BAYESIAN CHAIN LADDER MODELS

BY

GREG TAYLOR

ABSTRACT

The literature on Bayesian chain ladder models is surveyed. Both Mack and cross-classified forms of the chain ladder are considered. Both cases are examined in the context of error terms distributed according to a member of the exponential dispersion family. Tweedie and over-dispersed Poisson errors follow as special cases. Bayesian cross-classified chain ladder models may randomise row, column or diagonal parameters. Column and diagonal randomisation has been largely absent from the literature until recently. The present paper allows randomisation of row and column parameters. The Bayes estimator, the linear Bayes (credibility) estimator, and the MAP estimator are shown to be identical in the Mack case, and in the cross-classified case provided that the error terms are Tweedie distributed. In the Mack case the variance structure is generalised considerably from the existing literature. In the cross-classified case the model structure differs somewhat from the existing literature, and a comparison is made between the two. MAP estimators for the cross-classified case are often given by implicit equations that require numerical solution. Recursive formulas are given for these in the general case of error terms from the exponential dispersion family. The connection between the cross-classified case and Bornhuetter-Ferguson prediction is explored.

KEYWORDS

Bayesian chain ladder, Bornhuetter-Ferguson, chain ladder, column effects, exponential dispersion family, Mack model, cross-classified model, over-dispersed Poisson, row effects, Tweedie family.

1. INTRODUCTION

The chain ladder loss reserving algorithm has existed for many years. Stochastic models that yield this algorithm as forecasts have also existed for some time. Hachemeister and Stanard (1975) described a parametric model, and Mack

Astin Bulletin 45(1), 75-99. doi: 10.1017/asb.2014.25 © 2014 by Astin Bulletin. All rights reserved.

https://doi.org/10.1017/asb.2014.25 Published online by Cambridge University Press

(1993) a non-parametric one. A rather different formulation, a log normal parametric model, was given by Hertig (1985).

In more recent years there has been some emphasis on identifying the family of models that yield this algorithm. Taylor (2011) considered two categories of model, generically referred to as Mack models and cross-classified models. Each category was investigated as to whether its maximum likelihood estimators (**"MLEs"**) of parameters coincided with the algorithm in the case that all observations in the claim triangle were subject to a distribution from the exponential dispersion family (**"EDF"**). The **Tweedie** and over-dispersed Poisson (**"ODP"**) sub-families of the EDF were also considered.

All of these models were **fixed effect models**, i.e. their parameters were assumed to be fixed but unknown real numbers. However, claim data sets are often encountered in practice in which, with statistical significance, parameters do not remain stationary over time, e.g. chain ladder age-to-age factors may display significant differences as between accident years. One means of accounting for this is to regard the model parameters as subject to random variation over time.

The literature of recent years contains a number of **Bayesian models**, in which some or all of the model parameters are assumed subject to a prior distribution (Verrall, 2000, 2004; England and Verrall, 2002; Gisler and Müller, 2007; Wüthrich, 2007; Gisler and Wüthrich, 2008; Wüthrich and Merz, 2008; England *et al.*, 2012).

The purpose of the present paper is to consider these Bayesian models and the degree to which they may be extended. It is noteworthy that, while the cross-classified models include both row and column effects, the column effects have been randomised in only a limited way in the literature on Bayesian models.

2. Framework and Notation

2.1. Claims data

Consider a $K \times J$ rectangle of claims observations Y_{kj} with:

- accident periods represented by rows and labelled $k = 1, 2, \ldots, K$;
- development periods represented by columns and labelled by $j = 1, 2, \ldots, J \leq K$.

Within the rectangle identify a **development trapezoid** of **past** observations

$$\mathfrak{D}_K = \left\{ Y_{kj} : 1 \leq k \leq K \text{ and } 1 \leq j \leq \min(J, K-k+1) \right\}$$

The complement of this subset, representing **future** observations is

$$\begin{aligned} \mathfrak{D}_K^c &= \left\{ Y_{kj} : 1 \leq k \leq K \text{ and } \min\ (J, K-k+1) < j \leq J \right\} \\ &= \left\{ Y_{kj} : K-J+1 < k \leq K \text{ and } K-k+1 < j \leq J \right\} \end{aligned}$$

https://doi.org/10.1017/asb.2014.25 Published online by Cambridge University Press

Also let

$$\mathfrak{D}_K^+ = \mathfrak{D}_K \cup \mathfrak{D}_K^c$$

In general, the problem is to predict $\mathfrak{D}_K^c$ on the basis of observed $\mathfrak{D}_K$.

The usual case in the literature (though often not in practice) is that in which $J = K$, so that the trapezoid becomes a triangle. The more general trapezoid will be retained throughout the present paper.

Define the **cumulative row sums**

$$X_{kj} = \sum_{i=1}^{j} Y_{ki} \tag{2.1}$$

Let $\Sigma_{j\in\mathcal{R}(k)}$ denote summation over the entire row k of $\mathfrak{D}_K$, i.e. $\Sigma_{j=1}^{\min(J,K-k+1)}$ for fixed k.

Similarly, let $\Sigma_{k\in\mathcal{C}(j)}$ denote summation over the entire column of $\mathfrak{D}_K$, i.e. $\Sigma_{k=1}^{K-j+1}$ for fixed j.

2.2. Families of distributions

2.2.1. *Exponential dispersion family.* The **exponential dispersion family** (Nelder and Wedderburn, 1972) consists of those variables Y with log-likelihoods of the form

$$\ell(y|\theta,\phi) = [y\theta - \kappa(\theta)]/a(\phi) + \lambda(y,\phi) \tag{2.2}$$

for parameters θ (canonical parameter) and ϕ (scale parameter) and suitable functions a, κ and λ, with a continuous, κ twice differentiable, and λ such as to produce a unit total probability mass. It will be further assumed in the following that the derivative $\kappa'(\theta)$ maps one-one onto the strictly positive half-line.
For Y so distributed,

$$E[Y] = \kappa'(\theta) \tag{2.3}$$

$$Var[Y] = a(\phi)\,\kappa''(\theta) \tag{2.4}$$

If $\mu(\theta)$ denotes $E[Y]$, then (2.3) establishes the relation between μ and θ, and so (2.4) may be expressed in the form

$$Var[Y] = a(\phi)\,V(\mu) \tag{2.5}$$

for some function V, referred to as the **variance function**.

Henceforth only the special case in which

$$a(\phi) = \phi \tag{2.6}$$

will be considered, and the notation $Y \sim EDF(\theta,\phi;\kappa,\lambda)$ will be used to mean that a random variable Y is subject to the EDF likelihood (2.2) and (2.6).

https://doi.org/10.1017/asb.2014.25 Published online by Cambridge University Press

TABLE 1

PROPERTIES OF TWEEDIE AND ODP FAMILIES.

Quantity	Tweedie	ODP
$\kappa(\theta)$	$(2-p)^{-1}[(1-p)\theta]^{(2-p)/(1-p)}$	$exp\theta$
$\mu(\theta) = E[Y]$	$[(1-p)\theta]^{1/(1-p)}$	$exp\theta$
$Var[Y]$	$\phi[\mu(\theta)]^p$	$\phi\mu(\theta)$
$\ell(y;\mu,\phi)$	$\left[y\mu^{1-p}/(1-p) - \mu^{2-p}/(2-p)\right]/\phi + \lambda(y,\phi)$	$[yln\mu - \mu]/\phi + \lambda(y,\phi)$
$\partial\ell/\partial\mu$	$\left(y\mu^{-p} - \mu^{1-p}\right)/\phi$	$(y-\mu)/\phi\mu$

The **Tweedie family** (Tweedie, 1984) is the sub-family of the EDF for which (2.6) holds and

$$V(\mu) = \mu^p,\ p \le 0\ or\ p \ge 1 \tag{2.7}$$

The **over-dispersed Poisson** family is the Tweedie sub-family with $p = 1$.

Table 1 displays major properties of these two sub-families. Some of the entries in the table do not exist when $p = 1$ or 2. Such entries need to be replaced by their limits as $p \to 1$ or 2.

The notation $Y \sim Tw(\mu, \phi; p, \lambda)$ will be used to mean that a random variable Y is subject to the Tweedie likelihood with parameters μ, ϕ, p and normalising function λ.

2.2.2. *Natural conjugate prior.* The natural conjugate prior log-likelihood associated with the EDF member (2.2) is (Landsman and Makov, 1998)

$$\ell(\theta; m, \psi) = \frac{[m\theta - \kappa(\theta)]}{\psi} + \lambda^*(m, \psi) \tag{2.8}$$

for specific constants m, ψ and functions $\kappa(.)$, $\lambda^*(.)$. If Θ denotes the domain of θ, it will be assumed that the likelihood $L(\theta; m, \psi) = exp\ell(\theta; m, \psi)$ and its derivative $\partial L/\partial\theta$ vanish on $\partial\Theta$. Let this distribution be denoted $EDF^*(m, \psi; \kappa, \lambda^*)$.

