ASTIN, AFIR/ERM and IACA Colloquia
Innovation & Invention
23-27 August 2015 | Amora Hotel Jamison Sydney

Actuaries Institute

Existence and Uniqueness of Chain Ladder Solutions

Prepared by Greg Taylor

Presented to the Actuaries Institute
ASTIN, AFIR/ERM and IACA Colloquia
23-27 August 2015
Sydney

Institute of Actuaries of Australia
ABN 69 000 423 656

Level 2, 50 Carrington Street, Sydney NSW Australia 2000
t +61 (0) 2 9233 3466 **f** +61 (0) 2 9233 3446
e actuaries@actuaries.asn.au **w** www.actuaries.asn.au

Existence and uniqueness of chain ladder solutions

Greg Taylor

UNSW Business School
Level 6, West Lobby, UNSW Business School Building E12
UNSW Sydney 2052
Australia

Phone: +61 (0) 421 338 448

gregory.taylor@unsw.edu.au

ASTIN Colloquium, August 2015
Sydney, Australia

Abstract

The cross-classified chain ladder has a number of versions, depending on the distribution to which observations are subject. The simplest case is that of Poisson distributed observations, and then maximum likelihood estimates of parameters are explicit.

Most other cases, however, lead to implicit solutions of these parameter estimates, raising questions as to their existence and uniqueness. The present paper investigates these questions in the case where observations are distributed according to some member of the exponential dispersion family.

It is shown that a solution always exists provided only that the data array meets a regularity condition concerned with which (if any) cells are missing. A condition for uniqueness of that solution is also found. It is shown that this condition reduces to a simple form in the case that the EDF distribution falls within the Tweedie sub-family.

It is further found that uniqueness always occur when the Tweedie dispersion index lies in the interval [1,2]. Uniqueness is not established for larger values of the dispersion index, but a condition for uniqueness is found. It depends on the closeness of the array to “proportionality”, i.e. all rows (or columns) proportional.

The investigation is then widened to Bayesian models, in which the parameters defining the data likelihoods described above are randomised with conjugate priors. Again, it is shown that a solution always exists in the general case (subject to some technical conditions) and conditions for uniqueness are found, which largely parallel those for the non-Bayesian model.

A numerical example is analysed, in which the solutions of the non-Bayesian maximum likelihood equations are found analytically. The behaviour of the log-likelihood is found to be complex even for a very simple data structure, and it is found that multiple solutions can indeed occur.

Keywords: Bayesian chain ladder, cross-classified chain ladder, EDF chain ladder, existence, loss reserving, maximum likelihood estimate, Tweedie chain ladder, uniqueness.

1. Introduction

This paper is concerned with cross-classified (sometimes known as ANOVA) chain ladder models in which the cell mean μ_{kj} for the (k,j) cell of the data array is the product of a row effect and column effect: $\mu_{kj} = \alpha_k \beta_j$.

Such stochastic chain ladder models have been in use for many years, having been introduced by Hachemeister & Stanard (1975). Their model assumed a Poisson distribution in each cell, and this assumption was retained in the literature for some time subsequently.

More recently, other distributions have been considered. For example, England & Verrall (2002) subjected the Poisson distribution to over-dispersion. Taylor (2009) generalised to the Tweedie family of distributions, and Taylor (2011) generalised further to the exponential dispersion family **("EDF").**

The Poisson case provides explicit maximum likelihood estimates **("MLEs")** of model parameters but in this it is unique in the EDF. For other members of the EDF, solutions of the maximum likelihood **("ML")** equations are implicit. In consequence, their existence is not obvious and, if they exist, their uniqueness is not obvious.

A common approach to the establishment of uniqueness of an MLE is application of the Lehmann-Scheffé theorem, which provides that, under regularity conditions, an unbiased MLE based on a complete sufficient statistic is unique and minimum-variance unbiased. It is known that a specific linear combination of observations is a sufficient statistic for the location parameter of a member of the EDF.

This might raise the hope that this approach might provide a simple means of establishing existence and uniqueness of parameter estimates in the EDF cross-classified chain ladder model. Such an approach is certainly useful in the Poisson case. Indeed, Kuang, Nielsen & Nielsen (2009) used it to prove uniqueness of the MLE, and Taylor (2011) to prove minimum variance properties.

However, as will be shown in Section 3.2.4, the same approach does not work for members of the EDF other than Poisson. In fact, Taylor (2011, Theorem 5.2) showed that, for such models, there is no minimal sufficient statistic for any of the parameters that is a proper subset of the full data set. In short questions of existence and uniqueness of MLEs remain open.

In the following, after a brief consideration of the mathematical set-up of the EDF cross-classified chain ladder model in Section 3, the existence (Section 4) and uniqueness (Section 5) of MLEs is considered. Section 6 then extends to a Bayesian version of the EDF cross-classified chain ladder model, while Section 7 examines a numerical example in which multiple solutions of the ML equations are found.

2. Framework and notation

Consider an array $\mathfrak{D}$ of claim observations $Y_{kj} > 0$ with:

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

The nature of these observations is unspecified. They may be paid losses, reported claim counts, claim finalisation counts, or any other quantities that satisfy the conditions prescribed in Section 4.

The dimensions K and J are arbitrary natural numbers, but conditions will be placed on which observations Y_{kj} can be absent from the array.

Consider the array as an undirected graph $\Gamma(\mathfrak{D})$, with the observations as vertices, and define an edge as existing between two observations if and only if they are either from the same row of $\mathfrak{D}$ in adjacent columns, or from the same column of $\mathfrak{D}$ (not necessarily adjacent rows).

Consider an array $\mathfrak{D}$ satisfying the following three requirements:
(A1) It contains a subset of precisely $K+J-1$ observations such that, if all other observations were deleted, the subset would form a sub-array $\mathcal{S}$ of $\mathfrak{D}$.
(A2) Each row of $\mathcal{S}$ contains at least one observation, and similarly each column.
(A3) $\Gamma(\mathcal{S})$ is connected.
Such an array will be called **regular**. It may be noted that a regular array must contain at least $K+J-1$ observations. The sub-array $\mathcal{S}$ will be called a **core** of $\mathfrak{D}$. The core need not be unique.

A regular array $\mathfrak{D}$ may, in general, be of a considerably more general structure than the typical claims triangle included in the loss reserving literature.

A form of $\mathfrak{D}$ of special interest is

$$\mathfrak{D} = \{Y_{kj} : k = 1,2, \dots, K, j = 1,2, \dots \min(J, K-k+1), K \geq J\}$$

An array of this form will be called **trapezoidal**. It includes the case of a triangular array ($K = J$) that occurs widely in the literature. A trapezoidal array is trivially regular.

Kuang, Nielsen & Nielsen (2008) discussed arrays that were more general than trapezoidal, but less general than the regular arrays defined above. These were rectangular arrays, possibly with some upper and some lower diagonals deleted, called **generalized trapezoids**. Such arrays are automatically regular if conditions (A1) and (A2) are satisfied.

Let $\mathcal{R}(k)$ denote the k-th row and $\mathcal{C}(j)$ the j-th column of $\mathfrak{D}$. Let $\sum_{\mathcal{R}(k)}$ denote summation over the entire row $\mathcal{R}(k)$, and similarly $\sum_{\mathcal{C}(j)}$ denote summation over the column $\mathcal{C}(j)$.

Fundamental quantities derived from $\mathfrak{D}$ will prove to be the following:

$$m(\mathfrak{D}) = \min_{\mathfrak{D}} Y_{kj}$$
$$M(\mathfrak{D}) = \max_{\mathfrak{D}} Y_{kj}$$
$$\rho(\mathfrak{D}) = M(\mathfrak{D})/m(\mathfrak{D}) = max\{Y_{k_1 j_1}/Y_{k_2 j_2} ; Y_{k_1 j_1}, Y_{k_2 j_2} \in \mathfrak{D}\}$$

Thus, $\rho(\mathfrak{D})$ is a **measure of inequality** of observations within the same rows or columns of array $\mathfrak{D}$.

Consider $Y_{rs} \notin \mathcal{S}$. By definition, $\mathcal{S}$ contains an observation Y_{rj_r} in row r and an observation $Y_{k_s s}$ in column s. The choice of these observations may not be unique. In that case it is arbitrary.

Also by definition, there exists a path γ_{rs} from Y_{rj_r} to $Y_{k_s s}$. Denote this path $\{Y_{r_i s_i}, i = 1,2, \dots, m\}$ where $(r_1, s_1) = (r, j_r)$ and $(r_m, s_m) = (k_s, s)$, and further:

- in the case $s > j_r$, $(r_{i+1}, s_{i+1}) = (r_i, s_i + 1)$ or (r_{i+1}, s_i);

- in the case $s < j_r$, $(r_{i+1}, s_{i+1}) = (r_i, s_i - 1)$ or (r_{i+1}, s_i);

and where, in the latter case of each alternative, (r_{i+1}, s_i) is just any other index in the same column as (r_i, s_i).

Consider only the subset $\eta_{rs} \subset \gamma_{rs}$ of edges between adjacent columns in γ_{rs}, i.e. the above cases $(r_{i+1}, s_{i+1}) = (r_i, s_i + 1)$ or $(r_{i+1}, s_{i+1}) = (r_i, s_i - 1)$, and define

$$\begin{aligned}
\pi_{rs} &= \frac{Y_{rs}}{Y_{rj_r}} \prod_{\eta_{rs}} \frac{Y_{r_i s_i}}{Y_{r_{i+1} s_{i+1}}} - 1 \\
\bar{\pi}_{rs} &= \max_{j_r, k_s, \gamma_{rs}} [\pi_{rs}]_+ \\
\underline{\pi}_{rs} &= \max_{j_r, k_s, \gamma_{rs}} [-\pi_{rs}]_+
\end{aligned} \tag{2.1}$$

where the maximum is taken over all possible choices of j_r, k_s and all possible paths γ_{rs} from Y_{rj_r} to $Y_{k_s s}$, and $[.]_+$ denotes the non-negative part of the argument.

Now define

$$\begin{aligned}
\bar{\pi}(\mathfrak{D}) &= \max_{\mathcal{S}} \max_{\{r,s: Y_{rs} \notin \mathcal{S}\}} \bar{\pi}_{rs} \\
\underline{\pi}(\mathfrak{D}) &= \max_{\mathcal{S}} \max_{\{r,s: Y_{rs} \notin \mathcal{S}\}} \underline{\pi}_{rs}
\end{aligned} \tag{2.2}$$

where the maxima are taken over all possible choices of $\mathcal{S}$ as core of $\mathfrak{D}$.

Finally, define

$$\xi(\mathfrak{D}) = \frac{1 + \bar{\pi}(\mathfrak{D})}{1 - \underline{\pi}(\mathfrak{D})} \geq 1 \tag{2.3}$$

Consider an array $\mathfrak{D}$ that is perfectly **proportional** in the sense that $Y_{kj} = \alpha_k \beta_j$. It may be remarked that the ratios $Y_{r_i s_i}/Y_{r_{i+1} s_{i+1}}$ in (2.1) all take the form $Y_{r_i s_i}/Y_{r_i, s_i+1} = \beta_{s_i}/\beta_{s_i+1}$ or $Y_{r_i s_i}/Y_{r_i, s_i-1} = \beta_{s_i}/\beta_{s_i-1}$, and these ratios proceed by single steps from column j_r to column s. It follows that $\pi_{rs} = 0$, and hence $\bar{\pi}(\mathfrak{D}) = \underline{\pi}(\mathfrak{D}) = 0, \xi(\mathfrak{D}) = 1$. Thus, $\bar{\pi}(\mathfrak{D})$, $\underline{\pi}(\mathfrak{D})$ and $\xi(\mathfrak{D})$ are **measures of the non-proportionality** of the array $\mathfrak{D}$.

These measures could be defined more simply in the case of a trapezoidal array. There one might define, again for fixed j_r, k_s,

$$\begin{aligned}
\pi_{rs}^{trap} &= \frac{Y_{rs}}{Y_{rj_s}} \frac{Y_{k_r j_s}}{Y_{k_r s}} - 1 \\
\bar{\pi}_{rs}^{trap} &= \max_{j_r, k_s} [\pi_{rs}]_+ \\
\underline{\pi}_{rs}^{trap} &= \max_{j_r, k_s} [-\pi_{rs}]_+ \\
\bar{\pi}^{trap}(\mathfrak{D}) &= \max_{r,s} \bar{\pi}_{rs} \\
\underline{\pi}^{trap}(\mathfrak{D}) &= \max_{r,s} \underline{\pi}_{rs}
\end{aligned} \tag{2.4}$$

Further comment on the definition of the non-proportionality measures will be made in Section 5.4.

3. Chain ladder models

3.1 Mack models

The literature identifies two main types of chain ladder model (Taylor, 2011) more recently referred to as:

- **Mack models**; and
- **cross-classified models**.

The former category consists of two further types of model:

- the **distribution-free**, or **non-parametric**, Mack model (Mack, 1993); and
- the EDF Mack model and its special cases.

The present paper will be concerned with the existence and uniqueness of MLEs of the parameters of chain ladder models.

The literature gives explicit MLEs in the case of Mack models, and so existence and uniqueness is trivially established. The paper will therefore be concerned with just cross-classified models.

3.2 Cross-classified models

3.2.1 Exponential dispersion family

Consider the model defined by the following conditions:

(EDFCC1) The array $\mathfrak{D}$ is regular.

(EDFCC2) The random variables $Y_{kj} \in \mathfrak{D}$ are stochastically independent.

(EDFCC3) For each $k = 1,2, \ldots, K$ and $j = 1,2, \ldots, J$,

(a) Y_{kj} is distributed according to a member of the EDF, specifically with log-likelihood of $Y_{kj} = y$ as follows:

$$\ell_{kj}(y|\theta,\phi) = [y\theta_{kj} - \kappa(\theta_{kj})]/a(\phi_{kj}) + \lambda(y,\phi_{kj}) \qquad (3.1)$$

for parameters $\theta_{kj}, \phi_{kj}(\phi_{kj} > 0)$, and for functions a, κ, λ that do not depend on k, j, with a continuous, κ twice differentiable, and λ such as to produce a unit total probability mass. It will be further assumed that the derivative $\kappa'(.)$ maps one-one onto the strictly positive half-line, and that $\kappa'' > 0$.

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

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

Henceforth the function $a(.)$ in (3.1) will be restricted to the case

$$a(\phi) = \phi \qquad (3.2)$$

Condition (EDFCC3c) 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.

This model is referred to as the **EDF cross-classified model**. It consists of cross-classified multiplicative mean structure, as in (EDFCC3b), supplemented by an EDF distribution in (EDFCC3a).

It will be convenient to express the log-likelihood (3.1) in a different representation, as follows.

Let μ_{kj} denote $E[Y_{kj}] = \alpha_k \beta_j$. It is known (McCullagh & Nelder, 1989) that

$$\mu_{kj} = \kappa'(\theta_{kj}) \tag{3.3}$$

whence the following expressions:

$$\theta_{kj} = c(\mu_{kj}) \tag{3.4}$$
$$\kappa(\theta_{kj}) = d(\mu_{kj}) \tag{3.5}$$

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

Then (3.1) may be re-written, taking account of (3.2), in the form

$$\ell_{kj}(y|\mu, \phi) = [yc(\mu_{kj}) - d(\mu_{kj})]/\phi_{kj} + \lambda(y, \phi_{kj}) \tag{3.6}$$

or

$$\ell_{kj}(y|\alpha, \beta, \phi) = [yc(\alpha_k \beta_j) - d(\alpha_k \beta_j)]/\phi_{kj} + \lambda(y, \phi_{kj}) \tag{3.7}$$

where μ on the left side is now the vector of values μ_{kj}, and similarly α, β are vectors of the α_k, β_j respectively.

3.2.2 Tweedie family

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

$$\kappa(\theta) = \frac{1}{2-p}[(1-p)\theta]^{(2-p)/(1-p)} \tag{3.8}$$

or equivalently

$$c(\mu) = \mu^{1-p}/(1-p) \tag{3.9}$$
$$d(\mu) = \mu^{2-p}/(2-p) \tag{3.10}$$

The parameter p will be referred to as the **Tweedie index**.

