CHAIN LADDER WITH RANDOM EFFECTS

Greg Taylor
Taylor Fry Consulting Actuaries
Level 11, 55 Clarence Street
Sydney NSW 2000
Australia

Phone:  61 2 9249 2901
Fax:  61 2 9249 2999
greg.taylor@taylorfry.com.au

February 2013
Abstract

The literature on chain ladder models with random effects is surveyed. Both Mack and cross-classified forms of the chain ladder are considered, particularly with a view to extending the models in the literature.

The literature on the EDF Mack model with random effects covers only the variance structure found in the well-known Mack fixed effects model. Models with more general variance structures are considered here.

Cross-classified models include both row and column effects. However, the literature on EDF cross-classified models with random effects includes randomisation of only the row effects. The present paper extends this to randomisation of column effects also.

While the existing EDF cross-classified model with random row effects could be easily extended in this way, the extended model presented in this paper is somewhat different. The essential difference between the two approaches is the use of different prior distributions on the row parameters.

All results obtained here for the chain ladder models with random effects are expressed in credibility form.

Keywords: chain ladder, column effects, exponential dispersion family, fixed effects, Mack model, cross-classified model, over-dispersed Poisson, random effects, 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 & Stanard (1975) described a parametric model, and Mack (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. The literature of recent years contains a number of random effects models, in which some or all of the model parameters are assumed subject to a prior distribution (England & Verrall, 2002; England, Verrall & Wüthrich, 2012; Gisler & Müller, 2007; Gisler & Wüthrich, 2008; Verrall, 2000, 2004; Wüthrich, 2007; Wüthrich & Merz, 2008).
The purpose of the present paper is to consider these random effects 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, only the row effects have been randomised in the literature on random effects 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

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

The complement of this subset, representing future observations is

$$\mathcal{D}_K^c = \{Y_{kj}: 1 \leq k \leq K \text{ and } \min(J, K - k + 1) < j \leq J\}$$

$$= \{Y_{kj}: K - J + 1 < k \leq K \text{ and } k - K + 1 < j \leq J\}$$

Also let

$$\mathcal{D}_K^c = \mathcal{D}_K \cup \mathcal{D}_K^c$$

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

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

$$X_{jk} = \sum_{i=1}^{j} Y_{ki}$$

(2.1)

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

Similarly, let $\Sigma^C(j)$ denote summation over the entire column of $\mathcal{D}_K$, 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 & Wedderburn, 1972) consists of those variables $Y$ with log-likelihoods of the form

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

for parameters $\theta$ (canonical parameter) and $\phi$ (scale parameter) and suitable functions $a, \kappa$ and $\lambda$, with $a$ continuous, $\kappa$ differentiable and one-one, and $\lambda$ such as to produce a total probability mass of unity.

For $Y$ so distributed,

$$E[Y] = \kappa(\theta) \quad (2.3)$$
$$Var[Y] = a(\phi)\kappa(\theta) \quad (2.4)$$

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

$$Var[Y] = a(\phi)V(\mu) \quad (2.5)$$

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

Henceforth only the special case in which

$$a(\phi) = \phi \quad (2.6)$$

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).

2.2.2 Tweedie family

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

$$V(\mu) = \mu^p, p \leq 0 \text{ or } p \geq 1 \quad (2.7)$$

For this family,

$$\kappa(\theta) = (2 - p)^{-1}[(1 - p)\theta]^{(2-p)/(1-p)} \quad (2.8)$$

$$\mu(\theta) = [(1 - p)\theta]^{1/(1-p)} \quad (2.9)$$

$$\ell(y; \mu, \phi) = \left[y\mu^{1-p}/(1 - p) - \mu^{2-p}/(2 - p)\right]/\phi + \lambda(y, \phi) \quad (2.10)$$

$$\frac{\partial \ell}{\partial \mu} = (y\mu^{1-p} - \mu^{1-p})/\phi \quad (2.11)$$
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 $\mu, \phi, p, \lambda$.

### 2.2.3 Over-dispersed Poisson family

The **over-dispersed Poisson** family is the Tweedie sub-family with $p = 1$. The limit of (2.9) as $p \to 1$ gives

$$E[Y] = \mu(\theta) = \exp \theta$$

(2.12)

By (2.5) – (2.7),

$$Var[Y] = \phi \mu(\theta)$$

(2.13)

By (2.11),

$$\frac{\partial \ell}{\partial \mu} = \frac{y - \mu}{\phi \mu}$$

(2.14)

### 2.2.4 Natural conjugate prior

The natural conjugate prior log-likelihood associated with the EDF member (2.2) is (Landsman & Makov, 1998)

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

(2.15)

for specific constants $m, \psi$ and functions $\kappa(.), \lambda^*(.).$ If $\Theta$ denotes the domain of $\theta$, 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 $\theta$ and $\kappa(\theta)$ in (2.15) according to (2.8) and (2.9), giving $Tw^*(m, \psi; p, \lambda^*)$ with log-likelihood

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

(2.16)

One may calculate $E[\kappa(\theta)]$ and $E[\kappa^2(\theta)]$ from (2.15). Begin by noting that

$$\frac{\partial L}{\partial \theta} = \frac{m - \kappa(\theta)}{\psi} L(\theta)$$

(2.17)

$$\frac{\partial^2 L}{\partial \theta^2} = \left\{\frac{[m - \kappa(\theta)]^2}{\psi^2} - \frac{\kappa(\theta)}{\psi}\right\} L(\theta)$$

(2.18)

With recognition of the above boundary condition, integration of $\partial L/\partial \theta$ over $\Theta$ yields
\[ 0 = L(\theta)|_{\theta \in \theta_0} = \int_{\theta} \frac{m - \kappa(\theta)}{\psi} L(\theta) d\theta = \frac{m - E[\kappa(\theta)]}{\psi} \]

from which follows

\[ E[\kappa(\theta)] = m \quad (2.19) \]

This result was quoted by Landsman & Makov (1998).