The natural conjugate prior to the Tweedie $Tw(\mu, \phi; p, \lambda)$ may be obtained by substituting for θ and $\kappa(\theta)$ in (2.8) according to Table 1, giving $Tw^*(m, \psi; p, \lambda^*)$ with log-likelihood

$$\ell(\mu^{1-p}/(1-p); m, \psi) = \left[\frac{m\mu^{1-p}}{1-p} - \frac{\mu^{2-p}}{2-p}\right]/\psi + \lambda^*(m, \psi) \tag{2.9}$$

One may calculate $E[\kappa'(\theta)]$ and $E[\kappa''(\theta)]$ from (2.8). This is done by integrating $\partial L/\partial\theta$ and $\partial^2 L/\partial\theta^2$ respectively over the support of θ, with due recognition of the above boundary condition. This yields (see e.g. Landsman and Makov, 1998)

$$E\left[\kappa'(\theta)\right] = m \tag{2.10}$$

$$E\left[\kappa''(\theta)\right] = Var\left[\kappa'(\theta)\right]/\psi \tag{2.11}$$

https://doi.org/10.1017/asb.2014.25 Published online by Cambridge University Press

Bayes estimation of θ

The above prior converts (2.2) into the joint log-likelihood

$$\ell(y, \theta; m, \phi, \psi) = (y\phi^{-1} + m\psi^{-1})\theta - (\phi^{-1} + \psi^{-1})\kappa(\theta) + \textit{normalising term} \tag{2.12}$$

where the normalising term is independent of θ.

Hence, the posterior log-likelihood takes the form

$$\ell(\theta|y; m, \phi, \psi) = (y\phi^{-1} + m\psi^{-1})\theta - (\phi^{-1} + \psi^{-1})\kappa(\theta) + \textit{normalising term} \tag{2.13}$$

If this is re-written as

$$\ell(\theta|y; m, \phi, \psi) = \left[(y\phi^{-1} + m\psi^{-1})(\phi^{-1} + \psi^{-1})^{-1}\theta - \kappa(\theta)\right]/(\phi^{-1} + \psi^{-1})^{-1} + \textit{normalising term} \tag{2.14}$$

then comparison with (2.8), with (2.10) taken into account, leads to the conclusion that

$$E[\kappa'(\theta)|y; m, \phi, \psi] = (y\phi^{-1} + m\psi^{-1})(\phi^{-1} + \psi^{-1})^{-1} \tag{2.15}$$

$$= zy + (1 - z)m \tag{2.16}$$

with

$$z = \phi^{-1}/(\phi^{-1} + \psi^{-1}) = 1/(1 + \phi/\psi) \tag{2.17}$$

This is essentially the result obtained by Landsman and Makov (1998, Theorem 1).

Linear Bayes (credibility) estimation of θ

Result (2.16) is also a credibility estimator. Consider the linear Bayes estimator

$$\tilde{L}(y) = \underset{L}{arg\,min}\, E_{\theta,y}[L(y) - \mu(\theta)]^2 \tag{2.18}$$

where minimisation is carried out over all functions L such that $L(y)$ is linear in y and independent of θ (i.e. minimisation with respect to a and b in $L(y) = a + by$), and the likelihood of Y and the prior on θ are not necessarily EDF but are unrestricted. The quantity $\tilde{L}(y)$ is known as a **credibility estimator** (or **linear Bayes estimator**) of $\mu(\theta)$ (Bühlmann, 1970).

It is known (Jewell, 1974, 1975) that $\tilde{L}(y)$ is given by the posterior mean $E[\mu(\theta)|y]$ when this linear in y. Thus (2.16), being linear in y, is the credibility estimator of $\mu(\theta)$ in the case of a general likelihood, and of $\mu(\theta) = \kappa'(\theta)$ in the case of an EDF likelihood with natural conjugate prior.

When the Bayes estimator is also linear Bayes, it is sometimes referred to an **exact credibility estimator**.

MAP estimation of θ

To obtain the **maximum a posteriori (MAP)** estimator of θ, differentiate (2.13):

$$\partial\ell(\theta|y; m, \phi, \psi)/\partial\theta = (y\phi^{-1} + m\psi^{-1}) - (\phi^{-1} + \psi^{-1})\kappa'(\theta) \tag{2.19}$$

https://doi.org/10.1017/asb.2014.25 Published online by Cambridge University Press

which is set to zero for MAP estimation. Thus the MAP estimator of $\mu(\theta) = \kappa'(\theta)$ is given by

$$\hat{\mu}(\theta) = zy + (1 - z)m \tag{2.20}$$

In summary then, (2.16) is simultaneously the Bayes estimator, the linear Bayes estimator, and the MAP estimator of $\mu(\theta) = \kappa'(\theta)$ in the case of an EDF likelihood with natural conjugate prior.

Bayes estimation with multiple observations

Hitherto the present sub-section considers Bayes estimation of θ conditioned by a single observation $Y = y$. Consider a sample $\{Y_1, \ldots, Y_n\} = \{y_1, \ldots, y_n\}$, which will also be denoted temporarily by $Y = y$, and suppose that $Y_i|\theta \sim EDF(\theta, \phi_i; \kappa, \lambda)$ and that the Y_i are conditionally independent for given θ. Use (again temporarily) ϕ to denote $\{\phi_1, \ldots, \phi_n\}$ The posterior likelihood (2.13) becomes

$$\ell(\theta|y; m, \phi, \psi) = \left(\sum_{i=1}^{n} y_i \phi_i^{-1} + m\psi^{-1}\right)\theta - \left(\sum_{i=1}^{n} \phi_i^{-1} + \psi^{-1}\right)\kappa(\theta) \tag{2.21}$$

up to a normalising constant, and then (2.15)–(2.17) are replaced by the following:

$$E\left[\kappa'(\theta)|y; m, \phi, \psi\right] = z\bar{y} + (1 - z)m \tag{2.22}$$

with

$$\bar{y} = \sum_{i=1}^{n} y_i \phi_i^{-1} \Big/ \sum_{i=1}^{n} \phi_i^{-1} \tag{2.23}$$

$$z = 1 \Big/ \left[1 + \psi^{-1} \Big/ \sum_{i=1}^{n} \phi_i^{-1}\right] \tag{2.24}$$

The estimator on the right side of (2.22) continues to be simultaneously the Bayes estimator, the linear Bayes estimator, and the MAP estimator of $\mu(\theta) = \kappa'(\theta)$ in the case of an EDF likelihood with natural conjugate prior.

3. Chain Ladder Models

3.1. Mack models

Consider the model defined by the following conditions.

(EDFM1) Accident periods are stochastically independent, and the X_{kj} (j varying) form a Markov chain for each k. More precisely, $(X_{kj})_{j=1,\ldots,J}$ are independent (in $k = 1, 2, \ldots, K$) Markov chains (in j).

(EDFM2) For each $k = 1, 2, \ldots, K$ and $j = 1, 2, \ldots, J - 1$, define $G_{kj} = Y_{k,j+1}/X_{kj}$ and suppose that $G_{kj}|X_{kj} \sim EDF(\theta_j, \phi_{kj}(X_{kj}); \kappa, \lambda)$ for some

https://doi.org/10.1017/asb.2014.25 Published online by Cambridge University Press

functions κ, λ that do not depend on j and k, and with $E[X_{k,j+1}|X_{kj}] = f_j X_{kj}$ for some parameters $f_j > 1$.

This will be referred to subsequently as the **EDF Mack model**. The location parameter θ depends only on j but the dispersion parameter ϕ_{kj} may depend on X_{kj}. The parameters f_j from (EDFM2) are referred to as **age-to-age factors**. In this formulation the f_j are fixed parameters, and so the model is a fixed effects model that will be referred to as the **EDF Mack fixed effects model**.

If $\mu_{k.j+1}$ denotes $E[Y_{k,j+1}|X_{kj}] = X_{kj}E[G_{kj}|X_{kj}]$, then, by (EDFM2) and (2.3),

$$\mu_{k,j+1} = \kappa'(\theta_j) X_{kj} = g_j X_{kj} \tag{3.1}$$

with

$$g_j = f_j - 1 > 0 \tag{3.2}$$

Thus

$$\theta_j = c(\mu_{k,j+1}/ X_{kj}) = c(g_j) \tag{3.3}$$

with $c = (\kappa')^{-1}$.

By (2.4) and (2.6),

$$Var[Y_{k,j+1}|X_{kj}] = \phi_{kj}(X_{kj})\kappa''(\theta_j) X_{kj}^2 \tag{3.4}$$

The original Mack (1993) stochastic version of the chain ladder model included an assumption that $Var[X_{k,j+1}|X_{kj}] = Var[Y_{k,j+1}|X_{kj}]$ is proportional to X_{kj}. Comparison with (3.4) indicates that $\phi_{kj}(X_{kj})$ is proportional to X_{kj}^{-1} in this case.

The model defined in the present sub-section is somewhat different from the EDF Mack model defined in Taylor (2011). That model was subject to the above conditions (EDFM1) but (EDFM2) took a different form, subject to the following distribution:

(EDFM2′) $Y_{k,j+1}|X_{kj} \sim EDF(\theta_{kj}(X_{kj}), \phi_{kj}(X_{kj}); \kappa, \lambda)$ for some functions κ, λ that do not depend on j and k

still subject to the requirement that $E[X_{k,j+1}|X_{kj}] = f_j X_{kj}$.

In that case (3.1) and (3.4) are replaced by the following:

$$\mu_{k,j+1} = \kappa'(\theta_{kj}(X_{kj})) = g_j X_{kj} \tag{3.5}$$

$$Var[Y_{k,j+1}|X_{kj}] = \phi_{kj}(X_{kj})\kappa''(c(g_j X_{kj})) \tag{3.6}$$

Here the variance is related to X_{kj} in a more complex manner than in (3.4). The additional complexity is absent in the special case in which $Y_{k,j+1}|X_{kj} \sim Tw(p)$. In this case, (3.6) yields (making use of (3.5) and (2.7))

$$Var[Y_{k,j+1}|X_{kj}] = \phi_{kj}(X_{kj})[g_j X_{kj}]^p \tag{3.7}$$

https://doi.org/10.1017/asb.2014.25 Published online by Cambridge University Press

compared with

$$Var[Y_{k,j+1}|X_{kj}] = \phi_{kj}(X_{kj})g_j^p X_{kj}^{2+p} \quad \text{(different } \phi_{kj}(X_{kj})\text{)} \tag{3.8}$$

from (3.4).

