MAXIMUM LIKELIHOOD AND ESTIMATION EFFICIENCY OF THE CHAIN LADDER

BY

GREG TAYLOR

ABSTRACT

The chain ladder is considered in relation to certain recursive and non-recursive models of claim observations. The recursive models resemble the (distribution free) Mack model but are augmented with distributional assumptions. The non-recursive models are generalisations of Poisson cross-classified structure for which the chain ladder is known to be maximum likelihood. The error distributions considered are drawn from the exponential dispersion family.

Each of these models is examined with respect to sufficient statistics and completeness (Section 5), minimum variance estimators (Section 6) and maximum likelihood (Section 7). The chain ladder is found to provide maximum likelihood and minimum variance unbiased estimates of loss reserves under a wide range of recursive models. Similar results are obtained for a much more restricted range of non-recursive models.

These results lead to a full classification of this paper's chain ladder models with respect to the estimation properties (bias, minimum variance) of the chain ladder algorithm (Section 8).

KEYWORDS

Chain ladder, cross-classified model, completeness, exponential dispersion family, loss reserve, Mack model, maximum likelihood, minimum variance unbiased estimator, non-recursive model, over-dispersed Poisson, recursive model, sufficient statistic, Tweedie family.

1. INTRODUCTION

The chain ladder was originally introduced into actuarial practice as a heuristic loss reserving algorithm. Prior to Hachemeister & Stanard (1975) it did not even have a formulation as a stochastic model.

In more recent years some stochastic formulations have been achieved and some statistical properties of the models deduced. It is desirable to explore the major properties of these models as fully as possible.

Astin Bulletin 41(1), 131-155. doi: 10.2143/AST.41.1.2084389 © 2011 by Astin Bulletin. All rights reserved.

Such major properties include:

– whether the application of conventional statistical inference criteria, such as maximum likelihood, lead to the chain ladder algorithm as the formal estimator of loss reserve;
– whether the algorithm generates biased estimators; and
– the efficiency of the estimators, i.e. whether their variances are low or high.

Unbiasedness may not always be of central importance in statistical inference. In some fields of endeavour a biased but efficient estimator might be preferred to an unbiased but less efficient one. However, unbiasedness lies at the heart of much actuarial estimation.

Consider pricing and liability valuation as examples. Any premium is based on an attempted estimate of the long-run cost of the insured risk. If all such estimates are unbiased, then the total risk premium collected, taken over all the risks and over a sufficiently long period, will be correct.

Subject to this requirement, it is evident that an efficient estimate is preferable to an inefficient one, though with one qualification mentioned at the end of the present section.

Over recent years, two families of model underlying chain ladder estimates of loss reserves have been identified. These are (using the terminology of Verrall (2000)) the **recursive** models, in which the expectations of observations are conditioned by earlier observations, and the **non-recursive** models, in which no such conditioning occurs.

Some statistical properties of these models are known. For example, it has been known for many years that the non-recursive model of claim observations consisting of multiplicative row and column effects is maximum likelihood (ML) if the distribution of those observations is Poisson (Hachemeister & Stanard, 1975), or over-dispersed Poisson (ODP) (England & Verrall, 2002).

The recursive models were placed on a formal statistical footing by Mack (1993). He showed that the chain ladder was minimum variance unbiased among those algorithms that estimate age-to-age factors as linear combinations of empirical age-to-age factors. However, Mack's formulation was distribution-free and so relatively few statistical properties followed from it.

The present paper considers both forms of model with particular reference to the circumstances under which the chain ladder is maximum likelihood, unbiased and/or minimum variance.

In the case of a predictive model (such as the chain ladder), one is ultimately concerned with the size of prediction error rather than simply the error in the model parameter estimates. This issue is also discussed at the end of the paper.

2. Framework and notation

2.1. Claims data

It will be convenient to follow the framework and notation of Buchwalder, Bühlmann, Merz & Wüthrich (2006). They 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 $j = 1, 2, \ldots, J \le K$.

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

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

By convention, $\mathcal{D}_0$ denotes the empty set of observations.

The complement of $\mathcal{D}_K$, representing **future** observations, is

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

Also let

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

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**

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

And the full **row and column sums** (or horizontal and vertical sums)

$$\begin{aligned} H_k &= \sum_{j=1}^{\min(J, K-k+1)} Y_{kj} \\ V_j &= \sum_{k=1}^{K-j+1} Y_{kj} \end{aligned} \tag{2.2}$$

Also define, for $k = K - J + 2, \ldots, K$,

$$R_k = \sum_{j=K-k+2}^{J} Y_{kj} = X_{kJ} - X_{k,K-k+1} \tag{2.3}$$

$$R = \sum_{k=K-J+2}^{K} R_k \tag{2.4}$$

Note that R is the sum of the (future) observations in $\mathcal{D}_K^c$. It will be referred to as the total amount of **outstanding losses**. Likewise, R_k denotes the amount of outstanding losses in respect of accident period k. The objective stated earlier is to forecast the R_k and R.

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

Similarly, let $\sum^{C(j)}$ summation over the entire column j of $\mathcal{D}_K$, i.e. $\sum_{k=1}^{K-j+1}$ for fixed j. For example, (2.2) may be expressed as

$$V_j = \sum^{C(j)} Y_{kj}$$

Finally, let $\sum^{\mathcal{T}}$ denote summation over the entire trapezoid of (k,j) cells, i.e.

$$\sum^{\mathcal{T}} = \sum_{k=1}^{K} \sum_{j=1}^{\min(J,K-k+1)} = \sum_{k=1}^{K} \sum^{\mathcal{R}(k)}$$
$$= \sum_{j=1}^{J} \sum_{k=1}^{K-j+1} = \sum_{j=1}^{J} \sum^{C(j)}$$

2.2. Families of distributions

2.2.1. *Exponential dispersion family*

The **exponential dispersion family** (EDF) (Nelder & Wedderburn, 1972) consists of those variables Y with log-likelihoods of the form

$$\ell(y,\theta,\phi) = [y\theta - b(\theta)] / a(\phi) + c(y,\phi) \tag{2.5}$$

for parameters θ (canonical parameter) and ϕ (scale parameter) and suitable functions a, b and c, with a continuous, b differentiable and one-one, and c such as to produce a total probability mass of unity.

For Y so distributed,

$$E[Y] = b'(\theta) \tag{2.6}$$

$$\text{Var}[Y] = a(\phi)\, b''(\theta) \tag{2.7}$$

If μ denotes $E[Y]$, then (2.6) establishes a relation between μ and θ, and so (2.7) may be expressed in the form

$$\text{Var}[Y] = a(\phi)V(\mu) \tag{2.8}$$

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

The notation $Y \sim EDF(\theta, \phi; a, b, c)$ will be used to mean that a random variable Y is subject to the EDF likelihood (2.5).

