046	0.000666
9	1.000012	0.000936	0.000769	1.000012	0.000581	0.000703
10		0.000012	0.000011		0.000012	0.000011

the different development years. Just to exemplify, such proportions have been obtained by the "worst case" and "best case" link ratios, i.e. the highest and lowest ratios of successive cumulative payments in the specific portfolio, for any development year. The link ratios, the external estimates, the updated estimates that take account of the portfolio data are reported in Table 5; the reserves and the prediction errors are reported in Table 6.

Note that, with respect to the reserve estimates in Table 3, the estimates in Table 6 are closer to the reserves of Boucher and Davidov (2011), for the earlier origin years. In fact, the worst and best external proportions for such origin years are more in line with the portfolio estimates r_j^P in Table 1. The total reserve estimated in the quoted paper is intermediate between the worst and the best case estimates in Table 6. The same happens for the prediction errors.

Finally, we discuss the effect of different choices of the parameter p by using again the external data in Table 1. Preliminarily, we point out that for values of p close to 2, the convergence of the estimation algorithm is problematic, since the variance of the risk parameters for the origin years becomes almost zero. Possibly, in these cases, a large amount of the variability is captured by the variance of the conditional standardized payments and the risk parameters for the origin years could be omitted. By evaluating the adjusted profile likelihood (3.2), we find that it is increasing with p. In order to make some sensitivity analysis, we report in Table 7 the estimates of the parameters ϕ_j, γ_{λ_U} and γ_{λ_V}, the total reserve and the square root of the MSEP for some large values of the parameter p. We observe that the pattern of the estimates of the ϕ_j is preserved, whereas the values are quite different for the different values of p. The estimates of γ_{λ_V} remain stable, whereas the estimates of γ_{λ_U} are decreasing and imply that the

TABLE 6

RESERVE AND PREDICTION ERROR ESTIMATES.

Origin Year	Worst case Reserve	Worst case Prediction Error	Best case Reserve	Best case Prediction Error
1	318	778	318	776
2	21,274	28,897	19,447	26,320
3	42,157	37,878	36,085	32,946
4	95,496	62,629	77,135	51,281
5	153,576	77,124	125,932	64,052
6	216,606	84,513	186,698	72,564
7	390,315	109,125	345,093	96,505
8	622,368	124,398	571,100	112,798
Total	1,542,108	268,188	1,361,808	231,682

TABLE 7

MODEL ESTIMATES BY VARYING p.

p	1.8	1.85	1.865
$\hat{\phi}_0$	240	247	254
$\hat{\phi}_1$	403	438	458
$\hat{\phi}_2$	2,314	2,795	3,014
$\hat{\phi}_3$	6,422	8,155	8,929
$\hat{\phi}_4$	14,719	19,005	20,914
$\hat{\phi}_5$	24,082	32,588	36,370
$\hat{\phi}_6$	47,573	65,630	73,668
$\hat{\phi}_7$	62,799	86,831	97,528
$\hat{\phi}_8$	80,363	116,342	132,472
$\hat{\phi}_9$	114,789	170,768	195,763
$\hat{\phi}_{10}$	114,789	170,768	195,763
γ_{λ_U}	−8.220515	−9.172940	−10.395696
γ_{λ_V}	−7.156065	−7.175111	−7.179398
Reserve	1,597,066	1,637,210	1,651,221
Prediction Error	269,545	313,871	331,016

variance of the risk parameters related to the origin years becomes lower and lower. The estimated reserves show a moderate increase with p; more remarkable is the effect on the prediction errors.

Example 2. As a second example, we apply the model to a set of data from Australian Auto Bodily Injury claims. As a difference to the first example, such data come from a longer tailed line of business. The data are taken from the Appendix B in Taylor (2000), for origin years 1979 to 1995 and development

TABLE 8

MODEL ESTIMATES.