In the cases $p = 1,2$, (3.9) and (3.10) must be replaced by their limiting values as p approaches the relevant value:

$$\lim_{q \to 0} \mu^q/q = \ln \mu \tag{3.11}$$

When the distributions of the EDF cross-classified model are restricted to Tweedie, the model will be referred to as the **Tweedie cross-classified model**.

3.2.3 Poisson family

The **Poisson family** is the sub-family of the Tweedie family for which $p = 1$, so that, by (3.9)-(3.11),

$$c(\mu) = ln\,\mu \tag{3.12}$$
$$d(\mu) = \mu \tag{3.13}$$

Substitution in (3.6) yields

$$\ell_{kj}(y|\mu,\phi) = [y\,ln\,\mu_{kj} - \mu_{kj}]/\phi_{kj} \;+ \lambda(y,\phi_{kj}) \tag{3.14}$$

For the special case $\phi_{kj} = 1$,

$$\ell_{kj}(y|\mu,\phi) = ln\left[e^{-\mu_{kj}}{\mu_{kj}}^{y}\,exp\,\lambda(y,1)\right] \tag{3.15}$$

The normalising function can be recognised to be $\lambda(y,1) = -\,ln\,y!$, in which case (3.15) is seen to be the **Poisson** log-likelihood. For the case $\phi_{kj} \neq 1$, (3.14) is called the **over-dispersed Poisson ("ODP")** log-likelihood.

When the distributions of the EDF cross-classified model are restricted to ODP (or Poisson), the model will be referred to as the **ODP (or Poisson) cross-classified model**.

3.2.4 Sufficient statistics

When (EDFCC3b) is recognised, (3.14) becomes

$$\ell_{kj}(y|\mu,\phi) = [y\left(ln\,\alpha_k + ln\,\beta_j\right) - \alpha_k\beta_j]/\phi_{kj} \;+ \lambda(y,\phi_{kj}) \tag{3.16}$$

and summation over $\mathfrak{D}$ yields

$$\ell = \left[\sum_{k=1}^{K} ln\,\alpha_k \sum_{\mathcal{R}(k)} \frac{y_{kj}}{\phi_{kj}} + \sum_{j=1}^{J} ln\,\beta_j \sum_{\mathcal{C}(j)} \frac{y_{kj}}{\phi_{kj}} - \sum_{\mathfrak{D}} \frac{\alpha_k\beta_j}{\phi_{kj}}\right] + \sum_{\mathfrak{D}} \lambda(y,\phi_{kj}) \tag{3.17}$$

Application of the Fisher-Neyman theorem to this likelihood proves that $\sum_{\mathcal{R}(k)} y_{kj}/\phi_{kj}$ is a sufficient statistic for α_k and $\sum_{\mathcal{C}(j)} y_{kj}/\phi_{kj}$ for β_j.

As mentioned in Section 1, one might be encouraged to extend this result beyond the Poisson family, perhaps to the Tweedie family. In this case, (3.16) and (3.17) would be replaced by the following [by (3.6), (3.9) and (3.10)]:

$$\ell_{kj}(y|\mu,\phi) = [y\ \frac{(\alpha_k\beta_j)^{1-p}}{1-p} - \frac{(\alpha_k\beta_j)^{2-p}}{2-p}]/\phi_{kj}\ + \lambda(y,\phi_{kj}) \tag{3.18}$$

$$\ell = \left[\sum_{k=1}^{K} \alpha_k{}^{1-p} \sum_{\mathcal{R}(k)} \frac{y_{kj}\beta_j{}^{1-p}}{(1-p)\phi_{kj}} - \sum_{\mathfrak{D}} \frac{(\alpha_k\beta_j)^{2-p}}{(2-p)\phi_{kj}}\right] + \sum_{\mathfrak{D}} \lambda(y,\phi_{kj}) \tag{3.19}$$

The linear combination of observations that was previously a Neyman-Fisher factor, representing a sufficient statistic for α_k, now entangles the data with the set of parameters $\{\beta_j\}$. This does not produce a sufficient statistic. Indeed, as mentioned in Section 1, it has been shown by Taylor (2011) that there exists no minimal sufficient statistic that is a proper subset of $\mathfrak{D}$.

4. Existence of chain ladder solutions

It will be assumed that the observations $Y_{kj} \in \mathfrak{D}$ are compatible with any distribution subsequently imposed on them, e.g. $Y_{kj} > 0$ if subject to a gamma distribution.

4.1 Poisson family

It is known (Hachemeister & Stanard, 1975; Renshaw & Verrall, 1998; Taylor, 2000) that explicit MLEs exist for the parameters of the Poisson cross-classified model. Existence is therefore obvious.

4.2 Exponential dispersion family

Now consider the general EDF cross-classified model of Section 3.2.1 with joint likelihood function of $\mathfrak{D}$ equal to $\ell = \sum_{\mathfrak{D}} \ell_{kj}$. By (3.6) and (3.7), the maximum likelihood equations are:

$$\partial\ell/\partial\alpha_k = \textstyle\sum_{\mathcal{R}(k)}[Y_{kj}c'(\mu_{kj}) - d'(\mu_{kj})]\beta_j/\phi_{kj} = 0\,, k = 1,\dots,K \tag{4.1}$$
$$\partial\ell/\partial\beta_j = \textstyle\sum_{\mathcal{C}(j)}[Y_{kj}c'(\mu_{kj}) - d'(\mu_{kj})]\alpha_k/\phi_{kj} = 0\,, k = 1,\dots,J \tag{4.2}$$

These equations can be simplified slightly if it is noted that, by the definition of $c(\mu)$ and $d(\mu)$ in Section 3.2.1,

$$d'(\mu) = \kappa'\big(c(\mu)\big)\ c'(\mu) = \mu\ c'(\mu) \tag{4.3}$$

This converts (4.1) and (4.2) to the following:

$$\textstyle\sum_{\mathcal{R}(k)}[Y_{kj} - \mu_{kj}]c'(\mu_{kj})\beta_j/\phi_{kj} = 0 \tag{4.4}$$
$$\textstyle\sum_{\mathcal{C}(j)}[Y_{kj} - \mu_{kj}]c'(\mu_{kj})\alpha_k/\phi_{kj} = 0 \tag{4.5}$$

where the left side in each case can be recognised as a weighted sum of raw residuals $Y_{kj} - \mu_{kj}$.

It is noted for future reference that

$$c'(\mu) = 1/\kappa''\big(c'(\mu)\big) > 0 \tag{4.6}$$

by (EDFCC3b), and then

$$d'(\mu) > 0 \tag{4.7}$$

by (4.3).

Equations (4.4) and (4.5) do not yield explicit solutions for the α_k, β_j in general. An example occurs in Taylor (2009), who examines the Tweedie cross-classified model. Hence questions of existence and uniqueness of solutions arise.

Existence is fairly easily established, by means of an application of the Weierstrass theorem to the joint likelihood function of $\mathfrak{D}$. This requires continuity of the likelihood as a function of the parameter set $\{\alpha_k, \beta_j\}$, and compactness of the parameter space.

The first of these conditions amounts to continuity of κ in (3.1), and this is implied by its assumed differentiability in (EDFCC3a). Establishment of compactness of the parameter space requires slightly greater effort. Compactness is established in Lemma A.3 in the appendix. This justifies the following result.

Theorem 4.1. For the EDF cross-classified model defined in Section 3.2.1, at least one MLE of the parameter set α_k, β_j exists. ■

A useful limit on the MLE is the following. The proof is given in the appendix.

Lemma 4.2. Let the Y_{kj} and μ_{kj} be as in Theorem 5.1, i.e. $\mu_{kj} = \hat{\alpha}_k \hat{\beta}_j$, where the $\hat{\alpha}_k, \hat{\beta}_j$ are MLEs of parameters α_k, β_j. Then, for each k, j, the following two conditions hold:

$$\min_{\mathcal{R}(k)} Y_{ks} \le \mu_{kj} \le \max_{\mathcal{R}(k)} Y_{ks}$$
$$\min_{\mathcal{C}(j)} Y_{rj} \le \mu_{kj} \le \max_{\mathcal{C}(j)} Y_{rj}$$ ■

5. Uniqueness of chain ladder solutions

5.1 Mathematical preliminaries

Uniqueness of solutions will be proved by establishment of convexity of the log-likelihood function ℓ defined in Section 4.2. This will be done by reference to derivatives of that function for the EDF cross-classified model of Section 3.2.1. Hence it will be useful to record the following derivatives, derived by reference to (3.6) and (3.7), where $r_k = ln\,\alpha_k, s_j = ln\,\beta_j$:

$$\begin{aligned}\partial\ell/\partial r_k &= (\partial\ell/\partial\alpha_k)(\partial\alpha_k/\partial ln\,\alpha_k) = \alpha_k\,\partial\ell/\partial\alpha_k \\ &= \sum_{\mathcal{R}(k)} [Y_{kj}c'(\mu_{kj}) - d'(\mu_{kj})]\mu_{kj}/\phi_{kj}\end{aligned}$$

$$= \sum_{\mathcal{R}(k)} \mu_{kj} \frac{\partial \ell}{\partial \mu_{kj}} \tag{5.1}$$

$$\partial \ell / \partial s_j = \sum_{\mathcal{C}(j)} \mu_{kj} \frac{\partial \ell}{\partial \mu_{kj}} \tag{5.2}$$

$$\frac{\partial^2 \ell}{\partial r_k^2} = \sum\nolimits_{\mathcal{R}(k)} \frac{\partial}{\partial \mu_{kj}} \left(\mu_{kj} \frac{\partial \ell}{\partial \mu_{kj}} \right) \frac{\partial \mu_{kj}}{\partial r_k} = \sum\nolimits_{\mathcal{R}(k)} \mu_{kj} \frac{\partial}{\partial \mu_{kj}} \left(\mu_{kj} \frac{\partial \ell}{\partial \mu_{kj}} \right)$$
$$= \sum\nolimits_{\mathcal{R}(k)} \left[\mu_{kj}^2 \frac{\partial^2 \ell}{\partial \mu_{kj}^2} + \mu_{kj} \frac{\partial \ell}{\partial \mu_{kj}} \right] \tag{5.3}$$

$$\frac{\partial^2 \ell}{\partial s_j^2} = \sum\nolimits_{\mathcal{C}(j)} \left[\mu_{kj}^2 \frac{\partial^2 \ell}{\partial \mu_{kj}^2} + \mu_{kj} \frac{\partial \ell}{\partial \mu_{kj}} \right] \tag{5.4}$$

$$\frac{\partial^2 \ell}{\partial r_k \partial s_j} = \mu_{kj}^2 \frac{\partial^2 \ell}{\partial \mu_{kj}^2} + \mu_{kj} \frac{\partial \ell}{\partial \mu_{kj}} \tag{5.5}$$

Note that derivatives are taken here with respect to $ln\ \alpha_k, ln\ \beta_j$ rather than α_k, β_j. An attempt to prove Theorem 5.1 by means of derivatives with respect to α_k, β_j leads to hopeless mathematical entanglement. The choice of derivatives arises ultimately from the multiplicative nature of the model in (EDFCC3b). As noted in Section 3.2.1, this model structure contains a degree of redundancy in the absence of (EDFCC3c). As a result, certain variations of α_k, β_j lead to the same values of μ_{kj}. The condition for this is $\delta(\alpha_k \beta_j) = 0$, which takes the neat linear form $\delta(ln\ \alpha_k + ln\ \beta_j) = 0$ if log transforms of the parameters are used.

Let $v = (v_1, \ldots, v_K)^T$ and $w = (w_1, \ldots, w_J)^T$ be real-valued vectors, where the upper T denotes transposition. Let $Q(v, w)$ denote the quadratic form associated with the Hessian of ℓ, i.e.

$$Q(v, w) = \sum_{k=1}^{K} v_k^2 \frac{\partial^2 \ell}{\partial r_k^2} + \sum_{j=1}^{J} w_j^2 \frac{\partial^2 \ell}{\partial s_j^2} + 2 \sum_{\mathfrak{D}} v_k w_j \frac{\partial^2 \ell}{\partial r_k \partial s_j}$$
$$= \sum_{\mathfrak{D}} (v_k + w_j)^2 \left[\mu_{kj}^2 \frac{\partial^2 \ell}{\partial \mu_{kj}^2} + \mu_{kj} \frac{\partial \ell}{\partial \mu_{kj}} \right] \tag{5.6}$$

by (5.3)-(5.5).

5.2 Exponential dispersion family

The proof of the following theorem appears in the appendix.

Theorem 5.1. Let the array $\mathfrak{D}$ be regular, and suppose that the Y_{kj} are subject to the EDF cross-classified model of Section 3.2.1. Let R denote a compact set $\{\mu_{kj}: 0 < \underline{\mu}_{kj} < \mu_{kj} < \overline{\mu}_{kj}, (k,j)\ such\ that\ Y_{kj} \in \mathfrak{D}\}$, where the $\underline{\mu}_{kj}, \overline{\mu}_{kj}$ are known bounds. Then a necessary and sufficient condition for the log-likelihood ℓ to be convex upward over R is that, for all $\{\mu_{kj}; (k,j)\ such\ that\ Y_{kj} \in \mathfrak{D}\} \in R$, $\frac{Y_{kj}}{\mu_{kj}} \leq 1 + \frac{d'(\mu_{kj})}{\mu_{kj} d''(\mu_{kj})}$ whenever $d''(\mu_{kj}) > 0$

$$\frac{Y_{kj}}{\mu_{kj}} \geq 1 - \frac{d'(\mu_{kj})}{-\mu_{kj} d''(\mu_{kj})} \text{ whenever } d''(\mu_{kj}) < 0 \quad (5.7)$$

(note that $d' > 0$, from (4.7)) with strict inequality for at least one pair (k, j) (not necessarily the same (k, j) at each point of R).

Hence (5.7) is a sufficient condition for a unique MLE of the model parameters.■

Theorem 5.1 establishes a unique MLE on the set R, but gives no detail about the construction of that set. Corollary 5.2 below remedies this.

By Lemma A.3, it is known that any MLE of the model parameters must lie in a rectangle $\left\{0 < \underline{\alpha}_k \leq \alpha_k \leq \overline{\alpha}_k, 0 < \underline{\beta_j} \leq \beta_j \leq \overline{\beta}_j \; for \; all \; k, j\right\}$ and, by Theorem 4.1, such an MLE exists. Define

$$\underline{\mu}_{kj} = \underline{\alpha}_k \underline{\beta}_j, \overline{\mu}_{kj} = \overline{\alpha}_k \overline{\beta}_j \quad (5.8)$$

This leads immediately to the following corollary of Theorem 5.1.

Corollary 5.2. The compact set R in Theorem 5.1 may be defined by (5.8), where the bounds $\underline{\alpha}_k, \underline{\beta}_j, \overline{\alpha}_k, \overline{\beta}_j$ are defined as in the proof of Lemma A.3. ■

The binomial distribution is a member of the EDF that is not contained in the Tweedie family. It has

$$\kappa(\theta) = N \, ln \left(1 + e^{\theta}\right) \quad (5.9)$$

where N is the number of trials and θ the log odds ratio $p/(1-p)$, with p the probability of success.

The following corollary applies Theorem 5.1 to the case of binomial observations Y_{kj}.

Corollary 5.3. Consider the special case of Theorem 5.1 in which $Y_{kj} \sim Bin\left(N_{kj}, p_{kj}\right)$ with the N_{kj} known and the p_{kj} unknown parameters. Then condition (5.7) reduces to

$$\frac{Y_{kj}}{\mu_{kj}} \leq \frac{N_{kj}}{\mu_{kj}} \quad (5.10)$$

with strict inequality for at least one pairs k, j.

It then follows, other than in the case $Y_{kj} = N_{kj}$ for all pairs k, j, that the MLE of the model parameters is unique.

Proof. As a simple matter of calculation from (5.9),

$$d'(\mu_{kj}) = N_{kj}(N_{kj} - \mu_{kj})^{-1}$$
$$d''(\mu_{kj}) = N_{kj}(N_{kj} - \mu_{kj})^{-2} > 0$$

Substitution of these results into (5.7) yields (5.10). This is automatically satisfied except in the case $Y_{kj} = N_{kj}$ for all pairs k, j, and so uniqueness is assured other than in this exceptional case. ■

5.3 Tweedie family

The following corollary applies Theorem 5.1 to the Tweedie family.

