EXISTENCE AND UNIQUENESS OF CHAIN LADDER SOLUTIONS

BY

GREG TAYLOR

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, including Bayesian chain ladder models, lead to implicit MAP (Bayesian) or MLE (non-Bayesian) solutions for 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.

KEYWORDS

Bayesian chain ladder, cross-classified chain ladder, EDF chain ladder, existence, loss reserving, MAP estimator, 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 and 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 and Verrall (2002) subjected the Poisson distribution to over-dispersion. Wüthrich (2003) discussed the more general case of a distribution from the Tweedie family, and Wüthrich and Merz (2008) the even more general 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

Astin Bulletin 47(1), 1-41. doi: 10.1017/asb.2016.23

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 the 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 *et al*. (2009) used it to prove uniqueness of the MLE, and Taylor (2011) to prove minimum variance properties.

However, as will be shown in Section 3.1, 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.

Similarly, the MAP estimators of various Bayesian chain ladder models are seen to be solutions to implicit equations (Taylor, 2015), and parallel questions of existence and uniqueness arise there. So, in general, the "chain ladder solutions" mentioned in the title of the present paper refer to the solutions of MAP estimation or ML equations in the cases of Bayesian or non-Bayesian chain ladder models, respectively.

In the following, after a brief consideration of the mathematical setup of the Bayesian and non-Bayesian forms of the EDF cross-classified chain ladder model in Section 2, the existence (Section 3) and uniqueness (Section 4) of MAP estimators in the Bayesian model and MLEs in the non-Bayesian are considered. Section 5 examines a numerical example in which multiple solutions of the ML equations are found for the non-Bayesian model.

2. Chain ladder framework and notation

2.1. Data framework and notation

Consider an array $\mathfrak{D}$ of claim observations (random variables) $Y_{kj} > 0$ 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 3. Those conditions are anticipated slightly here by noting that they will require independence between all

https://doi.org/10.1017/asb.2016.23 Published online by Cambridge University Press

observations. This would usually limit them to incremental, rather than cumulative, observations.

The dimensions K and J are the 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

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

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 *et al.* (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 **generalised 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 $\sum_{j \in \mathcal{R}(k)}$ denote summation over the entire row $\mathcal{R}(k)$, and similarly $\sum_{k \in \mathcal{C}(j)}$ denote summation over the column $\mathcal{C}(j)$.

2.2. Non-Bayesian chain ladder model

Consider the model defined by the following conditions:

E1. The array $\mathfrak{D}$ is regular.
E2. The random variables $Y_{kj} \in \mathfrak{D}$ are stochastically independent.
E3. 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:

$$\ell(y|\theta_{kj}, \phi_{kj}) = \left[y\theta_{kj} - \kappa\left(\theta_{kj}\right)\right]/a\left(\phi_{kj}\right) + \lambda\left(y, \phi_{kj}\right), \quad (2.1)$$

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 $\kappa\prime$ (.) maps one-one onto the strictly positive half-line, and that $\kappa\prime\prime > 0$.

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

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

Henceforth, the function $a(.)$ in (2.1) will be restricted to the case

$$a\left(\phi\right) = \phi. \quad (2.2)$$

Condition (E3)(c) 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**, as in Taylor (2011). It consists of cross-classified multiplicative mean structure, as in (E3)(b), supplemented by an EDF distribution in (E3)(a).

It will be convenient to express the log-likelihood (2.1) in a different representation, as follows.

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

$$\mu_{kj} = \kappa\,'\left(\theta_{kj}\right), \quad (2.3)$$

whence the following expressions:

$$\theta_{kj} = c\left(\mu_{kj}\right), \quad (2.4)$$

$$\kappa\left(\theta_{kj}\right) = d\left(\mu_{kj}\right), \quad (2.5)$$

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

Then (2.1) may be re-written, taking account of (2.2), in the form

$$\ell(y|\mu_{kj}, \phi_{kj}) = \left[yc\left(\mu_{kj}\right) - d\left(\mu_{kj}\right)\right]/\phi_{kj} + \lambda\left(y, \phi_{kj}\right), \quad (2.6)$$

or

$$\ell(y|\alpha_k, \beta_j, \phi_{kj}) = \left[yc\left(\alpha_k\beta_j\right) - d\left(\alpha_k\beta_j\right)\right]/\phi_{kj} + \lambda\left(y, \phi_{kj}\right) \quad (2.7)$$

Remark 2.1. Mack models

In addition to the EDF cross-classified model just introduced, the literature identifies a different type of chain ladder, viz. the **Mack model**. This comes in two variants: the **distribution-free**, or **non-parametric**, form (Mack, 1993), and the **EDF Mack model** (Taylor, 2011). 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.

Remark 2.2. Tweedie family
The **Tweedie family** is the sub-family of the EDF for which (Tweedie, 1984)

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

or equivalently

$$c(\mu) = \mu^{1-p}/(1-p), \tag{2.9}$$

$$d(\mu) = \mu^{2-p}/(2-p), \tag{2.10}$$

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

In the cases $p = 1, 2$, (2.9) and (2.10) must be replaced by their limiting values as p approaches the relevant value:

$$\lim_{q\to 0} \mu^q/q = ln\ \mu. \tag{2.11}$$

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.

Remark 2.3. Poisson family
The **Poisson family** is the sub-family of the Tweedie family for which $p = 1$ so that, by (2.9)–(2.11),

$$c(\mu) = ln\ \mu, \tag{2.12}$$

$$d(\mu) = \mu. \tag{2.13}$$

Substitution in (2.6) yields

$$\ell(y|\mu_{kj}, \phi_{kj}) = \left[y\ ln\ \mu_{kj} - \mu_{kj}\right]/\phi_{kj}\ + \lambda\left(y, \phi_{kj}\right) \tag{2.14}$$

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

$$\ell(y|\mu_{kj}, \phi_{kj}) = ln\ \left[e^{-\mu_{kj}} {\mu_{kj}}^{y}\ exp\,\lambda\,(y, 1)\right]. \tag{2.15}$$

The normalizing function can be recognised to be $\lambda(y, 1) = -\,ln\ y!$, in which case (2.15) is seen to be the **Poisson** log-likelihood. For the case $\phi_{kj} \neq 1$, (2.14) 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**.

Remark 2.4. Sufficient statistics
When (E3)(b) is recognised, (2.1) becomes

$$\ell(y|\mu_{kj}, \phi_{kj}) = \left[y\ \left(ln\ \alpha_k + ln\ \beta_j\right) - \alpha_k\beta_j\right]/\phi_{kj}\ + \lambda\left(y, \phi_{kj}\right), \tag{2.16}$$

https://doi.org/10.1017/asb.2016.23 Published online by Cambridge University Press

and summation over $\mathfrak{D}$ yields

$$\ell = \left[\sum_{k=1}^{K} ln\ \alpha_k \sum_{j \in \mathcal{R}(k)} \frac{y_{kj}}{\phi_{kj}} + \sum_{j=1}^{J} ln\ \beta_j \sum_{k \in \mathcal{C}(j)} \frac{y_{kj}}{\phi_{kj}} - \sum_{(k,j) \in \mathfrak{D}} \frac{\alpha_k \beta_j}{\phi_{kj}} \right] + \sum_{(k,j) \in \mathfrak{D}} \lambda \left(y, \phi_{kj} \right), \tag{2.17}$$

where ℓ without arguments denotes the log-likelihood for the entire array $\mathfrak{D}$.

Application of the Fisher–Neyman theorem to this likelihood proves that $\sum_{j \in \mathcal{R}(k)} y_{kj}/\phi_{kj}$ is a sufficient statistic for α_k and $\sum_{k \in \mathcal{C}(j)} y_{kj}/\phi_{kj}$ 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, (2.16) and (2.17) would be replaced by the following [by (2.6), (2.9) and (2.10)]:

$$\ell(y|\mu_{kj}, \phi_{kj}) = \left[y \frac{\left(\alpha_k \beta_j\right)^{1-p}}{1-p} - \frac{\left(\alpha_k \beta_j\right)^{2-p}}{2-p} \right] / \phi_{kj} + \lambda \left(y, \phi_{kj} \right), \tag{2.18}$$

$$\ell = \left[\sum_{k=1}^{K} \alpha_k{}^{1-p} \sum_{j \in \mathcal{R}(k)} \frac{y_{kj} \beta_j{}^{1-p}}{(1-p)\,\phi_{kj}} - \sum_{(k,j) \in \mathfrak{D}} \frac{\left(\alpha_k \beta_j\right)^{2-p}}{(2-p)\,\phi_{kj}} \right] + \sum_{(k,j) \in \mathfrak{D}} \lambda \left(y, \phi_{kj} \right). \tag{2.19}$$

The linear combination of observations that was previously a Fisher–Neyman 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 $\mathfrak{D}$.

3. Existence of chain ladder solutions

3.1. Bayesian chain ladder model

Bayesian versions of EDF cross-classified model or special cases of it have been studied in the literature (Verrall, 2000, England and Verrall, 2002; Verrall, 2004; Gisler and Müller, 2007; Wüthrich, 2007, Gisler and Wüthrich, 2008; Wüthrich and Merz (2008); England *et al.*, 2012; Shi *et al.*, 2012; Wüthrich, 2012; Merz *et al.*, 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 results parallel to those in Section 4 but subject to different priors.

Taylor uses (2.1) as the conditional log-likelihood of $Y_{kj}|\alpha_k, \beta_j$, and the prior log-densities on $c(\alpha_k)$, $c(\beta_j)$ that appear in the model immediately below.

With the addition of those priors, the EDF cross-classified model of Section 2.2 becomes the **Bayesian EDF cross-classified model** described in full as follows.

https://doi.org/10.1017/asb.2016.23 Published online by Cambridge University Press

B1. The array $\mathfrak{D}$ is regular.

B2. $\alpha_k, k = 1, \ldots, K$; $\beta_j, j = 1, \ldots, J$ are stochastically independent, non-negative random parameters, subject to the prior log-densities on $c(\alpha_k)$, $c(\beta_j)$ (omitting terms that do not depend on α_k, β_j)

$$\ell_k^{(\alpha)\text{prior}}\left(c\left(\alpha_k\right)\right) = \left[c\left(\alpha_k\right) A_k - d\left(\alpha_k\right)\right] / \psi_k^{(\alpha)}, \tag{2.20}$$

$$\ell_j^{(\beta)\text{prior}}\left(c\left(\beta_j\right)\right) = \left[c\left(\beta_j\right) B_j - d\left(\beta_j\right)\right] / \psi_j^{(\beta)}, \tag{2.21}$$

where $c(.)$, $d(.)$ are as defined in (2.4) and (2.5), A_k, $\psi_k^{(\alpha)} > 0$ are location and dispersion parameters, and so are B_j, $\psi_j^{(\beta)} > 0$.

B3. The random variables $Y_{kj} \in \mathfrak{D}$ are stochastically independent, conditionally, of the parameter set $\{\alpha_k, k = 1, \ldots, K;\ \beta_j, j = 1, \ldots, J\}$.

B4. For each $k = 1, 2, \ldots, K$ and $j = 1, 2, \ldots, J$,

a. Y_{kj} is distributed conditionally on the parameters $\alpha_k,\ \beta_j$ according to a member of the EDF, specifically with log-likelihood of $Y_{kj} = y$ as follows:

$$\ell^{cond}(y|\alpha_k,\ \beta_j, \phi_{kj}) = \left[y\theta_{kj} - \kappa\left(\theta_{kj}\right)\right] / \phi_{kj} + \lambda\left(y, \phi_{kj}\right), \tag{2.22}$$

for parameters $\phi_{kj} > 0$ and θ_{kj}, defined in terms of $\alpha_k,\ \beta_j$ by (2.3) 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 $\kappa\prime\,(.)$ maps one-one onto the strictly positive half-line, and that $\kappa\prime\prime > 0$.

b. $E[Y_{kj}|\alpha_k,\ \beta_j] = \alpha_k\beta_j$.