Similarly, integration of \( \partial^2 L / \partial \theta^2 \) instead of \( \partial L / \partial \theta \) leads to

\[ \int_{\theta} \left\{ \frac{[m - \kappa(\theta)]^2}{\psi^2} - \frac{\kappa(\theta)}{\psi} \right\} L(\theta) d\theta = 0 \quad (2.20) \]

and hence, with recognition of (2.19),

\[ E_\theta[\kappa(\theta)] = \frac{\text{Var}_\theta[\kappa(\theta)]}{\psi} \quad (2.21) \]

**Bayes estimation of \( \theta \)**

This prior converts (2.2) into the compound log-likelihood

\[ \ell(y; \theta, m, \phi, \psi) = (y\phi^{-1} + m\psi^{-1})\theta - (\phi^{-1} + \psi^{-1})\kappa(\theta) \quad (2.22) \]

where terms independent of \( \theta \) have been omitted.

Hence the posterior likelihood takes the form

\[ \ell(\theta|y; m, \phi, \psi) = (y\phi^{-1} + m\psi^{-1})\theta - (\phi^{-1} + \psi^{-1})\kappa(\theta) \quad (2.23) \]

up to a normalising constant. If this is re-written as

\[ \ell(\theta|y; m, \phi, \psi) = [(y\phi^{-1} + m\psi^{-1})(\phi^{-1} + \psi^{-1})^{-1}\theta - \kappa(\theta)]/(\phi^{-1} + \psi^{-1})^{-1} \quad (2.24) \]

then comparison with (2.15), with (2.19) taken into account, leads to the conclusion that

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

\[ = zy + (1 - z)m \quad (2.25) \]

with

\[ z = \phi^{-1}/(\phi^{-1} + \psi^{-1}) = 1/(1 + \phi/\psi) \quad (2.27) \]

This is essentially the result obtained by Landsman & Makov (1998, Theorem 1).
**Linear Bayes (credibility) estimation of \( \theta \)**

Result (2.26) is also a credibility estimator. Consider the linear Bayes estimator

\[
\hat{L}(y) = \arg \min_L E_{\theta \mid y} [L(y) - \mu(\theta)]^2
\]

where \( L(y) \) is linear in \( y \) and independent of \( \theta \), and the likelihood of \( Y \) and prior on \( \theta \) are not necessarily EDF but are unrestricted. The quantity \( L(y) \) is known as a **credibility estimator** (or **linear Bayes estimator**) of \( \mu(\theta) \) (Bühlmann, 1970).

It is known (Jewell, 1974, 1975) that \( L(y) \) is given by the posterior mean \( E[\mu(\theta) \mid y] \) when this linear in \( y \). Thus (2.26), 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 linear also linear Bayes, it is sometimes referred to as an **exact credibility estimator**.

**MAP estimation of \( \theta \)**

To obtain the **maximum a posteriori (MAP)** estimator of \( \theta \), differentiate (2.23):

\[
\partial \ell(\theta \mid y; m, \phi, \psi) / \partial \theta = (y\phi^{-1} + m\psi^{-1}) - (\phi^{-1} + \psi^{-1}) \kappa(\theta)
\]

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

\[
\kappa(\theta) = zy + (1 - z)m
\]

In summary then, (2.26) 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 \( \theta \) conditioned by a single observation \( Y = y \). Consider a sample \( \{Y_1, ..., Y_n\} = \{y_1, ..., y_n\} \), which will also be denoted temporarily by \( Y = y \), and suppose that \( Y_i \sim EDF(\theta, \phi_i; \kappa, \lambda) \). Use (again temporarily) \( \phi \) to denote \( \{\phi_1, ..., \phi_n\} \)

The posterior likelihood (2.23) becomes

\[
\ell(\theta \mid y; m, \phi, \psi) = (\sum_{i=1}^n y_i \phi_i^{-1} + m\psi^{-1})\theta - (\sum_{i=1}^n \phi_i^{-1} + \psi^{-1})\kappa(\theta)
\]

up to a normalising constant, and then (2.25)-(2.27) are replaced by the following:

\[
E[\kappa(\theta) \mid y; m, \phi, \psi] = z\bar{y} + (1 - z)m
\]

with
\[ \bar{y} = \sum_{i=1}^{n} y_i \phi_i^{-1} / \sum_{i=1}^{n} \phi_i^{-1} \]  
(2.33)

\[ z = 1 \left[ 1 + \psi^{-1} / \sum_{i=1}^{n} \phi_i^{-1} \right] \]  
(2.34)

The estimator on the right side of (2.32) 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, i.e. \( Y_{k_{1,j_1}}, Y_{k_{2,j_2}} \) are stochastically independent if \( k_1 \neq k_2 \).

(EDFM2) For each \( k = 1,2,...,K \), the \( X_{k,j} \) (\( j \) varying) form a Markov chain.

(EDFM3) For each \( k = 1,2,...,K \) and \( j = 1,2,...,J-1 \), define \( G_{kj} = Y_{k,j+1}/X_{kj} \) and suppose that \( G_{kj} \sim EDF(\theta_j, \phi_{kj}(X_{kj}); \kappa, \lambda) \) for some functions \( \kappa, \lambda \) 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 \( \theta \) depends only on \( j \) but the dispersion parameter \( \phi_{kj} \) may depend on \( X_{kj} \). The parameters \( f_j \) from (EDFM3) 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 \( E[Y_{k,j+1}|X_{kj}] = X_{kj} E[G_{kj}] \), then, by (EDFM3) and (2.3),

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

with

\[ g_j = f_j - 1 > 0 \]  
(3.2)

Thus

\[ \theta_j = c(\mu_{k,j+1}/X_{kj}) = c(g_j) \]  
(3.3)
with \( c = (\kappa')^{-1} \).