2.2.2. *Tweedie family*

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

$$a(\phi) = \phi \tag{2.9}$$

$$V(\mu) = \mu^p,\ p \le 0 \text{ or } p \ge 1 \tag{2.10}$$

For this family,

$$b(\theta) = (2-p)^{-1}\left[(1-p)\theta\right]^{(2-p)/(1-p)} \tag{2.11}$$

$$\mu = \left[(1-p)\theta\right]^{1/(1-p)} \tag{2.12}$$

$$\ell(y;\mu,\phi) = \left[y\mu^{1-p}/(1-p) - \mu^{2-p}/(2-p)\right]/\phi + c(y,\phi) \tag{2.13}$$

$$\partial\ell/\partial\mu = \left(y\mu^{-p} - \mu^{1-p}\right)/\phi \tag{2.14}$$

The notation $Y \sim Tw(\mu, \phi, p)$ will be used to mean that a random variable Y is subject to the Tweedie likelihood with parameters μ, ϕ, p. The abbreviated form $Y \sim Tw(p)$ will mean that Y is a member of the sub-family with specific parameter p.

2.2.3. *Over-dispersed Poisson family*

The **over-dispersed Poisson** (ODP) family is the Tweedie sub-family with $p = 1$. The limit of (2.12) as $p \to 1$ gives

$$E[Y] = \mu = \exp\theta \tag{2.15}$$

By (2.8)–(2.10),

$$\mathrm{Var}[Y] = \phi\mu \tag{2.16}$$

By (2.14),

$$\partial\ell/\partial\mu = (y-\mu)/\phi\mu \tag{2.17}$$

The notation $Y \sim ODP(\mu,\phi)$ means $Y \sim Tw(\mu,\phi,1)$.

3. Chain ladder

The chain ladder was originally (pre-1975) devised as a heuristic algorithm for forecasting outstanding losses. It had no statistical foundation. The algorithm is as follows.

Define the following factors:

$$\hat{f}_j = \sum_{k=1}^{K-j} X_{k,j+1} \Big/ \sum_{k=1}^{K-j} X_{kj},\ j = 1, 2, ..., J-1 \tag{3.1}$$

Note that $\hat{f}_j$ can be expressed in the form

$$\hat{f}_j = \sum_{k=1}^{K-j} w_{kj}\left(X_{k,j+1}/X_{kj}\right) \tag{3.2}$$

with

$$w_{kj} = X_{kj} \Big/ \sum_{k=1}^{K-j} X_{kj} \tag{3.3}$$

i.e. as a weighted average of factors $X_{k,j+1}/X_{kj}$ for fixed j.

Then define the following forecasts of $Y_{kj} \in \mathcal{D}_K^c$:

$$\hat{Y}_{kj} = X_{k,K-k+1}\,\hat{f}_{K-k+1}\,\hat{f}_{K-k+2} \cdots \hat{f}_{j-2}\left(\hat{f}_{j-1} - 1\right) \tag{3.4}$$

Call these **chain ladder forecasts**. They yield the additional chain ladder forecasts:

$$\hat{X}_{kj} = X_{k,K-k+1}\,\hat{f}_{K-k+1} \cdots \hat{f}_{j-1} \tag{3.5}$$

$$\hat{R}_k = \hat{X}_{kJ} - \hat{X}_{k,K-k+1} \tag{3.6}$$

$$\hat{R} = \sum_{k=K-J+2}^{K} \hat{R}_k. \tag{3.7}$$

4. Models adapted to the chain ladder

The present section defines two families of models for which the maximum likelihood estimates (MLE) of outstanding losses are given by the chain ladder forecasts (3.6) and (3.7).

In the first of these families, the model, called a **recursive model**, takes the general form

$$E\left[X_{k,j+1}\middle|X_{kj}\right] = \text{function of } \mathcal{D}_{k+j-1} \text{ and some parameters} \tag{4.1}$$

where $\mathcal{D}_{k+j-1}$ is the data sub-array of $\mathcal{D}_K$ obtained by deleting diagonals on the right side of $\mathcal{D}_K$ until X_{kj} is contained in its right-most diagonal.

In the second family, the model, called a **non-recursive model**, takes the unconditional form

$$E\left[X_{kj}\right] = \text{function of some parameters.} \tag{4.2}$$

The terminology of recursive and non-recursive models is due to Verrall (2000).

4.1. Recursive models

Consider the following model, subsequently referred to as the **EDF Mack model**:

(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 $k = 1, 2, \ldots, K$, the X_{kj} (j varying) form a Markov chain.

(EDFM3) For each $k = 1, 2, \ldots, K$ and $j = 1, 2, \ldots, J-1$,

(a) $Y_{k,j+1} \,|\, X_{kj} \sim EDF(\theta_{kj}, \phi_{kj}; a, b, c)$ for some functions a, b, c that do not depend on j and k; and

(b) $E\left[X_{k,j+1}\middle|X_{kj}\right] = f_j X_{kj}$ for some parameters $f_j > 0$.

The parameters f_j in (EDFM3b) are referred to as **age-to-age factors**.

A special case of the EDF Mack model arises when a and b take the respective forms (2.9) and (2.11), i.e. (EDFM3a) takes the form:

$$Y_{k,j+1} \,|\, X_{kj} \sim Tw(p) \text{ for some } p \text{ that does not depend on } j \text{ and } k. \tag{4.3}$$

The resulting model will be referred to as the **Tweedie Mack model**.

A further special case is that of the Tweedie Mack model with $p = 1$, i.e. the member of the Tweedie family is ODP. This model will be referred to as the **ODP Mack model**.

Mack (1993) defined a model that included assumptions EDFM1, 2, and 3b. It did not include the distributional assumption EDFM3a, but did include a

further assumption concerning $Var\left[Y_{k,j+1} \middle| X_{kj} \right]$. That model, which was distribution free, has subsequently come to be known as "**the Mack model**".

Lemma 4.1. The following hold for any of the above Mack models:

(a) The estimators $\hat{f}_j$ are unbiased for f_j.

(b) The estimators $\hat{f}_j$ are uncorrelated.

(c) The estimators $\hat{X}_{kj}, \hat{R}_k$ and $\hat{R}$ defined by (3.5)–(3.7) are unbiased for $E\left[X_{kj} \middle| \mathcal{D}_k \right]$, $E\left[R_k \middle| \mathcal{D}_k \right]$ and $E\left[R \middle| \mathcal{D}_k \right]$ respectively.

Proof. These results were proved by Mack (1993) for the distribution free Mack model. The assumptions (EDFM1-2) and (EDFM3b) defining that model apply also to the EDF, Tweedie and ODP Mack models.

As mentioned above, the Mack model also included an assumption concerning $Var\left[Y_{k,j+1} \middle| X_{kj} \right]$. However, this was not used by Mack in the proof of propositions (a)–(c) for his model. These propositions therefore also hold for the EDF, Tweedie and ODP Mack models. □

Mack (1993) points out implicitly (p. 217) that $\hat{f}_j$ is the minimum variance estimator of f_j among those that are weighted averages of the empirical age-to-age factors $X_{k,j+1}/X_{kj}$, as in (3.1).