The Bayesian EDF cross-classified model is the earlier non-Bayesian model with its parameters $\alpha_k,\ \beta_j$ randomised according to priors that are conjugate to the conditional log-likelihood (2.22). Note that condition (E3)(c) of the non-Bayesian model is no longer required, as the imposition of a prior on the parameter set $\{\alpha_k,\ \beta_j\}$ means that these parameters can no longer be re-scaled at will.

Remark 2.5. The non-Bayesian model of Section 2.2 may be recovered by the selection of uninformative priors, i.e., by allowing all $\psi_k^{(\alpha)}, \psi_j^{(\beta)} \to \infty$.

It will be notationally convenient, henceforth, to represent the conditional log-likelihood ℓ^{cond} in the (k, j) cell by ℓ_{kj}^{cond}.

The **posterior log-likelihood** of the α_k, β_j (again omitting terms that do not depend on the α_k, β_j) is

$$\ell^{\text{post}}\left(c\left(\alpha\right), c\left(\beta\right)|\mathfrak{D};\ A, B, \phi, \psi^{(\alpha)}, \psi^{(\beta)}\right) = \ell^{\text{cond}} + \ell^{\text{prior}}, \tag{2.23}$$

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

$$\ell^{\text{cond}} = \sum_{(k,j)\in\mathfrak{D}} \ell_{kj}^{\text{cond}}, \tag{2.24}$$

$$\ell^{\text{prior}} = \sum_{k=1}^{K} \ell_k^{(\alpha)\text{prior}} + \sum_{j=1}^{J} \ell_j^{(\beta)\text{prior}}. \tag{2.25}$$

Thus, by (2.20)–(2.25), the posterior log-likelihood for the whole array $\mathfrak{D}$ takes the form

$$\ell^{\text{post}} = \sum_{(k,j)\in\mathfrak{D}} \left\{\left[y_{kj} c\left(\mu_{kj}\right) - d\left(\mu_{kj}\right)\right]/\phi_{kj} + \lambda\left(y, \phi_{kj}\right)\right\}$$
$$+ \sum_{k=1}^{K} \left[c\left(\alpha_k\right) A_k - d\left(\alpha_k\right)\right]/\psi_k^{(\alpha)} + \sum_{j=1}^{J} \left[c\left(\beta_j\right) B_j - d\left(\beta_j\right)\right]/\psi_j^{(\beta)}, \tag{2.26}$$

where (2.22) has been replaced by the more convenient form (2.6).

3.2. Mathematical preliminaries

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.

Sections 3 and 4 will examine existence and uniqueness of maximum *a posteriori* **(MAP)** estimates of the parameter set $\{\alpha_k, \ \beta_j\}$ in the case of the Bayesian EDF cross-classified model. Existence and uniqueness of MLE estimates of the same parameter set in the case of the non-Bayesian model will also be examined.

It is known (Hachemeister and Stanard, 1975; Renshaw and Verrall, 1998; Taylor, 2000) that explicit MLEs exist for the parameters of the non-Bayesian Poisson cross-classified model in the case where ϕ_{kj} is independent of k, j. Existence is, therefore, obvious.

However, now consider the general Bayesian EDF cross-classified model of Section 3.1 with joint posterior log-likelihood function over $\mathfrak{D}$ given by (2.23)–(2.25). The conditions for MAP estimation of α, β are

$$\frac{\partial\left(\ell^{\text{cond}} + \ell^{\text{prior}}\right)}{\partial\alpha_k} = \frac{\partial\left(\ell^{\text{cond}} + \ell^{\text{prior}}\right)}{\partial\beta_j} = 0, k = 1, \ldots, K; j = 1, \ldots, J. \tag{3.1}$$

Section 4.2 of Taylor (2015) derives the following solution $\{\hat{\alpha}_k, \hat{\beta}_j\}$ of the system (3.1):

$$\hat{\alpha}_k = z_k^{(\alpha)} \bar{Y}_k^{(\alpha)} + \left[1 - z_k^{(\alpha)}\right] A_k, \tag{3.2}$$

https://doi.org/10.1017/asb.2016.23 Published online by Cambridge University Press

with

$$z_k^{(\alpha)} = \sum_{j \in \mathcal{R}(k)} z_{kj}^{(\alpha)}, \tag{3.3}$$

$$\bar{Y}_k^{(\alpha)} = \sum_{j \in \mathcal{R}(k)} z_{kj}^{(\alpha)} \left[Y_{kj} / \hat{\beta}_j \right] / \sum_{j \in \mathcal{R}(k)} z_{kj}^{(\alpha)}, \tag{3.4}$$

$$z_{kj}^{(\alpha)} = \frac{\hat{\beta}_j^2 c' \left(\hat{\mu}_{kj} \right)}{\phi_{kj}} / \left[\sum_{j \in \mathcal{R}(k)} \frac{\hat{\beta}_j^2 c' \left(\hat{\mu}_{kj} \right)}{\phi_{kj}} + \frac{c' \left(\hat{\alpha}_k \right)}{\psi_k^{(\alpha)}} \right], \tag{3.5}$$

$$\hat{\beta}_j = z_j^{(\beta)} \bar{Y}_j^{(\beta)} + \left[1 - z_j^{(\beta)} \right] B_{j,} \tag{3.6}$$

with

$$z_j^{(\beta)} = \sum_{k \in \mathcal{C}(j)} z_{kj}^{(\beta)}, \tag{3.7}$$

$$\bar{Y}_j^{(\beta)} = \sum_{k \in \mathcal{C}(j)} z_{kj}^{(\beta)} \left[Y_{kj} / \hat{\alpha}_k \right] / \sum_{k \in \mathcal{C}(j)} z_{kj}^{(\beta)}, \tag{3.8}$$

$$z_{kj}^{(\beta)} = \frac{\hat{\alpha}_k^2 c' \left(\hat{\mu}_{kj} \right)}{\varphi_{kj}} / \left[\sum_{k \in \mathcal{C}(\mathrm{j})} \frac{\hat{\alpha}_k^2 c' \left(\hat{\mu}_{kj} \right)}{\varphi_{kj}} + \frac{c' \left(\hat{\beta}_j \right)}{\psi_j^{(\beta)}} \right]. \tag{3.9}$$

It may be noted immediately that the solution (3.2)–(3.9) is implicit because the solution $\hat{\alpha}_k$ requires knowledge of $\hat{\beta}_j$ and vice versa.

As noted in Remark 2.5, the non-Bayesian model of Section 2.2 may be recovered by allowing all $\psi_k^{(\alpha)}, \psi_j^{(\beta)} \to \infty$. This solution is also implicit.

Now expand to the MAP equations (3.1). Substitute (2.7) into (2.24), and differentiate to obtain

$$\partial \ell^{\text{cond}} / \partial \alpha_k = \sum_{j \in \mathcal{R}(k)} \left[Y_{kj} c' \left(\mu_{kj} \right) - d' \left(\mu_{kj} \right) \right] \beta_j / \phi_{kj}, \ k = 1, \ldots, K \tag{3.10}$$

$$\partial \ell^{\text{cond}} / \partial \beta_j = \sum_{k \in \mathcal{C}(j)} \left[Y_{kj} c' \left(\mu_{kj} \right) - d' \left(\mu_{kj} \right) \right] \alpha_k / \phi_{kj}, \ j = 1, \ldots, J \tag{3.11}$$

Also, substitute (2.20) and (2.21) into (2.25), and differentiate to obtain

$$\partial \ell^{\text{prior}} / \partial \alpha_k = \left[c' \left(\alpha_k \right) A_k - d' \left(\alpha_k \right) \right] / \psi_k^{(\alpha)}, \ k = 1, \ldots, K \tag{3.12}$$

$$\partial \ell^{\text{prior}} / \partial \beta_j = \left[c' \left(\beta_j \right) B_j - d' \left(\beta_j \right) \right] / \psi_j^{(\beta)}, \ j = 1, \ldots, J \tag{3.13}$$

These equations can be simplified slightly if it is noted that, by the definition of $c(\mu)$ and $d(\mu)$ in Section 2.2,

$$d' \left(\mu \right) = \kappa' \left(c \left(\mu \right) \right) \ c' \left(\mu \right) = \mu \ c' \left(\mu \right). \tag{3.14}$$

If this result is substituted into (3.10)–(3.13), then substitution of the results into (3.1) yields the final MAP estimation equations:

$$\sum_{j\in\mathcal{R}(k)} \left[Y_{kj}-\mu_{kj}\right] c'\left(\mu_{kj}\right)\beta_j/\phi_{kj} + c'\left(\alpha_k\right)\left[A_k-\alpha_k\right]/\psi_k^{(\alpha)} = 0, k=1,\ldots,K, \tag{3.15}$$

$$\sum_{k\in\mathcal{C}(j)} \left[Y_{kj}-\mu_{kj}\right] c'\left(\mu_{kj}\right)\alpha_k/\phi_{kj} + c'\left(\beta_j\right)\left[B_j-\beta_j\right]/\psi_j^{(\beta)} = 0, j=1,\ldots,J. \tag{3.16}$$

Remark 3.1. The first member of each of these equations case can be recognised as a weighted sum of raw residuals $Y_{kj}-\mu_{kj}$. The second member can be recognised as the deviation of the estimate α_k from its prior location parameter A_k (or β_j from B_j).

It is noted for future reference that

$$c'\left(\mu\right) = 1/\kappa''^{\left(c'(\mu)\right)} > 0, \tag{3.17}$$

by (B4)(a), and then

$$d'\left(\mu\right) > 0, \tag{3.18}$$

by (3.14).

Equations (3.15) and (3.16) do not yield explicit solutions for the parameter set $\{\alpha_k, \beta_j\}$ in general. An example occurs in Taylor (2009), which examines the non-Bayesian Tweedie cross-classified model. Hence, questions of existence and uniqueness of solutions arise.

3.3. Existence of solutions

Most proofs of the existence of ML or MAP estimators rely on compactness of the parameter space. In the case of the Bayesian EDF cross-classified model, there is no explicit upper limit on any parameter. In the non-Bayesian case, there is no upper limit on the parameters α_k.

However, with some mild regularity conditions imposed, upper limits can be shown to exist (see Lemma A.1 for the Bayesian case and Lemma A.2 for the non-Bayesian). Specifically, it will be supposed that the function c satisfies the following conditions.

R1. The function $q(\alpha,\beta) = c'(\alpha\beta)/c'(\alpha)c'(\beta) = 1$ (Tweedie), or else has the following properties.
   a. It is bounded within a fixed strictly positive finite interval [m, M] over all $\alpha, \beta > 0$.
   b. q$(\alpha,\beta) \to 1$ as $\beta \to 0$ or ∞, with uniform convergence over α.

R2. Either
   a. $\mu c'(\mu)$ is bounded away from zero and infinity for all $\mu \in [0,\infty)$;
   or

https://doi.org/10.1017/asb.2016.23 Published online by Cambridge University Press

b. $\mu^2 c'(\mu)$ is bounded away from zero and infinity for all $\mu \in [0, \infty)$; or

c. $\mu c'(\mu)$ is a monotone strictly decreasing function of μ, with the **tail convergence properties** that, for any $\mu > 0$, $(t\mu)c'(t\mu)/\mu c'(\mu) \to \infty$ *as* $t \to 0$ and $(t\mu)c'(t\mu)/\mu c'(\mu) \to 0$ *as* $t \to \infty$ (note that this implies $\mu c'(\mu) \to \infty$ *as* $\mu \to 0$ and $\mu c'(\mu) \to 0$ *as* $\mu \to \infty$), and $\mu^2 c'(\mu)$ is a function of μ with one of the following properties.

i. It is strictly decreasing, with $\mu^2 c'(\mu) \to \infty$ *as* $\mu \to 0$ and $\mu^2 c'(\mu) \to 0$ *as* $\mu \to \infty$;

ii. It is strictly increasing, with $\mu^2 c'(\mu) \to 0$ *as* $\mu \to 0$ and $\mu^2 c'(\mu) \to \infty$ *as* $\mu \to \infty$, and the function $[\mu c'(\mu)]^2 c'(\mu c'(\mu))$ is not of order μ^{-1} as $\mu \to 0$.