In the general case of $Y_{k,j+1}|X_{kj} \sim EDF$, the variance (3.6) does not necessarily factorise into separate terms involving g_j and X_{kj} respectively whereas (3.4) does so.

The definition of ϕ_{kj} in (EDFM2) as dependent on accident year and with unspecified dependence on prior cumulative claims experience is unusual in previous literature. It is well known that the definition $\phi_{kj}(X_{kj}) = \sigma_j^2 X_{kj}^{-2}$ (σ_j^2 constant) leads to parameter MLE for stochastic chain ladder models that are the same as for the commonly used heuristic chain ladder algorithm.

In this case (3.8) yields the simple variance structure $Var[Y_{k,j+1}|X_{kj}] = \sigma_j^2 g_j X_{kj}$ in the case $p = 1$. Allowance of a general form for $\phi_{kj}(X_{kj})$ provides for other variance structures. For example, data analysis such as residual plots might indicate an $X-$ dependent multiplier in place of σ_j^2.

Some claim triangles exhibit clear discontinuities with respect to row. These usually indicate a fundamental change in the claim settlement environment. Such changes may affect the dispersion parameter $\phi_{kj}(X_{kj})$, requiring a dependency on k.

3.2. Cross-classified models

The classification of chain ladder models given by Taylor (2011) included the **EDF cross-classified model** defined by the following conditions.

(EDFCC1) The random variables $Y_{kj} \in \mathfrak{D}_K^+$ are stochastically independent.
(EDFCC2) For each $k = 1, 2, \ldots, K$ and $j = 1, 2, \ldots, J$,
- (a) $Y_{kj} \sim EDF(\theta_{kj}, \phi_{kj}; \kappa, \lambda)$ for some functions κ, λ that do not depend on j and k;
- (b) $E[Y_{kj}] = \alpha_k \beta_j$ for some parameters $\alpha_k, \beta_j > 0$; and
- (c) $\sum_{j=1}^J \beta_j = 1$.

Condition (EDFCC2c) is required to remove one degree of redundancy from the parameter set $\{\alpha_k, \beta_j\}$. Alternative constraints on these parameter values produce an equivalent model.

By (EDFCC2b) and (2.3),

$$\mu_{kj} = \kappa'(\theta_{kj}) = \alpha_k \beta_j \tag{3.9}$$

Thus

$$\theta_{kj} = c(\mu_{kj}) = c(\alpha_k \beta_j) \tag{3.10}$$

where $c(.)$ has the same meaning as in Section 3.1.

https://doi.org/10.1017/asb.2014.25 Published online by Cambridge University Press

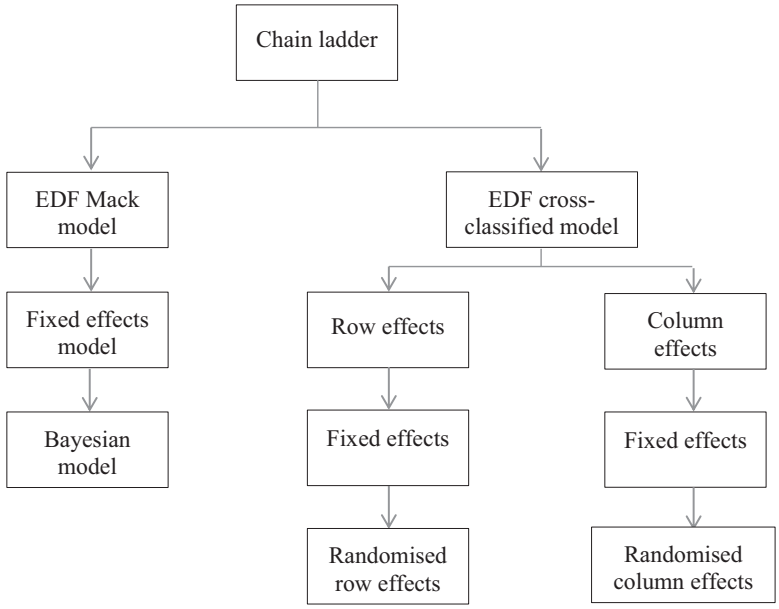


FIGURE 1: Schematic of Bayesian chain ladder models.

The parameters α_k, β_j in (EDFCC2b) are fixed parameters, and so the model is a fixed effects model that will be referred to as the **EDF cross-classified fixed effects model**.

4. BAYESIAN CHAIN LADDER MODELS

The fixed effects models defined in Section 3 may be converted to Bayesian models by randomisation of their parameters, the f_j in the EDF Mack model, and the α_k, β_j in the case of the EDF cross-classified model. Some of these Bayesian models appear in the prior literature.

A Bayesian structure was introduced into the EDF Mack model by Gisler and Müller (2007) and Gisler and Wüthrich (2008), and into the ODP cross-classified model's row effects by Verrall (2000, 2004), England and Verrall (2002) and England *et al.*, (2012). Wüthrich (2007) and Wüthrich and Merz (2008) extended the analysis to the EDF cross-classified model.

England *et al.*, (2012) introduced randomisation of column effects in cross-classified models but, for the most part, used uninformative priors. Until very recently, randomisation of column effects in cross-classified models was absent from the literature. Some recent papers have introduced simultaneous randomisation of row, column and diagonal effects (Shi *et al.*, 2012; Wüthrich, 2012; Merz *et al.*, 2013).

https://doi.org/10.1017/asb.2014.25 Published online by Cambridge University Press

4.1. Mack models

The EDF Mack fixed effects model can be converted to a Bayesian model by the retention of a conditioned form of (EDFM1–2) above, in which each $Y_{k,j+1}$ is now conditioned on $\{\theta_1, \ldots \theta_{J-1}\}$ (and is thus independent of $\theta_1, \ldots \theta_{j-1}, \theta_{j+1}, \ldots \theta_{J-1}$), and a further condition (EDFM3) is added, as follows:

(EDFM3) For $j = 1, 2, \ldots, J-1$, g_j is a random parameter subject to the distribution $\theta_j \sim EDF^*(m_j, \psi_j; \kappa, \lambda^*)$, where the $\theta_1, \ldots \theta_{J-1}$ are independent.

This will be referred to as **Bayesian EDF Mack model**. Note that, by (3.3), (EDFM3) amounts to a prior on g_j and, by (2.10) and (3.1), m_j is the unconditional expectation

$$m_j = E[\kappa'(\theta_j)] = E[g_j] \tag{4.1}$$

4.1.1. *Parameter estimation.* In view of (EDFM3), the conclusion at the end of Section 2.2.2 holds but in respect of observations G_{ij} (fixed j) in place of the y_i. The substitutions set out in Table 2 are made in (2.22)–(2.24) to yield the following Bayes estimator of g_j:

$$\tilde{g}_j = z_j \hat{g}_j + (1 - z_j) m_j \tag{4.2}$$

where

$$\hat{g}_j = \sum_{i=1}^{K-j} G_{ij}[\phi_{ij}(X_{ij})]^{-1} \Big/ \sum_{i=1}^{K-j} [\phi_{ij}(X_{ij})]^{-1} \tag{4.3}$$

$$z_j = 1 \Big/ \left[1 + \psi_j^{-1} \Big/ \sum_{i=1}^{K-j} [\phi_{ij}(X_{ij})]^{-1} \right] \tag{4.4}$$

The credibility coefficient (4.4) can be expressed in an alternative form familiar to credibility theorists. By (2.4) and (2.6),

$$Var[G_{kj}|\theta_j, X_{kj}] = \phi_{kj}(X_{kj}) \kappa''(\theta_j) \tag{4.5}$$

Take expectations with respect to θ_j (but not X_{kj}), through this last equality, and then rearrange slightly to obtain

$$[\phi_{kj}(X_{kj})]^{-1} = E[\kappa''(\theta_j)] / E_{\theta_j} Var[G_{kj}|\theta_j, X_{kj}] \tag{4.6}$$

Now substitute (4.6) into the left side of (2.11), and (4.1) into the right side, to obtain

$$\psi_j^{-1} = E[\kappa''(\theta_j)] / Var_{\theta_j}[\kappa'(\theta_j)] = E[\kappa''(\theta_j)] / Var_{\theta_j}[E[G_{kj}|\theta_j, X_{kj}]] \tag{4.7}$$

https://doi.org/10.1017/asb.2014.25 Published online by Cambridge University Press

TABLE 2

CORRESPONDENCES BETWEEN (2.22)–(2.24) AND (4.2)–(4.4).

Quantity in (2.22)–(2.24)	Replaced by
$E[\kappa'(\theta)\,\vert y; m, \phi, \psi]$	$E\left[g_j \vert \mathfrak{D}_K; \{m_j\}, \{\phi_{kj}\}, \{\psi_j\}\right]$
y_i	G_{ij}
ϕ_i	ϕ_{ij}
ψ	ψ_j

Substitution of (4.6) and (4.7) into (4.4) yields a credibility coefficient of familiar appearance:

$$z_j = 1 \Big/ \left[1 + 1 \Big/ \sum_{i=1}^{K-j} v_{ij}^2 \right] \tag{4.8}$$

where

$$v_{ij}^2 = \frac{[\phi_{ij}(X_{ij})]^{-1}}{\psi_j^{-1}} = \frac{Var_{\theta_j} E[G_{ij}|\theta_j,\, X_{ij}]}{\phi_{ij}(X_{ij})\, E_{\theta_j}[[\phi_{ij}(X_{ij})]^{-1}\, Var_{\theta_j}[G_{ij}|\theta_j,\, X_{ij}]]} \tag{4.9}$$

In conventional credibility spirit, the credibility coefficient is expressed in terms of two separate components of variance, one reflecting noise within the observation process, and the other variation between prior assessments of the mean.