i	$\hat{u}_i$	j	r_j^P	r_j	$\hat{v}_j$	$\hat{\phi}_j$
0	0.998072	0	0.040078	0.015	0.036639	39,376
1	1.010673	1	0.088686	0.070	0.086803	9,252
2	1.055404	2	0.114870	0.120	0.114909	9,088
3	1.056602	3	0.153609	0.150	0.152054	14,622
4	1.009720	4	0.143157	0.160	0.146702	21,069
5	1.063330	5	0.127131	0.150	0.134887	27,600
6	0.989020	6	0.111240	0.120	0.113858	40,116
7	0.991417	7	0.069049	0.070	0.068578	53,009
8	0.960769	8	0.053155	0.050	0.051091	78,191
9	0.936656	9	0.035340	0.035	0.034624	99,404
10	0.952385	10	0.028272	0.025	0.025806	193,134
11	0.950014	11	0.012658	0.015	0.013813	212,893
12	0.977967	12	0.008294	0.008	0.007871	335,916
13	1.033009	13	0.009028	0.007	0.007403	419,447
14	1.004634	14	0.005432	0.005	0.004962	419,447
15	1.004855					
16	1.005472					

years 0 to 14 (the last development year refers to the claims with development year 14 and later). For the incremental payments, we have taken the data from Table B.3.3 rounded to the thousand; For the claim numbers, we have used the numbers of claims finalized (Table B.3.7) and for the exposure measures the vehicle years in Table B.1.

For the external data, as a possible market development pattern, we have used the proportions r_j, reported in Table 8. In the same table, the proportions r_j^P have been obtained by the CL link ratios.

The model specifications are the same as in *Example 1* except for the parameter p of the Tweedie distribution that is $p = 1.9$. Such value has been obtained by maximizing the adjusted profile likelihood (3.2) on a set of finite points in the interval]1,2[.

As for the estimates of the dispersion parameters related to the risk parameters, now we get $\hat{\lambda}_U = 0.005405$ and $\hat{\lambda}_V = 0.002998$. The coefficients of variation of the risk parameters U_i (about 7%) are higher than those in *Example 1*, as one could expect due to the much more uncertainty involved in bodily injury claims. Along the development years, the coefficients of variation are high for $j = 0$ and $j \geq 9$, whereas they are lower for $1 \leq j \leq 6$ where most of the claim amounts are paid. The estimates of the risk parameters and of the dispersion parameters ϕ_j are reported in Table 8.

For the sake of comparison, the reserve and prediction error estimates in the mixture model and in a DGLM are reported in Table 9. The estimated reserves in the two models are rather close, although the mixture model takes account

TABLE 9

RESERVE AND PREDICTION ERROR ESTIMATES.

	Current Model		DGLM	
Origin Year	Reserve	Prediction Error	Reserve	Prediction Error
3	200,368	406,582	214,263	493,386
4	494,851	696,729	516,346	822,097
5	880,000	945,016	997,520	1,194,046
6	1,413,344	1,165,090	1,263,704	1,158,324
7	2,617,881	1,777,678	2,426,005	1,805,546
8	4,076,876	2,136,785	3,489,083	1,995,858
9	6,403,486	2,665,481	5,255,130	2,380,593
10	10,190,977	3,358,668	8,839,950	3,177,508
11	17,045,880	4,475,855	14,686,843	4,255,008
12	25,967,900	5,579,319	23,732,131	5,747,618
13	36,701,222	6,717,698	38,273,527	8,198,109
14	46,552,267	7,388,837	45,625,040	9,162,121
15	54,596,674	7,929,020	53,276,737	11,995,003
16	60,973,369	8,436,424	64,262,059	28,665,867
Total	268,115,094	21,749,162	262,858,337	38,306,064

of the external information, whereas the other one does not. On the contrary, the prediction errors are remarkably different even though the estimates of the dispersion parameters ϕ_j are very close in the two models, as happens in *Example 1*. We note that the main source of such difference can be ascribed to the last accident year. In fact, whereas the prediction errors of the mixture model are steadily increasing along the accident years, those of the DGLM are substantially increasing along the most recent accident years and the increment is particularly sharp in the last accident year. This is due, in part, because only one observation is available for the estimate of the last accident year parameter, and therefore a considerable amount of estimation uncertainty is involved. Note that the same does not happen in *Example 1* where for the last origin year more information is available. Possibly, when a random effect model is applied, part of such uncertainty could be smoothed, thanks to the random effects instead of fixed effects. Moreover, the mixture model incorporates also the external information and this could play a role when evaluating the prediction errors.