Corollary 5.4. Consider the special case of Theorem 5.1 in which the Y_{kj} are subject to a Tweedie distribution with index $p \geq 1$. Then condition (5.7) reduces to

$$\frac{Y_{kj}}{\overline{\mu}_{kj}} \geq \frac{p-2}{p-1} \tag{5.11}$$

Proof. It follows from (3.10) that

$$d'(\mu) = \mu^{1-p}$$
$$d''(\mu) = -(p-1)\mu^{-p} \leq 0$$

whence condition (5.7) becomes

$$\frac{Y_{kj}}{\mu_{kj}} \geq \frac{p-2}{p-1} \text{ for all } \underline{\mu}_{kj} \geq \mu_{kj} \geq \overline{\mu}_{kj}$$

A necessary and sufficient condition for this is (5.11). ■

Corollary 5.5. Consider the special case of Theorem 5.1 in which the Y_{kj} are subject to a Tweedie distribution with index $1 \leq p \leq 2$. Then there is a unique MLE of the model parameters.

Proof. Substitution of p in this range into the right side of (5.11) yields a negative result. Thus (5.11) is always satisfied since $Y_{kj}/\mu_{kj} > 0$. ■

Corollary 5.6. Consider the special case of Theorem 5.1 in which the Y_{kj} satisfy one of the following conditions:

(a) All are subject to an ODP distribution (which includes simple Poisson as a special case);
(b) All are subject to a gamma distribution;
(c) All are subject to a compound Poisson distribution with gamma severity distribution.

Then there is a unique MLE of the model parameters.

Proof. ODP and gamma are the cases $p = 1$ and 2 respectively, and the compound Poisson-gamma distributions occupy the region $1 \leq p \leq 2$, all of which are special cases of Corollary 5.4. ■

A "proportional" array, i.e. with $Y_{kj} = a_k b_j$, has the obvious MLE $\alpha_k = a_k, \beta_j = b_j$. it is not difficult to show that this solution is unique in the Tweedie case. One might conjecture a couple of things as a consequence of this:

(a) that, for arrays that are close to proportional in some sense, a cross-classified model fitted to any part of the array will be "close" to the model fitted to the entire array; and

(b) that an array will have a unique MLE chain ladder solution if it is "close enough"

The next two results address these conjectures. Proofs appear in the appendix.

Theorem 5.7. Consider a trapezoidal array $\mathfrak{D}$ that contains at least $K + J - 1$ observations. Let $\mathcal{S}$ be a core of $\mathfrak{D}$, and fit to $\mathcal{S}$ a Tweedie cross-classified model $Y_{kj} = \alpha_k^* \beta_j^* (= \mu_{kj}^*)$ say, subject to multiplicative weights $\phi_{kj} = (v_k w_j)^{-1}$ (Lemma A.4 guarantees the possibility of this). Now fit the Tweedie cross-classified model $\mu_{kj} = \alpha_k \beta_j$ to the entire array $\mathfrak{D}$, and subject to the same multiplicative system of weights. Then the following relations hold:

$$(1 + \underline{\pi})\xi^{-(|J_k - s| + 2)} \leq \frac{\alpha_k \beta_s}{\alpha_k^* \beta_s^*} \leq (1 + \overline{\pi})\xi^{|J_k - s| + 2} \tag{5.12}$$

where $J_k = min(J, K - k + 1) =$ maximum value of j for which an observation exists in row k. ■

Corollary 5.8. Consider the special case of Theorem 5.1 in which $\mathfrak{D}$ is a trapezoidal array, and the Y_{kj} are subject to a Tweedie distribution with index $p > 2$ and multiplicative weights $\phi_{kj} = (v_k w_j)^{-1}$. Then a sufficient condition for convexity of the log-likelihood ℓ is

$$\xi(\mathfrak{D})^{-(J+1)} \geq \frac{p-2}{p-1} \tag{5.13}$$

This is therefore a sufficient condition for uniqueness of the MLE of model parameters. ■

It was noted in Section 2 that $\xi(\mathfrak{D})$ is a measure of the non-proportionality of the array $\mathfrak{D}$. For a proportional array, $\xi(\mathfrak{D}) = 1$, and $\xi(\mathfrak{D})$ increases steadily with increasing non-proportionality. This leads immediately to the following result, already stated just above.

Corollary 5.9. Consider the special case of Theorem 5.1 in which $\mathfrak{D}$ is a **proportional** trapezoidal array, and the Y_{kj} are subject to a Tweedie distribution with index $p > 2$ and multiplicative weights $\phi_{kj} = (v_k w_j)^{-1}$. Then the log-likelihood ℓ is convex upward, and there is a unique MLE of model parameters.■

Condition (5.13) for a unique MLE may be re-stated as:

$$p \leq 2 + \frac{1}{\xi(\mathfrak{D})^{J+1}}$$

which decreases monotonically from ∞ to 2 as $\xi(\mathfrak{D})$ increases from 1 to ∞. Thus, Corollary 5.8 shows that as non-proportionality of $\mathfrak{D}$ increases (i.e. $\xi(\mathfrak{D})$ increases), satisfaction of the condition (5.13) for a unique MLE requires a steadily decreasing Tweedie index p.

5.4 Discussion

Consider Tweedie distributions with index parameter $p > 2$. The necessary and sufficient condition for convexity of the log-likelihood ℓ is still (5.11) but now $\frac{p-2}{p-1} > 0$, so is no longer automatically satisfied by the $Y_{kj}/\bar{\mu}_{kj}$. Indeed, sufficiently small $Y_{kj}/\bar{\mu}_{kj}$ will violate (5.11), a necessary and sufficient condition for convexity.

It follows that, in any such cases, ℓ is **not convex**. However, in none of the above results does non-convexity imply non-uniqueness. It is simply that uniqueness is not established in these non-convex cases. This issue will be explored further in Section 7.

A number of the results obtained in Section 5.3 apply just to trapezoidal arrays. The reason for this is mere simplicity. Extensions to non-trapezoidal arrays would be possible, but their statements tedious.

6. Bayesian chain ladder models

6.1 Bayesian framework

Bayesian versions of EDF cross-classified model, or special cases of it have been studied in the literature (Verrall, 2000, 2004; England & Verrall, 2002; Gisler & Müller, 2007; Wüthrich, 2007,2012; Gisler & Wüthrich, 2008; Wüthrich & Merz (2008); England, Verrall & Wüthrich, 2012; Shi, Basu & Meyers, 2012; Merz, Wüthrich & Hashorva, 2013; Taylor, 2015).

Different papers use different Bayesian structures. The present paper will rely on Taylor (2015). That paper notes that its framework differs from those used in the others. It may be of interest to others to investigate the possibility of result parallel to those in Section 6.2 but subject to different priors.

Taylor uses (3.1) as the conditional log-likelihood of $Y_{kj}|\alpha_k, \beta_j$, and the following prior log-densities on $c(\alpha_k)$ (omitting terms that do not depend on α_k):

$$\ell_k^{(\alpha)prior}\big(c(\alpha_k)\big) = [c(\alpha_k)A_k - d(\alpha_k)]/\psi_k^{(\alpha)} \tag{6.1}$$

where $c(.), d(.)$ are as defined in (3.4) and (3.5), and $A_k, \psi_k^{(\alpha)} > 0$ are location and dispersion parameters of the prior.

Similarly

$$\ell_j^{(\beta)prior}\left(c(\beta_j)\right) = \left[c(\beta_j)B_j - d(\beta_j)\right]/\psi_j^{(\beta)} \tag{6.2}$$

In summary, the EDF cross-classified model of Section 3.2.1 is replaced by a Bayesian model, the **Bayesian EDF cross-classified model**, defined by the following conditions:

(BEDFCC1) As for (EDFCC1).

(BEDFCC2) As for (EDFCC2) except that the independence is now conditional on the parameter set $\{\alpha_k, \beta_j\}$.

(BEDFCC3)

(a) As just discussed, the likelihood (3.1) is retained but now becomes conditional on the parameter set, with priors (6.1) and (6.2) on those parameters. The conditions placed on the parameters in (EDFCC3a) continue to hold.

(b) $E[Y_{kj}|\alpha_k, \beta_j] = \alpha_k\beta_j$ (in parallel with (EDFCC3b)).

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

Note that the condition (EDFCC3c) is no longer required.

If the conditional log-likelihood is re-labelled ℓ_{kj}^{cond}, then the posterior likelihood of the α_k, β_j (again omitting terms that do not depend on the α_k, β_j) is;

$$\ell^{post}\left(c(\alpha), c(\beta)|\mathfrak{D}; A, B, \phi, \psi^{(\alpha)}, \psi^{(\beta)}\right) = \ell^{cond} + \ell^{prior} \tag{6.3}$$

where $c(\alpha), c(\beta)$ denote the vectors of $c(\alpha_k)$ and $c(\beta_j)$ values respectively, and $A, B, \phi, \psi^{(\alpha)}, \psi^{(\beta)}$ are similar parameter vectors, and with

$$\ell^{cond} = \textstyle\sum_{\mathfrak{D}} \ell_{kj}^{cond} \tag{6.4}$$

$$\ell^{prior} = \textstyle\sum_{k=1}^{K} \ell_k^{(\alpha)prior} + \sum_{j=1}^{J} \ell_j^{(\beta)prior} \tag{6.5}$$

The conditions for MAP estimation of α, β are:

$$\frac{\partial\left(\ell^{cond} + \ell^{prior}\right)}{\partial\alpha_k} = \frac{\partial\left(\ell^{cond} + \ell^{prior}\right)}{\partial\beta_j} = 0, k = 1, \ldots, K; j = 1, \ldots, J \tag{6.6}$$

Section 4.2 of Taylor (2015) derives the following solution $\{\hat{\alpha}_k, \hat{\beta}_j\}$ of the system (6.6):

$$\hat{\alpha}_k = z_k^{(\alpha)}\bar{Y}_k^{(\alpha)} + \left[1 - z_k^{(\alpha)}\right]A_k \tag{6.7}$$

with

$$z_k^{(\alpha)} = \sum_{j\in\mathcal{R}(k)} z_{kj}^{(\alpha)} \tag{6.8}$$

$$\bar{Y}_k^{(\alpha)} = \sum_{j\in\mathcal{R}(k)} z_{kj}^{(\alpha)} \left[Y_{kj}/\hat{\beta}_j\right] \Big/ \sum_{j\in\mathcal{R}(k)} z_{kj}^{(\alpha)} \tag{6.9}$$

$$z_{kj}^{(\alpha)} = \frac{\hat{\beta}_j^2 c'(\hat{\mu}_{kj})}{\phi_{kj}} \Bigg/ \left[\sum_{j\in\mathcal{R}(k)} \frac{\hat{\beta}_j^2 c'(\hat{\mu}_{kj})}{\phi_{kj}} + \frac{c'(\hat{\alpha}_k)}{\psi_k^{(\alpha)}}\right] \tag{6.10}$$

$$\hat{\beta}_j = z_j^{(\beta)} \bar{Y}_j^{(\beta)} + \left[1 - z_j^{(\beta)}\right] B_j \tag{6.11}$$

with

$$z_j^{(\beta)} = \sum_{k\in\mathcal{C}(j)} z_{kj}^{(\beta)} \tag{6.12}$$

$$\bar{Y}_j^{(\beta)} = \sum_{k\in\mathcal{C}(j)} z_{kj}^{(\beta)} \left[Y_{kj}/\hat{\alpha}_k\right] \Big/ \sum_{k\in\mathcal{C}(j)} z_{kj}^{(\beta)} \tag{6.13}$$

$$z_{kj}^{(\beta)} = \frac{\hat{\alpha}_k^2 c'(\hat{\mu}_{kj})}{\varphi_{kj}} \Bigg/ \left[\sum_{k\in\mathcal{C}(\mathrm{j})} \frac{\hat{\alpha}_k^2 c'(\hat{\mu}_{kj})}{\varphi_{kj}} + \frac{c'(\hat{\beta}_j)}{\psi_j^{(\beta)}}\right] \tag{6.14}$$

It may be noted immediately that the solution (6.7)-(6.14) is implicit because the solution $\hat{\alpha}_k$ requires knowledge of the $\hat{\beta}_j$ and vice versa.

6.2 Existence and uniqueness of MAP estimators

6.2.1 Existence

Commence by defining

$$\zeta(\alpha,\beta) = \beta^2 c'(\alpha\beta)/c'(\alpha) \text{ for } 0 < \alpha, \beta < \infty \tag{6.15}$$

Suppose that $\zeta(\alpha,\beta)$ satisfies the following conditions:

(TC1) For numbers $\underline{t}, \bar{t}$, arbitrarily small and large respectively, with $0 < \underline{t} < \bar{t} < \infty$, the set $\{c'(tu)/c'(u) : u \in (0,\infty), \underline{t} \le t \le \bar{t}\}$ is bounded above and below, with strictly positive lower bound.

(TC2) For numbers $\underline{q}, \overline{q}$, arbitrarily small and large respectively, with $0 < \underline{q} < \overline{q} < \infty$, and for any $0 < \alpha < \infty$, $\zeta(\alpha,\beta)$ is bounded away from zero and infinity for $\underline{q} \le \beta \le \overline{q}$.

(TC3) If $\zeta(\alpha,\beta) \to \infty$ as $\beta \to 0$, then $\zeta(\alpha,\beta)$ is bounded as $\beta \to \infty$; and the bounds are uniform with respect to α.

(TC4) If $\zeta(\alpha,\beta) \to \infty$ as $\beta \to \infty$, then $\beta^{-1}\zeta(\alpha,\beta) \times \zeta(\beta, C/\zeta(\alpha,\beta)) \to 0$ for any constant, $C > 0$ as $\beta \to \infty$.

When $\zeta(\alpha, \beta)$ satisfies these conditions, the Bayesian model will be said to be subject to the **tail convergence property**.

Remark 6.1. For the Tweedie sub-family of the EDF, (3.9) gives $c'(\mu) = \mu^{-p}$, and so $\zeta(\alpha, \beta) = \beta^{2-p}$. It may be checked that $\beta^{-1}\zeta(\alpha, \beta) \times \zeta(\beta, C/\zeta(\alpha, \beta)) = C^{2-p}\beta^{(2-p)-(2-p)^2-1} \to 0$ as $\beta \to \infty$. Thus, any member of the Tweedie sub-family has the tail convergence property.

Theorem 6.2. Suppose the Bayesian EDF cross-classified model defined in Section 6.1 is subject to the tail convergence property. Then at least one MAP estimator of the parameter set α_k, β_j exists.

Proof. The proof follows exactly the same type of argument as that of Theorem 4.1 (see appendix), by application of the Weierstrass theorem to a compact set in which any MAP estimator must lie. The existence of the compact set is guaranteed by Lemma A.9. ■

6.2.2 Uniqueness

Just as in Theorem 5.1, uniqueness of solutions will be proved by establishment of convexity of the log-likelihood function ℓ^{post}, defined by (6.3), with respect to $ln\ \alpha, ln\ \beta$. By (6.3), $\ell^{post} = \ell^{cond} + \ell^{prior}$ and, as convexity of $Th\ell^{cond}$ is proven in Theorem 5.1, a proof of the convexity of ℓ^{prior}, defined by (6.1), (6.2) and (6.5) will be sufficient to establish convexity of ℓ^{post}.