4.1.2. *Special case: Tweedie likelihood.* Consider the special case of (EDFM2) $G_{kj}|\theta_j,\, X_{kj} \sim Tw(g_j, \phi_{kj}(X_{kj});\, p, \lambda)$, and suppose that (EDFM3) still holds, i.e. stipulates a natural conjugate prior to Tweedie. Then substitution of (2.5)–(2.7) into (4.9) yields

$$v_{ij}^2 = \frac{Var_{\theta_j} E[G_{ij}|\theta_j,\, X_{ij}]}{E_{\theta_j} E[G_{ij}^p|\theta_j,\, X_{ij}]} \frac{1}{\phi_{ij}(X_{ij})} \tag{4.10}$$

Further restriction to the ODP case ($p = 1$) gives

$$v_{ij}^2 = \frac{Var_{\theta_j} E[G_{ij}|\theta_j,\, X_{ij}]}{m_j \phi_{kj}(X_{ij})} \tag{4.11}$$

when (2.3) and (2.10) are taken into account.

4.1.3. *Special case: original Mack model.* As mentioned in Section 3.1, the original Mack (1993) model required that

$$Var[Y_{k,j+1}|\, X_{kj}] = \sigma_j^2 X_{kj} \tag{4.12}$$

But this variance is given by (3.4) for the EDF Mack fixed effects model, whence

$$\phi_{kj}(X_{kj}) = \tau_j^2 X_{kj}^{-1} \tag{4.13}$$

https://doi.org/10.1017/asb.2014.25 Published online by Cambridge University Press

with

$$\tau_j^2 = \sigma_j^2/\kappa''(\theta_j) \tag{4.14}$$

and (4.9) now reads

$$v_{ij}^2 = \frac{\psi_j}{\tau_j^2} X_{ij} \tag{4.15}$$

and (4.8) becomes

$$z_j = 1 \Big/ \left[1 + \frac{\tau_j^2}{\psi_j} \Big/ \sum_{i=1}^{K-j} X_{ij} \right] \tag{4.16}$$

Moreover, (4.3)becomes

$$\hat{g}_j = \sum_{i=1}^{K-j} Y_{i,j+1} \Big/ \sum_{i=1}^{K-j} X_{ij} \tag{4.17}$$

and $1 + \hat{g}_j$ can be recognised as the conventional chain ladder estimator of an age-to-age factor.

These results were obtained by Gisler and Wüthrich (2008, Theorem 6.4). The result (4.9) is a little more general because it applies to variance structures other than Mack's (4.12).

The results of the present section may be summarised in the following theorem.

Theorem 4.1. *(a) Consider the Bayesian EDF Mack model defined by (EDFM1–3). The Bayes estimator, the linear Bayes estimator, and the MAP estimator of g_j are all the same and are given by (4.2)–(4.4). There is an equivalent form in which (4.4) is replaced by (4.8)–(4.9).*

(b) In the special case of the Mack variance structure (4.12), (4.8) takes the form (4.16). □

The equality of the Bayes estimator, the linear Bayes estimator, and the MAP estimator is established in Section 2.2.2.

4.2. Cross-classified models

4.2.1. *Model definition.* The EDF cross-classified fixed effects model can be converted to a Bayesian model by the retention of a conditioned form of (EDFCC1,2a-b) above, in which each Y_{kj} is now conditioned on the relevant parameters α_k and β_j, and further conditions (EDFCC3a-c) added as below. The full description of the model is as follows.

(EDFCC1) The random variables $Y_{kj} \in \mathfrak{D}_K^+$ now depend on random parameters $\alpha_k > 0, k = 1, \ldots, K$ and, $\beta_j > 0$, $j = 1, \ldots, J$, and are conditionally independent given the $\{\alpha_1, \ldots, \alpha_K, \beta_1, \ldots, \beta_J\}$.

(EDFCC2) For each $k = 1, 2, \ldots, K$ and $j = 1, 2, \ldots, J$,

https://doi.org/10.1017/asb.2014.25 Published online by Cambridge University Press

(a) $Y_{kj} \sim EDF(\theta_{kj}, \phi_{kj}; \kappa, \lambda)$ for some functions κ, λ that do not depend on j and k;
(b) $E[Y_{kj}|\alpha_1, \ldots, \alpha_K, \beta_1, \ldots, \beta_J] = \alpha_k \beta_j$;

(EDFCC3a) For each $k = 1, 2, \ldots, K$, α_k is subject to the distribution $c(\alpha_k) \sim EDF^*(a_k, \psi_k^{(\alpha)}; \kappa, \lambda^{*(\alpha)})$.

(EDFCC3b) For each $j = 1, 2, \ldots, J$, β_j is subject to the distribution $c(\beta_j) \sim EDF^*(b_j, \psi_j^{(\beta)}; \kappa, \lambda^{*(\beta)})$.

(EDFCC3c) The random parameters $\alpha_1, \ldots, \alpha_K, \beta_1, \ldots, \beta_J$ are independent.

This model will be referred to as **Bayesian EDF cross-classified model**, with (EDFCC3a,b) amounting to priors on α_k, β_j respectively. Note that this model does not include the condition (EDFCC2c). The random parameters α_k, β_j are distributed according to (EDFCC3a,b) and no longer contain a degree of redundancy.

Consider now the interpretation of the a_k, b_j. By (3.10) and the definition of c,

$$E[\kappa'(\theta_{kj})] = E[c^{-1}(\theta_{kj})] = E[\alpha_k \beta_j] = E[\alpha_k]E[\beta_j] \tag{4.18}$$

the last equality following from (EDFCC3c).

Further interpretation of the a_k, b_j is not obvious in this general case but will be discussed in Section 4.2.3 in the special case of the Tweedie distribution.

Alternative model forms

A somewhat different model is considered by Wüthrich (2007) and Wüthrich and Merz (2008, Section 4.3.2) in which only row effects are randomised. Then assumptions (EDFCC3b,c) may be dropped. In their formulation (EDFCC2,3a) are replaced by the following:

(EDFCC2*) For each $k = 1, 2, \ldots, K$ and $j = 1, 2, \ldots, J$, $Y_{kj}/a_k b_j \sim EDF(\theta_k, \phi_{kj}; \kappa, \lambda)$ for some parameters $a_k, b_j > 0$ and some functions κ, λ that do not depend on j and k; and

(EDFCC3*) For each $k = 1, 2, \ldots, K$, θ_k is a random parameter subject to the distribution $\theta_k \sim EDF^*(1, \psi; \kappa, \lambda^*)$;

(EDFCC4*) The random parameters $\theta_k, k = 1, \ldots, K$ are independent.

There are differences between the two models. On the one hand, the Bayesian EDF cross-classified model (EDFCC1,2,3a-c) directly randomises the row and column parameters α_k, β_j, which are thus assumed unknown.

The Wüthrich model (EDFCC1,2*,3*,4*), on the other hand, incorporates a multiplicative random row effect θ_k, which is applied to the product of the (assumed known) prior row and column parameters. The practical estimation of these parameters is discussed in this case by Wüthrich (2007).

It is interesting to compare the two models when the EDF is reduced to the Tweedie sub-family. Here (EDFCC2*) takes the form $Y_{kj}/a_k b_j \sim Tw(b'(\theta_k), \phi_{kj}; p, \lambda) = Tw([(1-p)\theta_k]^{\frac{1}{1-p}}, \phi_{kj}; p, \lambda)$ with the usual limiting case for $p = 1$. Since members of this sub-family satisfy the scale transformation property (Jorgensen and Paes de Souza, 1994), (EDFCC2*) may be re-expressed

https://doi.org/10.1017/asb.2014.25 Published online by Cambridge University Press

in that case in the form

$$Y_{kj} \sim Tw\left([(1-p)\theta_k]^{\frac{1}{1-p}} a_k b_j, \phi_{kj}[(1-p)\theta_k]^{\frac{p}{1-p}} (a_k b_j)^2; p, \lambda\right) \tag{4.19}$$

In the case $p = 2$ in this model, the quantity $[(1-p)\theta_k]^{\frac{1}{1-p}} a_k = -a_k/\theta_k$ may be viewed as a single randomised row parameter, much the same as in (EDFCC3a). This not the case for $p \neq 2$, rendering the model an essentially different row effect randomisation from that described by (EDFCC1–3).