By (2.4) and (2.6),

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

The original Mack (1993) stochastic version of the chain ladder model included an assumption that \( \text{Var}\left[X_{k,j+1} | X_{kj}\right] = \text{Var}\left[Y_{k,j+1} | X_{kj}\right] \) 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 (EDFM1-2) but (EDFM3) took a different form, subject to the following distribution:

(EDFM3) \( Y_{k,j+1} | X_{kj} \sim \text{EDF}(\theta_{kj}(X_{kj}), \phi_{kj}(X_{kj}); \kappa, \lambda) \) for some functions \( \kappa, \lambda \) 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}

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

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 \( Y_{k,j+1} | X_{kj} \sim \text{Tw}(p) \). In this case, (3.6) yields (making use of (3.5) and (2.7))

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

compared with

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

from (3.4).

In the general case of \( Y_{k,j+1} | X_{kj} \sim \text{EDF} \), the variance (3.6) does not necessarily factor into separate terms involving \( g_j \) and \( X_{kj} \) respectively whereas (3.4) does so.
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 $Y_{kj} \in \mathcal{D}^+_k$ are stochastically independent.

(EDFCC2) For each $k = 1, 2, \ldots, K$ and $j = 1, 2, \ldots, J - 1$,
   
   (a) $Y_{kj} \sim \text{EDF} \left( \theta_{kj}, \varphi_{kj}; \kappa, \lambda \right)$ for some functions $\kappa, \lambda$ 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) $\sum_{j=1}^{J-1} \beta_j = 1$.

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),

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

Thus

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

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

The parameters $\alpha_k, \beta_j$ 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. Chain ladder models with random effects

The fixed effects models defined in Section 3 may be converted to random effects models by randomisation of their parameters, the $f_j$ in the EDF Mack model, and the $\alpha_k, \beta_j$ in the case of the EDF cross-classified model.

Some of these random effects models appear in the prior literature, as indicated by Figure 4-1.

England, Verrall & Wüthrich (2012) introduce randomisation of column effects in cross-classified models but, for the most part, use uninformative priors. Otherwise, randomisation of column effects in cross-classified models appears absent from the literature. Hence their omission from Figure 4-1.

4.1 Mack models

The EDF Mack fixed effects model can be converted to a random effects model by the retention of (EDFM1-3) above and the addition of the further condition:

(EDFM4) For each $k = 1,2,..,K$ and $j = 1,2,..,l - 1$, $g_j$ is a random effect subject to the distribution $\theta_{kj} \sim EDF^*(m_{kj}, \psi_j; \kappa, \lambda^*)$. 
This will be referred to as **EDF random effects Mack model**. Note that, by (3.3), (EDFM4) amounts to a prior on \( g_j \) and, by (2.19) 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 (EDFM4), the conclusion at the end of Section 2.2.4 holds but in respect of observations \( G_{ij} \) (fixed \( j \)) in place of the \( y_i \). In particular, by (2.32), (3.1) and (4.1), the Bayes estimator of \( g_j \) is

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

where

\[
\hat{g}_j = \sum_{i=1}^{K-j} G_{ij} \phi_{ij}^{-1} \left/ \sum_{i=1}^{K-j} \phi_{ij}^{-1} \right. = \sum_{i=1}^{K-j} (Y_{i,j+1}/X_{ij}) \phi_{ij}^{-1} \left/ \sum_{i=1}^{K-j} \phi_{ij}^{-1} \right. \tag{4.3}
\]

\[
z_j = 1 \left[ 1 + \psi_j^{-1} \sum_{i=1}^{K-j} \phi_{ij}^{-1} \right]^{-1} \tag{4.4}
\]

the last of these relations from (2.34).

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

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

and so

\[
E_{\theta_j} Var[G_{kj} | \theta_j, X_{kj}] = \phi_{kj}(X_{kj}) E[\kappa (\theta_j)] \tag{4.6}
\]

Now substitute (4.6) into the left side of (2.21), and (4.1) into the right side, to obtain

\[
E_{\theta_j} Var[G_{kj} | \theta_j, X_{kj}] / \phi_{kj}(X_{kj}) = Var_{\theta_j} E[G_{kj} | \theta_j] / \psi_j \tag{4.7}
\]

Substitution of this result into (4.4) yields a credibility coefficient of familiar appearance:

\[
z_j = 1 \left[ 1 + 1 / \sum_{i=1}^{K-j} v_{ij}^2 \right] \tag{4.8}
\]
4.1.2 Special case: Tweedie likelihood

Consider the special case of (EDFM3) $G_{kj} \sim Tw\left(g_j, \phi_{kj}(X_{kj}); p, \lambda \right)$, and suppose that (EDFM4) still holds, i.e. natural conjugate prior to Tweedie. Then substitution of (2.5)-(2.7) into (4.9) yields

$$v_{ij}^2 = \frac{\text{var}_{\theta_j} E[G_{kj}|\theta_j]}{E_{\theta_j} \text{var}[G_{kj} | \theta_j X_{kj}]}$$

(4.10)

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

$$v_{ij}^2 = \frac{\text{var}_{\theta_j} E[G_{kj}|\theta_j]}{m_j \phi_{kj}(x_{kj})}$$

(4.11)

when (2.3) and (2.19) 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

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

(4.12)

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

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

(4.13)

with

$$\tau_j^2 = \frac{\sigma_j^2}{\kappa} \left(\theta_j\right)$$

(4.14)

and (4.9) now reads

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

(4.15)

and (4.8) becomes

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

where

$$v_{ij}^2 = \frac{\psi_j}{\phi_{kj}(X_{kj})} = \frac{\text{var}_{\theta_j} E[G_{kj}|\theta_j]}{E_{\theta_j} \text{var}[G_{kj} | \theta_j X_{kj}]}$$

(4.9)
Moreover, (4.3) becomes
\[ \hat{g}_j = \sum_{i=1}^{K-j} \frac{Y_{i,j+1}}{\sum_{i=1}^{K-j} X_{ij}} \]
(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 & 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 EDF Mack random effects model defined by (EDFM1-4). The Bayes estimator, the linear Bayes estimator, and the MAP estimator of \( \hat{g}_j \) 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 2.2.4.