Remark 3.2. By (2.9), any member of the Tweedie family with $p \geq 1$ satisfies conditions (R1)–(R2).

Theorem 3.3 assures the existence of an MAP estimate under these conditions. The proof appears in the appendix.

Theorem 3.3. *For the Bayesian EDF cross-classified model, subject to conditions (R1)–(R2), there exists an MAP estimate for the parameter set* $\{\alpha_k, \beta_j : k = 1, \ldots, K; j = 1, \ldots, J\}$*, solving the system of equations (3.15) and (3.16). This estimate lies within a closed* $(K + J)$ *-dimensional co-ordinate rectangle* R_+^{K+J}*, in the positive orthant, with all boundary planes bounded away from zero and infinity.*

Remark 3.4. For future reference, let the co-ordinate rectangle R_+^{K+J} be defined as

$$R_+^{K+J} = \left\{\alpha_k, \beta_j : k = 1, \ldots, K; j = 1, \ldots, J; \underline{\alpha}_k \leq \alpha_k \leq \bar{\alpha}_k, \underline{\beta}_j \leq \beta_j \leq \bar{\beta}_j\right\}, \tag{3.19}$$

for fixed $0 < \underline{\alpha}_k, \bar{\alpha}_k, \underline{\beta}_j, \bar{\beta}_j < \infty$.

Then define $\underline{\mu}_{kj} = \underline{\alpha}_k \underline{\beta}_j$ and $\bar{\mu}_{kj} = \bar{\alpha}_k \bar{\beta}_j$ so that $0 < \underline{\mu}_{kj} \leq \bar{\mu}_{kj} < \infty$.

Theorem 3.5 states the result corresponding to Theorem 3.3 for the non-Bayesian case. Again, the proof appears in the appendix.

Theorem 3.5. *For the non-Bayesian EDF cross-classified model, subject to conditions (R1) and (R2), there exists an MLE for the parameter set* $\{\alpha_k, \beta_j : k = 1, \ldots, K; j = 1, \ldots, J\}$*, solving the system of equations (3.15) and (3.16) with the terms involving* $\psi_k^{(\alpha)}$ *and* $\psi_j^{(\beta)}$ *deleted. This estimate lies within a closed* $(K + J)$ *-dimensional co-ordinate rectangle* R_+^{K+J}*, in the positive orthant, with all boundary planes bounded away from zero and infinity.*

4. Uniqueness of chain ladder solutions

4.1. Exponential dispersion family

Uniqueness of MAP estimators will be proved by establishment of convexity of the posterior log-likelihood function ℓ^{post} defined by (2.20)–(2.25). This will be done by reference to derivatives of that function for the Bayesian EDF cross-classified model of Section 3.1 with respect to $r_k = ln\ \alpha_k$, $s_j = ln\ \beta_j$.

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 the results of the present sub-section 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 (E3)(b).

The following lemma establishes conditions for convexity for the Bayesian EDF cross-classified model. The proof appears in the appendix.

Lemma 4.1. *For the Bayesian EDF cross-classified model, subject to conditions (R1) and (R2), define the rectangle R_+^{K+J} as in Remark 3.4. Then a sufficient condition for the log-likelihood ℓ to be strictly convex upward over R_+^{K+J} is that, for all $(\alpha_1, \ldots, \alpha_K, \beta_1, \ldots, \beta_J) \in R_+^{K+J}$,*

$$\frac{A_k}{\alpha_k} < 1 + \frac{d'(\alpha_k)}{\alpha_k d''(\alpha_k)} \text{ if } d''(\alpha_k) > 0,$$

or

$$\frac{A_k}{\alpha_k} > 1 + \frac{d'(\alpha_k)}{\alpha_k d''(\alpha_k)} \text{ if } d''(\alpha_k) < 0, \tag{4.1}$$

and

$$\frac{B_k}{\beta_j} < 1 + \frac{d'(\beta_j)}{\beta_j d''(\beta_j)} \text{ if } d''(\beta_j) > 0,$$

or

$$\frac{B_j}{\beta_j} > 1 + \frac{d'(\beta_j)}{\beta_j d''(\beta_j)} \text{ if } d''(\beta_j) < 0, \tag{4.2}$$

and

$$\frac{Y_{kj}}{\mu_{kj}} \leq 1 + \frac{d'(\mu_{kj})}{\mu_{kj} d''(\mu_{kj})} \text{ whenever } d''(\mu_{kj}) > 0,$$

or

$$\frac{Y_{kj}}{\mu_{kj}} \geq 1 - \frac{d'(\mu_{kj})}{-\mu_{kj} d''(\mu_{kj})} \text{ whenever } d''(\mu_{kj}) < 0. \tag{4.3}$$

Theorem 4.2 follows immediately from the lemma.

Theorem 4.2. *For the Bayesian EDF cross-classified model, subject to conditions (R1) and (R2), define the rectangle R_+^{K+J} as in Remark 3.4. Then a sufficient condition for the existence of a unique MAP estimate of $(\alpha_1, \ldots, \alpha_K, \beta_1, \ldots, \beta_J)$ is that conditions (4.1)–(4.3) hold.*

https://doi.org/10.1017/asb.2016.23 Published online by Cambridge University Press

One may also consider the question of uniqueness for the non-Bayesian case. In this case, it turns out within the proof of Lemma 4.1 that the log-likelihood ℓ^{cond} is not strictly convex. However, when the constraint (E3)(c) is imposed, ℓ^{cond} is found to be strictly convex on the admissible subspace of MAP estimators. Lemma 4.3 states the relevant result, proven in the appendix.

Lemma 4.3. *For the non-Bayesian EDF cross-classified model, subject to conditions (R1) and (R2), define the rectangle R_+^{K+J} as in Remark 3.4. Then a necessary and sufficient condition for the log-likelihood ℓ^{cond} to be strictly convex upward over R_+^{K+J} is that (4.3) holds for all $(\alpha_1, \ldots, \alpha_K, \beta_1, \ldots, \beta_J) \in R_+^{K+J}$.*

Theorem 4.4 follows immediately from the lemma.

Theorem 4.4. *For the non-Bayesian EDF cross-classified model, subject to conditions (R1) and (R2), define the rectangle R_+^{K+J} as in Remark 3.4. Then a necessary and sufficient condition for the existence of a unique MLE of $(\alpha_1, \ldots, \alpha_K, \beta_1, \ldots, \beta_J)$ is that condition (4.3) holds.*

4.2. Tweedie family

The following corollary applies Theorem 4.2 to the Tweedie family (proof in the appendix).

Corollary 4.5. *Consider the special case of Theorem 4.2 in which Y_{kj} are subject to a Tweedie distribution with index $p \geq 1$. Then conditions (4.1)–(4.3) reduce to*

$$\frac{A_k}{\alpha_k} > \frac{p-2}{p-1}, \tag{4.4}$$

and

$$\frac{B_j}{\beta_j} > \frac{p-2}{p-1}, \tag{4.5}$$

and

$$\frac{Y_{kj}}{\bar{\mu}_{kj}} \geq \frac{p-2}{p-1}, \tag{4.6}$$

with $\bar{\mu}_{kj}$ as defined in Remark 3.4.

Note that $(p-2)/(p-1) \leq 0$ when $1 \leq p \leq 2$, and so conditions (4.4)–(4.6) are necessarily satisfied in this case, leading to the following corollary.

Corollary 4.6. *Consider the special case of Theorem 4.4 in which Y_{kj} are subject to a Tweedie distribution with index $p \geq 1$. Then condition (4.3) reduces to (4.6).*

Corollary 4.7. *Consider the special case of Theorem 4.2 in which Y_{kj} are subject to a Tweedie distribution with index $1 \leq p \leq 2$. Then there is a unique MAP estimate of $(\alpha_1, \ldots, \alpha_K, \beta_1, \ldots, \beta_J)$.*

The same argument, applied to Theorem 4.4, yields the following result.

https://doi.org/10.1017/asb.2016.23 Published online by Cambridge University Press

Corollary 4.8. *Consider the special case of Theorem 4.4 in which Y_{kj} are subject to a Tweedie distribution with index $1 \leq p \leq 2$. Then there is a unique MLE of $(\alpha_1, \ldots, \alpha_K, \beta_1, \ldots, \beta_J)$.*

ODP and gamma are the cases $p = 1$ and 2, respectively, and the compound Poisson–gamma distributions occupy the interval $1 \leq p \leq 2$, all of which are special cases of Corollaries 4.7 and 4.8. Hence, the following result.

Corollary 4.9. *Consider the special case of Theorem 4.2 in which 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.

Similarly, there is a unique MLE of the model parameters in the special case of Theorem 4.4 in which Y_{kj} satisfy (a), (b) or (c).

4.3. Tweedie sub-family with $p > 2$

Consider Tweedie distributions with index parameter $p > 2$. The necessary and sufficient conditions for convexity of the posterior log-likelihood ℓ^{post} are still (4.4)–(4.6), but now $(p-2)/(p-1) > 0$, so these conditions are no longer automatically satisfied. Indeed, sufficiently small $Y_{kj}/\bar{\mu}_{kj}$ will violate (4.6), a necessary condition for convexity.

It follows that, in any such cases, ℓ^{post} is **not convex**. However, in none of the above results does non-convexity necessarily imply non-uniqueness. It is simply that uniqueness is not established in these non-convex cases. The present subsection will, therefore, examine some additional conditions that lead to convexity and uniqueness in the case $p > 2$. The issue of uniqueness will be explored further in Section 5.

4.3.1. *Non-Bayesian model.* To commence, consider an array $\mathfrak{D}$ that is perfectly **proportional** in the sense that $Y_{kj} = a_k b_j$. The non-Bayesian EDF cross-classified model, applied to this array, has the obvious MLE $\alpha_k = a_k$, $\beta_j = b_j$, and it is unique. This follows from Lemma 4.3, together with (A41), (A42), and the fact that $Y_{kj}/\mu_{kj} = 1$.

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" to proportional.

https://doi.org/10.1017/asb.2016.23 Published online by Cambridge University Press

The results stated in the remainder of the current sub-section formalise these intuitive ideas.