Let r_k, s_j be defined as in Section 5.1, and recall from there that

$$\partial/\partial r_k = \alpha_k\, \partial/\partial \alpha_k\, , \partial/\partial s_j = \beta_j\, \partial/\partial \beta_j$$

Now

$$\ell^{prior} = \sum_{k=1}^{K} [c(\alpha_k)A_k - d(\alpha_k)]/\psi_k^{(\alpha)} + \sum_{j=1}^{J} [c(\beta_j)B_j - d(\beta_j)]/\psi_j^{(\beta)}$$

Straightforward calculation, taking (4.3) into account, yields

$$\frac{\partial \ell^{prior}}{\partial r_k} = \alpha_k \frac{\partial \ell^{prior}}{\partial \alpha_k} = (A_k - \alpha_k)d'(\alpha_k)/\psi_k^{(\alpha)} \tag{6.16}$$

$$\frac{\partial^2 \ell^{prior}}{\partial r_k^2} = \alpha_k \frac{\partial}{\partial \alpha_k}\left(\alpha_k \frac{\partial \ell^{prior}}{\partial \alpha_k}\right) = \alpha_k[(A_k - \alpha_k)d''(\alpha_k) - d'(\alpha_k)]/\psi_k^{(\alpha)} \tag{6.17}$$

Similarly

$$\frac{\partial^2 \ell^{prior}}{\partial s_j^2} = \beta_j[(B_j - \beta_j)d''(\beta_j) - d'(\beta_j)]/\psi_j^{(\beta)} \tag{6.18}$$

All other second derivatives of ℓ^{prior} are zero.

Now ℓ^{prior} will be convex upward if and only if all derivatives $\partial^2 \ell^{prior}/\partial r_k^2, \partial^2 \ell^{prior}/\partial s_j^2 < 0$. Consider the first of these requirements. It requires that

$(A_k - \alpha_k) d''(\alpha_k) - d'(\alpha_k) < 0$
i.e.
$\frac{a_k}{\alpha_k} < 1 + \frac{d'(\alpha_k)}{\alpha_k d''(\alpha_k)}$ if $d''(\alpha_k) > 0$
or
$\frac{a_k}{\alpha_k} > 1 + \frac{d'(\alpha_k)}{\alpha_k d''(\alpha_k)}$ if $d''(\alpha_k) < 0$ (6.19)

The following theorem, parallel to Theorem 5.1, results.

Theorem 6.3. Let the array $\mathfrak{D}$ be regular, and suppose that the Y_{kj} are subject to the Bayesian EDF cross-classified model of Section 6.1 and subject to the tail convergence property. Let R_μ denote a compact set $\left\{\mu_{kj}: 0 < \underline{\mu}_{kj} < \mu_{kj} < \overline{\mu}_{kj}, (k,j) \; such \; that \; Y_{kj} \in \mathfrak{D}\right\}$, where the $\underline{\mu}_{kj}, \overline{\mu}_{kj}$ are known bounds. Let $R_{\alpha,\beta}$ denote a compact set $\left\{\alpha_k, \beta_j: 0 < \underline{\alpha}_k < \alpha_k < \overline{\alpha}_k, 0 < \underline{\beta_j} < \beta_j < \overline{\beta}_j, k = 1, \dots, K, j = 1, \dots, J\right\}$, where the $\underline{\alpha}_k, \overline{\alpha}_k, \underline{\beta_j}, \overline{\beta}_j$ are known bounds.

Then a sufficient condition for the posterior log-likelihood ℓ^{post} to be convex upward over R is that, for all $\left\{\mu_{kj}, (k,j) \; such \; that \; Y_{kj} \in \mathfrak{D}\right\} \in R_\mu$ and for all $\left\{\alpha_k, \beta_j: k = 1, \dots, K, j = 1, \dots, J\right\} \in R_{\alpha,\beta}$, all of the following hold:
$\frac{Y_{kj}}{\mu_{kj}} \leq 1 + \frac{d'(\mu_{kj})}{\mu_{kj} d''(\mu_{kj})}$ whenever $d''(\mu_{kj}) > 0$
$\frac{Y_{kj}}{\mu_{kj}} \geq 1 - \frac{d'(\mu_{kj})}{-\mu_{kj} d''(\mu_{kj})}$ whenever $d''(\mu_{kj}) < 0$

with strict inequality for at least one pair (k,j) (not necessarily the same (k,j) at each point of R_μ);

$\frac{A_k}{\alpha_k} \leq 1 + \frac{d'(\alpha_k)}{\alpha_k d''(\alpha_k)}$ whenever $d''(\alpha_k) > 0$
$\frac{A_k}{\alpha_k} \geq 1 + \frac{d'(\alpha_k)}{\alpha_k d''(\alpha_k)}$ whenever $d''(\alpha_k) < 0$
$\frac{B_j}{\beta_j} \leq 1 + \frac{d'(\beta_j)}{\beta_j d''(\beta_j)}$ whenever $d''(\beta_j) > 0$
$\frac{B_j}{\beta_j} \geq 1 + \frac{d'(\beta_j)}{\beta_j d''(\beta_j)}$ whenever $d''(\beta_j) < 0$

with strict inequality for at least one value of k or j (not necessarily the same k or j at each point of $R_{\alpha,\beta}$).

This is therefore a sufficient condition for a unique MAP estimator of the model parameters. ■

Corollary 6.4. The compact set R_μ in Theorem 6.3 may be defined by (5.8), where the bounds $\underline{\alpha}_k, \underline{\beta}_j, \overline{\alpha}_k, \overline{\beta}_j$ are defined as in the proof of Lemma A.3. The compact set $R_{\alpha,\beta}$ may be defined by the bounds $\underline{\alpha}_k, \underline{\beta}_j, \overline{\alpha}_k, \overline{\beta}_j$ from Lemma A.9. ■

Corollary 6.5. Consider the special case of Theorem 6.3 in which the Y_{kj} are subject to a Tweedie distribution with index $p \geq 1$. Then conditions for uniqueness of MAP estimators reduce to the following:

$$\frac{Y_{kj}}{\overline{\mu}_{kj}} \geq \frac{p-2}{p-1} \tag{6.20}$$

with strict inequality for one pair (k, j);

$$\frac{A_k}{\alpha_k} > \frac{p-2}{p-1} \tag{6.21}$$

$$\frac{B_j}{\beta_j} > \frac{p-2}{p-1} \tag{6.22}$$

with strict inequality for one value of k or j. ■

Proof. See proof of Corollary 5.4.

Corollary 6.6. Consider the special case of Theorem 6.3 in which the Y_{kj} are subject to a Tweedie distribution with index $1 \leq p \leq 2$. Then there is a unique MLE of the model parameters.

Proof. Substitution of p in this range into the right sides of (6.20)-(6.22) yields a negative result. Thus each of these three inequalities is always satisfied since $Y_{kj}/\mu_{kj}, A_k/\alpha_k, B_j/\beta_j > 0$. ■

Corollary 6.7. Consider the special case of Theorem 6.3 in which the Y_{kj} satisfy one of the following conditions:

(a) All are subject to an ODP distribution (which includes simple Poisson as a special case);
(b) All are subject to a gamma distribution;
(c) All are subject to a compound Poisson distribution with gamma severity distribution.

Then there is a unique MAP estimator of the model parameters.

Proof. See proof of Corollary 5.6. ■

7. Multiple solutions of maximum likelihood equations

The foregoing sections established upward convexity of the relevant log-likelihood function under certain circumstances, and uniqueness of chain ladder solutions, whether MLEs or MAP estimators, followed in these cases.

Section 5, dealing with non-Bayesian chain ladder models, also demonstrated that, under certain circumstances, the log-likelihood function is not convex. The proofs of uniqueness therefore do not follow in this cases. As pointed out in Section 5.4, this does not prove non-uniqueness, but simply leaves the question of uniqueness open.

One is left to consider whether uniqueness might always occur, but has simply not been proven here. The present section investigates this question by examining a specific simple data set, and searching for the existence of multiple MLEs.

The simplest form of regular array occurs in the case $K = J = 2$. By simply re-scaling the array, which would simply re-scale the MLE and so not affect the question of uniqueness, one of the elements of the array may be set to unity. Therefore consider an array of the form

$$\mathfrak{D} = \begin{bmatrix} Y_{11} & Y_{12} \\ Y_{21} & 1 \end{bmatrix} \tag{6.23}$$

In fact, even this degree of generality will not be required for present purpose, and it will be sufficient to consider the more specific form

$$\mathfrak{D} = \begin{bmatrix} 1 & Y \\ Y & 1 \end{bmatrix} \tag{6.24}$$

with inverse Gaussian (Tweedie, p=3) error terms.

Remark 7.1. An extremely simple situation is under consideration. The data set (6.24), containing 4 observations, has only one free variable, and the cross-classified chain ladder model, when applied to it, has only three free parameters.

It is also possible to draw some inferences about the solutions of its ML equations from the symmetry of the array (6.24). The array is symmetric under interchange or rows, accompanied by interchange of columns. The conclusions from this are:

(a) There must be a solution according to which $\alpha_1 = \alpha_2, \beta_1 = \beta_2$.
(b) If there is a solution for which $\alpha_1 \neq \alpha_2$ (or $\beta_1 \neq \beta_2$), then there must be further solution in which the values of α_1, α_2 (or β_1, β_2) are interchanged.

■

The ML equations for the case of general p, K, J are obtained from (3.9), (4.4) and (4.5), as the following:

$$\alpha_k = \frac{\sum_{\mathcal{R}(k)} Y_{kj} {\beta_j}^{1-p} / \phi_{kj}}{\sum_{\mathcal{R}(k)} {\beta_j}^{2-p} / \phi_{kj}}$$

$$\beta_j = \frac{\sum_{\mathcal{C}(j)} Y_{kj} {\alpha_k}^{1-p} / \phi_{kj}}{\sum_{\mathcal{C}(j)} {\alpha_k}^{2-p} / \phi_{kj}}$$

The parameter set for this array is $\{\alpha_k, \beta_j : k = 1,2; j = 1,2\}$. According to (EDFCC3c), this parameter set can be subjected to the constraint $\alpha_1 + \alpha_2 = 1$, but an equivalent model can be obtained by setting $\alpha_1 = 1$ instead. It will also be assumed that $\phi_{kj} = 1$.

The ML equations for the remaining parameters $\alpha_2, \beta_1, \beta_2$ when $p = 3$ are then

$$b_1 = \frac{1+a}{1+Ya^2}, b_2 = \frac{1+a}{Y+a^2}, a = \frac{b_1+b_2}{Yb_1^2+b_2^2} \tag{6.25}$$

where $b_j = \beta_j^{-1}, a = \alpha_2^{-1}$.

The symmetry of $\mathfrak{D}$ suggest the solution $\alpha_1 = \alpha_2, b_1 = b_2$, and it may indeed be checked that a solution of (6.25) is $\alpha_1 = \alpha_2 = 1, b_1 = b_2 = 2/(1+Y)$. However, it is possible to search for other admissible solutions.

Substitution for b_1 and b_2 in the expression for a, and slight rearrangement, yields

$$Ya^5 - (1-Y+Y^2)a^4 + 2Ya^3 - 2Ya^2 + (1-Y+Y^2)a - Y = 0$$

As noted just above, this must have a root of $a = 1$, and so a factor of $a - 1$ may be removed from the left side, leaving

$$Ya^4 - (1-Y)^2a^3 - (1-4Y+Y^2)a^2 - (1-Y)^2a + Y = 0 \tag{6.26}$$

It is evident that, if $a = x$ is a root of this equation, then so is $a = x^{-1}$. So the left side of (6.26) must contain a factor of $(a-x)(a-x^{-1}) = a^2 - Xa + 1$, where $X = x + x^{-1}$.

Factorise the left side of (6.26) by equating coefficient of a^4, a^3, a^0 in the following:

$$Ya^4 - (1-Y)^2a^3 - (1-4Y+Y^2)a^2 - (1-Y)^2a + Y = (a^2 - Xa + 1)[Ya^2 + (XY - (1-Y)^2)a + Y] \tag{6.27}$$

The factorisation also requires equation of coefficient of a^2, a^1, which yields the following additional conditions:

$$Y - X[XY - (1-Y)^2] + Y = -(1-4Y+Y^2) \tag{6.28}$$
$$XY - (1-Y)^2 - XY = -(1-Y)^2 \tag{6.29}$$

Now (6.29) is an identity, and so adds no information. However, (6.28) requires that

$$YX^2 - (1-Y)^2X - (1-Y)^2 = 0$$

which yields the strictly positive solution

$$\begin{aligned} X &= Y - 1 \text{ } if \text{ } Y > 1 \\ &= 1/Y - 1 \text{ } if \text{ } Y < 1 \end{aligned} \tag{6.30}$$

Note that, for given Y, both Y and $1/Y$ lead to the same value of X, and hence the same values of x.

Now not all of these solutions in X lead to a solution in x since, for positive x, X has a minimum value of 2 (at x=1). Therefore, a solution in x is obtained only if $X \geq 2$, i.e., by (6.30), only if $Y \geq 3$ or $Y \leq {}^1/_3$. This result is unsurprising, because

Corollaries 5.8 and 5.9 guaranteed a convex log-likelihood unless array $\mathfrak{D}$ was sufficiently "non-proportional".

When Y satisfies this condition, the solutions $a = x$ and $a = 1/x$ of the ML equations are given by

$$x = ½\left[(Y-1) + \sqrt{(Y-1)^2 - 4}\right] \tag{6.31}$$

For the special cases $Y = 3$ or $Y = {}^1\!/_3$ this yields no more than two additional solutions $a = 1$, so distinct multiple solutions will be found if and only if $Y > 3$ or $Y < {}^1\!/_3$. Then the following three distinct solutions of the ML equations exist:

$$a = 1, ½\left[(Y-1) \pm \sqrt{(Y-1)^2 - 4}\right] \tag{6.32}$$

Note that the last two of these are reciprocals of each other. A statement of this analysis is as follows.

Result 7.2. Consider the special case of Theorem 5.1 in which the Y_{kj} are subject to an inverse Gaussian distribution (Tweedie with index $p = 3$), and the array $\mathfrak{D}$ is given by (6.24). The ML equations (4.4) and (4.5) reduce to (6.25) and:

(a) have unique solution $a = 1$ if ${}^1\!/_3 < Y < 3$; and
(b) otherwise have the multiple solution $a = 1, ½\left[(Y-1) \pm \sqrt{(Y-1)^2 - 4}\right]$.

■

Remark 7.3. These solutions are consistent with Remark 7.1. ■

A study of the convexity of the log-likelihood as Y increases through the value 3 is of interest. Let H denote the Hessian matrix of ℓ with respect to the r_k and s_j, evaluated at the solution corresponding to $a = 1$ by reference to (5.3)-(5.5). Upward convexity of ℓ occurs if H is negative definite, and this is the case if and only if all leading principal minors of $-H$ are strictly positive.

For the present example, the dimension of $-H$ is 4×4. Denote the $q \times q$ leading principal minor by Δ_q. Table 7.1 displays the values of these minor for a sample of values of Y increases through the value 3.

Table 7.1 Values of negative Hessian minors for varying Y

q	Value of Δ_q for						
	$Y = 2.5$	$Y = 2.9$	$Y = 2.99$	$Y = 3$	$Y = 3.01$	$Y = 3.1$	$Y = 3.5$
1	+1.143	+1.026	+1.003	+1	+0.998	+0.976	+0.889
2	+1.306	+1.052	+1.005	+1	+0.995	+0.952	+0.790
3	+1.198	+0.027	+0.003	0	-0.002	-0.023	-0.082
4	0	0	0	0	0	0	0

The table shows that $\Delta_4 = 0$ throughout. This simply reflects the one degree of redundancy in the parameter set $\{r_1, r_2, s_1, s_2\}$. This was removed in the above calculations by the constraint $a_1 = 1$, but the Hessian is free of any constraint.

If this constraint is applied for $Y < 3$, $-H$ is seen to positive definite. However, an interesting phenomenon occurs as $Y \to 3$. The Hessian tends to semi-definiteness, which it attains at $Y = 3$. For $Y \to 3$, it is not even positive semi-definite. The interpretation of this is that the stationary point of the log-likelihood ℓ corresponding to $a = 1$ changes from a maximum when $Y < 3$ to a saddle point as one passes to $Y > 3$.

It is also of interest to enquire into the properties of ℓ in the vicinity of its other stationary points, seen to occur in the case $Y > 3$. This is done in the following example.

Example 7.2. As an example, set $Y = 3.5$. This generates the solutions $a = 1,2,½$. Table 7.2 expresses these three solutions in their α, β parameterisations.