4.2 Cross-classified models

4.2.1 Model definition
The EDF cross-classified fixed effects model can be converted to a random effects model by the retention of (EDFCC1,2a-b) above and the addition of the further conditions:

(EDFCC3a) For each \( k = 1, 2, \ldots, K \), \( \alpha_k \) is a random effect subject to the distribution \( c(\alpha_k) \sim EDF^*(a_k, \psi^{(\alpha)}_k; \kappa, \lambda^{(\alpha)}) \).
(EDFCC3b) For each \( j = 1, 2, \ldots, J \), \( \beta_j \) is a random effect subject to the distribution \( c(\beta_j) \sim EDF^*(b_j, \psi^{(\beta)}_j; \kappa, \lambda^{(\beta)}) \).
(EDFCC3c) The variates \( \alpha_k \) and \( \beta_j \) are stochastically independent.

This model will be referred to as **EDF cross-classified random effects model**, with (EDFCC3a,b) amounting to priors on \( \alpha_k, \beta_j \) respectively. Note that this model does not include the condition (EDFCC2c). The variable parameters \( \alpha_k, \beta_j \) 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 $c$,

$$E[\kappa'(\theta_{kj})] = E[c^{-1}(\theta_{kj})] = E[a_k \beta_j] = E[a_k]E[\beta_j]$$  \hspace{1cm} (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 & 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 $j = 1,2,\ldots,J-1$, $Y_{kj}/a_k \beta_j \sim EDF(\theta_k, \varphi_{kj}; \kappa, \lambda)$ for some parameters $a_k, \beta_j > 0$ and some functions $\kappa, \lambda$ that do not depend on $j$ and $k$; and

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

The EDF cross-classified random effects model (EDFCC1,2,3a-c) directly randomises the row and column parameters $a_k, \beta_j$, which are thus assumed unknown. The Wüthrich model (EDFCC1,2*,3*), on the other hand, incorporates a multiplicative random row effect $\theta_k$, 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).

This is not a great difference when the EDF is reduced to the Tweedie sub-family. Here (EDFCC2*) takes the form $Y_{kj}/a_k \beta_j \sim Tw(\theta_k, \varphi_{kj}; \psi^*, \lambda^*)$. Since members of this sub-family satisfy the scale transformation property (Jorgensen & Paes de Souza, 1994), (EDFCC2*) may be re-expressed in that case in the form

$$Y_{kj} \sim Tw(\theta_k a_k \beta_j, \varphi_{kj}(a_k \beta_j)^p; \psi^*, \lambda^*)$$  \hspace{1cm} (4.19)

In this model, the quantity $\theta_k a_k$ may be viewed as a single randomised row parameter, much the same as in (EDFCC3a).

A more significant difference between the two models is that, whereas the random parameter $\theta_k a_k$ in (4.19) is assumed to have a prior distribution that is natural conjugate to the Tweedie, the model (EDFCC1,2,3a-c), specifically (EDFCC3a), assumes $c(\alpha_k)$ to be subject to this natural conjugate.

The choice between this model and that of Wüthrich thus provides a choice of prior distribution. In the Tweedie special case of ODP ($p = 1$), for example, the natural conjugate prior is gamma and $c(\alpha_k) = \ln \alpha_k$. This means that model (EDFCC1,2,3a-c) assigns a log gamma distribution to $\alpha_k$, whereas the
Wüthrich model assigns a gamma distribution to its corresponding parameter $\theta_k \alpha_k$.

### 4.2.2 Parameter estimation

**General EDF**

Let $R(k)$ and $C(j)$ denote the $k$-th row and $j$-th column respectively of $D_K$. If $\alpha, \beta$ denote the vectors of values of $\alpha_k, \beta_j$ respectively; $a, b$ the vectors of values of $a_k, b_j$; $\psi^{(a)}, \psi^{(b)}$ the vectors of values of $\psi_k^{(a)}, \psi_j^{(b)}$; and $\phi$ the trapezoids of values $\varphi_k, j$; then, by (2.2), (2.6) and (2.15), the posterior log-likelihood for $\alpha$, conditional on $D_K, \beta$ is

$$\ell'(\alpha|D_K, \beta; a, b, \phi, \psi^{(a)}, \psi^{(b)}) = \sum_{R(k)} [Y_{kj} c'(\mu_{kj}) - d'(\mu_{kj})] \beta_j / \varphi_{kj} + \sum_{k=1}^{K} [c(\alpha_k) a_k - d(\alpha_k)] / \psi_k^{(a)}$$

(4.20)

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

$$d(u) = \kappa(c(u))$$

(4.21)

Bayes, linear Bayes, and MAP estimation of the $\alpha_k, \beta_j$ will be considered. Little progress appears possible with the first two of these in the case of general $\kappa$. Some progress may be made, however, with MAP estimation.

Differentiate (4.20):

$$\partial \ell / \partial \alpha_k = \sum_{R(k)} [Y_{kj} c'(\mu_{kj}) - d'(\mu_{kj})] \beta_j / \varphi_{kj} + [c'(\alpha_k) a_k - d'(\alpha_k)] / \psi_k^{(a)}$$

(4.22)

$$= \sum_{R(k)} \beta_j c'(\mu_{kj}) [Y_{kj} - \mu_{kj}] / \varphi_{kj} + c'(\alpha_k) [a_k - \alpha_k] / \psi_k^{(a)}$$

since, by (4.21),

$$d'(\mu) = \kappa'(c(\mu)) c'(\mu) = \mu c'(\mu)$$

(4.23)

where the last equality derives from the definition of $c$.