For a regular 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 $\gamma_{r,j_r:k_s,s}$ from Y_{rj_r} to $Y_{k_s s}$. Denote this path $\{Y_{r_i s_i}, i = 1, 2, \ldots, m\}$, 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 $\eta_{r,j_r:k_s,s} \subset \gamma_{r,j_r:k_s,s}$ of edges between adjacent columns in $\gamma_{r,j_r:k_s,s}$, 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

$$\pi_{r,j_r:k_s,s} = \frac{Y_{rs}}{Y_{rj_r}} \prod_{\eta_{r,j_r:k_s,s}} \frac{Y_{r_i s_i}}{Y_{r_{i+1}s_{i+1}}} - 1 \; \bar{\pi}_{rs} = \max_{j_r,k_s,\gamma_{r,j_r:k_s,s}} \left[\pi_{r,j_r:k_s,s}\right]_+ \underline{\pi}_{rs}$$
$$= \max_{j_r,k_s,\gamma_{r,j_r:k_s,s}} \left[-\pi_{r,j_r:k_s,s}\right]_+, \tag{4.7}$$

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

Now define

$$\bar{\pi}(\mathfrak{D}) = \max_{\mathcal{S}} \max_{\{r,s:\, Y_{rs} \notin \mathcal{S}\}} \bar{\pi}_{rs} \; \underline{\pi}(\mathfrak{D}) = \max_{\mathcal{S}} \max_{\{r,s:\, Y_{rs} \notin \mathcal{S}\}} \underline{\pi}_{rs}, \tag{4.8}$$

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

Finally, define

$$\xi(\mathfrak{D}) = \frac{1 + \bar{\pi}(\mathfrak{D})}{1 - \underline{\pi}(\mathfrak{D})} \geq 1. \tag{4.9}$$

It may be remarked that, for a proportional array, the ratios $Y_{r_i s_i}/Y_{r_{i+1}s_{i+1}}$ in (4.7) all take the form $Y_{r_i s_i}/Y_{r_i, s_i+1} = \beta_{s_i}/\beta_{s_i+1}$ or $Y_{r_i s_i}/Y_{r_i, s_i-1} = \beta_{s_i}/\beta_{s_i-1}$, independent of row r_i, and these ratios proceed by single steps from column j_r to column s. It follows that $\pi_{r,j_r:k_s,s} = 0$, 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 a general regular array $\mathfrak{D}$.

https://doi.org/10.1017/asb.2016.23 Published online by Cambridge University Press

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:

$$\pi^{\text{trap}}_{r,j_r:k_s,s} = \frac{Y_{rs}}{Y_{rj_r}}\frac{Y_{k_s j_r}}{Y_{k_s s}} - 1 \; \bar{\pi}^{\text{trap}}_{rs} = \max_{j_r,k_s}\left[\pi_{r,j_r:k_s,s}\right]_+ \underline{\pi}^{\text{trap}}_{rs}$$
$$= \max_{j_r,k_s}\left[-\pi_{r,j_r:k_s,s}\right]_+ \; \bar{\pi}^{\text{trap}}(\mathfrak{D}) = \max_{r,s}\bar{\pi}_{rs} \; \underline{\pi}^{\text{trap}}(\mathfrak{D}) = \max_{r,s}\underline{\pi}_{rs}. \quad (4.10)$$

The results below are formulated as applying to just trapezoidal arrays. The reason for this is mere simplicity. Extensions to non-trapezoidal arrays would be possible, but their statements are tedious.

The results immediately below address conjectures (a) and (b) from earlier in the present sub-section. Proofs appear in the appendix.

Theorem 4.10. *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 non-Bayesian Tweedie cross-classified model $Y_{kj} = \alpha^*_k\beta^*_j (= \mu^*_{kj})$ say, subject to multiplicative weights $\phi_{kj} = (\upsilon_k w_j)^{-1}$ (Lemma A.3 guarantees the possibility of this). Now fit the non-Bayesian 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:*

$$[1 + \underline{\pi}(\mathfrak{D})][\xi(\mathfrak{D})]^{-(|J_k-s|+2)} \le \frac{\alpha_k\beta_s}{\alpha^*_k\beta^*_s} \le [1 + \bar{\pi}(\mathfrak{D})][\xi(\mathfrak{D})]^{|J_k-s|+2}, \quad (4.11)$$

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

Corollary 4.11. *Consider the special case of Theorem 4.4 in which $\mathfrak{D}$ is a trapezoidal array, and Y_{kj} are subject to a Tweedie distribution with index $p > 2$ and multiplicative weights $\phi_{kj} = (\upsilon_k w_j)^{-1}$. Then a sufficient condition for convexity of the log-likelihood ℓ^{cond} is*

$$[\xi(\mathfrak{D})]^{-(J+2)} \ge \frac{p-2}{p-1}. \quad (4.12)$$

This is, therefore, a sufficient condition for the 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. Then Corollary 4.12 leads immediately to the following result, already stated earlier in the present sub-section.

Corollary 4.12. *Consider the special case of Theorem 4.4 in which $\mathfrak{D}$ is a* ***proportional*** *trapezoidal array, and Y_{kj} are subject to a Tweedie distribution with index $p > 2$ and multiplicative weights $\phi_{kj} = (\upsilon_k w_j)^{-1}$. Then the log-likelihood ℓ^{cond} is convex upward, and there is a unique MLE of model parameters.*

Condition (4.12) for a unique MLE may be re-stated as follows:

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

which decreases monotonically from 3 to 2 as $\xi(\mathfrak{D})$ increases from 1 to ∞. Thus, Corollary 4.11 shows that, as non-proportionality of $\mathfrak{D}$ increases (i.e. $\xi(\mathfrak{D})$ increases), satisfaction of the condition (4.12) for a unique MLE requires a steadily decreasing Tweedie index p.

4.3.2. *Bayesian model.* Certain of the results of Section 4.3.1 are easily extended to the Bayesian Tweedie cross-classified model. Theorem 4.13 gives the formal statement, with proof given in the appendix.

Theorem 4.13. *Corollaries 4.11 and 4.12 also apply to the Bayesian Tweedie cross-classified model, with ℓ^{cond} replaced by ℓ^{post} and the MLE replaced by the MAP estimators.*

5. 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 4, 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 these cases. As pointed out in Section 4.3, 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

$$\mathfrak{D} = \begin{bmatrix} Y_{11} & Y_{12} \\ Y_{21} & 1 \end{bmatrix}. \tag{5.1}$$

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

$$\mathfrak{D} = \begin{bmatrix} 1 & Y \\ Y & 1 \end{bmatrix}, \tag{5.2}$$

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

https://doi.org/10.1017/asb.2016.23 Published online by Cambridge University Press

Remark 5.1. An extremely simple situation is under consideration. The data set (5.2), containing four 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 (5.2). The array is symmetric under interchange or rows, accompanied by interchange of columns. The conclusions from this are as follows:

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 (3.15) and (3.16) with all $\psi_k^{(\alpha)}$, $\psi_j^{(\beta)}$ set to infinity, as the following:

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

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

The parameter set for this array is $\{\alpha_k, \beta_j : k = 1, 2;\ j = 1, 2\}$. According to (E3)(c), 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

$$b_1 = \frac{1+a}{1+Ya^2}, b_2 = \frac{1+a}{Y+a^2}, a = \frac{b_1+b_2}{Yb_1^2+b_2^2}, \tag{5.3}$$

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 (5.3) 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

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

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

$$Ya^4 - (1-Y)^2a^3 - \left(1 - 4Y + Y^2\right)a^2 - (1-Y)^2a + Y = 0. \tag{5.4}$$

It is evident that, if $a = x$ is a root of this equation, then so is $a = x^{-1}$. So the left side of (5.4) must contain a factor of $(a-x)(a-x^{-1}) = a^2 - Xa + 1$, where $X = x + x^{-1}$.

Factorise the left side of (5.4) by equating coefficients of $a^4,\ a^3,\ a^0$ in the following:

$$Ya^4 - (1-Y)^2 a^3 - \left(1 - 4Y + Y^2\right) a^2 - (1-Y)^2 a + Y$$
$$= \left(a^2 - Xa + 1\right)\left[Ya^2 + \left(XY - (1-Y)^2\right) a + Y\right]. \tag{5.5}$$

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

$$Y - X\left[XY - (1-Y)^2\right] + Y = -\left(1 - 4Y + Y^2\right), \tag{5.6}$$

$$XY - (1-Y)^2 - XY = -(1-Y)^2. \tag{5.7}$$

Now (5.7) is an identity, and so adds no information. However, (5.6) requires that

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

which yields the strictly positive solution

$$X = Y - 1\ if\ Y > 1 = 1/Y - 1\ if\ Y < 1. \tag{5.8}$$

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 (5.8), only if $Y \geq 3$ or $Y \leq 1/3$. This result is unsurprising, because Corollaries 4.11 and 4.12 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

$$x = 1/2\left[(Y-1) + \sqrt{(Y-1)^2 - 4}\right]. \tag{5.9}$$

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:

$$a = 1,\ 1/2\left[(Y-1) \pm \sqrt{(Y-1)^2 - 4}\right]. \tag{5.10}$$

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

Result 5.2. Consider the special case of Theorem 4.4 in which Y_{kj} are subject to an inverse Gaussian distribution (Tweedie with index $p = 3$), and the array $\mathfrak{D}$ is given by (5.2). The ML equations reduce to (5.3) and

https://doi.org/10.1017/asb.2016.23 Published online by Cambridge University Press

TABLE 1

VALUES OF NEGATIVE HESSIAN MINORS FOR VARYING Y.

	Value of Δ_q for						
q	$Y = 2.5$	$Y = 2.9$	$Y = 2.99$	$Y = 3$	$Y = 3.01$	$Y = 3.1$	$Y = 3.5$
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

a. have unique solution $a = 1$ if $1/3 < Y < 3$; and

b. otherwise have the multiple solution $a = 1, \frac{1}{2}[(Y-1) \pm \sqrt{(Y-1)^2 - 4}]$.

Remark 5.3. These solutions are consistent with Remark 5.1.

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^{cond} with respect to r_k and s_j, evaluated at the solution corresponding to $a = 1$ by reference to (A25)–(A27). Upward convexity of ℓ^{cond} 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 1 displays the values of these minor for a sample of values of Y increases through the value 3.

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^{cond} 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^{cond} in the vicinity of its other stationary points, seen to occur in the case $Y > 3$. This is done in the following example.

Example 5.2. As an example, set $Y = 3.5$. This generates the solutions $a = 1, 2, \frac{1}{2}$. Table 2 expresses these three solutions in their α, β parameterisations.

As is apparent from 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,

https://doi.org/10.1017/asb.2016.23 Published online by Cambridge University Press

TABLE 2

MULTIPLE SOLUTIONS OF ML EQUATIONS.

	Value of			
Solution a=	α_1	α_2	β_1	β_2
1	1	1	2.25	2.25
2	1	0.5	5	2.5
0.5	1	2	1.25	2.5

TABLE 3

VALUES OF NEGATIVE HESSIAN MINORS FOR MULTIPLE SOLUTIONS AT $Y = 3.5$.

	Value of Δ_q for solution at		
q	$a = 1$	$a = 2$	$a = 0.5$
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

is found from (A25)–(A27) 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 3 displays the leading principal minors of these matrices, found from (A25)–(A27), just as in Table 1. Once again, the zero values of Δ_q merely reflect the one degree of parameter redundancy.

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.

TABLE 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

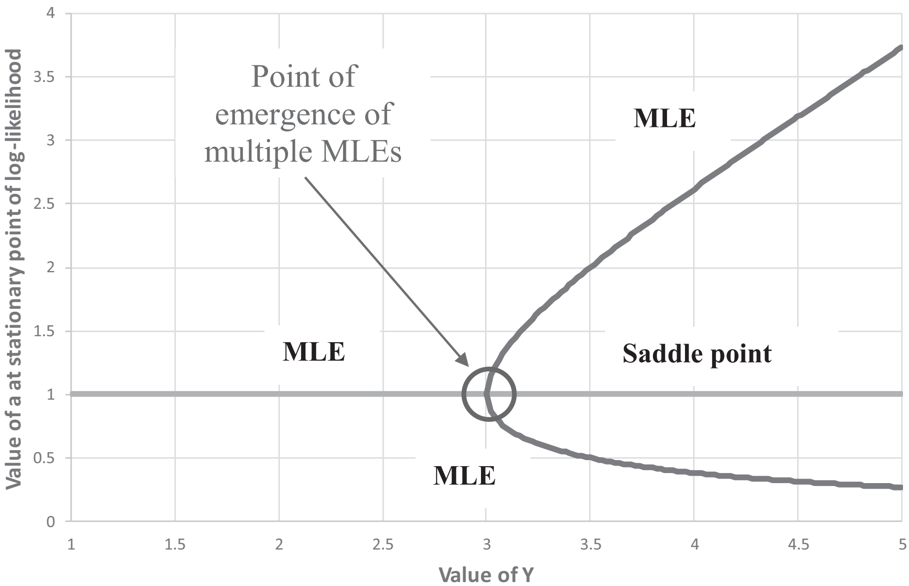

FIGURE 1: Loci of stationary points of ℓ^{cond} with varying Y.

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 4 checks this by tabulating values of the (partial) log-likelihood ℓ^{cond} and finds it not to be the case. The two solutions at $a = 2,\ 0.5$ are global maxima.