4.2.2. *Parameter estimation.*

General EDF

Let $\mathcal{R}(k)$ and $\mathcal{C}(j)$ denote the k-th row and j-th column respectively of $\mathfrak{D}_K$. If α, β denote the vectors of values of α_k, β_j respectively; a, b the vectors of values of a_k, b_j; $\psi^{(\alpha)}, \psi^{(\beta)}$ the vectors of values of $\psi_k^{(\alpha)}, \psi_j^{(\beta)}$; and ϕ the trapezoid of values ϕ_{kj}; then, by (2.2), (2.6) and (2.8), the posterior log-likelihood for $\{c(\alpha_k); k = 1, 2, \ldots, K\}$, denoted $c(\alpha)$, conditional on $\mathfrak{D}_K, \beta$ is

$$\begin{aligned}\ell(c(\alpha)|\mathfrak{D}_K, \beta; a, b, \phi, \psi^{(\alpha)}, \psi^{(\beta)}) = &\sum_{\mathfrak{D}_K} [Y_{kj} c(\alpha_k \beta_j) - d(\alpha_k \beta_j)]/\phi_{kj} \\ &+ \sum_{k=1}^{K} [c(\alpha_k) a_k - d(\alpha_k)]/\psi_k^{(\alpha)}\end{aligned} \tag{4.20}$$

where $c(.)$ is defined as in Section 3.1, and

$$d(u) = \kappa(c(u)) \tag{4.21}$$

Bayes, linear Bayes, and MAP estimation of the α_k, β_j will be considered. Little progress appears possible with the first two of these in the case of general κ. Some progress may be made, however, with MAP estimation. Differentiate (4.20):

$$\begin{aligned}\partial\ell/\partial\alpha_k &= \sum_{j \in \mathcal{R}(k)} [Y_{kj} c'(\mu_{kj}) - d'(\mu_{kj})]\beta_j/\phi_{kj} + [c'(\alpha_k) a_k - d'(\alpha_k)]/\psi_k^{(\alpha)} \\ &= \sum_{j \in \mathcal{R}(k)} \beta_j c'(\mu_{kj})[Y_{kj} - \mu_{kj}]/\phi_{kj} + c'(\alpha_k)[a_k - \alpha_k]/\psi_k^{(\alpha)}\end{aligned} \tag{4.22}$$

since, by (4.21),

$$d'(\mu) = \kappa'(c(\mu))c'(\mu) = \mu c'(\mu) \tag{4.23}$$

where the last equality derives from the definition of c.

https://doi.org/10.1017/asb.2014.25 Published online by Cambridge University Press

Re-insert (3.9) into (4.22) to obtain

$$\partial \ell / \partial \alpha_k = \sum_{j \in \mathcal{R}(k)} \frac{\beta_j c'(\mu_{kj})}{\phi_{kj}} Y_{kj} + \frac{c'(\alpha_k)}{\psi_k^{(\alpha)}} a_k - \alpha_k \left[\sum_{j \in \mathcal{R}(k)} \frac{\beta_j c'(\mu_{kj})}{\phi_{kj}} \beta_j + \frac{c'(\alpha_k)}{\psi_k^{(\alpha)}} \right] \tag{4.24}$$

By the assumption in Section 2.2.1 that $c^{-1}(.) = \kappa'(.)$ is one-one, setting (4.24) equal to zero yields the following MAP estimator for α_k:

$$\hat{\alpha}_k = \frac{\sum_{j \in \mathcal{R}(k)} \frac{\beta_j c'(\hat{\mu}_{kj})}{\phi_{kj}} Y_{kj} + \frac{c'(\hat{\alpha}_k)}{\psi_k^{(\alpha)}} a_k}{\sum_{j \in \mathcal{R}(k)} \frac{\beta_j c'(\hat{\mu}_{kj})}{\phi_{kj}} \beta_j + \frac{c'(\hat{\alpha}_k)}{\psi_k^{(\alpha)}}} \tag{4.25}$$

where $\hat{\mu}_{kj} = \hat{\alpha}_k \beta_j$ is the estimator of μ_{kj}.

This reasoning applies to known β. If in fact β requires estimation, then (4.25) is replaced by

$$\hat{\alpha}_k = \frac{\sum_{j \in \mathcal{R}(k)} \frac{\hat{\beta}_j c'(\hat{\mu}_{kj})}{\phi_{kj}} Y_{kj} + \frac{c'(\hat{\alpha}_k)}{\psi_k^{(\alpha)}} a_k}{\sum_{j \in \mathcal{R}(k)} \frac{\hat{\beta}_j c'(\hat{\mu}_{kj})}{\phi_{kj}} \hat{\beta}_j + \frac{c'(\hat{\alpha}_k)}{\psi_k^{(\alpha)}}} \tag{4.26}$$

with $\hat{\mu}_{kj} = \hat{\alpha}_k \hat{\beta}_j$ now the estimator of μ_{kj}.

Similar reasoning produces the following estimator for β_j:

$$\hat{\beta}_j = \frac{\sum_{k \in \mathcal{C}(j)} \frac{\hat{\alpha}_k c'(\hat{\mu}_{kj})}{\phi_{kj}} Y_{kj} + \frac{c'(\hat{\beta}_j)}{\psi_j^{(\beta)}} b_j}{\sum_{k \in \mathcal{C}(j)} \frac{\hat{\alpha}_k c'(\hat{\mu}_{kj})}{\phi_{kj}} \hat{\alpha}_k + \frac{c'(\hat{\beta}_j)}{\psi_j^{(\beta)}}} \tag{4.27}$$

The estimate (4.26) is implicit since it involves the estimand $\hat{\alpha}_k$ on the right side also. Similarly the estimate $\hat{\beta}_j$, and so both require iterative solution in the form

$$\hat{\alpha}_k^{(n+1)} = \frac{\sum_{j \in \mathcal{R}(k)} \frac{\hat{\beta}_j^{(n)} c'(\hat{\mu}_{kj}^{(n)})}{\phi_{kj}} Y_{kj} + \frac{c'(\hat{\alpha}_k^{(n)})}{\psi_k^{(\alpha)}} a_k}{\sum_{j \in \mathcal{R}(k)} \frac{\hat{\beta}_j^{(n)} c'(\hat{\mu}_{kj}^{(n)})}{\phi_{kj}} \hat{\beta}_j^{(n)} + \frac{c'(\hat{\alpha}_k^{(n)})}{\psi_k^{(\alpha)}}} \tag{4.28}$$

$$\hat{\beta}_j^{(n+1)} = \frac{\sum_{k \in \mathcal{C}(j)} \frac{\hat{\alpha}_k^{(n)} c'(\hat{\mu}_{kj}^{(n)})}{\phi_{kj}} Y_{kj} + \frac{c'(\hat{\beta}_j^{(n)})}{\psi_j^{(\beta)}} b_j}{\sum_{k \in \mathcal{C}(j)} \frac{\hat{\alpha}_k^{(n)} c'(\hat{\mu}_{kj}^{(n)})}{\phi_{kj}} \hat{\alpha}_k + \frac{c'(\hat{\beta}_j^{(n)})}{\psi_j^{(\beta)}}} \tag{4.29}$$

where $\hat{\alpha}_k^{(n)}, \hat{\beta}_j^{(n)}, \hat{\mu}_{kj}^{(n)}$ denote the n-th iterates of $\hat{\alpha}_k, \hat{\beta}_j, \hat{\mu}_{kj}$ respectively with $\hat{\mu}_{kj}^{(n)} = \hat{\alpha}_k^{(n)} \hat{\beta}_j^{(n)}$.

The implicit nature of (4.26) and (4.27) raises a question about the uniqueness of their solutions. In common with a number of other contributions to the

https://doi.org/10.1017/asb.2014.25 Published online by Cambridge University Press

literature, it has simply been assumed for present purposes that the solutions of (4.26) and (4.27) exist and are unique.

Naturally, it is necessary to check for convergence of these iterates. However, convergence of iterative solutions of chain ladder systems of equations has not been problematic in the past (see e.g. Taylor, 2009, Section 4).

These estimators of α_k, β_j may be represented in the familiar credibility format, though it may be noted that the credibility factors $z_{kj}^{(\alpha)}$ and $z_{kj}^{(\beta)}$ depend on the parameter estimates $\hat{\beta}_j$ and $\hat{\alpha}_k$ respectively:

$$\hat{\alpha}_k = z_k^{(\alpha)} \bar{Y}_k^{(\alpha)} + \left[1 - z_k^{(\alpha)}\right] a_k \tag{4.30}$$

with

$$z_k^{(\alpha)} = \sum_{j \in \mathcal{R}(k)} z_{kj}^{(\alpha)} \tag{4.31}$$

$$\bar{Y}_k^{(\alpha)} = \sum_{j \in \mathcal{R}(k)} z_{kj}^{(\alpha)} [Y_{kj} / \hat{\beta}_j] / \sum_{j \in \mathcal{R}(k)} z_{kj}^{(\alpha)} \tag{4.32}$$

$$z_{kj}^{(\alpha)} = \frac{\hat{\beta}_j^2 c'(\hat{\mu}_{kj})}{\phi_{kj}} \Bigg/ \left[\sum_{j \in \mathcal{R}(k)} \frac{\hat{\beta}_j^2 c'(\hat{\mu}_{kj})}{\phi_{kj}} + \frac{c'(\hat{\alpha}_k)}{\psi_k^{(\alpha)}} \right] \tag{4.33}$$

$$\hat{\beta}_j = z_j^{(\beta)} \bar{Y}_j^{(\beta)} + \left[1 - z_j^{(\beta)}\right] b_j \tag{4.34}$$

with

$$z_j^{(\beta)} = \sum_{k \in \mathcal{C}(j)} z_{kj}^{(\beta)} \tag{4.35}$$

$$\bar{Y}_j^{(\beta)} = \sum_{k \in \mathcal{C}(j)} z_{kj}^{(\beta)} \left[Y_{kj} / \hat{\alpha}_k \right] / \sum_{k \in \mathcal{C}(j)} z_{kj}^{(\beta)} \tag{4.36}$$

$$z_{kj}^{(\beta)} = \frac{\hat{\alpha}_k^2 c'(\hat{\mu}_{kj})}{\varphi_{kj}} \Bigg/ \left[\sum_{k \in \mathcal{C}(\mathrm{j})} \frac{\hat{\alpha}_k^2 c'(\hat{\mu}_{kj})}{\varphi_{kj}} + \frac{c'(\hat{\beta}_j)}{\psi_j^{(\beta)}} \right] \tag{4.37}$$

Example

A **non-Tweedie** example is the binomial distribution for which $\kappa(\theta) = N ln(1 + e^{\theta})$. From this it follows that $c'(u) = 1/[u(1 - u/N)]$, which may be substituted in (4.25)–(4.37).