Table 7.2 Multiple solutions of ML equations

Solution a=	**Value of**			
	$\boldsymbol{\alpha_1}$	$\boldsymbol{\alpha_2}$	$\boldsymbol{\beta_1}$	$\boldsymbol{\beta_2}$
1	1	1	0.444	0.444
2	1	2	0.444	0.444
0.5	1	0.5	0.444	0.444

As is apparent form the reasoning leading to these three solutions, they are stationary points of the log-likelihood, but the nature of the stationary points is unknown at this stage. It is determined by reference to the Hessian matrix. The negative of this matrix, evaluated at the three respective stationary points, is found from (5.3)-(5.5) to be as follows:

$$-H(a=1) = \begin{bmatrix} +0.889 & 0 & -0.049 & +0.938 \\ 0 & +0.889 & +0.938 & -0.049 \\ -0.049 & +0.938 & +0.889 & 0 \\ +0.938 & -0.049 & 0 & +0.889 \end{bmatrix}$$

$$-H(a=2) = \begin{bmatrix} +0.60 & 0 & -0.12 & +0.72 \\ 0 & +1.20 & +0.72 & +0.48 \\ -0.12 & +0.72 & +0.60 & 0 \\ +0.72 & +0.48 & 0 & +1.20 \end{bmatrix}$$

$$-H(a=0.5) = \begin{bmatrix} +1.20 & 0 & +0.48 & +0.72 \\ 0 & +0.60 & +0.72 & -0.12 \\ +0.48 & +0.72 & +1.20 & 0 \\ +0.72 & -0.12 & 0 & +0.60 \end{bmatrix}$$

Table 7.3 displays the leading principal minors of these matrices, found from (5.3)-(5.5), just as in Table 7.1. Once again, the zero values of Δ_q merely reflect the one degree of parameter redundancy.

Table 7.3 Values of negative Hessian minors for multiple solutions at $\boldsymbol{Y = 3.5}$

$\boldsymbol{q}$	**Value of $\boldsymbol{\Delta_q}$ for solution at**

	$a = 1$	$a = 2$	$a = 0.5$
1	+0.889	+0.600	+1.20
2	+0.790	+0.720	+0.720
3	-0.082	+0.104	+0.104
4	0	0	0

The stationary point of the log-likelihood at $a = 1$ was already known to be a saddle point, but the other two stationary points are seen here to be maxima.

Although the two solutions at $a = 2, 0.5$ are local maxima, it is conceivable that they are dominated by the saddle point at $a = 1$. Table 7.4 checks this by tabulating values of the (partial) log-likelihood ℓ, and finds it not to be the case. The two solutions at $a = 2, 0.5$ are global maxima.

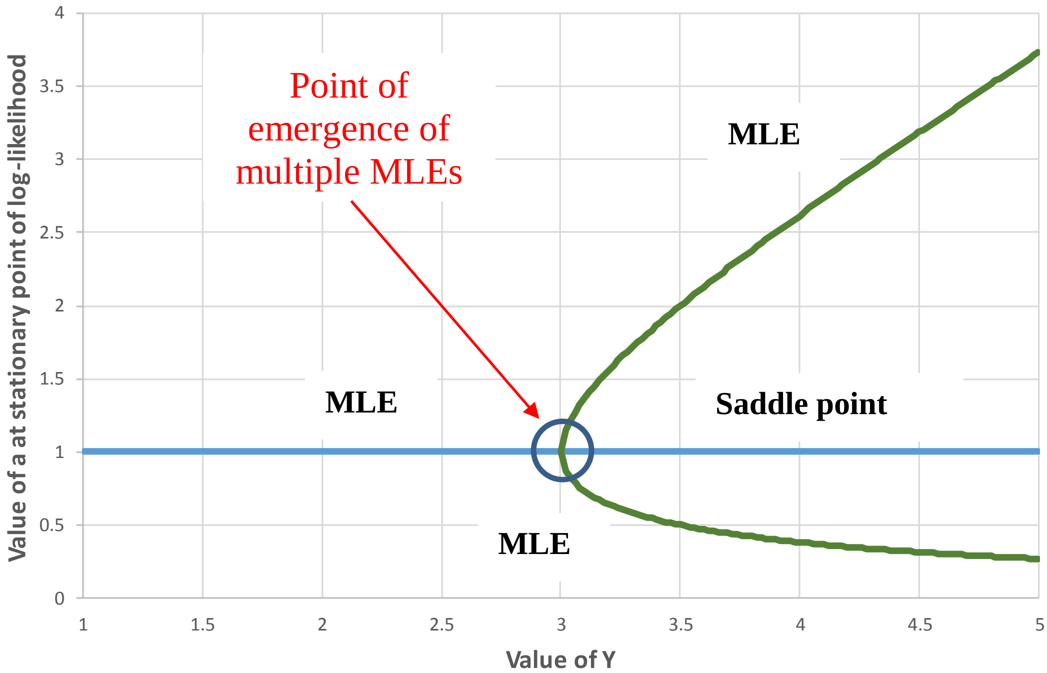
Table 7.4 Values of (partial) log-likelihood for multiple solutions at $Y = 3.5$

Solution of ML equations	Value of partial log-likelihood
$a = 1$	0.889
$a = 2$	0.900
$a = 0.5$	0.900

The difference between the global maxima of ℓ and its value at the saddle point is very small in this case, raising a question about rounding error, but further numerical investigation reveals that the difference between maxima and saddle point grows systematically as Y increases. For example, at $Y = 10$, the maximum is 0.61, compared with a value of 0.36 at the saddle point.

Empirically, then, the loci of stationary of ℓ with varying Y are as illustrated in Figure 7-1.

Figure 7-1 Loci of stationary points of ℓ with varying Y



8. Conclusion

The ML equations (4.1) and (4.2) for the EDF cross-classified chain ladder model are implicit equations. In the absence of explicit solutions, a question arises about the existence and uniqueness of solutions. Similarly for the Bayesian version of this model, whose solution is given by (6.7)-(6.14).

Some papers in the literature (e.g. Taylor (2009, 2015)) derive iterative solutions of these systems of equations. This assumes, in effect, that a unique solution exists in the case of each problem considered. These numerical solutions would be informed by a knowledge of the circumstances in which this assumption holds.

Sections 4 to 6 obtain certain results in this direction. Subject to some mild regularity conditions, existence always occurs. The results in relation to uniqueness are clearest for Tweedie error distributions, and may be summarised this way. A unique MLE exists provided that either:

(a) the error distribution is sufficiently well-behaved (Tweedie index $1 \le p \le 2$); or
(b) the data array is sufficiently well-behaved (close enough to proportional).

If both of these conditions are breached, i.e. both $p > 2$ and data array sufficiently non-proportional, then the results of the cited sections do not establish uniqueness. Indeed, it is shown in Section 7 that multiple solutions can occur.

The data array in the example of Section 7 has been deliberately chosen to be as simple as possible. It contains only 4 observations, has only one free variable, and the cross-classified chain ladder model, when applied to it, has only three free parameters. An "ill behaved" inverse Gaussian ($p = 3$) error distribution is assumed.

One might be forgiven for expecting the likelihood for this array to exhibit reasonably bland behaviour. Not so, however. As mentioned just above, uniqueness of MLE is not guaranteed as the single free parameter in the array is varied to produce increasing non-proportionality. Beyond a critical point for this parameter, the following occur:

(a) the stationary point of the likelihood that was indeed an MLE in the well behaved region continues to be a stationary point, but becomes a saddle point instead of an MLE.
(b) two new stationary points emerge, coincident with the original one, diverging from the original as non-proportionality of the array increases further, and these stationary points both become MLEs, with equal likelihood values, exceeding the likelihood at the saddle point.

Acknowledgement

This research was supported under Australian Research Council's Linkage Projects funding scheme (project number LP130100723). The views expressed herein are those of the author and are not necessarily those of the Australian Research Council.

Appendix
Proof of Theorem 4.1

Lemma A.1. For the EDF cross-classified model defined in Section 3.2.1, any MLE of the parameter set α_k, β_j must have the property

$$\rho^{-2}(\mathfrak{D}) \leq \alpha_{k_1}/\alpha_{k_2}, \beta_{j_1}/\beta_{j_2} \leq \rho^2(\mathfrak{D}), k_1, k_2 = 1, \dots, K; j_1, j_2 = 1, \dots, J \tag{A.1}$$

Proof. First note that

$$c'(\mu) = 1/\kappa'\big(c(\mu)\big) > 0 \tag{A.2}$$

according to the condition on κ' in (EDFCC3a).

Now substitute $\alpha_k \beta_j$ for the first μ_{kj} in (4.4) to yield

$$\alpha_k = \frac{\sum_{\mathcal{R}(k)} [Y_{kj}/\beta_j] c'(\mu_{kj}) \beta_j^2/\phi_{kj}}{\sum_{\mathcal{R}(k)} c'(\mu_{kj}) \beta_j^2/\phi_{kj}} \tag{A.3}$$

a weighted average of the values Y_{kj}/β_j over row k, where the weights are all strictly positive by virtue of (A.2). Note that the summations in (A.3) are guaranteed non-null by (EDFCC1). It follows that

$$\min_{\mathcal{R}(k)} Y_{kj}/\beta_j \leq \alpha_k \leq \max_{\mathcal{R}(k)} Y_{kj}/\beta_j \tag{A.4}$$

Then

$$\frac{\min_{\mathcal{R}(k_1)} Y_{k_1 j}/\beta_j}{\max_{\mathcal{R}(k_2)} Y_{k_2 j}/\beta_j} \leq \frac{\alpha_{k_1}}{\alpha_{k_2}} \leq \frac{\max_{\mathcal{R}(k_1)} Y_{k_1 j}/\beta_j}{\min_{\mathcal{R}(k_2)} Y_{k_2 j}/\beta_j} \tag{A.5}$$

Consider the inequality on the right, and suppose that

$$\max_{\mathcal{R}(k_1)} Y_{k_1 j}/\beta_j = Y_{k_1 j_1}/\beta_{j_1} \tag{A.6}$$

$$\min_{\mathcal{R}(k_2)} Y_{k_2 j}/\beta_j = Y_{k_2 j_2}/\beta_{j_2} \tag{A.7}$$

so that

$$\frac{\alpha_{k_1}}{\alpha_{k_2}} \leq \frac{Y_{k_1 j_1}/\beta_{j_1}}{Y_{k_2 j_2}/\beta_{j_2}} \tag{A.8}$$

Now, by (A.6),

$$Y_{k_1 j_1}/\beta_{j_1} \geq Y_{k_1 j_2}/\beta_{j_2} = \big(Y_{k_1 j_1}/\beta_{j_1}\big)\big(Y_{k_1 j_2}/Y_{k_1 j_1}\big)\big(\beta_{j_1}/\beta_{j_2}\big)$$

whence

$$\big(\beta_{j_1}/\beta_{j_2}\big) \leq Y_{k_1 j_1}/Y_{k_1 j_2} \leq \rho(\mathfrak{D}) \tag{A.9}$$

Then, by (A.8) and (A.9),
$$\frac{\alpha_{k_1}}{\alpha_{k_2}} \leq \rho^2(\mathfrak{D})$$

This proves one of the four inequalities contained in (A.1). The remaining three are similarly proved. ■

Remark A.2. While the bounds (A.1) are adequate for current purposes, it may be noted that they are not tight, at least in some special cases. For example, consider the Poisson chain ladder with $\phi_{kj} = 1$. Substitution of (3.12) in (4.4) and (4.5) yields

$$\sum\nolimits_{\mathcal{R}(k)} [Y_{kj} - \mu_{kj}] = 0$$
$$\sum\nolimits_{\mathcal{C}(j)} [Y_{kj} - \mu_{kj}] = 0$$

from which the following conventional chain ladder results are obtained:

$$\alpha_k = \frac{\sum_{\mathcal{R}(k)} Y_{kj}}{\sum_{\mathcal{R}(k)} \beta_j} \tag{A.10}$$
$$\beta_j = \frac{\sum_{\mathcal{C}(j)} Y_{kj}}{\sum_{\mathcal{C}(j)} \alpha_k} \tag{A.11}$$

In the case where all observations are present in $\mathfrak{D}$ it then follows that

$$\alpha_k = \sum\nolimits_{\mathcal{R}(k)} Y_{kj}$$
and so

$$\rho^{-1}(\mathfrak{D}) \leq \frac{\alpha_{k_1}}{\alpha_{k_2}} \leq \rho(\mathfrak{D})$$

tighter than (A.1). ■

Lemma A.3. For the EDF cross-classified model defined in Section 3.2.1, any MLE of the parameter set α_k, β_j must lie within a co-ordinate rectangle of the form $\left\{0 < \underline{\alpha}_k \leq \alpha_k \leq \overline{\alpha}_k, 0 < \underline{\beta_j}_j \leq \beta_j \leq \overline{\beta}_j; k = 1, \ldots, K, j = 1, \ldots, J\right\}$ for constants $\underline{\alpha}_k, \overline{\alpha}_k, \underline{\beta_j}, \overline{\beta}_j$.

Proof. Recall from (EDFCC3c) that $\sum_{j=1}^{J} \beta_j = 1$. Consider an arbitrary β_i, which may be expressed as

$$\beta_i = \frac{1}{1 + \sum_{\substack{j=1 \\ j \neq i}}^{J} \frac{\beta_j}{\beta_i}}$$

Combination of this result with (A.1) yields

$$\frac{1}{1+(J-1)\rho^2(\mathfrak{D})} \leq \beta_i \leq \frac{1}{1+(J-1)\rho^{-2}(\mathfrak{D})} \tag{A.12}$$

To establish bounds on α_k for a specific k, commence with an assumption that

$$\alpha_k < \frac{\min_{\mathcal{R}(k)} Y_{kj}}{\max_{\mathcal{R}(k)} \beta_j} \tag{A.13}$$

It then follows that

$$Y_{kj} - \mu_{kj} = Y_{kj} - \alpha_k \beta_j > 0 \text{ for all } j \in \mathcal{R}(k)$$

whence the left side of (4.4) is strictly positive, by virtue of (A.2).

This contradicts the required equality (4.4), from which it must be concluded that

$$\alpha_k \geq \frac{\min_{\mathcal{R}(k)} Y_{kj}}{\max_{\mathcal{R}(k)} \beta_j} \tag{A.14}$$

A similar argument establishes that

$$\alpha_k \leq \frac{\max_{\mathcal{R}(k)} Y_{kj}}{\min_{\mathcal{R}(k)} \beta_j} \tag{A.15}$$

and the lemma is proved. ■

Proof of Theorem 5.1

Assume that $\mathfrak{D}$ is regular. From (5.6),

$$Q(v,w) = \sum_{\mathfrak{D}} (v_k + w_j)^2 \left[\mu_{kj}^2 \frac{\partial^2 \ell}{\partial \mu_{kj}^2} + \mu_{kj} \frac{\partial \ell}{\partial \mu_{kj}}\right] \tag{A.16}$$

The second summand of the square bracket may be interpreted by reference to (5.1)

$$\begin{aligned} \mu_{kj} \frac{\partial \ell}{\partial \mu_{kj}} &= [Y_{kj} c'(\mu_{kj}) - d'(\mu_{kj})]\mu_{kj}/\phi_{kj} \\ &= \phi_{kj}^{-1}(Y_{kj} - \mu_{kj})\, d'(\mu_{kj}) \end{aligned} \tag{A.17}$$

by (4.3).