4.2. Non-recursive models

Consider the following model, subsequently referred to as the **EDF cross-classified model**:

(EDFCC1) The random variables $Y_{kj} \in \mathcal{D}_K^+$ are stochastically independent.

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

(a) $Y_{kj} \sim EDF\left(\theta_{kj}, \phi_{kj}; a, b, c\right)$ for some functions a, b, c that do not depend on j and k;

(b) $E\left[Y_{kj} \right] = \alpha_k \beta_j$ for some parameters $\alpha_k, \beta_j > 0$; and

(c) $\sum_{j=1}^{J} \beta_j = 1$.

The EDF cross-classified model may be specialised to the **Tweedie cross-classified model** and the **ODP cross-classified model** in the same way as for the Mack model of Section 4.1.

A version of the Tweedie cross-classified model appears in Wüthrich & Merz (2008, p. 191). Here a member of the Tweedie family is generated by the assumption that Y_{kj} represents claim payments and is compound Poisson with gamma severity distribution. Such compound Poisson distributions are known to coincide with the Tweedie sub-family for which $1 < p < 2$ (Jorgensen & Paes de Souza, 1994).

The following result is due to Taylor (2009). A similar result appears in Peters, Shevchenko & Wüthrich (2009).

Lemma 4.2. In the case when $\phi_{kj} = \phi$, independent of k and j, the maximum likelihood (ML) equations for the Tweedie cross-classified model are:

$$\sum^{R(k)} \mu_{kj}^{1-p} \left(Y_{kj} - \mu_{kj} \right) = 0, \quad k = 1, 2, ..., K \tag{4.4}$$

$$\sum^{C(j)} \mu_{kj}^{1-p} \left(Y_{kj} - \mu_{kj} \right) = 0, \quad j = 1, 2, ..., J \tag{4.5}$$

where $Y_{kj} \sim Tw(p)$ and

$$\mu_{kj} = \alpha_k \beta_j \tag{4.6}$$

□

A special case of Lemma 4.2 that can be found in Hachemeister & Stanard (1975), Renshaw & Verrall (1998) and Schmidt & Wünsche (1998) is as follows, though these authors considered only Poisson, rather than ODP, Y_{kj}.

Lemma 4.3. In the case when $\phi_{kj} = \phi$, independent of k and j, the ML equations for the ODP cross-classified model are the marginal sum equations

$$\sum^{R(k)} \left(Y_{kj} - \mu_{kj} \right) = 0 \tag{4.7}$$

$$\sum^{C(j)} \left(Y_{kj} - \mu_{kj} \right) = 0 \tag{4.8}$$

The solution of these equations is the chain ladder algorithm set out in Section 3. □

4.3. Relation between recursive and non-recursive models

The models described in Sections 4.1 and 4.2 are represented diagrammatically in Figure 4.1.

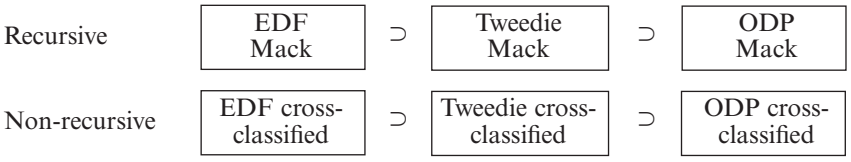

FIGURE 4.1: Families of recursive and non-recursive models.

None of the Mack models except distribution free appears to have been discussed in the literature previously, although Mack (1999) did consider the special case of a gamma distribution in (EDFM3a). Neither does the general EDF cross-classified case appear in the literature.

Wüthrich (2003) introduced the Tweedie cross-classified model, which was elaborated by Taylor (2009).

As indicated in Sections 4.1 and 4.2, the ODP cross-classified and distribution free Mack models have been discussed by a number of authors.

The relation between these last two categories of model was discussed by Verrall (2000), Mack & Venter (2000) and Verrall & England (2000). The present sub-section is not an attempt to reproduce the whole of that discussion but concentrates of the part that considers how the forecasts are conditioned by the data.

The Mack model has an inherently conditional form (see (EDFM3)) whereas the ODP cross-classified model does not (see (EDFCC2)). However, Verrall (2000) points out that, when conditioned on all past observations, the ODP cross-classified forecasts have the same dependency on those observations as do the Mack forecasts.

The specifics of this follow.

By Lemma 4.1(c), the Mack model estimator of $E\left[X_{kj}\middle|\mathcal{D}_K\right]$ is

$$\hat{X}_{kj} = X_{k,K-k+1}\,\hat{f}_{K-k+1}\,\hat{f}_{K-k+2}\cdots\hat{f}_{j-1} \tag{4.9}$$

for $K-k+1 < j \le J$. Note that $X_{k,K-k+1}$ is the sum of all observations in row k of $\mathcal{D}_k$.

Now, by (EDFCC2b), in the absence of data,

$$E\left[X_{kj}\middle|\mathcal{D}_0\right] = \alpha_k \sum_{i=1}^{j} \beta_i \tag{4.10}$$

However, Verrall (2000) shows that the MLE of $E\left[X_{k,K-k+1}\middle|\mathcal{D}_K\right]$ is $X_{k,K-k+1}$, and that the MLE of $E\left[X_{kj}\middle|\mathcal{D}_K\right]$ is $\hat{X}_{kj}$ for $K-k+1 < j \le J$.

Thus, while (4.10) gives the expectation of X_{kj} prior to the collection of any data, the MLE of that quantity changes as it is conditioned by accumulating data. This occurs in such a way that, while the distribution free Mack and ODP cross-classified models are quite distinct, their forecasts (MLE in the latter case) are identical.

5. Sufficiency and completeness

5.1. Sufficiency

5.1.1. *Recursive models*

Let $\mathcal{D}_K^+$ be subject to an EDF Mack model. By (EDFM2-3), the only information on parameter f_j will be contained in the observations in column $j+1$ of $\mathcal{D}_K$, conditional on column j.

Denote the partial conditional log-likelihood for column $j+1$ by $\ell_{j+1}(y, \theta, \phi, X)$ where y, θ, ϕ, X are the following vectors:

$$\begin{aligned} y &= y_{1,j+1}, \ldots, y_{K-j,j+1} \\ \theta &= \theta_{1,j+1}, \ldots, \theta_{K-j,j+1} \\ \phi &= \phi_{1,j+1}, \ldots, \phi_{K-j,j+1} \\ X &= X_{1j}, \ldots, X_{K-j,j} \end{aligned}$$

Then, by (2.5),

$$\begin{aligned} \ell_{j+1}(y;\theta,\phi,X) = &\sum_{k=1}^{K-j} \left[y_{k,j+1}\theta_{k,j+1} - b(\theta_{k,j+1}) \right] / a(\phi_{k,j+1}) \\ &+ \text{terms independent of } \theta \end{aligned} \tag{5.1}$$

By (2.6) and (EDFM3b),

$$\theta_{k,j+1} = (b')^{-1}(\mu_{k,j+1}) = (b')^{-1}(X_{kj} f_j) \tag{5.2}$$

where $\mu_{k,j+1}$ has been used to denote $E\left[X_{k,j+1} \middle| X_{kj}\right]$.