The Wüthrich model (EDFCC1,2*,3*) produces the different estimator:

$$\hat{\theta}_k = z_k^{(\theta)} \bar{Y}_k^{(\theta)} + \left[1 - z_k^{(\theta)}\right] 1 \tag{4.38}$$

https://doi.org/10.1017/asb.2014.25 Published online by Cambridge University Press

with

$$z_k^{(\theta)} = \sum_{j \in \mathcal{R}(k)} z_{kj}^{(\theta)} \tag{4.39}$$

$$\bar{Y}_k^{(\theta)} = \sum_{j \in \mathcal{R}(k)} z_{kj}^{(\theta)} \left[Y_{kj} / \alpha_k \beta_j \right] / \sum_{j \in \mathcal{R}(k)} z_{kj}^{(\theta)} \tag{4.40}$$

$$z_{kj}^{(\theta)} = \frac{\psi}{\phi_{kj}} \Bigg/ \left[1 + \sum_{j \in \mathcal{R}(k)} \frac{\psi}{\phi_{kj}} \right] \tag{4.41}$$

It would be possible to extend the Wüthrich model by randomising both row and column parameters. Column parameter estimators corresponding to (4.38)–(4.41) would be obtainable from those results by row-column symmetry.

4.2.3. *Tweedie family.* Consider (4.20) for the Tweedie family. The model will then be referred to as the **Bayesian Tweedie cross-classified model**. In this case, Table 1 and (4.21) yield

$$\begin{aligned} c(u) &= u^{1-p}/(1-p),\ p \neq 1 \\ &= lnu,\ p = 1 \end{aligned} \tag{4.42}$$

$$\begin{aligned} d(u) &= u^{2-p}/(2-p),\ p \neq 2 \\ &= lnu,\ p = 2 \end{aligned} \tag{4.43}$$

Therefore

$$c(\alpha_k \beta_j) = c(\alpha_k) \beta_j^{1-p} = \theta_k^{(\alpha)} \beta_j^{1-p} \tag{4.44}$$

with

$$\theta_k^{(\alpha)} = c(\alpha_k) = \alpha_k^{1-p}/(1-p) \tag{4.45}$$

Similarly

$$d(\alpha_k \beta_j) = d(\alpha_k) \beta_j^{2-p} = \kappa\left(\theta_k^{(\alpha)}\right) \beta_j^{2-p} \tag{4.46}$$

These relations convert (4.20) to the following, reduced to the likelihood of just $\theta_k^{(\alpha)}$ instead of the full $c(\alpha)$:

$$\begin{aligned} \ell(\theta_k^{(\alpha)} | \mathfrak{D}_K, \beta; a, b, \phi, \psi^{(\alpha)}, \psi^{(\beta)}) = &\sum_{j \in \mathcal{R}(k)} \left[Y_{kj} \theta_k^{(\alpha)} \beta_j^{1-p} - \kappa(\theta_k^{(\alpha)}) \beta_j^{2-p} \right] / \phi_{kj} \\ &+ \left[\theta_k^{(\alpha)} a_k - \kappa(\theta_k^{(\alpha)}) \right] / \psi_k^{(\alpha)} \end{aligned} \tag{4.47}$$

which may be further converted thus:

$$\begin{aligned} \ell(\alpha_k | \mathfrak{D}_K, \beta; a, b, \phi, \psi^{(\alpha)}, \psi^{(\beta)}) = &\sum_{j \in \mathcal{R}(k)} \left[(Y_{kj}/\beta_j) \theta_k^{(\alpha)} - \kappa(\theta_k^{(\alpha)}) \right] / \phi_{kj} \beta_j^{p-2} \\ &+ \left[\theta_k^{(\alpha)} a_k - \kappa(\theta_k^{(\alpha)}) \right] / \psi_k^{(\alpha)} \end{aligned} \tag{4.48}$$

https://doi.org/10.1017/asb.2014.25 Published online by Cambridge University Press

TABLE 3

CORRESPONDENCES BETWEEN (2.21) AND (4.48).

(2.21)	(4.48)
y_i	Y_{kj}/β_j
ϕ_i	$\phi_{kj}\beta_j^{p-2}$
m	a_k
ψ	$\psi_k^{(\alpha)}$
θ	$\theta_k^{(\alpha)}$
$\kappa(\theta)$	$\kappa\left(\theta_k^{(\alpha)}\right)$
$\kappa'(\theta)$	$\kappa'\left(\theta_k^{(\alpha)}\right) = \alpha_k$

This is now recognisable as of the same form as (2.21) with the correspondences set out in Table 3.

The results of Section 2.2.2 become applicable to the present case with these correspondences. In particular, in parallel with (2.10)

$$E\left[\kappa'\left(\theta_k^{(\alpha)}\right)\right] = a_k \tag{4.49}$$

But $\kappa'(.)$ is obtainable from Table 1:

$$\kappa'(\theta) = [(1-p)\theta]^{1/(1-p)} \tag{4.50}$$

Substitution of (4.50) and (4.45) into (4.49) yields

$$E[\alpha_k] = a_k \tag{4.51}$$

Similarly,

$$E[\beta_j] = b_j \tag{4.52}$$

Further, in parallel with (2.22)–(2.24),

$$E[\alpha_k|\mathfrak{D}_K, \beta; a, b, \phi, \psi^{(\alpha)}, \psi^{(\beta)}] = z_k^{(\alpha)}\bar{Y}_k^{(\alpha)} + (1 - z_k^{(\alpha)})a_k \tag{4.53}$$

with

$$\bar{Y}_k^{(\alpha)} = \sum_{j\in\mathcal{R}(k)} Y_{kj}\beta_j^{1-p}\phi_{kj}^{-1} \Big/ \sum_{j\in\mathcal{R}(k)} \beta_j^{2-p}\phi_{kj}^{-1} \tag{4.54}$$

$$z_k^{(\alpha)} = 1 \Big/ \left[1 + [\psi_k^{(\alpha)}]^{-1} \Big/ \sum_{j\in\mathcal{R}(k)} \beta_j^{2-p}\phi_{kj}^{-1}\right] \tag{4.55}$$

Similarly, the posterior likelihood of β_j can be constructed in parallel with (4.48) leading to the following parallel of (4.53)–(4.55):

$$E[\beta_j|\mathfrak{D}_K, \alpha; a, b, \phi, \psi^{(\alpha)}, \psi^{(\beta)}] = z_j^{(\beta)}\bar{Y}_j^{(\beta)} + \left(1 - z_j^{(\beta)}\right) b_j \tag{4.56}$$

https://doi.org/10.1017/asb.2014.25 Published online by Cambridge University Press

with

$$\bar{Y}_j^{(\beta)} = \sum_{k\in\mathcal{C}(j)} Y_{kj}\alpha_k^{1-p}\phi_{kj}^{-1} \Big/ \sum_{k\in\mathcal{C}(j)} \alpha_k^{2-p}\phi_{kj}^{-1} \tag{4.57}$$

$$z_j^{(\beta)} = 1 \Big/ \left[1 + [\psi_j^{(\beta)}]^{-1} \Big/ \sum_{k\in\mathcal{C}(j)} \alpha_k^{2-p}\phi_{kj}^{-1}\right] \tag{4.58}$$

Estimator (4.53)–(4.55) applies to the case of known β. Just as was the case for (2.22), it is are simultaneously the Bayes estimator, the linear Bayes estimator, and the MAP estimator of α_k for known β. Similarly for (4.56)–(4.58) as estimator of β_j in the case of known α.

One may form the **empirical Bayes** estimators

$$\hat{\alpha}_k = z_k^{(\alpha)}\,\bar{Y}_k^{(\alpha)} + \big(1 - z_k^{(\alpha)}\big)a_k \tag{4.59}$$

with

$$\bar{Y}_k^{(\alpha)} = \sum_{j\in\mathcal{R}(k)} Y_{kj}\hat{\beta}_j^{1-p}\phi_{kj}^{-1} \Big/ \sum_{j\in\mathcal{R}(k)} \hat{\beta}_j^{2-p}\phi_{kj}^{-1} \tag{4.60}$$

$$z_k^{(\alpha)} = 1 \Big/ \left[1 + [\psi_k^{(\alpha)}]^{-1} \Big/ \sum_{j\in\mathcal{R}(k)} \hat{\beta}_j^{2-p}\phi_{kj}^{-1}\right] \tag{4.61}$$

and

$$\hat{\beta}_j = z_j^{(\beta)}\,\bar{Y}_j^{(\beta)} + \big(1 - z_j^{(\beta)}\big)b_j \tag{4.62}$$

with

$$\bar{Y}_j^{(\beta)} = \sum_{k\in\mathcal{C}(j)} Y_{kj}\hat{\alpha}_k^{1-p}\phi_{kj}^{-1} \Big/ \sum_{k\in\mathcal{C}(j)} \hat{\alpha}_k^{2-p}\phi_{kj}^{-1} \tag{4.63}$$

$$z_j^{(\beta)} = 1 \Big/ \left[1 + \big[\psi_j^{(\beta)}\big]^{-1} \Big/ \sum_{k\in\mathcal{C}(j)} \hat{\alpha}_k^{2-p}\phi_{kj}^{-1}\right] \tag{4.64}$$

These are also the MAP estimators of α_k, β_j.