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

$$\partial \ell / \partial \alpha_k = \sum_{R(k)} \frac{\beta_j c'(\mu_{kj})}{\varphi_{kj}} Y_{kj} + \frac{c'(\alpha_k)}{\psi_k^{(a)}} a_k - \alpha_k \left[ \sum_{R(k)} \frac{\beta_j c'(\mu_{kj})}{\varphi_{kj}} \beta_j / \psi_k^{(a)} \right]$$

(4.24)

Setting (4.24) equal to zero yields the following MAP estimator for $\alpha_k$:
\[
\hat{\alpha}_k = \frac{\sum_{(k)} \beta_j c'(\mu_{kj}) Y_{kj} + c'(\hat{\alpha}_k) \alpha_k}{\sum_{(k)} \beta_j c'(\mu_{kj}) + c'(\alpha_k) \psi_k^{(\alpha)}}
\]

where \( \tilde{\mu}_{kj} = \hat{\alpha}_k \beta_j \) is the estimator of \( \mu_{kj} \).

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

\[
\hat{\alpha}_k = \frac{\sum_{(k)} \beta_j c'(\tilde{\mu}_{kj}) Y_{kj} + c'(\hat{\alpha}_k) \alpha_k}{\sum_{(k)} \beta_j c'(\tilde{\mu}_{kj}) + c'(\alpha_k) \psi_k^{(\alpha)}}
\]

with \( \tilde{\mu}_{kj} = \hat{\alpha}_k \hat{\beta}_j \) now the estimator of \( \mu_{kj} \).

Similar reasoning produces the following estimator for \( \beta_j \):

\[
\hat{\beta}_j = \frac{\sum_{(j)} \hat{\alpha}_k c'(\tilde{\mu}_{kj}) Y_{kj} + c'(\hat{\beta}_j) b_j}{\sum_{(j)} \hat{\alpha}_k c'(\tilde{\mu}_{kj}) + c'(\beta_j) \psi_j^{(\beta)}}
\]

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

\[
\hat{\alpha}_k^{(n+1)} = \frac{\sum_{(k)} \beta_j^{(n)} c'(\tilde{\mu}_{kj}^{(n)}) Y_{kj} + c'(\hat{\alpha}_k^{(n)}) \alpha_k}{\sum_{(k)} \beta_j^{(n)} c'(\tilde{\mu}_{kj}^{(n)}) + c'(\alpha_k^{(n)}) \psi_k^{(\alpha)}}
\]

\[
\hat{\beta}_j^{(n+1)} = \frac{\sum_{(j)} \hat{\alpha}_k^{(n)} c'(\tilde{\mu}_{kj}^{(n)}) Y_{kj} + c'(\hat{\beta}_j^{(n)}) b_j}{\sum_{(j)} \hat{\alpha}_k^{(n)} c'(\tilde{\mu}_{kj}^{(n)}) + c'(\beta_j^{(n)}) \psi_j^{(\beta)}}
\]

where \( \hat{\alpha}_k^{(n)} \), \( \hat{\beta}_j^{(n)} \), \( \tilde{\mu}_{kj}^{(n)} \) denote the n-th iterates of \( \hat{\alpha}_k \), \( \hat{\beta}_j \), \( \tilde{\mu}_{kj} \) respectively with \( \tilde{\mu}_{kj}^{(n)} = \hat{\alpha}_k^{(n)} \hat{\beta}_j^{(n)} \).
These estimators of $\alpha_k, \beta_j$ may be represented in the familiar credibility format:

$$\hat{\alpha}_k = \sum_{R(k)} z_{kj}^{(\alpha)} \left[ Y_{kj} / \hat{\beta}_j \right] + \left[ 1 - \sum_{R(k)} z_{kj}^{(\alpha)} \right] a_k$$

(4.30)

with

$$z_{kj}^{(\alpha)} = \frac{\hat{\beta}_j^2 c'(\hat{\mu}_{kj})}{\varphi_{kj}} \left[ \sum_{R(k)} \frac{\hat{\beta}_j^2 c'(\hat{\mu}_{kj})}{\varphi_{kj}} + \frac{c'(\hat{\alpha}_k)}{\psi_k^{(\alpha)}} \right]$$

(4.31)

$$\hat{\beta}_j = \sum_{c(i)} z_{kj}^{(\beta)} \left[ Y_{kj} / \hat{\alpha}_k \right] + \left[ 1 - \sum_{c(i)} z_{kj}^{(\beta)} \right] b_j$$

(4.32)

with

$$z_{kj}^{(\beta)} = \frac{\hat{\alpha}_k^2 c'(\hat{\mu}_{kj})}{\varphi_{kj}} \left[ \sum_{R(k)} \frac{\hat{\alpha}_k^2 c'(\hat{\mu}_{kj})}{\varphi_{kj}} + \frac{c'(\hat{\beta}_j)}{\psi_j^{(\beta)}} \right]$$

(4.33)

**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)-(4.31).

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

$$\hat{\theta}_k = \sum_{R(k)} z_{kj}^{(\theta)} \left[ Y_{kj} / \alpha_k \beta_j \right] + \left[ 1 - \sum_{R(k)} z_{kj}^{(\theta)} \right] 1$$

(4.34)

with

$$z_{kj}^{(\alpha)} = \frac{\psi}{\varphi_{kj}} \left[ 1 + \sum_{R(k)} \frac{\psi}{\varphi_{kj}} \right]$$

(4.35)

It would be possible to extend the Wüthrich model by randomising both row and column parameters. Column parameter estimators corresponding to (4.34) and (4.35) 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 Tweedie cross-classified random effects model. In this case, (2.8) and (4.21) yield

$$c(u) = u^{1-p} / (1 - p)$$

(4.36)

$$d(u) = u^{2-p} / (2 - p)$$

(4.37)
Therefore

\[ c(\alpha_k \beta_j) = c(\alpha_k) \beta_j^{1-p} = \theta_k^{(\alpha)} \beta_j^{1-p} \tag{4.38} \]

with

\[ \theta_k^{(\alpha)} = c(\alpha_k) = \alpha_k^{1-p}/(1-p) \tag{4.39} \]