Substitution of (5.2) into (5.1) yields

$$\begin{aligned} \ell_{j+1} = &\sum_{k=1}^{K-j} y_{k,j+1}(b')^{-1}(X_{kj} f_j) / a(\phi_{k,j+1}) \\ &- \sum_{k=1}^{K-j} b\left[(b')^{-1}(X_{kj} f_j)\right] / a(\phi_{k,j+1}) \\ &+ \text{terms independent of } f_j \end{aligned} \tag{5.3}$$

Now call $b(.)$ **separable** if it satisfies the following two conditions:

(i) $(b')^{-1}(Xf) = \kappa_1(X)\lambda_1(f)$

(ii) $b\left[(b')^{-1}(Xf)\right] = \kappa_0(X)\lambda_0(f)$

for some functions κ_0, κ_1, λ_0, λ_1.

If these conditions hold, then (5.3) may be re-written as

$$\begin{aligned} \ell_{j+1} = &\lambda_1(f_j)\, g_{1j}(y,\phi,X) - \lambda_0(f_j)\, g_{0j}(y,\phi,X) \\ &+ \text{terms independent of } f_j \end{aligned} \tag{5.4}$$

where

$$g_{1j}(y, \phi, X) = \sum_{k=1}^{K-j} y_{k,j+1} \kappa_1(X_{kj}) / a(\phi_{k,j+1}) \tag{5.5}$$

$$g_{0j}(y, \phi, X) = \sum_{k=1}^{K-j} \kappa_0(X_{kj}) / a(\phi_{k,j+1}) \tag{5.6}$$

Application of the Fisher-Neyman factorisation theorem to (5.4) shows that $g_{1j}(Y, \phi, X)$ is a sufficient statistic for f_j.

Theorem 5.1. Let $\mathcal{D}_K^+$ be subject to an EDF Mack model. Then $g_{1j}(Y, \phi, X)$ is sufficient for f_j if any one of the following conditions holds:

(a) $Y_{k,j+1} \sim Tw(p)$;

(b) $Y_{k,j+1} \sim Bin(X_{kj}, f_j)$;

(c) $Y_{k,j+1} \sim NegBin(X_{kj}, \alpha_j)$, i.e. $\text{Prob}\left[Y_{k,j+1} = y\right] = \binom{y + X_{kj} - 1}{y} (1 - \alpha_j)^{X_{kj}}$, $y = 0, 1, 2$, etc. with $\alpha_j / (1 + \alpha_j) = f_j$.

Proof. Sufficiency will be proven if $b(.)$ is shown to be separable in each case.

(a) By (2.6) and (2.12),

$$(b')^{-1}(\mu) = \mu^{1-p} / (1 - p) \tag{5.7}$$

Then

$$(b')^{-1}(X_{kj} f_j) = X_{kj}^{1-p} \left[f_j^{1-p} / (1 - p) \right] \tag{5.8}$$

By (2.11) and (5.7),

$$b\left[(b')^{-1}(\mu)\right] = \mu^{2-p} / (2 - p)$$

Then

$$b\left[(b')^{-1}(X_{kj} f_j)\right] = X_{kj}^{2-p} \left[f_j^{2-p} / (2 - p) \right]$$

which proves separability of $b(.)$.

(b) In this case

$$b(\theta) = X_{kj} \ln(1 + e^{\theta})$$

from which

$$(b')^{-1}(\mu) = \ln\frac{\mu / X_{kj}}{1-\mu / X_{kj}}$$

Then

$$(b')^{-1}(X_{kj} f_j) = \ln\left[f_j / (1-f_j)\right] \tag{5.9}$$

which proves separability.

(c) In this case,

$$b(\theta) = -X_{kj}\ln(1-e^{\theta}) \tag{5.10}$$

with $\theta = \ln\alpha$

Then

$$(b')^{-1}(\mu) = \ln\frac{\mu / X_{kj}}{1+\mu / X_{kj}}$$

and

$$(b')^{-1}(X_{kj} f_j) = \ln\left[f_j / (1+f_j)\right] \tag{5.11}$$

proving separability. □

5.1.2. *Non-recursive models*

Let $\mathcal{D}_K^+$ be subject to a Tweedie cross-classified model. Condition EDFCC2(a) then becomes

$$Y_{kj} \sim Tw(p) \tag{5.12}$$

and the log-likelihood for $\mathcal{D}_K$ is, by (2.13),

$$\begin{aligned}\ell(y;\mu,\phi,p) = \sum^{\mathcal{T}}\left[y_{kj}\,\mu_{kj}^{1-p}/(1-p) - \mu_{kj}^{2-p}/(2-p)\right]/\phi_{kj} \\ + \text{terms independent of } \mu\end{aligned} \tag{5.13}$$

where y, μ, ϕ denote vectors in the same way as in Section 5.1.1, and p is constant over $\mathcal{D}_K^+$.

Substitute for μ_{kj} according to (EDFCC2b):

$$\begin{aligned}\ell = \sum^{\tau} y_{kj}\,\alpha_k^{1-p}\beta_j^{1-p}/(1-p)\,\phi_{kj} \\ + \text{terms independent of } y + \text{terms independent of } \mu\end{aligned}$$

With the last two members on the right side ignored, ℓ may be written in the alternative forms

$$\ell = \sum_{k=1}^{K} \alpha_k^{1-p} / (1-p) \sum^{\mathcal{R}(k)} y_{kj} \beta_j^{1-p} / \phi_{kj} \tag{5.14a}$$

$$\ell = \sum_{j=1}^{J} \beta_j^{1-p} / (1-p) \sum^{C(j)} y_{kj} \alpha_k^{1-p} / \phi_{kj} \tag{5.14b}$$

A special case arises when $p = 1$ (ODP case). Since $\lim_{p \to 1} \mu_{kj}^{1-p} / (1-p) = \ln \mu_{kj}$, (5.13) becomes

$$\begin{aligned} \ell &= \sum^{\tau} y_{kj} (\ln \alpha_k + \ln \beta_j) / \phi_{kj} \\ &= \sum_{k=1}^{K} \ln \alpha_k \sum^{\mathcal{R}(k)} y_{kj} / \phi_{kj} + \sum_{j=1}^{J} \ln \beta_j \sum^{C(j)} y_{kj} / \phi_{kj} \end{aligned} \tag{5.15}$$

where only the terms of interest in (5.13) have been retained here.

Theorem 5.2. Let $\mathcal{D}_K^+$ be subject to a Tweedie cross-classified model. Then

(a) If $p = 1$ (ODP case), $\sum^{\mathcal{R}(k)} Y_{kj} / \phi_{kj}$ is a sufficient statistic for α_k, and $\sum^{C(j)} Y_{kj} / \phi_{kj}$ for β_j.