From this result, differentiate $\partial \ell / \partial \mu_{kj}$ a second time to obtain:

$$\mu_{kj}^2 \frac{\partial^2 \ell}{\partial \mu_{kj}^2} = \phi_{kj}^{-1}[\mu_{kj}(Y_{kj} - \mu_{kj}) d''(\mu_{kj}) - Y_{kj} d'(\mu_{kj})] \tag{A.18}$$

By substitution of (A.17) and (A.18) in (A.16),

$$-Q(v,w) = \sum_{\mathfrak{D}} \left(v_k + w_j\right)^2 \left[\mu_{kj} d'\left(\mu_{kj}\right) - \mu_{kj}\left(Y_{kj} - \mu_{kj}\right) d''\left(\mu_{kj}\right)\right]$$
$$= \sum_{\mathfrak{D}} \left(v_k + w_j\right)^2 \mu_{kj}^2 d''\left(\mu_{kj}\right) \left[\frac{d'\left(\mu_{kj}\right)}{\mu_{kj} d''\left(\mu_{kj}\right)} - \left(\frac{Y_{kj}}{\mu_{kj}} - 1\right)\right] \tag{A.19}$$

It follows that, when (5.7) holds, $Q(v,w) \leq 0$ for all selections of v, w with equality if and only if

$$v_k + w_j = 0 \text{ for all pairs } k, j \tag{A.20}$$

Now, for $\|v\| = \|w\| = 1$, $Q(v,w)$ is the second derivative of ℓ taken in the direction of vector (v,w). So second derivatives in all directions are non-positive over R, and in fact are strictly negative except along directions for which (A.20) holds.

Consider the interpretation of (A.20). It is seen that $w_j = -v_1$ for all j and $v_k = -w_1$ for all k. It follows that $v_k = v_1$ for all k. Thus (A.20) represents the direction described by a shift of all $r_k = ln\, \alpha_k$ by the same increment, and a shift of all $s_j = ln\, \beta_j$ by an equal and opposite increment, i.e. $\mu_{kj} = \alpha_k \beta_j$.

This shows that the second derivative of ℓ is zero in any direction in which all μ_{kj} are invariant. This is obvious from the fact that derivatives of ℓ depend on only the μ_{kj} (in fact, by this reasoning, the first derivative of ℓ is also zero in this direction). The second derivative is strictly negative in any other direction.

By compactness of R, it contains at least one maximum of log-likelihood ℓ (just as in the proof of Theorem 4.1). Moreover, convexity over compact R also guarantees uniqueness of the values $\{\mu_{kj}\}$ at which the maximum occurs.

This maximum will **not** correspond to a unique point $\{\alpha_k, \beta\}$. Suppose the maximum occurs at $\alpha_k = \alpha_k^*, \beta_j = \beta_j^*$. Then the same maximum also occurs at $\alpha_k = \gamma\alpha_k^*, \beta_j = \beta_j^*/\gamma$ for any $\gamma > 0$. However, there is a unique maximum satisfying (EDFCC3c), namely $\gamma = \sum_{j=1}^{J} \beta_j^*$.

In summary, when (5.7) holds, R contains a unique point $\{\mu_{kj}\}$ that maximises the log-likelihood ℓ, and this corresponds to a unique point $\{\alpha_k, \beta\}$ that satisfies (EDFCC3c). ■

Proof of Lemma 4.2

Proof. The proof of **Error! Reference source not found.** will be given; the proof of **Error! Reference source not found.** is similar.

Commence with the MLE equation (4.4), where $c'(\mu) > 0$ [by (4.6)]. Suppose that $\mu_{kj} < \min_{\mathcal{R}(k)} Y_{ks}$, in contradiction of **Error! Reference source not found.**. Then all summands on the right side of (4.4) are strictly positive, and (4.4) cannot hold.

Similarly if $\mu_{kj} > \max_{\mathcal{R}(k)} Y_{ks}$. Thus, (4.4) cannot hold unless **Error! Reference source not found.** holds. ■

Proof of Theorem 5.7

Lemma A.4. Suppose the array $\mathfrak{D}$ is regular and contains exactly $K + J - 1$ observations, subject to the EDF cross-classified model defined by (EDFCC1-3). Then a perfect fit of model to observations (i.e. $Y_{kj} = \mu_{kj} = \alpha_k \beta_j$) can be achieved, with explicit calculation of the α_k, β_j. This is the unique MLE.

Proof. Commence by setting $\beta_1 = 1$. This value will be re-scaled later in accordance with (EDFCC3c).

Since $\mathfrak{D}$ is regular and contains exactly $K + J - 1$ observations, it is uniquely its own core. It is therefore possible to select a path γ in $\Gamma(\mathfrak{D})$ that connects an element of the first column with an element of the last column of $\mathfrak{D}$. The existence of such a path is guaranteed by the connectedness of $\Gamma(\mathfrak{D})$. By definition of the edges of the graph, γ consists of a sequence of edges of the form $(Y_{kj}, Y_{k,j+1})$ or $(Y_{kj}, Y_{k+1,j})$.

Consider just the first of these forms, and set

$$\beta_{j+1}/\beta_j = Y_{k,j+1}/Y_{kj} \tag{A.21}$$

For fixed j, the estimator on the right exists for unique k, according to thr following argument. If it were not the case, one could find $k_1 \neq k$ such that the foursome of observations $Y_{kj}, Y_{k,j+1}, Y_{k_1j}, Y_{k_1,j+1}$ exists. That is, 4 observations would be concentrated in 2 rows and 2 columns. This would leave only $K + J - 5$ observations to account for $(K - 2) + (J - 2) = K + J - 4$ rows and columns, recalling that each row and each column must contain at least one observation.

Assume, without loss of generality, that the 4 observations named above occur as the top left 2×2 sub-array of $\mathfrak{D}$, so that $\mathfrak{D}$ takes the form

$$\begin{bmatrix} * & * & \\ * & * & \\ & & S \end{bmatrix}$$

with S a $(K - 2) \times (J - 2)$ sub-array containing at least one observation per row and column. This requires that all available $K + J - 5$ observations occur in S, in which case S is not connected to the top left sub-array, contradicting the regularity of $\mathfrak{D}$.

By the above argument, unique values of β_{j+1}/β_j can be found explicitly for all $j = 1, \dots J - 1$. This, together with the initial assumption that $\beta_1 = 1$, leads to calculation of all $\beta_j, j = 1, \dots, J$. These may now be re-scaled in accordance with (EDFCC3c). The ratios β_{j+1}/β_j are unchanged by this operation.

Now calculate the α_k as follows. Select an observation Y_{kj} in row k and calculate

$$\alpha_k = Y_{kj}/\beta_j \tag{A.22}$$

It remains to prove that this value of α_k is independent of the observation Y_{kj} in row k, and therefore unique. Suppose that there is a second observation Y_{kj_1} in the row. Choose a path γ in $\Gamma(\mathfrak{D})$ that connects Y_{kj} and Y_{kj_1}. This path will take the form $\{Y_{r_1s_1}, \dots, Y_{r_ms_m}\}$ where $(r_1, s_1) = (k, j), (r_m, s_m) = (k, j_1)$ and $(r_{i+1}, s_{i+1}) = (r_i \pm 1, s_i)$ or $(r_i, s_i + 1)$.

Now express

$$\frac{Y_{kj_1}}{Y_{kj}} = \frac{Y_{r_2s_2}}{Y_{r_1s_1}} \cdots \frac{Y_{r_ms_m}}{Y_{r_{m-1}s_{m-1}}} \tag{A.23}$$

Now each ratio on the right side of (A.23) must take the form α_{i+1}/α_i , α_{i-1}/α_i or β_{i+1}/β_i. The α ratios will cancel out, and the β ratios include all cases between j and j_1, so that (A.23) becomes

$$\frac{Y_{kj_1}}{Y_{kj}} = \frac{\beta_{j_1}}{\beta_j}$$

Rearranged, this is

$$\frac{Y_{kj_1}}{\beta_{j_1}} = \frac{Y_{kj}}{\beta_j}$$

which shows that (A.23) yields the same value of α_k if based on Y_{kj_1} instead of Y_{kj}.

Since this solution is a perfect fit to the data, the associated likelihood assumes its maximum possible value, and so the solution is unique. ■

Remark A.5. The result of Lemma A.4 applies more generally than most results used in the proof of Theorem 5.7. Although that corollary relates to the Tweedie cross-classified model with $p > 2$, Lemma A.4 applies to the more general EDF cross-classified model.

Lemma A.6. In the case of a Tweedie cross-classified model applied to a trapezoidal array, and subject to the multiplicative weights $\phi_{kj} = \left(v_k w_j\right)^{-1}$ the following relations hold:

$$\frac{\beta_{s+1}^{2-p} w_{s+1}}{\sum_{j=1}^{s} \beta_j^{2-p} w_j} = \frac{\sum_{k=1}^{K-s} Y_{k,s+1} \mu_{k,s+1}^{1-p} / \phi_{k,s+1}}{\sum_{j=1}^{s} \sum_{k=1}^{K-s} Y_{kj} \mu_{kj}^{1-p} / \phi_{kj}}, s = 1,2, \dots, J-1 \tag{A.24}$$

$$\frac{\alpha_{r+1}^{2-p} v_{r+1}}{\sum_{k=1}^{r} \alpha_k^{2-p} v_k} = \frac{\sum_{j=1}^{J_{r+1}} Y_{r+1,j} \mu_{r+1,j}^{1-p} / \phi_{r+1,j}}{\sum_{k=1}^{r} \sum_{j=1}^{J_{r+1}} Y_{kj} \mu_{kj}^{1-p} / \phi_{kj}}, r = 1,2, \ldots, K-1 \tag{A.25}$$

where $J_k = min(J, K-k+1) =$ maximum value of j for which an observation exists in row k.

Proof. Recall the MLE equations (4.4) and (4.5) for the EDF cross-classified model, reproduced here in a slightly modified form:

$$\sum_{\mathcal{R}(k)} [Y_{kj} - \mu_{kj}] c'(\mu_{kj}) \mu_{kj} / \phi_{kj} = 0 \tag{A.26}$$

$$\sum_{\mathcal{C}(j)} [Y_{kj} - \mu_{kj}] c'(\mu_{kj}) \mu_{kj} / \phi_{kj} = 0 \tag{A.27}$$

The array $\mathfrak{D}$ is equal to the union of all rows or all columns, and so

$$\sum_{\mathfrak{D}} [Y_{kj} - \mu_{kj}] c'(\mu_{kj}) \mu_{kj} / \phi_{kj} = \sum_k \sum_{\mathcal{R}(k)} [Y_{kj} - \mu_{kj}] c'(\mu_{kj}) \mu_{kj} / \phi_{kj} = \sum_j \sum_{\mathcal{C}(j)} [Y_{kj} - \mu_{kj}] c'(\mu_{kj}) \mu_{kj} / \phi_{kj} = 0 \tag{A.28}$$

It is evident that, if one deletes any number of complete rows and columns from $\mathfrak{D}$, the sum in (A.28) remains zero. If columns beyond the s-th, and rows beyond the $(K-s)$-th, are deleted, the result is:

$$\sum_{k=1}^{K-s} \sum_{j=1}^{s} [Y_{kj} - \mu_{kj}] c'(\mu_{kj}) \mu_{kj} / \phi_{kj} = 0 \tag{A.29}$$

Now express (4.5) for the $(s+1)$-th column with explicit summation limits:

$$\sum_{k=1}^{K-s} [Y_{k,s+1} - \mu_{k,s+1}] c'(\mu_{k,s+1}) \mu_{k,s+1} / \phi_{k,s+1} = 0 \tag{A.30}$$

For the special case of the Tweedie cross-classified model, $c(.)$ is given by (3.9) . With this substitution, with the special form of the weights recognised, and slight rearrangement, (A.29) and (A.30) become (for fixed s):

$$\frac{\sum_{k=1}^{K-s} Y_{k,s+1} \mu_{k,s+1}^{1-p} / \phi_{k,s+1}}{\sum_{j=1}^{s} \sum_{k=1}^{K-s} Y_{kj} \mu_{kj}^{1-p} / \phi_{kj}} = \frac{\sum_{k=1}^{K-s} \alpha_k^{2-p} \beta_{s+1}^{2-p} v_k w_{s+1}}{\sum_{j=1}^{s} \sum_{k=1}^{K-s} \alpha_k^{2-p} \beta_j^{2-p} v_k w_j} \tag{A.31}$$

Factorisation occurs in both numerator and denominator on the right side of this equation, to yield

$$\frac{\sum_{k=1}^{K-s} Y_{k,s+1} \mu_{k,s+1}^{1-p} / \phi_{k,s+1}}{\sum_{j=1}^{s} \sum_{k=1}^{K-s} Y_{kj} \mu_{kj}^{1-p} / \phi_{kj}} = \frac{\beta_{s+1}^{2-p} w_{s+1} \sum_{k=1}^{K-s} \alpha_k^{2-p} v_k}{\left[\sum_{j=1}^{s} \beta_j^{2-p} w_j\right] \left[\sum_{k=1}^{K-s} \alpha_k^{2-p} v_k\right]} \tag{A.32}$$

from which (A.24) follows. The proof of (A.25) is essentially the same with the roles of rows and columns interchanged. ■

Lemma A.7. Consider a trapezoidal array $\mathfrak{D}$ that contains more than $K+J-1$ observations. Let $\mathcal{S}$ be a core of $\mathfrak{D}$, and fit to $\mathcal{S}$ a Tweedie cross-classified model $Y_{kj} = \alpha_k^* \beta_j^* (= \mu_{kj}^*)$ say, subject to the multiplicative weights $\phi_{kj} = (v_k w_j)^{-1}$

(Lemma A.4 guarantees the possibility of this). Now fit the Tweedie cross-classified model $\mu_{kj} = \alpha_k \beta_j$ to the entire array $\mathfrak{D}$, and subject to the same multiplicative system of weights. Then the following relations hold:

$$\xi^{-1} \frac{\sum_{j=1}^{J_k-1} \beta_j^* \beta_j^{1-p} w_j}{\beta_{J_k}^* \beta_{J_k}^{1-p} w_{J_k}} \le \frac{\sum_{j=1}^{J_k-1} \beta_j^{2-p} w_j}{\beta_{J_k}^{2-p} w_{J_k}} \le \xi \frac{\sum_{j=1}^{J_k-1} \beta_j^* \beta_j^{1-p} w_j}{\beta_{J_k}^* \beta_{J_k}^{1-p} w_{J_k}} \tag{A.33}$$

$$\xi^{-1} \frac{\sum_{j=1}^{J_k-1} \beta_j^* \beta_j^{1-p} w_j}{\sum_{j=1}^{J_k} \beta_j^* \beta_j^{1-p} w_j} \le \frac{\sum_{j=1}^{J_k-1} \beta_j^{2-p} w_j}{\sum_{j=1}^{J_k} \beta_j^{2-p} w_j} \le \xi \frac{\sum_{j=1}^{J_k-1} \beta_j^* \beta_j^{1-p} w_j}{\sum_{j=1}^{J_k} \beta_j^* \beta_j^{1-p} w_j} \tag{A.34}$$

Proof. Select an arbitrary $Y_{rs} \notin \mathcal{S}$, and a path γ_{rs} from Y_{rj_r} to $Y_{k_s s}$, as defined in Section 2. It will be assumed that $s > j_r$; a similar argument deals with the case $s < j_r$. The path γ_{rs} involves all the observations appearing in (2.1), and so π_{rs} is defined.