Similarly

\[ d(\alpha_k \beta_j) = d(\alpha_k) \beta_j^{2-p} = \kappa(\theta_k^{(\alpha)}) \beta_j^{2-p} \tag{4.40} \]

These relations convert (4.20) to the following, reduced to the likelihood of just \( \alpha_k \) instead of the full \( \alpha \):

\[
\ell(\alpha_k|\mathcal{D}_K, \beta; a, b, \phi, \psi(\alpha), \psi(\beta)) = \sum_{R(k)} \left[ Y_{kj} \theta_k^{(\alpha)} \beta_j^{1-p} - \kappa(\theta_k^{(\alpha)}) \beta_j^{2-p} \right] / \phi_{kj} \\
+ \left[ \theta_k^{(\alpha)} \alpha_k - \kappa(\theta_k^{(\alpha)}) / \psi^{(\alpha)} \right] / \psi^{(\alpha)} \tag{4.41}
\]

which may be further converted thus:

\[
\ell(\alpha_k|\mathcal{D}_K, \beta; a, b, \phi, \psi(\alpha), \psi(\beta)) = \sum_{R(k)} \left[ (Y_{kj}/\beta_j) \theta_k^{(\alpha)} - \kappa(\theta_k^{(\alpha)}) \right] / \phi_{kj} \beta_j^{p-2} \\
+ \left[ \theta_k^{(\alpha)} \alpha_k - \kappa(\theta_k^{(\alpha)}) / \psi^{(\alpha)} \right] / \psi^{(\alpha)} \tag{4.42}
\]

This is now recognisable as of the same form as (2.31) with the correspondences set out in Table 4-1.

<table>
<thead>
<tr>
<th>(2.31)</th>
<th>(4.42)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( y_i )</td>
<td>( Y_{kj}/\beta_j )</td>
</tr>
<tr>
<td>( \phi_i )</td>
<td>( \phi_{kj} \beta_j^{p-2} )</td>
</tr>
<tr>
<td>( m )</td>
<td>( \alpha_k )</td>
</tr>
<tr>
<td>( \psi )</td>
<td>( \psi_k^{(\alpha)} )</td>
</tr>
<tr>
<td>( \theta )</td>
<td>( \theta_k^{(\alpha)} )</td>
</tr>
<tr>
<td>( \kappa(\theta) )</td>
<td>( \kappa(\theta_k^{(\alpha)}) )</td>
</tr>
<tr>
<td>( \kappa(\theta) )</td>
<td>( \kappa(\theta_k^{(\alpha)}) = \alpha_k )</td>
</tr>
</tbody>
</table>

The results of Section 2.2.4 become applicable to the present case with these correspondences. In particular, in parallel with

\[ E\left[ \kappa(\theta_k^{(\alpha)}) \right] = \alpha_k \tag{4.43} \]

But \( \kappa(\cdot) \) is obtainable from (2.8):

\[ \kappa(\theta) = [(1-p)\theta]^{1/(1-p)} \tag{4.44} \]
Substitution of (4.44) and (4.39) into (4.43) yields
\[ E[\alpha_k] = a_k \] (4.45)

Similarly,
\[ E[\beta_j] = b_j \] (4.46)

Further, in parallel with (2.32)-(2.34),
\[ E[\alpha_k | \mathcal{D}_K, \beta; a, b, \phi, \psi^{(\alpha)}, \psi^{(\beta)}] = z_k^{(\alpha)} y_k^{(\alpha)} + \left[1 - z_k^{(\alpha)}\right] a_k \] (4.47)

with
\[ \tilde{\bar{y}}_k^{(\alpha)} = \sum_{\mathcal{R}(k)} y_{kj} \beta_j^{1-p} \phi_k^{-1} \left/ \sum_{\mathcal{R}(k)} \beta_j^{2-p} \phi_k^{-1} \right. \] (4.48)

\[ z_k^{(\alpha)} = \frac{1}{1 + \left[ \psi_k^{(\alpha)} \right]^{-1} \sum_{\mathcal{R}(k)} \beta_j^{2-p} \phi_k^{-1}} \] (4.49)

Similarly, the posterior likelihood of \( \beta_j \) can be constructed in parallel with (4.42) leading to the following parallel of (4.47)-(4.49):
\[ E[\beta_j | \mathcal{D}_K, \alpha; a, b, \phi, \psi^{(\alpha)}, \psi^{(\beta)}] = z_j^{(\beta)} y_j^{(\beta)} + \left[1 - z_j^{(\beta)}\right] b_j \] (4.50)

with
\[ \tilde{\bar{y}}_j^{(\beta)} = \sum_{\mathcal{C}(j)} y_{kj} \alpha_k^{1-p} \phi_k^{-1} \left/ \sum_{\mathcal{C}(j)} \alpha_k^{2-p} \phi_k^{-1} \right. \] (4.51)

\[ z_j^{(\beta)} = \frac{1}{1 + \left[ \psi_j^{(\beta)} \right]^{-1} \sum_{\mathcal{C}(j)} \alpha_k^{2-p} \phi_k^{-1}} \] (4.52)

Estimator (4.47)-(4.49) applies to the case of known \( \beta \). Just as was the case for (2.32), it is are simultaneously the Bayes estimator, the linear Bayes estimator, and the MAP estimator of \( \alpha_k \) for known \( \beta \). Similarly for (4.50)-(4.52) as estimator of \( \beta_j \) in the case of known \( \alpha \).