(b) If $p \neq 1$, there is no minimal statistic for any of the parameters α_k, β_j that is a proper subset of $\mathcal{D}_K$.

Proof.

(a) This follows immediately from (5.15) and the Fisher-Neyman factorisation theorem.

(b) Consider the parameter α_k for arbitrary k. The partial log-likelihood involving this parameter is, from (5.14a),

$$\alpha_k^{1-p} / (1-p) \sum^{\mathcal{R}(k)} y_{kj} \beta_j^{1-p} / \phi_{kj} \tag{5.16}$$

It is evident that $\mathcal{D}_K$ is a sufficient statistic for α_k.

If a sufficient statistic of $\hat{\alpha}_k(Y)$ for α_k existed for Y a strict subset of $\mathcal{D}_K$, then, again by the factorisation theorem, it would be possible to express (5.16) in the form

$$\kappa(\hat{\alpha}_k; \alpha_k) + \lambda(\mathcal{D}_K)$$

for suitable functions κ, λ and specifically with κ independent of the parameters β_j.

This is evidently not the case since every y_{kj} appearing in (5.16) has a multiplier that depends on β_j. It follows that there is no minimal sufficient statistic for α_k that is a proper subset of $\mathcal{D}_K$. By a similar argument, there is also no minimal sufficient statistic for any β_j that is a proper subset of $\mathcal{D}_K$. □

A Tweedie cross-classified model contains $K + J$ parameters, but subject to constraint (EDFCC2c). Hence there are $K + J - 1$ independent parameters.

Theorem 5.2 identified $K + J$ sufficient statistics, one for each parameter. However, the fact that there are only $K + J - 1$ independent parameters suggests that not all of the $K + J$ sufficient statistics are required. Theorem 5.3 confirms this.

First, however, some notation. Let s denote the column vector of dimension $K + J$ consisting of the row sums identified in Theorem 5.2(a), followed by the column sums.

Theorem 5.3. Let $\mathcal{D}_K^+$ be subject to an ODP cross-classified model. Let s^* be the $(K+J-1)$-vector obtained by deleting an arbitrary single component of s, and let

$$s_{\min} = As^*$$

where A is any non-singular $(K + J - 1) \times (K + J - 1)$ matrix. Then $s_{\min}$ is minimal sufficient for the parameter set $\{\alpha_1, \ldots, \alpha_K, \beta_1, \ldots, \beta_J\}$.

Proof. By Theorem 5.2, s is sufficient for the parameter set $\{\alpha_1, \ldots, \alpha_K, \beta_1, \ldots, \beta_J\}$. Let the components of s be $s_1 \ldots, s_K, s_{K+1} \ldots, s_{K+J}$. By the relations at the end of Section 2.1,

$$s_1 + \ldots + s_K = s_{K+1} + \ldots + s_{K+J} \tag{5.17}$$

whence any component of s can be expressed in terms of the other components. This means that s^* contains the same information as s and is therefore sufficient for $\{\alpha_1, \ldots, \alpha_K, \beta_1, \ldots, \beta_J\}$

Since A is a one-one transformation, $s_{\min}$ is also a sufficient statistic. Its minimality is indicated by its dimension $K + J - 1$. Note that $\{\alpha, \beta\}$ has dimension $K + J - 1$, but after allowance for constraint (EDFCC2c), contains only $K + J - 1$ independent parameters.

In any event, $s_{\min}$ is shown to be a complete sufficient statistic in Theorem 5.5 and so is immediately minimal sufficient (Lehmann & Casella, 1998). □

Examples of minimal sufficient statistics

Example 5.1. Construct s^* by deleting s_{K+1} from s and select $A = I$. Then $s_{\min}$ consists of all row sums and all column sums except the first.

Example 5.2. Construct s^* by deleting s_1 from s and select $A = I$. Then $s_{\min}$ consists of all row sums except the first and all column sums.

Example 5.3. Construct s^* as in Example 5.1 and select

$$A = \left[\begin{array}{c|c|c} \underset{1\times 1}{1} & \underset{1\times(K-1)}{1} & \\ \hline & \underset{(K-1)\times(K-1)}{I} & \\ \hline & & \underset{(J-1)\times(J-1)}{I} \end{array}\right]$$

Then $s_{\min}$ consists of all row sums except the first, all column sums except the first, and the sum of all elements of $\mathcal{D}_K$. The last of these is given by the first row of A which yields the sum of all row sums.

This particular version of the minimal sufficient statistic can be found in Kuang, Nielsen & Neilsen (2009).

5.2. Completeness

5.2.1. *Recursive models*

Theorem 5.4. The statistic $g_{1j}(Y, \phi, X)$ is complete under the hypotheses of Theorem 5.1.

Proof. Note that the subject distribution is a member of the EDF and g_{1j} is sufficient for f_j, by Theorem 5.1. Note also that $\dim(g_j) = \dim(f_j) = 1$. These are necessary and sufficient conditions for g_{1j} to be complete (Cox & Hinkley, 1974, p. 31). □

5.2.2. *Non-recursive models*

Theorem 5.5. The statistic $s_{\min}$ is complete under the hypotheses of Theorem 5.3.

Proof. Note that the subject distribution is a number of the EDF and $s_{\min}$ is sufficient for $\{\alpha, \beta\}$, by Theorem 5.3. Note also that $\dim(s_{\min}) = \dim\{\alpha, \beta\} = K + J - 1$, also demonstrated in the proof of Theorem 5.3. These are necessary and sufficient conditions for $s_{\min}$ to be complete. □

6. Minimum variance estimators

6.1. Recursive models

Consider the three families of models that form the subject of Theorem 5.1 and the MLEs that derive from them.

6.1.1. *Tweedie family*

By (5.8) and the definition of κ_1, λ_1,

$$\kappa_1(X_{kj}) = X_{kj}^{1-p} \tag{6.1}$$

$$\lambda_1(f_j) = f_j^{1-p}/(1-p) \tag{6.2}$$

Then, by (5.5),

$$g_{1j}(Y, \phi, X) = \sum_{k=1}^{K-j} X_{kj}^{1-p} Y_{k,j+1} / a(\phi_{k,j+1}) \tag{6.3}$$

6.1.2. *Binomial family*

By (5.9) and the definition of κ_1, λ_1,

$$\kappa_1(X_{kj}) = 1 \tag{6.4}$$

$$\lambda_1(f_j) = \ln[f_j/(1-f_j)] \tag{6.5}$$

Moreover

$$\alpha(\phi_{k,j+1}) = 1 \tag{6.6}$$

Then, by (5.5),

$$g_{1j}(Y, \phi, X) = \sum_{k=1}^{K-j} Y_{k,j+1} \tag{6.7}$$

6.1.3. *Negative binomial family*

By (5.11) and the definition of κ_1, λ_1,

$$\kappa_1(X_{kj}) = 1 \tag{6.8}$$

$$\lambda_1(f_j) = \ln[f_j/(1+f_j)] \tag{6.9}$$

Moreover (6.6) holds here also. Then, by (5.5), (6.7) also holds for this case.