By definition of π_{rs} in Section 2,

$$Y_{rs} = (1 + \pi_{rs}) Y_{rj_r} \prod_{\eta_{rs}} \frac{Y_{r_i s_i}}{Y_{r_{i+1} s_{i+1}}} \tag{A.35}$$

The observations Y appearing in the product lie within the core, to which the multiplicative model $Y_{kj} = \alpha_k^* \beta_j^*$ has been applied. It is remarked at the end of Section 2 that, in this case, $Y_{r_i s_i} / Y_{r_i, s_i+1} = \beta_{s_i}^* / \beta_{s_i+1}^*$ or $\beta_{s_i}^* / \beta_{s_i-1}^*$, with s_i running from j_r to s. Thus, (A.35) simplifies to the following:

$$Y_{rs} = (1 + \pi_{rs}) Y_{rj_r} \frac{\beta_s^*}{\beta_{j_r}^*} = (1 + \pi_{rs}) \alpha_r^* \beta_s^* \tag{A.36}$$

By (A.24),

$$\frac{\sum_{j=1}^{J_k-1} \beta_j^{2-p} w_j}{\beta_{J_k}^{2-p} w_{J_k}} = \frac{\sum_{j=1}^{J_k-1} \sum_{k=1}^{K-J_k+1} Y_{kj} \mu_{kj}^{1-p} / \phi_{kj}}{\sum_{k=1}^{K-J_k+1} Y_{k,J_k} \mu_{k,J_k}^{1-p} / \phi_{k,J_k}} \tag{A.37}$$

$$= \frac{\sum_{j=1}^{J_k-1} \sum_{k=1}^{K-J_k+1} (1 + \pi_{kj}) \alpha_k^* \beta_j^* \alpha_k^{1-p} \beta_j^{1-p} v_k w_j}{\sum_{k=1}^{K-J_k+1} (1 + \pi_{k,J_k}) \alpha_k^* \beta_{J_k}^* \alpha_k^{1-p} \beta_{J_k}^{1-p} v_k w_{J_k}}$$

[by (A.36)]

$$\le \xi \frac{\sum_{j=1}^{J_k-1} \sum_{k=1}^{K-J_k+1} \alpha_k^* \beta_j^* \alpha_k^{1-p} \beta_j^{1-p} v_k w_j}{\sum_{k=1}^{K-J_k+1} \alpha_k^* \beta_{J_k}^* \alpha_k^{1-p} \beta_{J_k}^{1-p} v_k w_{J_k}}$$

[by definition of ξ]

$$= \xi \frac{\left[\sum_{k=1}^{K-J_k+1} \alpha_k^* \alpha_k^{1-p} v_k\right] \left[\sum_{j=1}^{J_k-1} \beta_j^* \beta_j^{1-p} w_j\right]}{\left[\sum_{k=1}^{K-J_k+1} \alpha_k^* \alpha_k^{1-p} v_k\right] \left[\beta_{J_k}^* \beta_{J_k}^{1-p} w_{J_k}\right]}$$

$$= \xi \frac{\sum_{j=1}^{J_k-1} \beta_j^* \beta_j^{1-p} w_j}{\beta_{J_k}^* \beta_{J_k}^{1-p} w_{J_k}}$$

(A.38)

A parallel argument produces a corresponding inequality in the opposite direction, so that (A.33) holds.

In order to prove (A.34), note that (A.24) leads easily to the following relation, corresponding to (A.37) above:

$$\frac{\sum_{j=1}^{J_k-1} \beta_j^{2-p} w_j}{\sum_{j=1}^{J_k} \beta_j^{2-p} w_j} = \frac{\sum_{j=1}^{J_k-1} \sum_{k=1}^{K-J_k+1} Y_{kj} \mu_{kj}^{1-p} / \phi_{kj}}{\sum_{j=1}^{J_k} \sum_{k=1}^{K-J_k+1} Y_{kj} \mu_{kj}^{1-p} / \phi_{kj}}$$

(A.39)

An argument may now be applied precisely parallel to that which led from (A.37) to (A.38), and (A.34) follows. ■

Proof of Theorem 5.7. There are two cases to be considered:
Case I: The array $\mathfrak{D}$ contains exactly $K + J - 1$ observations.
Case II: The array $\mathfrak{D}$ contains more than $K + J - 1$ observations.

Case I. By Lemma A.4, a unique solution exists, and it is the proportional solution $Y_{kj} = \alpha_k \beta_j$. As remarked at the end of Section 2, $\pi(\mathfrak{D}) = 0$ in this case. It follows that (5.13) is satisfied.

Case II. This is the case of Lemma A.7, in whose proof (A.36) is established. Substitute this into (A.26) to obtain:

$$\sum_{\mathcal{R}(k)} [(1 + \pi_{rs}) \alpha_r^* \beta_s^* - \alpha_k \beta_j] \mu_{kj} c'(\mu_{kj}) / \phi_{kj} = 0 \tag{A.40}$$

Now substitute (3.9) into (A.40), recognise the special case of the weights, $\phi_{kj}^{-1} = v_k w_j$, and rearrange slightly, to obtain

$$\sum_{\mathcal{R}(k)} [(1 + \pi_{kj}) \beta_j^* - \frac{\alpha_k}{\alpha_k^*} \beta_j] \beta_j^{1-p} w_j = 0 \tag{A.41}$$

It follows that

$$\frac{\alpha_k}{\alpha_k^*} = \frac{\sum_{\mathcal{R}(k)} (1 + \pi_{kj}) \beta_j^* \beta_j^{1-p} w_j}{\sum_{\mathcal{R}(k)} \beta_j^{2-p} w_j}$$

and then

$$\frac{\alpha_k \beta_s}{\alpha_k^* \beta_s^*} = \frac{\sum_{\mathcal{R}(k)} (1 + \pi_{kj}) \beta_j^* \beta_j^{1-p} w_j}{\sum_{\mathcal{R}(k)} \beta_j^{2-p} w_j} \times \frac{\beta_s}{\beta_s^*} \leq (1 + \overline{\pi}) \frac{\sum_{\mathcal{R}(k)} \beta_j^* \beta_j^{1-p} w_j}{\sum_{\mathcal{R}(k)} \beta_j^{2-p} w_j} \times \frac{\beta_s}{\beta_s^*}$$

(A.42)

by definition of $\overline{\pi}$.

For the trapezoidal array under consideration, the maximum value of j for which Y_{kj} exists in row k is defined in Lemma A.6 as J_k. It will be convenient to re-express (A.42) in the form:

$$\frac{\alpha_k\beta_s}{\alpha_k^*\beta_s^*} \leq (1+\overline{\pi})\frac{\sum_{j=1}^{J_k-1}\beta_j^*\beta_j^{1-p}w_j}{\beta_{J_k}^*\beta_{J_k}^{1-p}w_{J_k}}\Bigg/\frac{\sum_{j=1}^{J_k-1}\beta_j^{2-p}w_j}{\beta_{J_k}^{2-p}w_{J_k}} \times \frac{\beta_{J_k}^*}{\beta_s^*}\Big/\frac{\beta_{J_k}}{\beta_s} \tag{A.43}$$

Now note that one of the members on the right is of the same form (subject to reciprocation) as the middle member of (A.33), whence Lemma A.7 yields

$$\xi^{-1}\frac{\sum_{j=1}^{J_k-1}\beta_j^*\beta_j^{1-p}w_j}{\beta_{J_k}^*\beta_{J_k}^{1-p}w_{J_k}} \leq \frac{\sum_{j=1}^{J_k-1}\beta_j^{2-p}w_j}{\beta_{J_k}^{2-p}w_{J_k}} \leq \xi\frac{\sum_{j=1}^{J_k-1}\beta_j^*\beta_j^{1-p}w_j}{\beta_{J_k}^*\beta_{J_k}^{1-p}w_{J_k}} \tag{A.44}$$

Substitution of this result in (A.43), together with the simple extension of (A.43) to a two-sided inequality, yields

$$(1+\underline{\pi})\xi^{-1}\frac{\beta_{J_k}^*}{\beta_s^*}\Big/\frac{\beta_{J_k}}{\beta_s} \leq \frac{\alpha_k\beta_s}{\alpha_k^*\beta_s^*} \leq (1+\overline{\pi})\xi\frac{\beta_{J_k}^*}{\beta_s^*}\Big/\frac{\beta_{J_k}}{\beta_s} \tag{A.45}$$

Inequalities can be placed on the ratio that appears on left and right of this result, as follows. There are two cases to be considered: $s \leq J_k$ and $s \geq J_k$ respectively.

Case II(a): $s = J_k$
In this case (A.45) reduces trivially to the following:

$$(1+\underline{\pi})\xi^{-1} \leq \frac{\alpha_k\beta_s}{\alpha_k^*\beta_s^*} \leq (1+\overline{\pi})\xi \tag{A.46}$$

Case II(b): $s < J_k$
Commence by expressing the final ratio in (A.45) in the form:

$$\begin{aligned}\frac{\beta_{J_k}^{2-p}w_{J_k}}{\beta_s^{2-p}w_s} &= \left(\frac{\sum_{j=1}^{s}\beta_j^{2-p}w_j}{\beta_s^{2-p}w_s}\right)\left(\frac{\sum_{j=1}^{s+1}\beta_j^{2-p}w_j}{\sum_{j=1}^{s}\beta_j^{2-p}w_j}\right)\cdots\left(\frac{\sum_{j=1}^{J_k-1}\beta_j^{2-p}w_j}{\sum_{j=1}^{J_k-2}\beta_j^{2-p}w_j}\right)\left(\frac{\beta_{J_k}^{2-p}w_{J_k}}{\sum_{j=1}^{J_k-1}\beta_j^{2-p}w_j}\right)\\ &= \left(1+\frac{\sum_{j=1}^{s-1}\beta_j^{2-p}w_j}{\beta_s^{2-p}w_s}\right)\left(\frac{\sum_{j=1}^{s+1}\beta_j^{2-p}w_j}{\sum_{j=1}^{s}\beta_j^{2-p}w_j}\right)\cdots\left(\frac{\sum_{j=1}^{J_k-1}\beta_j^{2-p}w_j}{\sum_{j=1}^{J_k-2}\beta_j^{2-p}w_j}\right)\left(\frac{\beta_{J_k}^{2-p}w_{J_k}}{\sum_{j=1}^{J_k-1}\beta_j^{2-p}w_j}\right)\end{aligned} \tag{A.47}$$

All of the ratios on the right are of the form (subject to reciprocation) of the subject quantities in inequalities (A.33) and (A.34). it therefore follows from these inequalities that

$$\frac{\beta_{J_k}^{2-p}w_{J_k}}{\beta_s^{2-p}w_s} \leq \xi^{J_k-s}\left(1+\xi\frac{\sum_{j=1}^{s-1}\beta_j^*\beta_j^{1-p}w_j}{\beta_s^*\beta_s^{1-p}w_s}\right)\left(\frac{\sum_{j=1}^{s+1}\beta_j^*\beta_j^{1-p}w_j}{\sum_{j=1}^{s}\beta_j^*\beta_j^{1-p}w_j}\right)\cdots$$

$$\ldots\left(\frac{\sum_{j=1}^{J_k-1}\beta_j^*\beta_j^{1-p}w_j}{\sum_{j=1}^{J_k-2}\beta_j^*\beta_j^{1-p}w_j}\right)\left(\frac{\beta_{J_k}^*\beta_{J_k}^{1-p}w_{J_k}}{\sum_{j=1}^{J_k-1}\beta_j^*\beta_j^{1-p}w_j}\right)$$

$$\leq \xi^{J_k-s+1}\left(1+\frac{\sum_{j=1}^{s-1}\beta_j^*\beta_j^{1-p}w_j}{\beta_s^*\beta_s^{1-p}w_s}\right)\left(\frac{\sum_{j=1}^{s+1}\beta_j^*\beta_j^{1-p}w_j}{\sum_{j=1}^{s}\beta_j^*\beta_j^{1-p}w_j}\right)\ldots$$
$$\ldots\left(\frac{\sum_{j=1}^{J_k-1}\beta_j^*\beta_j^{1-p}w_j}{\sum_{j=1}^{J_k-2}\beta_j^*\beta_j^{1-p}w_j}\right)\left(\frac{\beta_{J_k}^*\beta_{J_k}^{1-p}w_{J_k}}{\sum_{j=1}^{J_k-1}\beta_j^*\beta_j^{1-p}w_j}\right)$$

$$= \xi^{J_k-s+1}\frac{\beta_{J_k}^*\beta_{J_k}^{1-p}w_{J_k}}{\beta_s^*\beta_s^{1-p}w_s}$$

(A.48)

Then

$$\frac{\beta_{J_k}}{\beta_s}\Big/\frac{\beta_{J_k}^*}{\beta_s^*}\leq \xi^{J_k-s+1}$$

A lower bound can also be found by the same argument, yielding

$$\xi^{-(J_k-s+1)}\leq\frac{\beta_{J_k}}{\beta_s}\Big/\frac{\beta_{J_k}^*}{\beta_s^*}\leq \xi^{J_k-s+1}$$

(A.49)

Substitution of (A.49) into (A.45) yields

$$(1+\underline{\pi})\xi^{-(J_k-s+2)}\leq\frac{\alpha_k\beta_s}{\alpha_k^*\beta_s^*}\leq(1+\overline{\pi})\xi^{J_k-s+2}$$

(A.50)

Case II(c): $s > J_k$

The argument runs parallel to that of Case II(b). It commences with the following relation in place of (A.47):

$$\frac{\beta_{J_k}^{2-p}w_{J_k}}{\beta_s^{2-p}w_s}=\left(\frac{\beta_{J_k}^{2-p}w_{J_k}}{\sum_{j=1}^{J_k}\beta_j^{2-p}w_j}\right)\left(\frac{\sum_{j=1}^{J_k}\beta_j^{2-p}w_j}{\sum_{j=1}^{J_k+1}\beta_j^{2-p}w_j}\right)\ldots\left(\frac{\sum_{j=1}^{s-2}\beta_j^{2-p}w_j}{\sum_{j=1}^{s-1}\beta_j^{2-p}w_j}\right)\left(\frac{\sum_{j=1}^{s-1}\beta_j^{2-p}w_j}{\beta_s^{2-p}w_s}\right)$$

and then continues in parallel with Case II(b), leading ultimately to the following result in place of (A.50):

$$(1+\underline{\pi})\xi^{-(s-J_k+2)}\leq\frac{\alpha_k\beta_s}{\alpha_k^*\beta_s^*}\leq(1+\overline{\pi})\xi^{s-J_k+2}$$

(A.51)■

Proof of Corollary 5.8

Use (A.36) to express $Y_{kj}/\overline{\mu}_{kj}$ in the form:

$$\frac{Y_{kj}}{\overline{\mu}_{kj}} = \frac{(1+\pi_{kj})\alpha_k^*\beta_j^*}{\overline{\mu}_{kj}} \tag{A.52}$$

Now

$$\underline{\pi} \leq \pi_{kj} \leq \overline{\pi} \qquad \text{[by definition of } \overline{\pi}] \tag{A.53}$$

so that

$$(1+\underline{\pi})\frac{\alpha_k^*\beta_j^*}{\overline{\mu}_{kj}} \leq \frac{Y_{kj}}{\overline{\mu}_{kj}} \leq (1+\overline{\pi})\frac{\alpha_k^*\beta_j^*}{\overline{\mu}_{kj}} \tag{A.54}$$

By Theorem 5.7,

$$\alpha_k^*\beta_j^*(1+\underline{\pi})\xi^{-(|J_k-j|+2)} \leq \alpha_k\beta_j \leq \alpha_k^*\beta_j^*(1+\overline{\pi})\xi^{|J_k-j|+2} \tag{A.55}$$

so set

$$\overline{\mu}_{kj} = \alpha_k^*\beta_j^*(1+\overline{\pi})\xi^{|J_k-j|+2}, \underline{\mu}_{kj} = \alpha_k^*\beta_j^*(1+\underline{\pi})\xi^{-(|J_k-j|+2)} \tag{A.56}$$

to convert (A.55) to the condition $\underline{\mu}_{kj} < \mu_{kj} < \overline{\mu}_{kj}$ required in Theorem 5.1.

Moreover, (A.56) reduces (A.54) to the following:

$$(1+\underline{\pi})\frac{\alpha_k^*\beta_j^*}{\overline{\mu}_{kj}} \leq \frac{Y_{kj}}{\overline{\mu}_{kj}} \leq (1+\overline{\pi})\frac{\alpha_k^*\beta_j^*}{\overline{\mu}_{kj}} \tag{A.57}$$

and substitution of (A.56) into this result yields

$$\xi^{-(|J_k-j|+2)} \leq \frac{Y_{kj}}{\overline{\mu}_{kj}} \leq \xi^{-(|J_k-j|+1)}$$

and, since $1 \leq J_k, j \leq J$,

$$\xi^{-(J+1)} \leq \frac{Y_{kj}}{\overline{\mu}_{kj}} \leq \xi^{-J} \tag{A.58}$$

If condition (5.13) holds, then combination of it with (A.58) yields condition (5.11) .

Now under the conditions of Corollary 5.8, those of Corollary 5.4 also hold, in which case (5.11) is a sufficient condition for a unique MLE of the model parameters. It then follows that, in the present case, (5.13) is a sufficient condition, and Corollary 5.8 is proved. ■

Proof of Theorem 6.2

The following lemma corresponds to Lemma A.1 for the EDF cross-classified model, and the proof follows exactly the same logic.