One may form the **empirical Bayes** estimators
\[ \tilde{\alpha}_k = z_k^{(\alpha)} \bar{Y}_k^{(\alpha)} + \left(1 - z_k^{(\alpha)}\right) \alpha_k \]  
(4.53)

with

\[ \bar{Y}_k^{(\alpha)} = \sum_{j \in (k)} Y_{kj} \hat{\beta}_j^{1-p} \phi_{kj}^{-1} / \sum_{j \in (k)} \hat{\beta}_j^{2-p} \phi_{kj}^{-1} \]  
(4.54)

\[ z_k^{(\alpha)} = 1 / \left[ 1 + \left[ \psi_k^{(\alpha)} \right]^{-1} / \sum_{j \in (k)} \hat{\beta}_j^{2-p} \phi_{kj}^{-1} \right] \]  
(4.55)

and

\[ \hat{\beta}_j = z_j^{(\beta)} \tilde{Y}_j^{(\beta)} + \left(1 - z_j^{(\beta)}\right) b_j \]  
(4.56)

with

\[ \tilde{Y}_j^{(\beta)} = \sum_{c \in (j)} Y_{kj} \tilde{\alpha}_k^{1-p} \phi_{kj}^{-1} / \sum_{c \in (j)} \tilde{\alpha}_k^{2-p} \phi_{kj}^{-1} \]  
(4.57)

\[ z_j^{(\beta)} = 1 / \left[ 1 + \left[ \psi_j^{(\beta)} \right]^{-1} / \sum_{c \in (j)} \tilde{\alpha}_k^{2-p} \phi_{kj}^{-1} \right] \]  
(4.58)

These are also the MAP estimators of \( \alpha_k, \beta_j \).

The estimators (4.53)-(4.58) are implicit in that the estimator of \( \alpha_k \) depends on the estimators of the \( \beta_j \), which in turn depend on the estimator of \( \alpha_k \). Their computation needs to be iterative, according to the following schema in which \( \tilde{\alpha}_k^{[n]}, \tilde{\beta}_j^{[n]} \) are the \( n \)-th iterates of \( \alpha_k, \beta_j, n = 1,2, etc \):

\[ \tilde{\alpha}_k^{[n+1]} = z_k^{(\alpha)[n]} \bar{Y}_k^{(\alpha)[n]} + \left(1 - z_k^{(\alpha)[n]}\right) \alpha_k \]  
(4.59)

with

\[ \bar{Y}_k^{(\alpha)[n]} = \sum_{j \in (k)} Y_{kj} [\tilde{\beta}_j^{[n]}]^{1-p} \phi_{kj}^{-1} / \sum_{j \in (k)} [\tilde{\beta}_j^{[n]}]^{2-p} \phi_{kj}^{-1} \]  
(4.60)

\[ z_k^{(\alpha)[n]} = 1 / \left[ 1 + \left[ \psi_k^{(\alpha)} \right]^{-1} / \sum_{j \in (k)} [\tilde{\beta}_j^{[n]}]^{2-p} \phi_{kj}^{-1} \right] \]
and

$$\hat{\beta}_{j}^{[n+1]} = z_{j}^{(\beta)[n]}\hat{y}_{j}^{(\beta)[n]} + \left(1 - z_{j}^{(\beta)[n]}\right)b_{j}$$  \hfill (4.62)

with

$$x_{j}^{(\beta)[n]} = \sum_{\ell(j)} Y_{kj}^{[n]} \left[\hat{\alpha}_{k}^{[n]}\right]^{1-p} \phi_{kj}^{-1} \left/ \sum_{\ell(j)} \left[\hat{\alpha}_{k}^{[n]}\right]^{2-p} \phi_{kj}^{-1} \right.$$  \hfill (4.63)

$$x_{j}^{(\beta)[n]} = 1 \left/ \left[1 + \left[\psi_{j}^{(\beta)}\right]^{-1} \left/ \sum_{\ell(j)} \left[\hat{\alpha}_{k}^{[n]}\right]^{2-p} \phi_{kj}^{-1} \right] \right.$$  \hfill (4.64)

### 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 ODP cross-classified random effects model. It was considered by England, Verrall & Wüthrich (2012). It may be noted that they adopt assumptions (EDFCC2,3a) as distinct from Wüthrich’s (2007) (EDFCC2*,3*)

The substitution \( p = 1 \) in the Tweedie Bayes estimators (4.53)-(4.56) gives

$$\bar{y}_{k}^{(\alpha)} = \sum_{\mathcal{R}(k)} Y_{kj}\phi_{kj}^{-1} \left/ \sum_{\mathcal{R}(k)} \hat{\beta}_{j}\phi_{kj}^{-1} \right.$$  \hfill (4.65)

$$z_{k}^{(\alpha)} = 1 \left/ \left[1 + \left[\psi_{k}^{(\alpha)}\right]^{-1} \left/ \sum_{\mathcal{R}(k)} \hat{\beta}_{j}\phi_{kj}^{-1} \right] \right.$$  \hfill (4.66)

$$\bar{y}_{j}^{(\beta)} = \sum_{\mathcal{C}(j)} Y_{kj}\phi_{kj}^{-1} \left/ \sum_{\mathcal{C}(j)} \hat{\alpha}_{k}\phi_{kj}^{-1} \right.$$  \hfill (4.67)

$$z_{j}^{(\beta)} = 1 \left/ \left[1 + \left[\psi_{j}^{(\beta)}\right]^{-1} \left/ \sum_{\mathcal{C}(j)} \hat{\alpha}_{k}\phi_{kj}^{-1} \right] \right.$$  \hfill (4.68)

Empirical Bayes estimators can be obtained from (4.53)-(4.58) by the substitution \( p = 1 \), and similarly MAP estimators from (4.59)-(4.64).
A case of particular interest arises if the following further assumption is made.