6.1.4. *Summary*

Note that for the special case of the Tweedie family in which

$$p = 1 \text{ (ODP) and} \tag{6.10}$$

$$\phi_{kj} = \phi_j \text{ (scale parameter dependent on only column)} \tag{6.11}$$

relation (6.3) reduces to

$$g_{1j}(Y, \phi, X) = \left[\sum_{k=1}^{K-j} Y_{k,j+1} \right] / a(\phi_{j+1}) \tag{6.12}$$

Lemma 6.1. Let $\mathcal{D}_K^+$ be subject to an EDF Mack model. If

(a) $Y_{k,j+1} \sim ODP(X_{kj} f_j, \phi_{k,j+1})$ subject to (6.11); or

(b) $Y_{k,j+1} \sim Bin(X_{kj}, f_j)$; or

(c) $Y_{k,j+1} \sim NegBin(X_{kj}, \alpha_j)$ with $\alpha_j / (1 + \alpha_j) = f_j$;

then there is a one-one correspondence between $g_{1j}(Y, \phi, X)$ and the chain ladder age-to-age factor estimator $\hat{f}_j$.

Proof. The form of $g_{1j}(Y, \phi, X)$ is given by (6.12), (6.7) and (6.7) for the respective cases (a)-(c). The form (6.7) is a special case of (6.12), so consider this latter, more general case.

Note that, by (3.1),

$$\begin{aligned} \hat{f}_j &= 1 + \sum_{k=1}^{K-j} Y_{k,j+1} / \sum_{k=1}^{K-j} X_{kj} \\ &= 1 + g_{1j}(Y, \phi, X)\, a(\phi_{j+1}) / \sum_{k=1}^{K-j} X_{kj} \end{aligned} \tag{6.13}$$

which is just a linear transformation of $g_{1j}(Y, \phi, X)$ depending on just the parameters ϕ_{j+1} and the Mack model's conditioning quantities X_{kj}. □

Theorem 6.2. Under the hypotheses of Lemma 6.1, the chain ladder estimates $\hat{f}_j, \hat{X}_{kj}, \hat{R}_k$ and $\hat{R}$ defined by (3.1) and (3.5)–(3.7) respectively are minimum variance unbiased estimators (MVUEs) of f_j, $E[X_{kj} | \mathcal{D}_K]$, $E[R_k | \mathcal{D}_K]$ and $E[R | \mathcal{D}_K]$.

Proof. Unbiasedness follows from Lemma 4.1. By Theorem 5.1, $g_{1j}(Y, \phi, X)$ is sufficient for f_j and so, by Lemma 6.1, $\hat{f}_j$ is sufficient for f_j. By Theorem 5.4, $g_{1j}(Y, \phi, X)$ is complete and therefore, by Lemma 6.1, so is $\hat{f}_j$.

In summary $\hat{f}_j$ is a complete sufficient statistic and an unbiased estimator of f_j. That $\hat{f}_j$ is an MVUE then follows from the Lehmann-Scheffé theorem.

Now consider $\hat{X}_{kj}$ defined by (3.5), i.e.

$$\hat{X}_{kj} = X_{k,K-k+1}\hat{f}_{K-k+1}\cdots\hat{f}_{j-1}$$

By (EDFM3b),

$$E\left[X_{kj}\middle|\mathcal{D}_K\right] = X_{k,K-k+1}f_{K-k+1}\cdots f_{j-1} \tag{6.14}$$

Thus, $\hat{X}_{kj}$ is a function of the conditioning information $X_{k,K-k+1}$ and the statistic $\{\hat{f}_{K-k+1},\ldots,\hat{f}_{j-1}\}$, while $E[X_{kj}|\mathcal{D}_K]$ is a function of the same conditioning information and the parameter $\{f_{K-k+1},\ldots,f_{j-1}\}$. As shown immediately above, $\{\hat{f}_{K-k+1},\ldots,\hat{f}_{j-1}\}$ is a complete sufficient statistic for $\{f_{K-k+1},\ldots,f_{j-1}\}$. A further application of the Lehmann-Scheffé theorem proves that $\hat{X}_{kj}$ is an MVUE of $E[X_{kj}|\mathcal{D}_K]$.

The proof that $\hat{R}_k$ and $\hat{R}$ are MVUEs of $E[R_k|\mathcal{D}_K]$ and $E[R|\mathcal{D}_K]$ are similar. □

6.2. Non-recursive models

One might consider attempting to show that MLEs are MVUEs for some of the EDF cross-classified models. However, the MLE for at least the ODP case (where the chain ladder is MLE (Lemma 4.3)) is known to be biased. Theorem 3 of Taylor (2003) shows that $\hat{X}_{kJ} \mid X_{k,K-k+1}$ is biased upward as an estimator of

$$X_{k,K-k+1}E\left[X_{kJ}\right] / E\left[X_{k,K-k+1}\right] = X_{k,K-k+1}\sum_{j=1}^{J}\beta_j \Big/ \sum_{j=1}^{K-k+1}\beta_j \tag{6.15}$$

Note that chain ladder estimates are unbiased for the recursive EDF Mack models but the **same estimators** are biased for the ODP cross-classified model.

Consider an EDF cross-classified model as defined in Section 4.2. Let $Z : \mathcal{D}_K^c \to \mathcal{R}$ and let $\hat{Z} : \mathcal{D}_K \to \mathcal{R}$ be some predictor of $Z \mid \mathcal{D}_K$. Define

$$\tilde{Z} = \hat{Z}\,E\left[Z\middle|\mathcal{D}_K\right] / E\left[\hat{Z}\,|\,\mathcal{D}_K\right] \tag{6.16}$$

Then

$$E\left[\tilde{Z}\,|\,\mathcal{D}_K\right] = E\left[Z\middle|\mathcal{D}_K\right] \tag{6.17}$$

and so $\tilde{Z}$ is a **bias corrected** form of the predictor $\hat{Z}$.

Theorem 6.3. Let $\mathcal{D}_K^+$ be subject to an ODP cross-classified model with

$$\phi_{kj} = \phi\text{, independent of } k, j \tag{6.18}$$

Then the bias corrected chain ladder estimates $\tilde{X}_{kj}$, $\tilde{R}_k$ and $\tilde{R}$ (derived from $\hat{X}_{kj}$, $\hat{R}_k$ and $\hat{R}$ defined by (3.5)-(3.7) respectively) are MVUEs of $E\big[X_{kj}\big|\mathcal{D}_K\big]$, $E\big[R_k\big|\mathcal{D}_K\big]$ and $E\big[R\big|\mathcal{D}_K\big]$.