Finally, we note that, in order to make some comparisons, we have applied in Example 2 the same regression structure as in Example 1, which results in a multiplicative CL-type structure for the expected values of the conditional responses. However, the second dataset is more critical and, as pointed out by Taylor (2000), it provides an example for which the CL model is unsuitable. Though the comparisons with the results of the DGLM show that the inclusion of random effects and external information could mitigate the consequences of such inadequacy.

6. Conclusions

This paper addresses a stochastic model for the evaluation of claims reserves that extends the compound Poisson model by Smyth and Jørgensen (2002) and Boucher and Davidov (2011) in the direction of HGLMs. By exploiting the *h*-likelihood approach by Lee *et al.* (2006), in addition to the fixed effects, random effects related to the origin and/or the development years are introduced. Moreover, the evaluation of the prediction errors is enhanced by allowing for structured dispersions. The parameters of the distributions of the random effects can be used to incorporate some external information into the model. The numerical examples explore the consequences of different external information on the claims development pattern, and of different data sets. As expected, it results that, if the external data are rather different from the portfolio data, they can remarkably affect the reserve evaluation. On the contrary, if the external data are quite close to the portfolio data, the reserve and prediction error evaluations are in line with those obtained by using only the run-off data.

Acknowledgements

The authors would like to thank two anonymous referees for their valuable insights and helpful suggestions. Patrizia Gigante and Liviana Picech acknowledge partial support by the FRA2013 grant "Models for Risk Evaluation, Uncertainty Measurement and Non-Life Insurance Applications".

References

Antonio, K. and Beirlant, J. (2007) Actuarial statistics with generalized linear mixed models. *Insurance: Mathematics and Economics,* **40**(1), 58–76.

Antonio, K., Beirlant, J., Hoedemakers, T. and Verlaak, R. (2006) Lognormal mixed models for reported claims reserves. *North American Actuarial Journal,* **10**(1), 30–48.

Booth, J.G. and Hobert, J.P. (1998) Standard errors of prediction in generalized linear mixed models. *Journal of the American Statistical Association,* **93**, 262–272.

Boucher, J.P. and Davidov, D. (2011) On the importance of dispersion modeling for claims reserving: An application with the Tweedie distribution. *Variance,* **5**(2), 158–172.

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

Bühlmann, H. and Moriconi, F. (2015) Credibility claims reserving with stochastic diagonal effects. *ASTIN Bulletin,* **45**(2), 309–353.

de Alba, E. (2002) Bayesian estimation of outstanding claim reserves. *North American Actuarial Journal,* **6**(4), 1–20.

England, P.D. and Verrall, R.J. (2002) Stochastic claims reserving in general insurance. *British Actuarial Journal,* **8**(3), 443–518.

ENGLAND, P.D. and VERRALL, R.J. (2006) Predictive distributions of outstanding liabilities in general insurance. *Annals of Actuarial Science,* **1**(2), 221–270.

ENGLAND P.D., VERRALL R.J. and WÜTHRICH, M.V. (2012). Bayesian over-dispersed Poisson model and the Bornhuetter & Ferguson claims reserving method. *Annals of Actuarial Science,* **6**(2), 258–283.

GIGANTE, P., PICECH, L. and SIGALOTTI, L. (2013a) Claims reserving in the hierarchical generalized linear models framework. *Insurance: Mathematics and Economics,* **52**, 381–390.

GIGANTE, P., PICECH, L. and SIGALOTTI, L. (2013b) Prediction error for credible claims reserves: an *h*-likelihood approach. *European Actuarial Journal,* **3**(2), 453–470.

GISLER, A. and WÜTHRICH, M. (2008) Credibility for the chain ladder reserving method. *ASTIN Bulletin,* **38**(2), 565–600.

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

JØRGENSEN, B. (1987) Exponential dispersion models. *Journal Royal Statistical Society* B, **49**, 127–162.

JØRGENSEN, B. (1997) *Theory of Dispersion Models*. London: Chapman & Hall.

JØRGENSEN, B. and DE SOUZA, M.C.P. (1994) Fitting Tweedie's compound Poisson model to insurance data. *Scandinavian Actuarial Journal*, **1994**(1), 69–93.

LEE, Y. and HA, I.D. (2010) Orthodox BLUP versus h-likelihood methods for inferences about random effects in Tweedie mixed models. *Statistics and Computing,* **20**(3), 295–303.