The estimators (4.59)–(4.64) are implicit in that the estimator of α_k depends on the estimators of the β_j, which in turn depend on the estimator of α_k. Their computation needs to be iterative, according to the following schema in which $\hat{\alpha}_k^{[n]}, \hat{\beta}_j^{[n]}$ are the n-th iterates of $\alpha_k, \beta_j, n = 1, 2, etc$:

$$\hat{\alpha}_k^{[n+1]} = z_k^{(\alpha)[n]}\,\bar{Y}_k^{(\alpha)[n]} + \big(1 - z_k^{(\alpha)[n]}\big)a_k \tag{4.65}$$

with

$$\bar{Y}_k^{(\alpha)[n]} = \sum_{j\in\mathcal{R}(k)} Y_{kj}\big[\hat{\beta}_j^{[n]}\big]^{1-p}\phi_{kj}^{-1} \Big/ \sum_{j\in\mathcal{R}(k)} \big[\hat{\beta}_j^{[n]}\big]^{2-p}\phi_{kj}^{-1} \tag{4.66}$$

https://doi.org/10.1017/asb.2014.25 Published online by Cambridge University Press

$$z_k^{(\alpha)[n]} = 1 \Big/ \left[1 + \left[\psi_k^{(\alpha)}\right]^{-1} \Big/ \sum_{j \in \mathcal{R}(k)} \left[\hat{\beta}_j^{[n]}\right]^{2-p} \phi_{kj}^{-1} \right] \tag{4.67}$$

and

$$\hat{\beta}_j^{[n+1]} = z_j^{(\beta)[n]} \bar{Y}_j^{(\beta)[n]} + \left(1 - z_j^{(\beta)[n]}\right) b_j \tag{4.68}$$

with

$$\bar{Y}_j^{(\beta)[n]} = \sum_{k \in \mathcal{C}(j)} Y_{kj} \left[\hat{\alpha}_k^{[n]}\right]^{1-p} \phi_{kj}^{-1} \Big/ \sum_{k \in \mathcal{C}(j)} \left[\hat{\alpha}_k^{[n]}\right]^{2-p} \phi_{kj}^{-1} \tag{4.69}$$

$$z_j^{(\beta)[n]} = 1 \Big/ \left[1 + \left[\psi_j^{(\beta)}\right]^{-1} \Big/ \sum_{k \in \mathcal{C}(j)} \left[\hat{\alpha}_k^{[n]}\right]^{2-p} \phi_{kj}^{-1} \right] \tag{4.70}$$

4.2.4. *ODP family.* The ODP family is obtained by the choice $p = 1$ in the Tweedie family. The model will then be referred to as the **Bayesian ODP cross-classified model**. It was considered by England *et al.*, (2012). It may be noted that, other than when considering uninformative uniform priors, they adopt assumptions (EDFCC2,3a) as distinct from Wüthrich's (2007) (EDFCC2*,3*)

The substitution $p = 1$ in the Tweedie Bayes estimators (4.59)–(4.64) gives

$$\bar{Y}_k^{(\alpha)} = \sum_{j \in \mathcal{R}(k)} Y_{kj} \phi_{kj}^{-1} \Big/ \sum_{j \in \mathcal{R}(k)} \hat{\beta}_j \phi_{kj}^{-1} \tag{4.71}$$

$$z_k^{(\alpha)} = 1 \Big/ \left[1 + \left[\psi_k^{(\alpha)}\right]^{-1} \Big/ \sum_{j \in \mathcal{R}(k)} \hat{\beta}_j \phi_{kj}^{-1} \right] \tag{4.72}$$

$$\bar{Y}_j^{(\beta)} = \sum_{k \in \mathcal{C}(j)} Y_{kj} \phi_{kj}^{-1} \Big/ \sum_{k \in \mathcal{C}(j)} \hat{\alpha}_k \phi_{kj}^{-1} \tag{4.73}$$

$$z_j^{(\beta)} = 1 \Big/ \left[1 + \left[\psi_j^{(\beta)}\right]^{-1} \Big/ \sum_{k \in \mathcal{C}(j)} \hat{\alpha}_k \phi_{kj}^{-1} \right] \tag{4.74}$$

Empirical Bayes estimators can be obtained from (4.59)–(4.64) by the substitution $p = 1$, and similarly MAP estimators from (4.65)–(4.70).

A case of particular interest arises if the following further assumption is made.

(EDFCC4) The dispersion parameters ϕ_{kj} are all equal: $\phi_{kj} = \phi$.

https://doi.org/10.1017/asb.2014.25 Published online by Cambridge University Press

In this case (4.71)–(4.74) simplify further:

$$\bar{Y}_k^{(\alpha)} = \sum_{j \in \mathcal{R}(k)} Y_{kj} \Big/ \sum_{j \in \mathcal{R}(k)} \hat{\beta}_j \tag{4.75}$$

$$z_k^{(\alpha)} = 1 \Big/ \left[1 + \left[\phi / \psi_k^{(\alpha)} \right] \Big/ \sum_{j \in \mathcal{R}(k)} \hat{\beta}_j \right] \tag{4.76}$$

$$\bar{Y}_j^{(\beta)} = \sum_{k \in \mathcal{C}(j)} Y_{kj} \Big/ \sum_{k \in \mathcal{C}(j)} \hat{\alpha}_k \tag{4.77}$$

$$z_j^{(\beta)} = 1 \Big/ \left[1 + \left[\phi / \psi_j^{(\beta)} \right] / \sum_{k \in \mathcal{C}(j)} \hat{\alpha}_k \right] \tag{4.78}$$

Uninformative priors

Consider the case of uninformative priors, for which $\psi_k^{(\alpha)}, \psi_j^{(\beta)} \to \infty$. Then, with assumption (EDFCC4) retained, by (4.76) and (4.78), $z_k^{(\alpha)}, z_j^{(\beta)} \to 1$, simplifying (4.59), together with (4.79), to the following:

$$\hat{\alpha}_k = \sum_{j \in \mathcal{R}(k)} Y_{kj} \Big/ \sum_{j \in \mathcal{R}(k)} \hat{\beta}_j \tag{4.79}$$

Similarly

$$\hat{\beta}_j = \sum_{k \in \mathcal{C}(j)} Y_{kj} \Big/ \sum_{k \in \mathcal{C}(j)} \hat{\alpha}_k \tag{4.80}$$

As MAP estimators in the case of uninformative priors, these estimators are the MLEs for the EDF cross-classified fixed effects model. This result was obtained by England *et al.*, (2012, Section 2.1) in the case of uninformative uniform priors.

It is pointed out by Wüthrich, (2013) that, if all $\psi_k^{(\alpha)}, \psi_j^{(\beta)} \to \infty$, then a proper posterior distribution of parameters is not obtained. A sufficient condition for the posterior distribution to be proper is that at least one of the $\psi_k^{(\alpha)}$ be finite. Nonetheless, unique Bayes estimators may be obtained, and these may be viewed as the limiting Bayes estimators as informative priors proceed towards uninformative.

The above MLEs are well known (Hachemeister and Stanard, 1975; Renshaw and Verrall, 1998; Taylor, 2000) and are also known to be equivalent to the age-to-age factor estimators (4.17) (Verrall, 2000, p.93) if one defines

$$\hat{f}_j = 1 + \hat{g}_j = \sum_{i=1}^{j+1} \hat{\beta}_i \Big/ \sum_{i=1}^{j} \hat{\beta}_i \tag{4.81}$$

https://doi.org/10.1017/asb.2014.25 Published online by Cambridge University Press

The results of the present section may be summarised in the following theorem.

Theorem 4.2. *Consider the Bayesian EDF cross-classified model defined by (EDFCC1–3).*

(a) The MAP estimators of the α_k, β_j are given by (4.26) and (4.27). These estimators may be represented in the credibility form (4.30)–(4.37).

(b) In the case in which observations have a distribution from the Tweedie family, an empirical Bayes estimator, linear empirical Bayes estimator, and the MAP estimator of the α_k, β_j are all the same and are given by (4.59)–(4.64).

(c) In the case in which observations have an ODP distribution, subject to the further assumption (EDFCC4), the estimators (4.59) and (4.62) continue to hold and, further, the compressed data points $\bar{Y}_k^{(\alpha)}$, $\bar{Y}_j^{(\beta)}$ in these equations are given by (4.75) and (4.77), and the credibility coefficients $z_k^{(\alpha)}$, $z_j^{(\beta)}$ by (4.76) and (4.78). The compressed data points here take the same form as the MLEs of α_k, β_j in the EDF cross-classified fixed effects model. □

4.2.5. *Bornhuetter-Ferguson estimators.* Let the (a_k, b_j)-parameterisation be chosen in such a way that a_k represents expected ultimate losses of accident period k, and the b_j represent the distribution of those losses over development periods j with $\sum_{j=1}^{J} b_j = 1$.

Consider the general case of the Bayesian EDF cross-classified model. MAP forecasts of future Y_{kj} take the form

$$\hat{Y}_{kj} = \hat{\alpha}_k \hat{\beta}_j \tag{4.82}$$

with $\hat{\alpha}_k$, $\hat{\beta}_j$ the MAP estimators (4.30)–(4.37). This forecast may be expanded to the following:

$$\hat{Y}_{kj} = \left\{ \sum_{j \in \mathcal{R}(k)} z_{kj}^{(\alpha)} [Y_{kj}/\hat{\beta}_j] + \left[1 - \sum_{j \in \mathcal{R}(k)} z_{kj}^{(\alpha)} \right] a_k \right\} \hat{\beta}_j \tag{4.83}$$

Now consider the special extreme cases of $\psi^{(\alpha)} \to 0, \infty$ for all k.