Proof. The estimators $\hat{X}_{kj}$, $\hat{R}_k$ and $\hat{R}$ are MLE for the ODP cross-classified model (Lemma 4.3). They are defined in terms of the statistic s, defined in Section 5.1.2. It is remarked in the proof of Theorem 5.3 that this is expressible in terms of $s_{\min}$, as defined in Theorem 5.3.

By Theorems 5.3 and 5.5, $s_{\min}$ is a complete sufficient statistic for the parameter set $\{\alpha_1,\ldots,\alpha_K,\beta_1,\ldots,\beta_J\}$ of the ODP cross-classified model. Thus, $\tilde{X}_{kj}$, $\tilde{R}_k$ and $\tilde{R}$ are unbiased estimators that are functions of a complete sufficient statistic. By the Lehmann-Scheffé theorem, they are MVUEs. □

The application of Theorem 6.3 is limited by the fact that the bias correction factors in $\tilde{X}_{kj}$, etc. would rarely be known practice. On the other hand, however, the biases contained in chain ladder estimates are tolerated in practice and, in this context, the theorem shows that the chain ladder provides a minimum variance estimate of whatever it is estimating.

When the chain ladder bias is small, it provides "minimum variance almost unbiased" estimators.

7. Maximum likelihood estimation

7.1. Exponential dispersion family

Consider a sample of stochastically independent observations $\{Y_1,\ldots,Y_n\}$ each subject to an EDF likelihood (2.5). For the moment, the Y_i are general observations not related to any development trapezoid. It is assumed that b and c are the same for all observations but θ,ϕ may vary.

Thus the total log-likelihood is

$$\ell = \sum_{i=1}^{n} \ell_i = \sum_{i=1}^{n} \big[y_i\theta_i - b(\theta_i)\big] / a(\phi_i) + c(y,\phi) \tag{7.1}$$

Let μ_i denote $E[Y_i]$. Then (2.6) gives

$$\mu_i = b'(\theta_i) \tag{7.2}$$

Suppose that μ_i depends on some vector of parameters $\alpha = (\alpha_1,\ldots,\alpha_p)^T$, and write $\mu_i = \mu_i(\alpha)$. For ML estimation with respect to α,

$$\frac{\partial \ell}{\partial \alpha_j} = \sum_{i=1}^{n} \frac{\partial \ell_i}{\partial \alpha_j} = \sum_{i=1}^{n} \frac{\partial \ell_i}{\partial \theta_i}\frac{\partial \theta_i}{\partial \mu_i}\frac{\partial \mu_i}{\partial \alpha_j} = 0 \tag{7.3}$$

By (7.2),

$$\partial\theta_i / \partial\mu_i = 1 / b''(\theta_i) = 1 / V(\mu_i) \tag{7.4}$$

by comparison of (2.7) and (2.8).

Evaluation of (7.3) with (7.1), (7.2) and (7.4) taken into account yields

$$\frac{\partial \ell}{\partial \alpha_j} = \sum_{i=1}^{n} \frac{y_i - \mu_i}{a(\phi_i) V(\mu_i)} \frac{\partial \mu_i}{\partial \alpha_j} = 0, \quad j = 1, ..., p \tag{7.5}$$

7.2. Recursive models

Theorem 7.1. Suppose that $\mathcal{D}_K^+$ is subject to an EDF Mack model and that, in addition,

$$Var\left[Y_{k,j+1} \middle| X_{kj} \right] = X_{kj} \sigma_{j+1}^2 \tag{7.6}$$

for fixed parameters $\sigma_2^2, ..., \sigma_J^2$. Then the chain ladder is ML.

Proof. Apply (7.5) to the sample $\{ Y_{k,j+1} \in \mathcal{D}_K \}$ for fixed j. By (EDFM3b),

$$E\left[Y_{k,j+1} \middle| X_{kj} \right] = (f_j - 1) X_{kj} \tag{7.7}$$

Note also that in (7.5) (see (2.8))

$$a(\phi_i) V(\mu_i) = Var\left[Y_i \right] \tag{7.8}$$

In the present context, Y_i is replaced by $Y_{k,j+1} \mid X_{kj}$ and so, by (7.6), $a(\phi_k) V(\mu_k)$ is replaced by $X_{kj} \sigma_{j+1}^2$.

There is only one parameter f_j corresponding to the vector α in (7.5). Thus (7.5) yields

$$\sum_{k=1}^{K-j} \frac{Y_{k,j+1} - (f_j - 1) X_{kj}}{X_{kj} \sigma_{j+1}^2} X_{kj} = 0$$

from which $f_j = \hat{f}_j$, as defined by (3.1). □

The condition (7.6) was imposed by Mack (1993). The model of Theorem 7.1 is that usually referred to in the literature as the Mack model (from Mack, 1993) equipped with the additional EDF distributional assumption. The theorem shows that, with this distributional extension to the Mack model, the chain ladder is ML.

7.3. Non-recursive models

As noted in Section 4.2, the chain ladder has been known for some time to be ML in the case of an ODP cross-classified model. This was stated formally in Lemma 4.3.

Consider other forms of EDF cross-classified model. Application of (7.5) yields

$$\sum^{C(j)} \frac{Y_{kj} - \alpha_k \beta_j}{a(\phi_{kj}) V(\alpha_k \beta_j)} \alpha_k = 0, \quad j = 1, ..., J \tag{7.9}$$

$$\sum^{\mathcal{R}(k)} \frac{Y_{kj} - \alpha_k \beta_j}{a(\phi_{kj}) V(\alpha_k \beta_j)} \beta_j = 0, \quad k = 1, ..., K \tag{7.10}$$

In the ODP case, $V(\alpha_k \beta_j) = \alpha_k \beta_j$, cancelling the multiplier α_k in (7.9) and the multiplier β_j in (7.10), and leading to the marginal sum equations (4.7) and (4.8), in the case when $\phi_{kj} = \phi$, independent of k and j. According to Lemma 4.3, these equations yield a chain ladder solution.

In EDF cases other than ODP, $V(\mu) \neq \mu$, and so these cancellations do not occur, the solution is not given by (4.7) and (4.8) and so is not chain ladder.

This reasoning allows the statement of Lemma 4.3 to be extended as follows.

Theorem 7.2. Suppose that $\mathcal{D}_K^+$ is subject to an EDF cross-classified model. Chain ladder estimates are ML if and only if the model is ODP with weights $\phi_{kj} = \phi$, independent of k and j. □

8. Parameter error and prediction error

The models of R_k considered in this paper are generally of the form

$$R_k = a(\theta) + \varepsilon \tag{8.1}$$

where

θ is a parameter vector;
$a(.)$ is some function; and
ε is a centred stochastic error.