LEE, Y. and NELDER, J.A. (1996) Hierarchical generalized linear models (with Discussion). *Journal of the Royal Statistical Society* B, **58**, 619–678.

LEE, Y. and NELDER, J.A. (2001) Hierarchical generalized linear models: A synthesis of generalized linear models, random-effect models and structured dispersion. *Biometrika,* **88**(4), 987–1006.

LEE, Y., NELDER, J.A. and PAWITAN, Y. (2006) *Generalized Linear Models with Random Effects. Unified Analysis via h-Likelihood*. Boca Raton: Chapman and Hall/CRC.

MACK, T. (2000) Credible claims reserves: The Benktander method. *ASTIN Bulletin,* **30**(2), 333–347.

MAITI, T., REN, H. and SINHA S. (2014) Prediction error of small area predictors shrinking both means and variances. *Scandinavian Journal of Statistics. Theory and Applications,* **41**(3), 775–790.

MIRANDA-MARTINEZ, M.D., NIELSEN, B., NIELSEN, J.P. and VERRALL, R.J. (2011) Cash flow simulation for a model of outstanding liabilities based on claim amounts and claim numbers. *ASTIN Bulletin,* **41**(1), 107–129.

MIRANDA-MARTINEZ, M.D., NIELSEN, J.P. and VERRALL, R.J. (2012) Double chain ladder. *ASTIN Bulletin,* **42**(1), 59–76.

MIRANDA-MARTINEZ, M.D., NIELSEN, J.P. and WÜTHRICH, M.V. (2012) Statistical modelling and forcasting of outstanding liabilities in non-life insurance. *SORT,* **36**(2), 195–218.

NELDER, J.A. and LEE, Y. (1991). Generalised linear models for the analysis of Taguchi-type experiments. *Applied Stochastic Models and Data Analysis,* **7**, 107–120.

NELDER, J.A. and PREGIBON, D. (1987) An extended quasi-likelihood function. *Biometrika,* **74**, 221–232.

NELDER, J.A. and VERRALL, R.J. (1997) Credibility theory and generalized linear models. *ASTIN Bulletin,* **27**(1), 71–82.

NTZOUFRAS, I. and DELLAPORTAS, P. (2002) Bayesian modelling of outstanding liabilities incorporating claim count uncertainty. *North American Actuarial Journal,* **6**(1), 113–128.

OHLSSON, E. (2008) Combining generalized linear models and credibility models in practice. *Scandinavian Actuarial Journal,* **4**, 301–314.

OHLSSON, E. and JOHANSSON, B. (2006) Exact credibility and Tweedie models. *ASTIN Bulletin,* **36**(1), 121–133.

SALUZ, A. (2015) Prediction uncertainties in the Cape Cod reserving method. *Annals of Actuarial Science*, **9**(2), 239–263.

SALUZ, A., GISLER, A. and WÜTHRICH, M.V. (2011) Development pattern and prediction error for the stochastic Bornhuetter-Ferguson claims reserving method. *ASTIN Bulletin,* **41**(2), 279–313.

SMYTH, G.K. (1989) Generalized linear models with varying dispersion. *Journal of the Royal Statistical Society* B, **51**, 47–60.
SMYTH, G.K. and JØRGENSEN, B. (2002) Fitting Tweedie's compound Poisson model to insurance claims data: Dispersion modelling. *ASTIN Bulletin,* **32**(1), 143–157.
SMYTH, G.K. AND VERBYLA, A.P. (1999) Adjusted likelihood methods for modelling dispersion in generalized linear models. *Environments,* **10**, 696–709.
TAYLOR, G. (2000) *Loss Reserving. An Actuarial Perspective.* Boston: Kluwer Academic Publishers.
TAYLOR, G. (2015) Bayesian chain ladder models. *ASTIN Bulletin,* **45**, 75–99, doi:10.1017/asb.2014.25.
VERRALL, R.J. (2004) A Bayesian generalized linear model for the Bornhuetter-Ferguson method of claims reserving. *North American Actuarial Journal,* **8**(3), 67–89.
VERRALL, R.J. and ENGLAND, P.D. (2005) Incorporating expert opinion into a stochastic model for the chain-ladder technique. *Insurance: Mathematics and Economics,* **37**(2), 355–370.
VERRALL, R.J., NIELSEN, J.P. and JESSEN, A. (2010) Including count data in claims reserving. *ASTIN Bulletin,* **40**(2), 871–887.
WÜTHRICH, M.V. (2003) Claims reserving using Tweedie's compound Poisson model. *ASTIN Bulletin,* **33**(2), 331–346.
WÜTHRICH, M.V. and MERZ, M. (2008) *Stochastic Claims Reserving Methods in Insurance.* Chichester: Wiley.