Special case: $\psi_k^{(\alpha)} \to 0$ ***for all*** k***.***

In this case, the α_k are known with certainty, and (4.33) yields

$$z_{kj}^{(\alpha)} = 0 \tag{4.84}$$

and so

$$\hat{Y}_{kj} = a_k \hat{\beta}_j \tag{4.85}$$

which is of the same form as a Bornhuetter-Ferguson forecast (Bornhuetter and Ferguson, 1972).

https://doi.org/10.1017/asb.2014.25 Published online by Cambridge University Press

Special case: $\psi_k^{(\alpha)} \to \infty$ ***for all*** k**.**
This is the case of uninformative priors on the α_k, and (4.33) yields

$$\sum_{j \in \mathcal{R}(k)} z_{kj}^{(\alpha)} = 1 \tag{4.86}$$

and so

$$\hat{Y}_{kj} = \left\{ \sum_{j \in \mathcal{R}(k)} z_{kj}^{(\alpha)} [Y_{kj} / \hat{\beta}_j] \right\} \hat{\beta}_j \tag{4.87}$$

which is independent of the prior mean a_k.

It is thus seen that, as the $\psi_k^{(\alpha)}$ vary from 0 to ∞**,** forecasts vary from the Bornhuetter-Ferguson form (4.85), whose row effect depends entirely on the α_k prior, to the form (4.87) whose estimated row effect is totally data-dependent.

Special case: ODP error structure subject to (EDFCC4) and uninformative priors on the β_j**.**
In this case the α_k are estimated by

$$\hat{\alpha}_k = z_k^{(\alpha)} \bar{Y}_k^{(\alpha)} + (1 - z_k^{(\alpha)}) a_k \tag{4.59}$$

subject to (from (4.75) and (4.76))

$$\bar{Y}_k^{(\alpha)} = \sum_{j \in \mathcal{R}(k)} Y_{kj} \Big/ \sum_{j \in \mathcal{R}(k)} \hat{\beta}_j \tag{4.88}$$

$$z_k^{(\alpha)} = 1 \Big/ \left[1 + \left[\phi / \psi_k^{(\alpha)} \right] \Big/ \sum_{j \in \mathcal{R}(k)} \hat{\beta}_j \right] \tag{4.89}$$

Since $\psi_j^{(\beta)} \to \infty$, the β_j are estimated by (4.80). In the event that the α_k are also subject to uninformative priors, i.e. $\psi_k^{(\alpha)} \to \infty$, the system of estimators (4.59), (4.88) and (4.89) reduces to just (4.79). This may be summarised in the following theorem.

Theorem 4.3. *Consider the Bayesian ODP cross-classified model subject to the additional assumption (EDFCC4) and with uninformative priors on the* β_j*. The MAP estimators, empirical Bayes estimators and linear empirical Bayes estimators of the* α_k, β_j *are given by those of the ODP cross-classified fixed effects model (conventional chain ladder estimators) but with the estimators for the* α_k *replaced by (4.59), (4.88) and (4.89).* □

This MAP result has been obtained previously by England *et al.*, (2012, Section 3.1) and a similar result by England and Verrall, (2002, Section 6).

https://doi.org/10.1017/asb.2014.25 Published online by Cambridge University Press

5. Conclusion

The beginning of Section 4 summarised the current state of the literature with respect to Bayesian chain ladder models.

There are two main forms of chain ladder model: the Mack form and the cross-classified form. The former includes only column effects but the latter includes both row and column effects. In the latter case, for the most part, the only effects randomised in the literature are essentially those relating to rows.

The present paper re-considers the Bayesian EDF Mack model (Section 4.1). The earlier results of Gisler and Wüthrich (2008) assume a variance structure on claim observations that follows the Mack (fixed effects) model. Their results are reproduced and extended to a framework that considers more general variance structures.

In the case of the EDF cross-classified model, Wüthrich (2007) introduced random row effects. This paper introduces simultaneous random column effects (Section 4.2).

The model used to do so is somewhat different from that of Wüthrich, as explained in Section 4.2.1. It does not reduce to the Wüthrich model when its column effects are fixed. The essential difference between the two approaches is the use of different prior distributions on the row parameters.

The model used here does, however, reduce to that of England and Verrall (2002) and England *et al*., (2012) in the event that an ODP distribution is selected from the EDF and random column effects are replaced by fixed effects.

It would not be difficult to extend the Wüthrich model by the randomisation of column effects, though this is not done here.

Credibility estimators are just linear Bayes estimators of mean parameters that are subject to prior distributions. All results obtained here for the Bayesian chain ladder models are expressed in credibility form.

References

Bornhuetter, R. L. and Ferguson, R. E. (1972) The actuary and IBNR. *Proceedings of the Casualty Actuarial Society*, **67**, 123–184.

Bühlmann (1970) *Mathematical Methods in Risk Theory*. Berlin: Springer-Verlag.

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

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), 258–281.

Gisler, A. and Müller, P. (2007) *Credibility for additive and multiplicative models*. Paper presented to the 37th ASTIN Colloquium, Orlando FL, USA. See http://www.actuaries.org/ASTIN/Colloquia/Orlando/Presentations/Gisler2.pdf.

Gisler, A. and Wüthrich, M. V. (2008) Credibility for the chain ladder reserving method. *Astin Bulletin*, **38**(2), 565–597.

Hachemeister, C. A. and Stanard, J. N. (1975) IBNR claims count estimation with static lag functions. Spring meeting of the Casualty Actuarial Society.

https://doi.org/10.1017/asb.2014.25 Published online by Cambridge University Press

HERTIG, J. (1985) A statistical approach to the IBNR-reserves in marine insurance. *Astin Bulletin*, **15**, 171–183.
JEWELL (1974) Credible means are exact Bayesian for exponential families. *Astin Bulletin*, **8**(1), 77–90.
JEWELL, W. S. (1975) Regularity conditions for exact credibility. *Astin Bulletin*, **8**(3), 336–341.
JORGENSEN, B. and PAES DE SOUZA, M. C. (1994) Fitting Tweedie's compound Poisson model to insurance claims data. *Scandinavian Actuarial Journal*, 69–93.
LANDSMAN, Z. and MAKOV, U. (1998) Exponential dispersion models and credibility. *Scandinavian Actuarial Journal*, 89–96.
MACK, T. (1993) Distribution-free calculation of the standard error of chain ladder reserve estimates. *Astin Bulletin*, **23**(2), 213–225.
MERZ, M., WÜTHRICH, M. V. and HASHORVA, E. (2013) Dependence modelling in multivariate claims run-off triangles. *Annals of Actuarial Science*, **7**(1), 3–25.
NELDER, J. A. and WEDDERBURN, R. W. M. (1972) Generalized linear models. *Journal of the Royal Statistical Society, Series A*, **135**, 370–384.
RENSHAW, A. E. and VERRALL, R. J. (1998) A stochastic model underlying the chain-ladder technique. *British Actuarial Journal*, **4**(iv), 903–923.
SHI, P., BASU, S. and MEYERS, P. P. (2012) A Bayesian log-normal model for multivariate loss reserving. *North American Actuarial Journal*, **16**(1), 29–51.
TAYLOR, G. (2000) *Loss Reserving: An Actuarial Perspective*. Kluwer Academic Publishers, Boston.
TAYLOR, G. (2009) The chain ladder and Tweedie distributed claims data. *Variance*, **3**, 96–104.
TAYLOR, G. (2011) Maximum likelihood and estimation efficiency of the chain ladder. *Astin Bulletin*, **41**(1), 131–155.
TWEEDIE, M. C. K. (1984) *An index which distinguishes between some important exponential families.* In *Statistics: Applications and New Directions.* Proceedings of the Indian Statistical Golden Jubilee International Conference (eds. J.K. Ghosh and J. Roy), pp. 579—604. Indian Statistical Institute, Kolkata, India.
VERRALL, R. J. (2000) An investigation into stochastic claims reserving models and the chain-ladder technique. *Insurance: Mathematics and Economics*, **26**(1), 91–99.
VERRALL, R. J. (2004) A Bayesian generalised linear model for the Bornhuetter-Ferguson method of claims reserving. *North American Actuarial Journal*, **8**(3), 67–89.
WÜTHRICH, M. V. (2007) Using a Bayesian approach for claims reserving. *Variance*, **1**(2), 292–301.
WÜTHRICH, M. V. and MERZ, M. (2008) *Stochastic Claims Reserving Methods in Insurance*, John Wiley & Sons Ltd, Chichester UK.
WÜTHRICH, M. V. (2012) Discussion of "A Bayesian log-normal model for multivariate loss reserving" by Shi-Basu-Meyers. *North American Actuarial Journal*, **16**(3), 398–401.
WÜTHRICH, M. V. (2013) Challenges with non-informative gamma priors in the Bayesian overdispersed Poisson reserving model. *Insurance: Mathematics and Economics*, **52**, 352–358.

GREG TAYLOR (Corresponding author)
School of Risk and Actuarial Studies
UNSW Business School
UNSW Australia
Level 6, West Lobby, UNSW Business School Building E12
UNSW Sydney 2052
Australia
Tel.: +61 421 338 448
Fax: +61 2 9869 4805
E-Mail: gregory.taylor@unsw.edu.au

https://doi.org/10.1017/asb.2014.25 Published online by Cambridge University Press