Lemma A.8. For the Bayesian EDF cross-classified model defined in Section 6.1, any MLE of the parameter set $\hat{\alpha}_k, \hat{\beta}_j$ must have the property

$$\rho^{-2}(\mathfrak{D}) \leq \hat{\alpha}_{k_1}/\hat{\alpha}_{k_2}, \hat{\beta}_{j_1}/\hat{\beta}_{j_2} \leq \rho^2(\mathfrak{D}), k_1, k_2 = 1, \dots, K; j_1, j_2 = 1, \dots, J \qquad \text{(A.59)}$$

■

The next lemma is then the strict Bayesian parallel of Lemma A.3.

Lemma A.9. Suppose the Bayesian EDF cross-classified model defined in Section 6.1 is subject to the tail convergence property. Then any MAP estimator of the parameter set α_k, β_j must lie within a co-ordinate rectangle of the form $\left\{0 < \underline{\alpha}_k \leq \alpha_k \leq \overline{\alpha}_k, 0 < \underline{\beta_j} \leq \beta_j \leq \overline{\beta}_j; k = 1, \dots, K, j = 1, \dots, J\right\}$ for constants $\underline{\alpha}_k, \overline{\alpha}_k, \underline{\beta_j}, \overline{\beta}_j$.

Remark A.10. The bounds $\underline{\alpha}_k, \overline{\alpha}_k, \underline{\beta_j}, \overline{\beta}_j$ are not the same as those in Lemma A.3.

Proof. By (6.8) and (6.10) and the positivity of c' (see (4.6)),

$$0 < z_k^{(\alpha)} < 1 \qquad \text{(A.60)}$$

whence (6.7) yields

$$\left[1 - z_k^{(\alpha)}\right] a_k < \hat{\alpha}_k \leq max\left(a_k, \bar{Y}_k^{(\alpha)}\right) \qquad \text{(A.61)}$$

Since $a_k > 0$ is fixed, $\hat{\alpha}_k$ will be bounded away from zero provided that $z_k^{(\alpha)}$ is bounded away from unity. So consider $z_k^{(\alpha)}$, given by (6.8) and (6.10). These relations give

$$\begin{aligned} z_k^{(\alpha)} &= \left[1 + \left(\sum_{j \in \mathcal{R}(k)} \frac{\phi_{kj}}{\psi_k^{(\alpha)}} \frac{\hat{\beta}_j^2 c'(\hat{\mu}_{kj})}{c'(\hat{\alpha}_k)}\right)^{-1}\right]^{-1} \\ &= \left[1 + \left(\sum_{j \in \mathcal{R}(k)} \frac{\phi_{kj}}{\psi_k^{(\alpha)}} \zeta(\hat{\alpha}_k, \hat{\beta}_j)\right)^{-1}\right]^{-1} \\ &\leq \left[1 + \left[\left(\sum_{j \in \mathcal{R}(k)} \frac{\phi_{kj}}{\psi_k^{(\alpha)}}\right)\left(\max_{j \in \mathcal{R}(k)} \zeta(\hat{\alpha}_k, \hat{\beta}_j)\right)\right]^{-1}\right]^{-1} \end{aligned} \qquad \text{(A.62)}$$

$$\frac{z_{kj}^{(\alpha)}}{\sum_{j \in \mathcal{R}(k)} z_{kj}^{(\alpha)}} = \left[\sum_{i \in \mathcal{R}(k)} \frac{\phi_{kj}}{\phi_{ki}} \frac{\hat{\beta}_i^2}{\hat{\beta}_j^2} \frac{c'(\hat{\alpha}_k \hat{\beta}_i)}{c'(\hat{\alpha}_k \hat{\beta}_j)}\right]^{-1} \qquad \text{(A.63)}$$

By Lemma A.8,

$$\rho^{-4}(\mathfrak{D}) \leq \frac{\hat{\beta}_i^2}{\hat{\beta}_j^2} \leq \rho^4(\mathfrak{D})$$

It then follows from (TC1) that

$$\frac{c'(\hat{\alpha}_k\hat{\beta}_i)}{c'(\hat{\alpha}_k\hat{\beta}_j)} = \frac{c'\left((\hat{\beta}_i/\hat{\beta}_j)\hat{\alpha}_k\hat{\beta}_j\right)}{c'(\hat{\alpha}_k\hat{\beta}_j)}$$

is bounded above and below, with strictly positive lower bound. Then, by (A.63), there exists a lower bound z:

$$0 < z \leq \frac{z_{kj}^{(\alpha)}}{\sum_{j\in\mathcal{R}(k)} z_{kj}^{(\alpha)}} \tag{A.64}$$

Similarly

$$0 < z \leq \frac{z_{kj}^{(\beta)}}{\sum_{k\in\mathcal{C}(j)} z_{kj}^{(\beta)}} \tag{A.65}$$

It follows from (A.62) that $z_k^{(\alpha)}$ is bounded away from unity unless $\zeta(\hat{\alpha}_k, \hat{\beta}_j)$ is without finite upper bound for some j.

There are three possibilities concerning the boundedness of $\zeta(\hat{\alpha}_k, \hat{\beta}_j)$, taking account of the conditions (TC2)-(TC4) imposed on the function ζ in the statement of the lemma:

(a) $\zeta(\hat{\alpha}_k, \hat{\beta}_j)$ is bounded away from infinity for all $\hat{\alpha}_k, \hat{\beta}_j$;
(b) $\zeta(\hat{\alpha}_k, \hat{\beta}_j) \to \infty$ for some $\hat{\alpha}_k, \hat{\beta}_j$ as $\hat{\beta}_j \to 0$;
(c) $\zeta(\hat{\alpha}_k, \hat{\beta}_j) \to \infty$ for some $\hat{\alpha}_k, \hat{\beta}_j$ as $\hat{\beta}_j \to \infty$.

Case (a). This case has been dealt with essentially in the brief argument just given, but the following adds a little detail.

It follows from (A.62) that $z_k^{(\alpha)} < z < 1$ for some z. Then, by (6.7), $\hat{\alpha}_k > [1-z]\alpha_k$, and one may set $\underline{\alpha}_k = [1-z]\alpha_k$.

Equations (6.12) and (6.14) yield the following result in parallel with (A.62)

$$z_j^{(\beta)} \geq \left[1 + \left[\left(\sum_{k\in\mathcal{C}(\mathrm{j})} \frac{\phi_{kj}}{\psi_j^{(\beta)}}\right)\left(\min_{k\in\mathcal{C}(\mathrm{j})} \zeta(\hat{\beta}_j, \hat{\alpha}_k)\right)\right]^{-1}\right]^{-1} \tag{A.66}$$

and an argument parallel to that above leading to the lower bound on $\hat{\alpha}_k$ establishes a lower bound $\underline{\beta}_j$ on $\hat{\beta}_j$.

If the $\hat{\beta}_j$ are bounded away from zero, then the quantities $Y_{kj}/\hat{\beta}_j$ are bounded away from infinity, and $\bar{Y}_k^{(\alpha)}$, being a weighted average of these quantities (see (6.9)), is also bounded away from infinity and, by (6.7), so is $\hat{\alpha}_k$. That is, there exists $\overline{\alpha}_k$ such that $\hat{\alpha}_k \leq \overline{\alpha}_k < \infty$.

A similar upper bound $\overline{\beta}_j$ for $\hat{\beta}_j$ may be found by a similar argument.

Case (b). Suppose that no lower bound $\underline{\beta_j}$ exists. Then it is possible to find a sequence of points $\left\{\left(\hat{\alpha}_1^{(n)},\dots,\hat{\alpha}_K^{(n)},\hat{\beta}_1^{(n)},\dots,\hat{\beta}_J^{(n)}\right);n=1,2,\dots\right\}$ in the admissible space of MAP solutions with $\hat{\beta}_j^{(n)}\to 0$. It will be convenient to write, with a slight abuse of notation, just $\hat{\beta}_j\to 0$. By the condition of case (b), $\zeta\left(\hat{\alpha}_k,\hat{\beta}_j\right)\to\infty$ as $\hat{\beta}_j\to 0$, and then (TC3) implies that $\zeta\left(\hat{\alpha}_k,\hat{\beta}_j\right)$ is bounded as $\hat{\beta}_j\to\infty$, equivalently $\zeta\left(\hat{\beta}_j,\hat{\alpha}_k\right)$ is bounded as $\hat{\alpha}_k\to\infty$.

These results lead to the following sequence of consequences as $\hat{\beta}_j\to 0$:

$z_k^{(\alpha)}\to 1$	[by (A.62)]	(A.67)
$Y_{kj}/\hat{\beta}_j\to\infty$		(A.68)
$\bar{Y}_k^{(\alpha)}\to\infty$	[by (6.8), (A.64) and (A.68)]	(A.69)
$\hat{\alpha}_k\to\infty$	[by (6.7), (A.67) and (A.69)]	(A.70)
$Y_{kj}/\hat{\alpha}_k\to 0$	[by (A.70)]	(A.71)
$\bar{Y}_j^{(\beta)}\to 0$	[by (6.13), (A.65) and (A.71)]	(A.72)
$\zeta\left(\hat{\beta}_j,\hat{\alpha}_k\right)$ bounded	[previous paragraph]	(A.73)
$z_k^{(\beta)}$bounded away from 1	[by (A.66) and (A.73)]	(A.74)
$\hat{\beta}_j$ bounded away from 0	[by (6.11) and (A.74)]	(A.75)

But then (A.75) contradicts the hypothesis of case (b), and it must be concluded that this is not a possible case for $z_k^{(\alpha)}$ to be bounded away from unity.

Case (c). Suppose that no upper bound $\overline{\beta}_j$ exists. Then it is possible to find a sequence of points $\left\{\left(\hat{\alpha}_1^{(n)},\dots,\hat{\alpha}_K^{(n)},\hat{\beta}_1^{(n)},\dots,\hat{\beta}_J^{(n)}\right);n=1,2,\dots\right\}$ in the admissible space of MAP solutions with $\hat{\beta}_j^{(n)}\to\infty$. It will be convenient to write, with the same abuse of notation as in (b), just $\hat{\beta}_j\to\infty$. By the condition of case (c), $\zeta\left(\hat{\alpha}_k,\hat{\beta}_j\right)\to\infty$ as $\hat{\beta}_j\to\infty$.

These results lead to the following sequence of consequences as $\hat{\beta}_j\to\infty$, where $\zeta\left(\hat{\alpha}_k,\hat{\beta}_j\right)$ is temporarily abbreviated to just ζ:

$1-z_k^{(\alpha)}=O(\zeta^{-1})$	[by (A.62)]	(A.76)
$Y_{kj}/\hat{\beta}_j=O(\zeta)$ for all j	[by Lemma A.8]	(A.77)
$\bar{Y}_k^{(\alpha)}=O(\zeta^{-1})$	[by (6.8), (A.64) and (A.77)]	(A.78)
$\hat{\alpha}_k=O(\zeta^{-1})$	[by (6.7), (A.67) and (A.78)]	(A.79)
$Y_{kj}/\hat{\alpha}_k=O(\zeta)$for all k	[by (A.79) and Lemma A.8]	(A.80)
$\bar{Y}_j^{(\beta)}=O(\zeta)$	[by (6.13), (A.65) and (A.80)]	(A.81)
$\zeta\left(\hat{\beta}_j,\hat{\alpha}_k\right)=\zeta\left(\hat{\beta}_j,O(\zeta^{-1})\right)\to 0$	[by (TC4)]	(A.82)

$$z_k^{(\beta)} = O\left(\zeta\left(\hat{\beta}_j, \hat{\alpha}_k\right)\right) = O\left(\zeta\left(\hat{\beta}_j, C/\zeta\left(\hat{\alpha}_k, \hat{\beta}_j\right)\right)\right) \text{ for some } C > 0 \quad \text{[by (A.66) and (A.82)]} \tag{A.83}$$

By (6.11),

$$\hat{\beta}_j = z_j^{(\beta)}\left[\bar{Y}_j^{(\beta)} - b_j\right] + b_j < z_j^{(\beta)}\bar{Y}_j^{(\beta)} \text{ as } \hat{\beta}_j \to \infty \tag{A.84}$$

since $\bar{Y}_j^{(\beta)} \to \infty$, by (A.81).

Now consider the sequence $\left\{\hat{\beta}_j^{(n)}\right\}$ explicitly, and also the sequences $\left\{z_j^{(\beta)(n)}\right\}, \left\{\bar{Y}_j^{(\beta)(n)}\right\}$ generated by it, as in (A.83) and (A.81) respectively. Now

$$\bar{Y}_j^{(\beta)(n)} < M_1\, \zeta\left(\hat{\alpha}_k, \hat{\beta}_j\right) \text{ for some } M_1 > 0 \quad \text{[by (A.81)]}$$
$$z_k^{(\beta)(n)} < M_2 \zeta\left(\hat{\beta}_j, C/\zeta\left(\hat{\alpha}_k, \hat{\beta}_j\right)\right) \text{ for some } M_2 > 0 \quad \text{[by (A.83)]}$$

Combination of these results with (A.84) yields

$$\hat{\beta}_j < M_1 M_2\, \zeta\left(\hat{\alpha}_k, \hat{\beta}_j\right) \times \zeta\left(\hat{\beta}_j, C/\zeta\left(\hat{\alpha}_k, \hat{\beta}_j\right)\right) \text{ as } \hat{\beta}_j \to \infty$$

equivalently

$$1 < M_1 M_2 \hat{\beta}_j^{-1}\, \zeta\left(\hat{\alpha}_k, \hat{\beta}_j\right) \times \zeta\left(\hat{\beta}_j, C/\zeta\left(\hat{\alpha}_k, \hat{\beta}_j\right)\right) \text{ as } \hat{\beta}_j \to \infty$$

But, by (TC4), the right side converges to zero. Thus, the hypothesis of case (c) leads to a contradiction, and it must be concluded that this is not a possible case for $z_k^{(\alpha)}$ to be bounded away from unity.

A consideration of all three cases (a)-(c) arrives at the conclusion that $\zeta\left(\hat{\alpha}_k, \hat{\beta}_j\right)$ must be bounded above for all j. By the argument given in case (a), this leads to the existence of the bounds $\underline{\alpha}_k, \overline{\alpha}_k, \underline{\beta}_j, \overline{\beta}_j$. ■

References

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

England P D, Verrall R J & Wüthrich M V (2012). Bayesian over-dispersed Poisson model and the Bornhuetter & Ferguson claims reserving method. **Annals of Actuarial Science**, 6(2), 258-281.

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

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

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

Kuang D, Nielsen B & Nielsen J P (2008). Identification of the age-period-cohort model and the extended chain ladder model. **Biometrika**, 95, 979-986.

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

McCullagh P & Nelder J A (1989). **Generalized linear models**, 2nd ed. Chapman & Hall, London, UK.

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

Merz M, Wüthrich M V & Hashorva E (2013). Dependence modelling in multivariate claims run-off triangles. **Annals of Actuarial Science**, 7(1), 3-25.

Renshaw AE & Verrall RJ (1998). A stochastic model underlying the chain-ladder technique. **British Actuarial Journal**, 4(iv), 903-923.

Shi P, Basu S & Meyers P P (2012). A Bayesian log-normal model for multivariate loss reserving. **North American Actuarial Journal**, 16(1), 29-51.

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

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

Taylor G (2011). Maximum likelihood and estimation efficiency of the chain ladder. **Astin Bulletin**, 41(1), 131-155.

Taylor G (2015). Bayesian chain ladder models. **Astin Bulletin**, 45(1) (2015), 75-99.

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 (2000). An investigation into stochastic claims reserving models and the chain-ladder technique. **Insurance: mathematics and economics**, 26(1), 91-99.

Verrall R J (2004). A Bayesian generalised linear model for the Bornhuetter-Ferguson method of claims reserving. **North American Actuarial Journal**, 8(3), 67-89.

Wüthrich M V (2007). Using a Bayesian approach for claims reserving. **Variance**, 1(2), 292-301.

Wüthrich M V (2012). Discussion of “A Bayesian log-normal model for multivariate loss reserving” by Shi-Basu-Meyers. **North American Actuarial Journal**, 16(3), 398-401.

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