PATRIZIA GIGANTE
Dipartimento di Scienze Economiche, Aziendali
Matematiche e Statistiche 'B. de Finetti'
University of Trieste
Piazzale Europa 1
34127 Trieste, Italy
E-Mail: patrizia.gigante@deams.units.it

LIVIANA PICECH (Corresponding author)
Dipartimento di Scienze Economiche, Aziendali
Matematiche e Statistiche 'B. de Finetti'
University of Trieste
Piazzale Europa 1
34127 Trieste, Italy
E-Mail: liviana.picech@deams.units.it

LUCIANO SIGALOTTI
Dipartimento di Scienze Economiche e Statistiche
University of Udine
Via Tomadini 30, 33100 Udine, Italy
E-Mail: luciano.sigalotti@uniud.it

APPENDIX A. TECHNICAL DETAILS

We report the technical details of the estimation procedure in Section 3.

A.1 *The derivatives of the approximate h-loglihood* $\bar{h}$

The derivatives of $\bar{h}$ (3.1) with respect to the fixed and random effects ($\boldsymbol{\beta}, \boldsymbol{w}$) are

$$\partial\bar{h}/\partial\beta_k = \sum_{i,j:i+j\leq t} x_{ij,k}\frac{\omega_{ij}}{\phi_{ij}}(y_{ij}-\mu_{ij})\frac{1}{g'(\mu_{ij})V(\mu_{ij})},$$

$$\partial\bar{h}/\partial w_{U,k} = \sum_{j=0}^{t-k}\frac{\omega_{kj}}{\phi_{kj}}(y_{kj}-\mu_{kj})\frac{1}{g'(\mu_{kj})V(\mu_{kj})} + \frac{1}{\lambda_{U,k}}(\psi_{U,k}-u_k)\frac{1}{g'_U(u_k)V_U(u_k)},$$

$$\partial\bar{h}/\partial w_{V,k} = \sum_{i=0}^{t-k}\frac{\omega_{ik}}{\phi_{ik}}(y_{ik}-\mu_{ik})\frac{1}{g'(\mu_{ik})V(\mu_{ik})} + \frac{1}{\lambda_{V,k}}(\psi_{V,k}-v_k)\frac{1}{g'_V(v_k)V_V(v_k)}.$$

They are equal to the derivatives of the h-loglihood (2.7).

The derivatives of $\bar{h}$ with respect to the regression parameters $\boldsymbol{\gamma}_\phi$ are

$$\partial\bar{h}/\partial\gamma_{\phi,k} = \sum_{i,j:i+j\leq t}\left\{-\frac{n_{ij}\phi_{ij}}{p-1} - \omega_{ij}\left[y_{ij}\theta_{ij} - b_p(\theta_{ij})\right]\right\}\frac{1}{\phi_{ij}^2 g'_\phi(\phi_{ij})}x_k^{(\phi_{ij})}. \tag{A.1}$$

Note that the observations n_{ij} of the payment numbers, that do not appear in the derivatives of $\bar{h}$ with respect to the fixed and random effects, intervene in (A.1). This additional information is important for improving the estimates of the dispersion parameters $\boldsymbol{\phi}$ related to the standardized payments.

Following Smyth and Jørgensen (2002), if we properly define weights and responses, the derivatives (A.1) can be seen, formally, as the log-likelihood derivatives of a GLM. In fact, if we let

$$\omega_{\phi,ij} = \frac{2\omega_{ij}\mu_{ij}^{2-p}}{(2-p)(p-1)\phi_{ij}}$$

and

$$\begin{aligned} d_{\phi,ij} &= -\frac{2}{\omega_{\phi,ij}}\left(\frac{n_{ij}\phi_{ij}}{p-1} + \omega_{ij}\left[y_{ij}\theta_{ij} - b_p(\theta_{ij})\right]\right) + \phi_{ij} \\ &= -\frac{2}{\omega_{\phi,ij}}\left(\frac{n_{ij}\phi_{ij}}{p-1} + \omega_{ij}\left[y_{ij}\frac{\mu_{ij}^{1-p}}{1-p} - \frac{\mu_{ij}^{2-p}}{2-p}\right]\right) + \phi_{ij}, \end{aligned}$$