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

In this case (4.65)-(4.68) simplify further:

$$
\bar{Y}_k^{(\alpha)} = \sum_{R(k)} Y_{kj} \bigg/ \sum_{R(k)} \hat{\beta}_j
$$

(4.69)

$$
\bar{z}_k^{(\alpha)} = \frac{1}{1 + \frac{\phi}{\psi_k^{(\alpha)}}} \left( \sum_{R(k)} \hat{\beta}_j \right)
$$

(4.70)

$$
\bar{Y}_j^{(\beta)} = \sum_{C(j)} Y_{kj} \bigg/ \sum_{C(j)} \hat{\alpha}_k
$$

(4.71)

$$
\bar{z}_j^{(\beta)} = \frac{1}{1 + \frac{\phi}{\psi_j^{(\beta)}}} \left( \sum_{C(j)} \hat{\alpha}_k \right)
$$

(4.72)

**Uninformative priors**

Consider the case of uninformative priors, for which $\psi_k^{(\alpha)}, \psi_j^{(\beta)} \to \infty$. Then, with assumption (EDFCC4) retained, by (4.70) and (4.72), $\bar{z}_k^{(\alpha)}, \bar{z}_j^{(\beta)} \to 1$, simplifying (4.53), together with (4.69), to the following:

$$
\bar{\alpha}_k = \sum_{R(k)} Y_{kj} \bigg/ \sum_{R(k)} \hat{\beta}_j
$$

(4.73)

Similarly

$$
\bar{\beta}_j = \sum_{C(j)} Y_{kj} \bigg/ \sum_{C(j)} \hat{\alpha}_k
$$

(4.74)

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, Verrall & Wüthrich, (2012, Section 2.1).

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

$$
\hat{f}_j = 1 + \hat{g}_j = \sum_{i=1}^{l+1} \hat{\beta}_i / \sum_{i=1}^{l} \hat{\beta}_i
$$

(4.75)
The results of the present section may be summarised in the following theorem.

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

(a) The MAP estimators of the $\alpha_k, \beta_j$ are given by (4.26) and (4.27). These estimators may be represented in the credibility form (4.30)-(4.33).

(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 $\alpha_k, \beta_j$ are all the same and are given by (4.53)-(4.58).

(c) In the case in which observations have an ODP distribution, subject to the further assumption (EDFCC4), the estimators (4.53) and (4.56) continue to hold and, further, the “observations” $\tilde{Y}_k^{(a)}, \tilde{r}_j^{(b)}$ in these equations are given by (4.69) and (4.71), and the credibility coefficients $z_k^{(a)}, z_j^{(b)}$ by (4.70) and (4.72). The “observations” here take the same form as the MLEs of $\alpha_k, \beta_j$ in the EDF cross-classified fixed effects model.

### 4.2.5 Bornhuetter-Ferguson estimators

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

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

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

(4.76)

with $\hat{\alpha}_k, \hat{\beta}_j$ the MAP estimators (4.30)-(4.33). This forecast may be expanded to the following:

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

(4.77)

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

**Special case:** $\psi^{(a)}_k \to 0$ for all $k$.

In this case, the $\alpha_k$ are known with certainty, and (4.31) yields

$$z_{kj}^{(a)} = 0$$

(4.78)

and so

$$\hat{Y}_{kj} = \alpha_k \hat{\beta}_j$$

(4.79)
which is of the same form as a Bornhuetter-Ferguson forecast (Bornhuetter & Ferguson, 1972).

**Special case:** \( \psi_k^{(\alpha)} \rightarrow \infty \) for all \( k \).

This is the case of unininformative priors on the \( \alpha_k \), and (4.31) yields

\[
\sum_{\mathcal{R}(k)} z_{kj}^{(\alpha)} = 1 \tag{4.80}
\]

and so

\[
\hat{Y}_{kj} = \left\{ \sum_{\mathcal{R}(k)} z_{kj}^{(\alpha)} \left[ Y_{kj} / \hat{\beta}_j \right] \right\} \hat{\beta}_j \tag{4.81}
\]

which is independent of the prior mean \( \alpha_k \).

It is thus seen that, as the \( \psi_k^{(\alpha)} \) vary from 0 to \( \infty \), forecasts vary from the Bornhuetter-Ferguson form (4.79), whose row effect depends entirely on the \( \alpha_k \) prior, to the form (4.81) whose estimated row effect is totally data-dependent.

**Special case:** ODP error structure subject to (EDFCC4) and unininformative priors on the \( \beta_j \).

In this case the \( \alpha_k \) are estimated by

\[
\hat{\alpha}_k = z_k^{(\alpha)} \hat{\psi}_k^{(\alpha)} + \left( 1 - z_k^{(\alpha)} \right) \alpha_k \tag{4.53}
\]

subject to (from (4.65) and (4.66))

\[
\hat{Y}_k^{(\alpha)} = \sum_{\mathcal{R}(k)} Y_{kj} / \sum_{\mathcal{R}(k)} \hat{\beta}_j \tag{4.82}
\]

\[
z_k^{(\alpha)} = 1 / \left[ 1 + \left( \phi / \psi_k^{(\alpha)} \right) / \sum_{\mathcal{R}(k)} \hat{\beta}_j \right] \tag{4.83}
\]

Since \( \psi_j^{(\beta)} \rightarrow \infty \), the \( \beta_j \) are estimated by (4.74). In the event that the \( \alpha_k \) are also subject to unininformative priors, i.e. \( \psi_k^{(\alpha)} \rightarrow \infty \), the system of estimators (4.53), (4.82) and (4.83) reduces to just (4.73). This may be summarised in the following theorem.

**Theorem 4.3.** Consider the ODP cross-classified random effects model subject to the additional assumption (EDFCC4) and with unininformative priors on the \( \beta_j \). The MAP estimators, empirical Bayes estimators and linear empirical Bayes estimators of the \( \alpha_k, \beta_j \) are given by those of the ODP cross-
classified fixed effects model (conventional chain ladder estimators) but with the estimators for the $\alpha_k$ replaced by (4.53), (4.82) and (4.83).

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

5. **Conclusion**

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

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, the only random effects treated in the literature are essentially those relating to rows.

The present paper re-considers the EDF Mack model with random effects (Section 4.1). The earlier results of Gisler & 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 & Verrall (2002) and England, Verrall & Wüthrich (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 random effects. All results obtained here for the chain ladder models with random effects are expressed in credibility form.
References