The difference between the global maxima of ℓ^{cond} 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^{cond} with varying Y are as illustrated in Figure 1.

https://doi.org/10.1017/asb.2016.23 Published online by Cambridge University Press

6. Conclusion

The MAP estimation equations (3.15) and (3.16) for the Bayesian 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 non-Bayesian version of this model, whose solution is given by the same equations with all $\psi_k^{(\alpha)}, \psi_j^{(\beta)}$ set to infinity.

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 3 and 4 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, where a unique MAP estimator exists in the Bayesian model if the Tweedie index p lies in the closed interval $[1, 2]$.

The corresponding result for the Bayesian model and may be summarised this way. A unique MLE exists provided that either:

a. the error distribution is sufficiently well-behaved (Tweedie index $1 \leq p \leq 2$); 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 5 that multiple solutions can occur in the non-Bayesian model.

The data array in the example of Section 5 has been deliberately chosen to be as simple as possible. It contains only four 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.

Acknowledgements

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.

I am also greatly indebted to Xinda Yang of UNSW Australia who provided valuable service in checking a number of proofs, and suggesting some improvements.

References

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

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

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

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

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

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

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

McCullagh, P. and Nelder, J.A. (1989). *Generalized Linear Models*, 2nd ed. London, UK: Chapman & Hall.

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

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

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

Taylor, G. (2000) *Loss Reserving: An Actuarial Perspective*. Boston: Kluwer Academic Publishers.

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), 75–99.

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

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. (2003) Wüthrich M V (2007). Claims reserving using Tweedie's compound Poisson model. *Astin Bulletin*, **33**(2), 331–346.

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. and MERZ, M. (2008). *Stochastic Claims Reserving Methods in Insurance*, Chichester, UK: John Wiley & Sons Ltd.

GREG TAYLOR
UNSW Business School
Level 6, West Lobby, UNSW Business School Building E12
UNSW Sydney 2052 Australia
E-Mail: gregory.taylor@unsw.edu.au

APPENDIX

Proof of Theorem 3.3

Lemma A.1. *For the Bayesian EDF cross-classified model, subject to conditions (R1)–(R2), there exists a $(K+J)$ -dimensional co-ordinate rectangle R_+^{K+J}, in the positive orthant, with all boundary planes bounded away from zero and infinity, and within which any MAP estimate of the parameter set $\{\alpha_k, \beta_j : k = 1, \ldots, K;\ j = 1, \ldots, J\}$ must lie.*

Proof. Recall the MAP estimation equations (3.15) and (3.16), and consider the first of these. Divide through by $c'(\alpha_k)$, and apply condition (R1), to obtain

$$\sum_{j\in\mathcal{R}(k)} \left[Y_{kj} - \alpha_k\beta_j\right] c'\left(\beta_j\right)\beta_j q\left(\alpha_k, \beta_j\right)/\phi_{kj} + [A_k - \alpha_k]/\psi_k^{(\alpha)} = 0, k = 1, \ldots, K, \tag{A1}$$

with

$$0 < m \le q\left(\alpha_k, \beta_j\right) \le M < \infty. \tag{A2}$$

An implicit solution for α_k, β_j is given by

$$\alpha_k = \frac{\sum_{j\in\mathcal{R}(k)} \frac{Y_{kj}}{\beta_j}\beta_j^2 c'\left(\beta_j\right) q\left(\alpha_k, \beta_j\right)/\phi_{kj} + A_k/\psi_k^{(\alpha)}}{\sum_{j\in\mathcal{R}(k)} \beta_j^2 c'\left(\beta_j\right) q\left(\alpha_k, \beta_j\right)/\phi_{kj} + 1/\psi_k^{(\alpha)}}, \tag{A3}$$

$$\beta_j = \frac{\sum_{k\in\mathcal{C}(j)} \frac{Y_{kj}}{\alpha_k}\alpha_k^2 c'\left(\alpha_k\right) q\left(\beta_j, \alpha_k\right)/\phi_{kj} + B_j/\psi_j^{(\beta)}}{\sum_{k\in\mathcal{C}(j)} \alpha_k^2 c'\left(\alpha_k\right) q\left(\beta_j, \alpha_k\right)/\phi_{kj} + 1/\psi_j^{(\beta)}}. \tag{A4}$$

The proof of the lemma considers four cases.

Case I: Condition (R2)(a) holds. By assumption, $0 < \underline{C} < \alpha_k c'(\alpha_k), \beta_j c'(\beta_j) < \bar{C} < \infty$ for constants $\underline{C}, \bar{C}$. It then follows from (A3) that

$$\alpha_k < \psi_k^{(\alpha)} \sum_{j\in\mathcal{R}(k)} Y_{kj} M\bar{C}/\phi_{kj} + A_k. \tag{A5}$$

https://doi.org/10.1017/asb.2016.23 Published online by Cambridge University Press

Similarly, using (A4),

$$\beta_j < \psi_j^{(\beta)} \sum_{k \in \mathcal{C}(j)} Y_{kj} M\bar{C}/\phi_{kj} + B_j. \tag{A6}$$

Substitution of (A6) into (A3) yields

$$\alpha_k > \frac{m\underline{C} \sum_{j \in \mathcal{R}(k)} Y_{kj}/\phi_{kj} + A_k/\psi_k^{(\alpha)}}{M\bar{C} \sum_{j \in \mathcal{R}(k)} \left[M\bar{C}\psi_j^{(\beta)} \sum_{k \in \mathcal{C}(j)} Y_{kj}/\phi_{kj} + B_j \right] /\phi_{kj} + 1/\psi_k^{(\alpha)}}. \tag{A7}$$

Thus, α_k is bounded away from zero and infinity. A similar calculation establishes the same property for β_j.

Case II: Condition (R2)(b) holds. The proof is similar to Case I. This time, it is assumed that $0 < \underline{C} < \alpha_k^2 c'(\alpha_k), \beta_j^2 c'(\beta_j) < \bar{C} < \infty$ for constants $\underline{C}$, $\bar{C}$. It then follows from (A3) and (A4) that

$$\alpha_k > \frac{A_k/\psi_k^{(\alpha)}}{\sum_{j \in \mathcal{R}(k)} M\bar{C}/\phi_{kj} + 1/\psi_k^{(\alpha)}}, \tag{A8}$$

$$\beta_j > \frac{B_j/\psi_j^{(\beta)}}{\sum_{k \in \mathcal{C}(j)} M\bar{C}/\phi_{kj} + 1/\psi_j^{(\beta)}}. \tag{A9}$$

Substitution of (A9) into (A3) yields

$$\alpha_k < \frac{M\bar{C} \sum_{j \in \mathcal{R}(k)} \frac{Y_{kj}}{\phi_{kj}} \frac{\psi_j^{(\beta)}}{B_j} \left[M\bar{C} \sum_{k \in \mathcal{C}(j)} 1/\phi_{kj} + 1/\psi_j^{(\beta)} \right] + A_k/\psi_k^{(\alpha)}}{m\underline{C} \sum_{j \in \mathcal{R}(k)} 1/\phi_{kj} + 1/\psi_k^{(\alpha)}}. \tag{A10}$$

As in Case I, α_k is bounded away from zero and infinity, and a similar calculation establishes the same property for β_j.

Cases III and IV will be proved by contradiction. Let α denote the vector $(\alpha_1, \ldots, \alpha_K)$ and β the vector $(\beta_1, \ldots, \beta_J)$. Suppose that the statement of the lemma is untrue so that a sequence of solutions $\{\alpha^{(m)}, \beta^{(m)}, m = 1, 2, \ldots\}$ with some component (not necessarily the same component for different m) approaching either zero or infinity as $m \to \infty$.

Since the number of components $K + J$ is finite, there is a sub-sequence for which $\alpha_k^{(m)} \to 0$ *or* ∞ or $\beta_j^{(m)} \to 0$ *or* ∞ as $m \to \infty$ with k, j now fixed as m varies. Consider such a subsequence $\{\beta^{(m)}\}$ for which $\beta_j^{(m)} \to 0$. There may be more than one such j, and $\beta_j^{(m)}$ may converge to zero at differing rates. However, since there is only a finite number of permutations of $\beta_j^{(m)}$ for fixed m, it is possible to find a sub-subsequence such that the ordering of the components of $\beta^{(m)}$ (in ascending order, say) is the same for all m.

With slight abuse of notation, let this sub-subsequence now be denoted simply $\{\beta^{(m)}\}$, and it is meaningful to speak of the subset of values $j \in \{1, \ldots, J\}$ for which $\beta_j^{(m)} \to 0$ most rapidly, i.e., become negligible relative to all other $\beta_j^{(m)}$, and the ratios of pairs of the most rapidly decreasing $\beta_j^{(m)}$ remain bounded, as $m \to \infty$. This subset will take the form $\{1, \ldots, \bar{j}\}$, where $\beta_{\bar{j}}^{(m)}, \beta_1^{(m)}$ denote the greatest and least of the most rapidly converging $\beta_j^{(m)}$, and then $\beta_{\bar{j}}^{(m)}/\beta_1^{(m)} < C$ for some constant $C > 0$.

https://doi.org/10.1017/asb.2016.23 Published online by Cambridge University Press

Case III: Condition (R2)(c)(i) holds. Consider a sequence $\{\beta_{j_1}^{(m)}\} \to 0$, as introduced immediately above. By condition (R2)(c)(i), $[\beta_{j_1}^{(m)}]^2 c'(\beta_{j_1}^{(m)}) \to \infty$. Henceforth, the index m will be omitted for brevity.

Let k be an arbitrary value in $\mathcal{C}(j_1)$, and note that (A3) expresses α_k as a weighted average of terms Y_{kj}/β_j and A_k, with weights $\beta_j^2 c'(\beta_j) q(\alpha_k, \beta_j)/\phi_{kj}$ and $1/\psi_k^{(\alpha)}$. These weights will be dominated by those involving $\beta_j \to 0$. Moreover, for these j, α_k must be at least $O(Y_{kj}/\beta_{j_{(z)}}) \to \infty$, where $j_{(z)}$ is the greatest of those $\beta_j \to 0$. It then follows that $\alpha_k c'(\alpha_k), \alpha_k^2 c'(\alpha_k) \to 0$.

Now consider β_{j_1}, as given by (A4). Since k was arbitrary within $\mathcal{C}(j_1)$, the summations asymptotically contribute nothing, and $\beta_{j_1} \to B_{j_1}$. This contradicts the assumption that $\beta_{j_1} \to 0$, which, therefore, cannot be a valid assumption.

Similar arguments dispose of the alternative assumptions that $\beta_j \to \infty$ or $\alpha_k \to 0$ or ∞, and so the assumption of the falsity of the lemma must itself be false.

Case IV: Condition (R2)(c)(ii) holds. Consider a sequence $\beta_{j_1} \to 0$, as in Case III, but now with j_1 such as to produce most rapid convergence to zero. By condition (R2)(c)(ii), $\beta_{j_1} c'(\beta_{j_1}) \to \infty$, $\beta_{j_1}^2 c'(\beta_{j_1}) \to 0$. Apply these conditions, together with the tail convergence property of (R2)(c)(ii), to (A3) to obtain $\alpha_k = O(\beta_{j_1} c'(\beta_{j_1})) \to \infty$ for any $k \in \mathcal{C}(j_1)$. This implies in turn that $\alpha_k c'(\alpha_k) \to 0$, $\alpha_k^2 c'(\alpha_k) \to \infty$.

Now apply these results to calculate β_{j_1} from (A4). The numerator converges to just $B_{j_1}/\psi_{j_1}^{(\beta)}$, and so $\beta_{j_1}^{-1} = O(\alpha_k^2 c'(\alpha_k)) = O([\beta_{j_1} c'(\beta_{j_1})]^2 c'(\beta_{j_1} c'(\beta_{j_1})))$. This contradicts condition (R2)(c)(ii), and so the premise of the argument, $\beta_{j_1} \to 0$, must be false.