we get

$$\partial\bar{h}/\partial\gamma_{\phi,k} = \sum_{i,j:i+j\leq t} x_k^{(\phi_{ij})}\frac{\omega_{\phi,ij}}{2}(d_{\phi,ij}-\phi_{ij})\frac{1}{g'_\phi(\phi_{ij})\phi_{ij}^2}. \tag{A.2}$$

If $\omega_{\phi,ij}$ and $d_{\phi,ij}$ are given, the derivatives (A.2) have the same expression as the derivatives of the log-likelihood of a GLM with gamma distributed responses, observations $d_{\phi,ij}$, covariates $\boldsymbol{x}^{(\phi_{ij})}$, weights $\omega_{\phi,ij}/2$ and link function g_ϕ.

The derivatives of $\bar{h}$ with respect to $\boldsymbol{\gamma}_{\lambda_U}$ and $\boldsymbol{\gamma}_{\lambda_V}$ are given by

$$\begin{aligned} \partial\bar{h}/\partial\gamma_{\lambda_U,k} &= \sum_{i=0}^{t} x_k^{(\lambda_{U,i})}\frac{1}{2}\left(d_{U,i}-\lambda_{U,i}\right)\frac{1}{g'_{\lambda_U}(\lambda_{U,i})\lambda_{U,i}^2}, \\ \partial\bar{h}/\partial\gamma_{\lambda_V,k} &= \sum_{j=0}^{t} x_k^{(\lambda_{V,j})}\frac{1}{2}\left(d_{V,j}-\lambda_{V,j}\right)\frac{1}{g'_{\lambda_V}(\lambda_{V,j})\lambda_{V,j}^2}, \end{aligned} \tag{A.3}$$

and, similarly to (A.2), they can be seen as the derivatives of the log-likelihoods of two GLMs with gamma distributed responses.

The parameter estimates are obtained by maximizing the function $\bar{h}$. The solutions can be found through an algorithm in which four interconnected GLMs are fitted iteratively, according to the following steps.

Step 1 Given the dispersion parameters $\boldsymbol{\phi}, \boldsymbol{\lambda}_U, \boldsymbol{\lambda}_V$, the fixed and random effects $\boldsymbol{\delta} = (\boldsymbol{\beta}^T, \boldsymbol{w}^T)^T$ can be estimated by means of an augmented GLM. The IWLS equations are in (2.9).

Step 2 Given $\boldsymbol{\delta}, \boldsymbol{\lambda}_U, \boldsymbol{\lambda}_V$, the parameters $\boldsymbol{\gamma}_\phi$, and hence $\boldsymbol{\phi}$, can be estimated by the GLM with score function in (A.2).

Step 3 Given $\boldsymbol{\delta}, \boldsymbol{\phi}$, the parameters $\boldsymbol{\gamma}_{\lambda_U}, \boldsymbol{\gamma}_{\lambda_V}$, and hence $\boldsymbol{\lambda}_U, \boldsymbol{\lambda}_V$, can be estimated by the two GLMs with score functions in (A.3).

At convergence, we can compute the standard errors of $(\tilde{\boldsymbol{\beta}}^T, (\tilde{\boldsymbol{w}} - \boldsymbol{W})^T)^T$ by means of the inverse $\mathcal{I}(\hat{\boldsymbol{\delta}})^{-1}$ of the Fisher information matrix of the augmented GLM in Step 1.