The vector θ is estimated by $\hat{\theta}$ and R_k by

$$\hat{R}_k = a(\hat{\theta}) \tag{8.2}$$

The mean square error of prediction of R_k is defined as

$$MSEP[\hat{R}_k] = E[R_k - \hat{R}_k]^2 \tag{8.3}$$

The MSEP may be decomposed as follows:

$$MSEP[\hat{R}_k] = E[a(\hat{\theta}) - a(\theta)]^2 + E[R_k - a(\theta)]^2 \tag{8.4}$$

where the two components on the right are referred to as **parameter error** and **process error** respectively (see e.g. Taylor, 2000, pp. 192-194).

Now the process error is independent of the estimator $\hat{\theta}$, and so minimisation of the parameter error is equivalent to minimisation of the MSEP. This type of proof leads to the following general result.

Theorem 9.1. If a forecast of some function of $\mathcal{D}_K^c$ is MVUE, then its MSEP is minimised among unbiased forecasts.

Functions of $\mathcal{D}_K^c$ include Y_{kj}, R_k and R.

9. Conclusion

This paper set out to consider the question of for which claim models the chain ladder is maximum likelihood or minimum variance. The models considered are the recursive and non-recursive models. In each case, stochastic error terms drawn from the EDF are considered.

The model that has come to be known as the Mack model (Mack, 1993) is distribution free. However, if it is equipped with EDF stochastic errors, the chain ladder becomes ML. This means that the chain ladder is ML for a wide range of distributions, eg Poisson, binomial, negative binomial, normal, gamma, etc.

The negative binomial case is of some interest here. Verrall (2000) considered the case of a Bayesian version of the Poisson Mack model (defined in Section 4.1) in which the chain ladder factors f_j are subject to a gamma prior. Verrall shows that the posterior distribution of $Y_{k,j+1} \mid X_{kj}$ is negative binomial.

The chain ladder also provides MVUEs of loss reserves under the same wide range of Mack models. In fact, Mack's variance assumption (7.6) is not required here. MVUE loss reserves are shown to be equivalent to MSEP among unbiased forecasts.

In the case of the ODP cross-classified model with uniform weights, the chain ladder is also found to be ML and to provide MVUEs. However, parallel results do **not** hold when the ODP distribution is replaced by any other member of the EDF.

These results provide a full classification of this paper's chain ladder models with respect to the main estimation properties of the chain ladder algorithm. The classification appears in the following table.

Classification of chain ladder models

Model	Chain ladder algorithm is	
	Unbiased?	Minimum variance?
Mack:		
Distribution free	Yes	N/A
EDF	Yes	Yes
EDF Cross-classified:		
ODP	No	Yes[(b)]
Other	No	Not necessarily[(a)]

[(a)] The statistics on which the chain ladder algorithm rests are not sufficient for the model parameters.
[(b)] Provided weights are uniform across $\mathcal{D}_K^+$.

The fact that the chain ladder possesses optimal estimation properties in a wider range of Mack models than cross-classified models derives from the greater parametric simplicity of the former family. In the Mack models each column is parametrically isolated from the others in the sense that it depends (conditionally on prior data) on just parameters (f_j, σ_j^2) specific to that column.

The cross-classified models, on the other hand, involve a more complex parametric structure that includes both row and column parameters. Consequently, each column depends on not only its column-specific parameters but also those of the rows involved.

10. Acknowledgements

Referees provided helpful comments on earlier versions of this paper.

References

Buchwalder, M., Bühlmann, H., Merz, M. and Wüthrich, M.V. (2006) The mean square error of prediction in the chain ladder reserving method (Mack and Murphy revisited). *Astin Bulletin*, **36(1)**, 521-542.

Cox, D.R. and Hinckley, D.V. (1974) *Theoretical Statistics*. Chapman and Hall, London UK.

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

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

Jorgensen, B. and Paes de Souza, M.C. (1994) Fitting Tweedie's compound Poisson model to insurance claims data. *Scandinavian Actuarial Journal*, 69-93.

Kuang, D., Nielsen, B. and Neilsen, J.P. (2009) Chain-ladder as maximum likelihood revisited. *Annals of Actuarial Science*, **4(1)**, 105-121.

Lehmann, E.L. and Casella, G. (1998) *Theory of point estimation (2nd edition)*. Springer.

Mack, T. (1991) A simple parametric model for rating automobile insurance or estimating IBNR claims reserves. *Astin Bulletin*, **21**, 93-109.

Mack, T. (1993) Distribution-free calculation of the standard error of chain ladder reserve estimates. *Astin Bulletin*, **23(2)**, 213-225.

Mack, T. and Venter, G. (2000) A comparison of stochastic models that reproduce chain ladder reserve estimates. *Insurance: mathematics and economics*, **26**, 101-107.

Nelder, J.A. and Wedderburn, R.W.M. (1972) Generalised linear models. *Journal of the Royal Statistical Society*, *Series A*, **135**, 370-384.

Peters, G.W., Shevchenko, P.V. and Wüthrich, M.V. (2009) Model uncertainty in claims reserving within Tweedie's compound Poisson models. *Astin Bulletin*, **39**, 1-33.

Renshaw, A.E. and Verrall, R.J. (1998) A stochastic model underlying the chain-ladder technique. *British Actuarial Journal*, **4(iv)**, 903-923.

Schmidt, K.D. and Wünsche, A. (1998) Chain ladder, marginal sum and maximum likelihood estimation. *Blätter der Versicherungsmathematiker*, **23**, 267-277.

Taylor, G. (2000) *Loss reserving: an actuarial perspective*. Kluwer Academic Publishers, Boston.

Taylor, G. (2003) Chain ladder bias. *Astin Bulletin*, **33(2)**, 313-330.

Taylor, G. (2009) The chain ladder and Tweedie distributed claims data. *Variance*, **3**, 96-104.

Tweedie, M.C.K. (1984) An index which distinguishes between some important exponential families, *Statistics: Applications and New Directions, Proceedings of the Indian Statistical Golden Jubilee International Conference*, J.K. Ghosh and J. Roy (Eds.), Indian Statistical Institute, 1984, 579-604.

Verrall, R.J. (1991) Chain ladder and maximum likelihood. *Journal of the Institute of Actuaries*, **118**, 489-499.

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. and ENGLAND, P.D. (2000) Comments on "A comparison of stochastic models that reproduce chain ladder reserve estimates", by Mack and Venter (Discussion). *Insurance: mathematics and economics*, **26**, 109-111.

WÜTHRICH, M.V. (2003) Claims reserving using Tweedie's compound Poisson model. *Astin Bulletin*, **33(2)**, 331-346.

WÜTHRICH, M.V. and MERZ, M. (2008) *Stochastic claims reserving methods in insurance*. John Wiley & Sons Ltd, Chichester UK.

GREG TAYLOR
Taylor Fry Consulting Actuaries
Level 8, 30 Clarence Street
Sydney NSW 2000
Australia

Professorial Associate,
Centre for Actuarial Studies
Faculty of Economics and Commerce
University of Melbourne
Parkville VIC 3052
Australia

Professorial Visiting Fellow in the School of Actuarial Studies
Australian School of Business
University of New South Wales
Kensington NSW 2033
Australia

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