Similar arguments dispose of the alternative assumptions that $\beta_j \to \infty$ or $\alpha_k \to 0$ or ∞, and so the assumption of the falsity of the lemma must itself be false. ■

Proof of Theorem 3.3.

Existence is established, by means of an application of the Weierstrass theorem to the posterior likelihood function (2.26) of $\mathfrak{D}$. This requires continuity of the likelihood as a function of the parameter set $\{\alpha_k, \beta_j\}$, and compactness of the admissible parameter space.

The first of these conditions amounts to continuity of $c(= (\kappa')^{-1})$ in (2.20) and (2.21), and this is implied by the assumption in (E3)(a) that κ is twice differentiable. Compactness is established in Lemma A.1. ■

Proof of Theorem 3.5

The following lemma is the non-Bayesian counterpart of Lemma A.1.

Lemma A.2. *The result of Lemma A.1 also holds for the non-Bayesian EDF cross-classified model of Section 2.2, subject to conditions (R1) and (R2).*

Proof. The proof commences as in the proof of Lemma A.1, except that all terms involving $\psi_k^{(\alpha)}$ are now absent, so that (A3) and (A4) become

$$\alpha_k = \frac{\sum_{j \in \mathcal{R}(k)} \frac{Y_{kj}}{\beta_j} \beta_j^2 c'\left(\beta_j\right) q\left(\alpha_k, \beta_j\right) / \phi_{kj}}{\sum_{j \in \mathcal{R}(k)} \beta_j^2 c'\left(\beta_j\right) q\left(\alpha_k, \beta_j\right) / \phi_{kj}}, \tag{A11}$$

https://doi.org/10.1017/asb.2016.23 Published online by Cambridge University Press

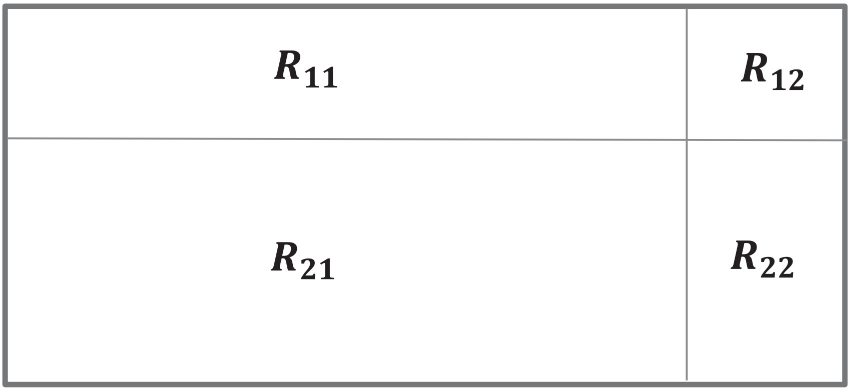

FIGURE A1: Sub-rectangle of infinite α's and zero β's.

$$\beta_j = \frac{\sum_{k \in \mathcal{C}(j)} \frac{Y_{kj}}{\alpha_k} \alpha_k^2 c'(\alpha_k) q(\beta_j, \alpha_k) / \phi_{kj}}{\sum_{k \in \mathcal{C}(j)} \alpha_k^2 c'(\alpha_k) q(\beta_j, \alpha_k) / \phi_{kj}}. \tag{A12}$$

The same four cases as in the proof of Lemma A.1 are considered.

Case I: Condition (R2)(a) holds. By (A11) and (A12)

$$\alpha_k < \frac{\sum_{j \in \mathcal{R}(k)} Y_{kj} M\bar{C}/\phi_{kj}}{\sum_{j \in \mathcal{R}(k)} \beta_j m \underline{C}/\phi_{kj}}, \tag{A13}$$

$$\beta_j > \frac{\sum_{k \in \mathcal{C}(j)} Y_{kj} m \underline{C}/\phi_{kj}}{\sum_{k \in \mathcal{C}(j)} \alpha_k M\bar{C}/\phi_{kj}}, \tag{A14}$$

where the constants m, M, $\underline{C}$, $\bar{C}$ are as in the proof of Case I of Lemma A.1.

Suppose that some $\beta_{j_1} \to 0$. By (A14), this can occur only if $\alpha_k \to \infty$ for all $k \in \mathcal{C}(j_1)$. Select $k_i \in \mathcal{C}(j_1)$, and note from (A13) that $\alpha_{k_i} \to \infty$ only if $\beta_j \to 0$ for all $j \in \mathcal{R}(k_i)$. This yields a sub-array of points (k_i, j_h) for which $\alpha_{k_i} \to \infty$, $\beta_{j_h} \to 0$. By permutation of rows and columns, the sub-array may be represented as a sub-rectangle in the top left corner of the rectangle consisting of K rows and J columns.

The alternate application of (A14) and (A13) may be continued, with each application possibly adding a row or column to sub-rectangle. At each stage, the sub-rectangle R_{11} appears within the full rectangle as illustrated by Figure A1. By condition (A2), there must be an observation in R_{12} or R_{22}.

Suppose the former. Then the column containing that observation may be added to R_{11}. Alternatively suppose that the observation is in R_{22}. By (A3), it must be connected to observations in R_{11}, and this implies an observation in R_{12} or R_{21}. Hence, another row or column can be added to R_{11}.

It follows that R_{11} will continue to expand until it coincides with the full rectangle. This means that, if any $\beta_{j_1} \to 0$, then all $\beta_j \to 0$ and all $\alpha_k \to \infty$. But then $\sum_{j=1}^{J} \beta_j \to 0$, in violation of (E3)(c). Hence, all $\beta_j \to 0$ must be bounded away from zero, and then all α_k must be bounded away from infinity. Also by condition (E3)(c), all β_j must be bounded away from infinity, and all α_k must be bounded away from zero.

https://doi.org/10.1017/asb.2016.23 Published online by Cambridge University Press

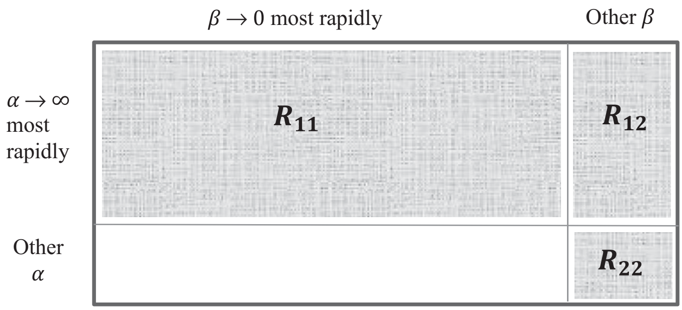

FIGURE A2: Arrangement of α's and β's.

Case II: Condition (R2)(b) holds. In this case, (A11) and (A12) are replaced by

$$\alpha_k < \frac{\sum_{j\in\mathcal{R}(k)} \left(Y_{kj}/\beta_j\right) M\bar{C}/\phi_{kj}}{\sum_{j\in\mathcal{R}(k)} m\underline{C}/\phi_{kj}}, \tag{A15}$$

$$\beta_j > \frac{\sum_{k\in\mathcal{C}(j)} \left(Y_{kj}/\alpha_k\right) m\underline{C}/\phi_{kj}}{\sum_{k\in\mathcal{C}(j)} M\bar{C}/\phi_{kj}}, \tag{A16}$$

where the constants $\underline{C}$, $\bar{C}$ are now as in the proof of Case II of Lemma A.1.

As in Case I, one commences with the supposition that some $\beta_{j_1} \to 0$. The proof then proceeds exactly as in that case.

Case III: Condition (R2)(c)(i) holds. Suppose that some $\beta_{j_1} \to 0$, with most rapid convergence in the sense used in the proof of Lemma A.1. Then the proof of Case III of Lemma A.1 continues to hold up to the point of demonstrating that $\alpha_k \to \infty$ for all $k \in \mathcal{C}(j_1)$. Hence, after suitable permutation of rows and columns, the full $K \times J$ rectangle may be placed in the form illustrated by Figure A2, where observations in $\mathfrak{D}$ are confined to the hatched regions.

Let $\underline{\mathcal{R}}(k)$ denote the set of $j \in \mathcal{R}(k)$ for which $\beta_j \to 0$ with most rapid convergence, and let $\bar{\mathcal{R}}(k)$ the set of $j \in \mathcal{R}(k)$ for which β_j converges to zero less rapidly or not at all. Suppose that $\beta_j = \varepsilon v_j(\varepsilon)$ for $j \in \underline{\mathcal{R}}(k)$, and $\beta_j = (1-\varepsilon)w_j(\varepsilon)$ for $j \in \bar{\mathcal{R}}(k)$, where $v_j(.)$ and $w_j(.)$ are weights (dependent on ε): $v_j, w_j > 0$, $\sum_{j\in\underline{\mathcal{R}}(k)} v_j = \sum_{j\in\bar{\mathcal{R}}(k)} w_j = 1$.

Now suppose that $\varepsilon \to 0$, with all v_j, w_j bounded away from zero. Expand the numerator of (A11) as follows:

$$\begin{aligned}
&\sum_{j\in\underline{\mathcal{R}}(k)} \frac{Y_{kj}}{\varepsilon v_j(\varepsilon)} \left[\varepsilon v_j(\varepsilon)\right]^2 c'\left(\varepsilon v_j(\varepsilon)\right) q\left(\alpha_k, \beta_j\right)/\phi_{kj} \\
&+ \sum_{j\in\bar{\mathcal{R}}(k)} Y_{kj}(1-\varepsilon) w_j(\varepsilon) c'\left((1-\varepsilon) w_j(\varepsilon)\right) q\left(\alpha_k, \beta_j\right)/\phi_{kj} \\
&= \varepsilon c'(\varepsilon) \sum_{j\in\underline{\mathcal{R}}(k)} \frac{Y_{kj}}{\phi_{kj} v_j(\varepsilon)} v_j^2(\varepsilon) c'\left(v_j(\varepsilon)\right) q\left(\alpha_k, \beta_j\right) q\left(v_j(\varepsilon), \varepsilon\right) \\
&+ \sum_{j\in\bar{\mathcal{R}}(k)} Y_{kj} w_j(\varepsilon) c'\left(w_j(\varepsilon)\right) (1-\varepsilon) c'(1-\varepsilon) q\left(\alpha_k, \beta_j\right) q\left(w_j(\varepsilon), 1-\varepsilon\right)/\phi_{kj},
\end{aligned} \tag{A17}$$

by condition (R1), and where $q(.,.)$ is the function introduced in (R1).

https://doi.org/10.1017/asb.2016.23 Published online by Cambridge University Press

Similarly, the denominator of (A11) may be expressed as

$$\varepsilon^2 c'(\varepsilon) \sum_{j \in \underline{\mathcal{R}}(k)} v_j^2(\varepsilon)\, c'\left(v_j(\varepsilon)\right) q\left(\alpha_k, \beta_j\right) q\left(v_j(\varepsilon), \varepsilon\right) / \phi_{kj}$$

$$+ \sum_{j \in \bar{\mathcal{R}}(k)} w_j^2(\varepsilon)\, c'\left(w_j(\varepsilon)\right)(1-\varepsilon)\, c'(1-\varepsilon)\, q\left(\alpha_k, \beta_j\right) q\left(w_j(\varepsilon), 1-\varepsilon\right) / \phi_{kj}. \quad \text{(A18)}$$