A.2 *The derivatives of the adjusted profile loglihood* $p_{\boldsymbol{\beta},\boldsymbol{w}}(\bar{h})$

Given the fixed and random effects, the derivatives of the adjusted profile loglihood (3.2) with respect to $\boldsymbol{\gamma}_\phi$, $\boldsymbol{\gamma}_{\lambda_U}$ and $\boldsymbol{\gamma}_{\lambda_V}$ give the following score equations

$$\sum_{i,j:i+j\le t} x_k^{(\phi_{ij})} \frac{\omega_{\phi,ij} - q_{ij}}{2} \left(d^*_{\phi,ij} - \phi_{ij}\right) \frac{1}{g'_\phi(\phi_{ij})\phi^2_{ij}} = 0, \tag{A.4}$$

$$\sum_{i=0}^{t} x_k^{(\lambda_{U,i})} \frac{1 - q_{U,i}}{2} \left(d^*_{U,i} - \lambda_{U,i}\right) \frac{1}{g'_{\lambda_U}(\lambda_{U,i})\lambda^2_{U,i}} = 0, \tag{A.5}$$

$$\sum_{j=0}^{t} x_k^{(\lambda_{V,j})} \frac{1 - q_{V,j}}{2} \left(d^*_{V,j} - \lambda_{V,j}\right) \frac{1}{g'_{\lambda_V}(\lambda_{V,i})\lambda^2_{V,j}} = 0, \tag{A.6}$$

where $d^*_{\phi,ij} = d_{\phi,ij}\omega_{\phi,ij}/(\omega_{\phi,ij} - q_{ij})$, $i + j \le t$, $d^*_{U,i} = d_{U,i}/(1 - q_{U,i})$, $i = 0, \dots, t$, $d^*_{V,j} = d_{V,j}/(1 - q_{V,j})$, $j = 0, \dots, t$ and q_{ij}, $q_{U,i}$ $q_{V,j}$ are the *leverages*, i.e. the diagonal elements of the hat matrix,

$$\hat{\boldsymbol{W}}_a^{1/2} \boldsymbol{T}(\boldsymbol{T}^T \hat{\boldsymbol{W}}_a \boldsymbol{T})^{-1} \boldsymbol{T}^T \hat{\boldsymbol{W}}_a^{1/2},$$

of the augmented GLM.

Note that (A.4) gives, formally, the score equations of a GLM for the response variables, say $D^*_{\phi,ij}$, with observed values $d^*_{\phi,ij} = d_{\phi,ij}\omega_{\phi,ij}/(\omega_{\phi,ij} - q_{ij})$, gamma distributed, with means and variances

$$E(D^*_{\phi,ij}) = \phi_{ij} = g_\phi^{-1}(\boldsymbol{x}^{(\phi_{ij})T}\boldsymbol{\gamma}_\phi), \qquad \mathrm{var}(D^*_{\phi,ij}) = \frac{2}{\omega_{\phi,ij} - q_{ij}}\phi^2_{ij},$$

where each $(\omega_{\phi,ij} - q_{ij})/2$ can be seen as a known weight, provided that it is positive; otherwise it is set equal to zero.

A similar argument applies to the score equations (A.5) and (A.6).

In the fitting algorithm the Steps 2 and 3 above have to be adapted accordingly.

APPENDIX B. SIMULATION STUDY

In order to empirically test the effect of the approximations in the evaluation of the conditional MSEP in Section 4, we have applied the estimation procedure to simulated data.

We have used the data of the first example in Section 5 as reference data. For the stochastic process

$$\{U_0, \ldots, U_8, V_0, \ldots, V_{10}, (N_{00}, Y_{00}), \ldots, (N_{8,10}, Y_{8,10})\},$$

we have assumed the distribution estimated in the example with expected values of the random effects U_i and V_j the estimates $\hat{u}_i$ and $\hat{v}_j$ in Table 2.

From this process, we have generated 20,000 (9×11) run-off tables of simulated pairs (n^*_{ij}, y^*_{ij}). The sum of the y^*_{ij} in the lower part of each table provides us with the simulated portfolio outstanding claims. Their average amounts to 1,599,167: very close to the reserve estimate in Table 3. Since our estimation algorithm applies only to positive incremental payments, we have kept only the tables with positive figures in the upper part, their number is $H = 15,637$. The corresponding simulated outstanding claims, $R\text{sim} = \frac{1}{H}\sum_{h=1}^{H} R\text{sim}_h$, where $R\text{sim}_h$ is the sum of the simulated payments in the lower table in the hth replication, amounts to 1,615,962.

Then, we have used the data in each upper table, (n^*_{ij}, y^*_{ij}), $i + j \leq 10$, to estimate the portfolio reserve and the conditional MSEP by using the estimators in Section 4 and the estimation algorithm with $\psi_{U_i} = 1$ and $\psi_{V_j} = r_j$ in Table 1, as in *Example 1*.

In Table 10, we report Rsim and the average of the predicted reserves $R\text{est} = \frac{1}{H}\sum_{h=1}^{H} \hat{R}_h$, where $\hat{R}_h$ is the reserve estimate in the hth replication. They are very close.

In the same table, RMSEPsim denotes the root of the MSEP estimated via simulation i.e. the average $\text{RMSEPsim} = (\frac{1}{H}\sum_{h=1}^{H}(R\text{sim}_h - \hat{R}_h)^2)^{1/2}$. The column RMSEPest is the root of the average of the estimated MSEPs, $\text{RMSEPest} = (\frac{1}{H}\sum_{h=1}^{H} \widehat{\text{MSEP}}_h)^{1/2}$, where $\widehat{\text{MSEP}}_h$ is the estimate of the conditional MSEP in the hth replication. The root mean square errors relative to the respective reserves are 19.15%, for the simulated reserve, and 19.95%, for the estimated reserve.

In order to quantify the performance of the MSEP estimator, we have also calculated the following statistics (see Maiti *et al.*, 2014):

- the empirical relative bias of the MSEP estimator,

$$RB = \frac{\text{MSEPest} - \text{MSEPsim}}{\text{MSEPsim}},$$

- the empirical relative root mean square error of the MSEP estimator,

$$\text{RRMSEP} = \frac{\left(\frac{1}{H}\sum_{h=1}^{H}(\widehat{\text{MSEP}}_h - \text{MSEPsim})^2\right)^{1/2}}{\text{MSEPsim}}.$$

TABLE 10

RESULTS OF THE SIMULATION STUDY

Rsim	Rest	RMSEPsim	RMSEPest	RB	RRMSEP
1,615,962	1,604,625	309,403	320,045	0.06997	0.62571

TABLE 11

SUMMARY STATISTICS FOR THE STANDARDIZED DIFFERENCES.

Mean	Std dev	Skew	$Q1$	Median	$Q3$
0.07876	1.02354	0.78902	−0.63749	−0.03048	0.67027

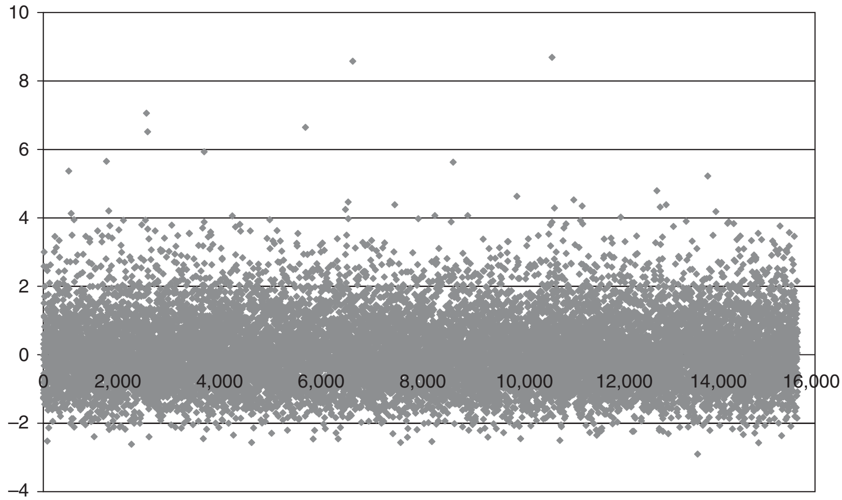

FIGURE 1: Standardized differences.

The relative bias, that quantifies the effect of the approximations in the estimator of the MSEP, is moderate, about 7%. However, it has to be remarked that the dispersion of the estimated errors measured by RRMSEP is quite high.

Finally, we have calculated the differences of simulated and estimated reserves, standardized with respect to the estimated root MSEP,

$$\frac{R\text{sim}_h - \hat{R}_h}{\widehat{\text{MSEP}}_h^{1/2}}.$$

The values are plotted in Figure 1 and some summary statistics are reported in Table 11. We note that the points spread around zero, show positive skew and almost all of them fall inside the interval [−3, +3]. However, a relatively small number of standardized differences are particularly high.