Note that the second summations in each of (A17) and (A18) are of order unity, whereas by condition (R2)(c)(i), $\varepsilon^2 c'(\varepsilon) \to \infty$, and therefore, $\varepsilon c'(\varepsilon) \to \infty$ also, but at a lesser rate. Therefore, substitution of (A17) and (A18) into (A11) allows α_k to be expressed in the form

$$\alpha_k = \varepsilon^{-1} D_1^{(k)}(\varepsilon) \left[1 + \frac{D_2^{(k)}(\varepsilon)}{\varepsilon^2 c'(\varepsilon)} + o\left(\frac{1}{\varepsilon^2 c'(\varepsilon)}\right)\right]^{-1}, \quad \text{(A19)}$$

where

$$D_1^{(k)}(\varepsilon) = \frac{\sum_{j \in \underline{\mathcal{R}}(k)} \frac{Y_{kj}}{v_j(\varepsilon)} v_j^2(\varepsilon)\, c'\left(v_j(\varepsilon)\right) q\left(\alpha_k, \beta_j\right) q\left(v_j(\varepsilon), \varepsilon\right) / \phi_{kj}}{\sum_{j \in \underline{\mathcal{R}}(k)} v_j^2(\varepsilon)\, c'\left(v_j(\varepsilon)\right) q\left(\alpha_k, \beta_j\right) q\left(v_j(\varepsilon), \varepsilon\right) / \phi_{kj}}, \quad \text{(A20)}$$

$$D_2^{(k)}(\varepsilon) = \frac{\sum_{j \in \bar{\mathcal{R}}(k)} w_j^2(0)\, c'\left(w_j(0)\right) c'(1)\, q\left(\alpha_k, \beta_j\right) / \phi_{kj}}{\sum_{j \in \underline{\mathcal{R}}(k)} v_j^2(\varepsilon)\, c'\left(v_j(\varepsilon)\right) q\left(\alpha_k, \beta_j\right) q\left(v_j(\varepsilon), \varepsilon\right) / \phi_{kj}} \geq 0. \quad \text{(A21)}$$

Note that $1/D_1^{(k)}(\varepsilon)$ is a weighted harmonic mean of the terms $v_j(\varepsilon)/Y_{kj}$, $j \in \underline{\mathcal{R}}(k)$, and so

$$1/D_1^{(k)}(\varepsilon) \geq \min_{j \in \underline{\mathcal{R}}(k)} v_j(\varepsilon) / Y_{kj} = v_{j_k}(\varepsilon) / Y_{kj_k} \text{ for some } j_k \in \underline{\mathcal{R}}(k). \quad \text{(A22)}$$

In the same way as (A11) was represented as (A19), β_j, as given by (A12), may be expressed as a weighted average, over $i \in \mathcal{C}(j)$, of terms

$$\frac{Y_{ij}}{\alpha_i} = \varepsilon Y_{ij}\left(1/D_1^{(i)}(\varepsilon)\right)\left[1 + \frac{D_2^{(i)}(\varepsilon)}{\varepsilon^2 c'(\varepsilon)} + o\left(\frac{1}{\varepsilon^2 c'(\varepsilon)}\right)\right] > \varepsilon Y_{ij}\left(1/D_1^{(i)}(\varepsilon)\right) \geq \varepsilon Y_{ij} \frac{v_{j_i}(\varepsilon)}{Y_{ij_i}}.$$

Set $j = j_i$ to obtain

$$\frac{Y_{ij_i}}{\alpha_i} > \varepsilon v_{j_i}(\varepsilon).$$

As already noted, β_{j_i} is a weighted average of these terms over $i \in \mathcal{C}(j_i)$, and so $\beta_{j_i} > \varepsilon v_{j_i}(\varepsilon)$. But, by definition, $\beta_{j_i} = \varepsilon v_{j_i}(\varepsilon)$, creating a contradiction.

This proves that no β_j can converge to zero. It is still necessary to consider the possibilities of $\beta_j \to \infty$, $\alpha_k \to 0$, $\alpha_k \to \infty$. The first of these cannot occur because of condition (E3)(c).

The second possibility, $\alpha_k \to 0$, can also be quickly eliminated. Suppose that this holds for some $k \in \mathcal{C}(j)$. Then, by (R2)(c)(i), β_j in (A12) is dominated by these k, for which $Y_{kj}/\alpha_k \to \infty$. In this case, $\beta_j \to \infty$, which, as just noted, violates (E3)(c).

For the third possibility, $\alpha_k \to \infty$, note that (A11) expresses α_k as a weighted average of terms Y_{kj}/β_j. These are all bounded, since β_j have been shown to be bounded, and so α_k must be bounded.

Case IV: Condition (R2)(c)(ii) holds. Suppose that some $\beta_{j_1} \to 0$, with most rapid convergence in the sense used in the proof of Lemma A.1. Then the proof of Case IV of Lemma A.1 continues to hold up to the point of demonstrating that $\alpha_k \to \infty$, and $\alpha_k c'(\alpha_k) \to 0$, $\alpha_k^2 c'(\alpha_k) \to \infty$ for all $k \in \mathcal{C}(j_1)$.

https://doi.org/10.1017/asb.2016.23 Published online by Cambridge University Press

Select $k_1 \in \mathcal{C}(j_1)$ and consider any $j_2 \in \mathcal{R}(k_1)$. Calculate β_{j_2} by means of (A12). Note that the denominator includes the term $\alpha_{k_1}^2 c'(\alpha_{k_1}) \to \infty$, whence $\beta_{j_2} \to 0$, with the possible exception of the case where $\alpha_{k_2} \to 0$ for some other $k_2 \in \mathcal{C}(j_2)$, for then the numerator includes the term $\alpha_{k_2} c'(\alpha_{k_2}) \to \infty$.

Now, by (A11), the case $\alpha_{k_2} \to 0$ can occur only if $\beta_j \to \infty$ for some $j \in \mathcal{R}(k_2)$, and this solution for β_j would then violate (E3)(c). Thus, it follows that $\beta_{j_2} \to 0$ for any $k_1 \in \mathcal{C}(j_1)$, $j_2 \in \mathcal{R}(k_1)$. This yields a sub-array of points (k_i, j_h) for which $\alpha_{k_i} \to \infty$, $\beta_{j_h} \to 0$, just as in Case I.

The proof then proceeds just as in Case I, with alternate applications of (A12) and (A11). This proves that there cannot exist $\beta_j \to 0$ nor $\alpha_k \to \infty$, and, in the process, that there cannot exist $\beta_j \to \infty$ nor $\alpha_k \to 0$. ■

The proof of Theorem 3.5 is then exactly as for Theorem 3.3.

Proof of Theorem 4.2

The following derivatives may be calculated, where r_k, s_j are defined at the commencement of Section 4.1

$$\partial \ell^{\text{cond}}/\partial r_k = \left(\partial \ell^{\text{cond}}/\partial \alpha_k\right)(\partial \alpha_k/\partial ln\ \alpha_k) = \alpha_k \partial \ell^{\text{cond}}/\partial \alpha_k$$

$$= \sum_{j\in\mathcal{R}(k)} \left[Y_{kj} c'\left(\mu_{kj}\right) - d'\left(\mu_{kj}\right)\right]\mu_{kj}/\phi_{kj} = \sum_{j\in\mathcal{R}(k)} \mu_{kj}\frac{\partial \ell^{\text{cond}}}{\partial \mu_{kj}}, \tag{A23}$$

$$\partial \ell^{\text{cond}}/\partial s_j = \sum_{\mathcal{C}(j)} \mu_{kj}\frac{\partial \ell^{\text{cond}}}{\partial \mu_{kj}}, \tag{A24}$$

$$\frac{\partial^2 \ell^{\text{cond}}}{\partial r_k^2} = \sum_{j\in\mathcal{R}(k)} \frac{\partial}{\partial \mu_{kj}}\left(\mu_{kj}\frac{\partial \ell^{\text{cond}}}{\partial \mu_{kj}}\right)\frac{\partial \mu_{kj}}{\partial r_k} = \sum_{j\in\mathcal{R}(k)} \mu_{kj}\frac{\partial}{\partial \mu_{kj}}\left(\mu_{kj}\frac{\partial \ell^{\text{cond}}}{\partial \mu_{kj}}\right)$$

$$= \sum_{j\in\mathcal{R}(k)}\left[\mu_{kj}^2\frac{\partial^2 \ell^{\text{cond}}}{\partial \mu_{kj}^2} + \mu_{kj}\frac{\partial \ell^{\text{cond}}}{\partial \mu_{kj}}\right], \tag{A25}$$

$$\frac{\partial^2 \ell^{\text{cond}}}{\partial s_j^2} = \sum_{k\in\mathcal{C}(j)}\left[\mu_{kj}^2\frac{\partial^2 \ell^{\text{cond}}}{\partial \mu_{kj}^2} + \mu_{kj}\frac{\partial \ell^{\text{cond}}}{\partial \mu_{kj}}\right], \tag{A26}$$

$$\frac{\partial^2 \ell^{\text{cond}}}{\partial r_k \partial s_j} = \mu_{kj}^2\frac{\partial^2 \ell^{\text{cond}}}{\partial \mu_{kj}^2} + \mu_{kj}\frac{\partial \ell^{\text{cond}}}{\partial \mu_{kj}}, \tag{A27}$$

$$\frac{\partial \ell^{\text{prior}}}{\partial r_k} = \alpha_k\frac{\partial \ell^{\text{prior}}}{\partial \alpha_k} = (A_k - \alpha_k)\, d'(\alpha_k)/\psi_k^{(\alpha)}, \tag{A28}$$

$$\frac{\partial^2 \ell^{\text{prior}}}{\partial r_k^2} = \alpha_k\frac{\partial}{\partial \alpha_k}\left(\alpha_k\frac{\partial \ell^{\text{prior}}}{\partial \alpha_k}\right) = \alpha_k\left[(A_k - \alpha_k)\, d''(\alpha_k) - d'(\alpha_k)\right]/\psi_k^{(\alpha)}. \tag{A29}$$

Similarly

$$\frac{\partial^2 \ell^{\text{prior}}}{\partial s_j^2} = \beta_j\left[\left(B_j - \beta_j\right) d''\left(\beta_j\right) - d'\left(\beta_j\right)\right]/\psi_j^{(\beta)}. \tag{A30}$$

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

https://doi.org/10.1017/asb.2016.23 Published online by Cambridge University Press

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 $Q^{\text{cond}}(v, w)$ denote the quadratic form associated with the Hessian of ℓ^{cond}, i.e.

$$\begin{aligned} Q^{\text{cond}}(v, w) &= \sum_{k=1}^{K} v_k^2 \frac{\partial^2 \ell^{\text{cond}}}{\partial r_k^2} + \sum_{j=1}^{J} w_j^2 \frac{\partial^2 \ell^{\text{cond}}}{\partial s_j^2} + 2 \sum_{\mathfrak{D}} v_k w_j \frac{\partial^2 \ell^{\text{cond}}}{\partial r_k \partial s_j} \\ &= \sum_{\mathfrak{D}} \left(v_k + w_j\right)^2 \left[\mu_{kj}^2 \frac{\partial^2 \ell^{\text{cond}}}{\partial \mu_{kj}^2} + \mu_{kj} \frac{\partial \ell^{\text{cond}}}{\partial \mu_{kj}}\right], \end{aligned} \tag{A31}$$

by (A25)–(A27).

The corresponding quadratic form associated with the Hessian of ℓ^{prior} can be calculated similarly. The prior is given by (2.20), (2.21) and (2.25):

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

Straightforward calculation, taking (3.14) into account, yields

$$\frac{\partial \ell^{\text{prior}}}{\partial r_k} = \alpha_k \frac{\partial \ell^{\text{prior}}}{\partial \alpha_k} = (A_k - \alpha_k) d'(\alpha_k) / \psi_k^{(\alpha)}, \tag{A32}$$

$$\frac{\partial^2 \ell^{\text{prior}}}{\partial r_k^2} = \alpha_k \frac{\partial}{\partial \alpha_k} \left(\alpha_k \frac{\partial \ell^{\text{prior}}}{\partial \alpha_k}\right) = \alpha_k \left[(A_k - \alpha_k) d''(\alpha_k) - d'(\alpha_k)\right] / \psi_k^{(\alpha)}. \tag{A33}$$

Similarly

$$\frac{\partial^2 \ell^{\text{prior}}}{\partial s_j^2} = \beta_j \left[\left(B_j - \beta_j\right) d''\left(\beta_j\right) - d'\left(\beta_j\right)\right] / \psi_j^{(\beta)}. \tag{A34}$$

Proof of Lemma 4.1.

The conditional and prior likelihoods are considered separately.

Conditional likelihood The second summand of the square bracket in (A31) may be interpreted by reference to (2.22) and (2.24):

$$\mu_{kj} \frac{\partial \ell^{\text{cond}}}{\partial \mu_{kj}} = \left[Y_{kj} c'\left(\mu_{kj}\right) - d'\left(\mu_{kj}\right)\right] \mu_{kj} / \phi_{kj} = \phi_{kj}^{-1} \left(Y_{kj} - \mu_{kj}\right) d'\left(\mu_{kj}\right), \tag{A35}$$

by (3.14).

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

$$\mu_{kj}^2 \frac{\partial^2 \ell^{\text{cond}}}{\partial \mu_{kj}^2} = \phi_{kj}^{-1} \left[\mu_{kj} \left(Y_{kj} - \mu_{kj}\right) d''\left(\mu_{kj}\right) - Y_{kj} d'\left(\mu_{kj}\right)\right]. \tag{A36}$$

By substitution of (A35) and (A36) into (A31)

$$\begin{aligned} &-Q^{\text{cond}}(v, w) = \textstyle\sum_{(k,j)\in\mathfrak{D}} \left(v_k + w_j\right)^2 \left[\mu_{kj} d'\left(\mu_{kj}\right) - \mu_{kj} \left(Y_{kj} - \mu_{kj}\right) d''\left(\mu_{kj}\right)\right] \\ &= \textstyle\sum_{(k,j)\in\mathfrak{D}} \left(v_k + w_j\right)^2 \mu_{kj}^2 d''\left(\mu_{kj}\right) \left[\frac{d'\left(\mu_{kj}\right)}{\mu_{kj} d''\left(\mu_{kj}\right)} - \left(\frac{Y_{kj}}{\mu_{kj}} - 1\right)\right]. \end{aligned} \tag{A37}$$

https://doi.org/10.1017/asb.2016.23 Published online by Cambridge University Press

It follows that, when (4.3) holds, $Q^{\text{cond}}(v, w) \leq 0$ for all selections of v, w with equality if and only if

$$v_k + w_j = 0 \text{ for all pairs } k, j. \tag{A38}$$

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

$$(A_k - \alpha_k)\, d''(\alpha_k) - d'(\alpha_k) < 0,$$

i.e.

$$\frac{A_k}{\alpha_k} < 1 + \frac{d'(\alpha_k)}{\alpha_k d''(\alpha_k)} \text{ if } d''(\alpha_k) > 0,$$

or

$$\frac{A_k}{\alpha_k} > 1 + \frac{d'(\alpha_k)}{\alpha_k d''(\alpha_k)} \text{ if } d''(\alpha_k) < 0. \tag{A39}$$

Similarly, the requirement that $\partial^2 \ell^{\text{prior}} / \partial s_j^2 < 0$ is equivalent to

$$\frac{B_k}{\beta_j} < 1 + \frac{d'\left(\beta_j\right)}{\beta_j d''\left(\beta_j\right)} \text{ if } d''\left(\beta_j\right) > 0,$$

or

$$\frac{B_k}{\beta_j} > 1 + \frac{d'\left(\beta_j\right)}{\beta_j d''\left(\beta_j\right)} \text{ if } d''\left(\beta_j\right) < 0. \tag{A40}$$

Thus, if all three conditions (4.3), (A39) and (A40) hold, then the log-likelihood $\ell^{\text{post}}(v, w) = \ell^{\text{cond}}(v, w) + \ell^{\text{prior}}(v, w)$ is strictly convex on R_+^{K+J}.

Proof of Theorem 4.4

Proof of Lemma 4.3. Recall from the proof of Lemma 4.1 that, when (4.3) holds, $Q^{\text{cond}}(v, w) \leq 0$ for all selections of v, w with equality if and only if (A38) holds. Now, for $v = w = 1$, $Q^{\text{cond}}(v, w)$ 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 (A38) holds.

Consider the interpretation of (A38). 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, (A38) 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^{cond} is zero in any direction in which all μ_{kj} are invariant. This is obvious from the fact that derivatives of ℓ^{cond} depend on only μ_{kj} (in fact, by this reasoning, the first derivative of ℓ^{cond} is also zero in this direction). The second derivative is strictly negative in any other direction.

By compactness of R_+^{K+J}, it contains at least one maximum of log-likelihood ℓ^{cond} (just as in the proof of Theorem 3.3). Moreover, convexity over compact R_+^{K+J} also guarantees uniqueness of the values $\{\mu_{kj}\}$ at which the maximum occurs.

This maximum will **not** correspond to a unique point $(\alpha_1, \ldots, \alpha_K, \beta_1, \ldots, \beta_J)$. 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 (E3)(c), namely $\gamma = \sum_{j=1}^{J} \beta_j^*$.

https://doi.org/10.1017/asb.2016.23 Published online by Cambridge University Press

In summary, when and only when (4.3) holds, R_+^{K+J} contains a unique point $\{\mu_{kj}\}$ that maximises the log-likelihood ℓ^{cond}, and this corresponds to a unique point $(\alpha_1, \ldots, \alpha_K, \beta_1, \ldots, \beta_J)$ that satisfies (E3)(c). ■

Proof of results in Section 4.2

Proof of Corollary 4.5. It follows from (2.10) that

$$d'(\mu) = \mu^{1-p} > 0, \tag{A41}$$

$$d''(\mu) = -(p-1)\,\mu^{-p} \leq 0, \tag{A42}$$

whence (4.1) and (4.2) are immediately converted to (4.4) and (4.5). In addition, condition (4.3) becomes

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

A necessary and sufficient condition for this is (4.6). ■

Proof of Theorem 4.10

Lemma A.3. *Suppose the array $\mathfrak{D}$ is regular and contains exactly $K+J-1$ observations, subject to the non-Bayesian EDF cross-classified model defined by (E1)–(E3). 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 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 (E3)(c).

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

$$\beta_{j+1}/\beta_j = Y_{k,j+1}/Y_{kj.} \tag{A43}$$

For fixed j, the estimator on the right exists for unique k, according to the following argument. If it were not the case, one could find $k_1 \neq k$ such that the foursome of observations $Y_{kj}, Y_{k,j+1}, Y_{k_1 j}, Y_{k_1,j+1}$ exists. That is, four observations would be concentrated in two rows and two 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 four observations named above occur as the top left 2×2 sub-array of $\mathfrak{D}$ so that $\mathfrak{D}$ takes the form

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

https://doi.org/10.1017/asb.2016.23 Published online by Cambridge University Press

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 $j = 1, \ldots, J-1$. This, together with the initial assumption that $\beta_1 = 1$, leads to calculation of all β_j, $j = 1, \ldots, J$. These may now be re-scaled in accordance with (E3)(c). 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

$$\alpha_k = Y_{kj}/\beta_j. \tag{A44}$$

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 $\{Y_{r_1s_1}, \ldots, Y_{r_ms_m}\}$, 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

$$\frac{Y_{kj_1}}{Y_{kj}} = \frac{Y_{r_2s_2}}{Y_{r_1s_1}} \cdots \frac{Y_{r_ms_m}}{Y_{r_{m-1}s_{m-1}}}. \tag{A45}$$

Each ratio on the right side of (A45) 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 (A45) 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 (A45) 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.4. The result of Lemma A.3 applies more generally than most results used in the proof of Theorem 4.10. Although that corollary relates to the Tweedie cross-classified model with $p > 2$, Lemma A.3 applies to the more general non-Bayesian EDF cross-classified model.

Lemma A.5. *In the case of a non-Bayesian Tweedie cross-classified model applied to a trapezoidal array, and subject to the multiplicative weights* $\phi_{kj} = (v_k w_j)^{-1}$ *the following relations hold:*

$$\frac{\beta_{s+1}^{2-p} w_{s+1}}{\sum_{j=1}^{s} \beta_j^{2-p} w_j} = \frac{\sum_{k=1}^{K-s} Y_{k,s+1}\mu_{k,s+1}^{1-p}/\phi_{k,s+1}}{\sum_{j=1}^{s}\sum_{k=1}^{K-s} Y_{kj}\mu_{kj}^{1-p}/\phi_{kj}}, s = 1, 2, \ldots, J-1, \tag{A46}$$

$$\frac{\alpha_{r+1}^{2-p} v_{r+1}}{\sum_{k=1}^{r} \alpha_k^{2-p} v_k} = \frac{\sum_{j=1}^{J_{r+1}} Y_{r+1,j}\mu_{r+1,j}^{1-p}/\phi_{r+1,j}}{\sum_{k=1}^{r}\sum_{j=1}^{J_{r+1}} Y_{kj}\mu_{kj}^{1-p}/\phi_{kj}}, r = 1, 2, \ldots, K-1. \tag{A47}$$

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

https://doi.org/10.1017/asb.2016.23 Published online by Cambridge University Press

Proof. Recall (3.15) and (3.16) with all $\psi_k^{(\alpha)}$, $\psi_j^{(\beta)}$ set to infinity to obtain the MLE equations for the non-Bayesian EDF cross-classified model, reproduced here in a slightly modified form:

$$\sum_{j\in\mathcal{R}(k)} \left[Y_{kj} - \mu_{kj}\right] c'\left(\mu_{kj}\right) \mu_{kj}/\phi_{kj} = 0, \tag{A48}$$

$$\sum_{k\in\mathcal{C}(j)} \left[Y_{kj} - \mu_{kj}\right] c'\left(\mu_{kj}\right) \mu_{kj}/\phi_{kj} = 0. \tag{A49}$$

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

$$\sum_{(k,j)\in\mathfrak{D}} \left[Y_{kj} - \mu_{kj}\right] c'\left(\mu_{kj}\right) \mu_{kj}/\phi_{kj} = \sum_k \sum_{j\in\mathcal{R}(k)} \left[Y_{kj} - \mu_{kj}\right] c'\left(\mu_{kj}\right) \mu_{kj}/\phi_{kj}$$
$$= \sum_j \sum_{k\in\mathcal{C}(j)} \left[Y_{kj} - \mu_{kj}\right] c'\left(\mu_{kj}\right) \mu_{kj}/\phi_{kj} = 0. \tag{A50}$$

It is evident that, if one deletes any number of complete rows and columns from $\mathfrak{D}$, the sum in (A50) remains zero. If columns beyond the s th, and rows beyond the $(K-s)$ -th, are deleted, the result is

$$\sum_{k=1}^{K-s}\sum_{j=1}^{s} \left[Y_{kj} - \mu_{kj}\right] c'\left(\mu_{kj}\right) \mu_{kj}/\phi_{kj} = 0. \tag{A51}$$

Now express (A49) for the $(s+1)$ -th column with explicit summation limits

$$\sum_{k=1}^{K-s} \left[Y_{k,s+1} - \mu_{k,s+1}\right] c'\left(\mu_{k,s+1}\right) \mu_{k,s+1}/\phi_{k,s+1} = 0. \tag{A52}$$

For the special case of the Tweedie cross-classified model, $c(.)$ is given by (2.9). With this substitution, with the special form of the weights recognised, and slight rearrangement, (A51) and (A52) become (for fixed s):

$$\frac{\sum_{k=1}^{K-s} Y_{k,s+1}\mu_{k,s+1}^{1-p}/\phi_{k,s+1}}{\sum_{j=1}^{s}\sum_{k=1}^{K-s} Y_{kj}\mu_{kj}^{1-p}/\phi_{kj}} = \frac{\sum_{k=1}^{K-s} \alpha_k^{2-p}\beta_{s+1}^{2-p} v_k w_{s+1}}{\sum_{j=1}^{s}\sum_{k=1}^{K-s} \alpha_k^{2-p}\beta_j^{2-p} v_k w_j}. \tag{A53}$$

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

$$\frac{\sum_{k=1}^{K-s} Y_{k,s+1}\mu_{k,s+1}^{1-p}/\phi_{k,s+1}}{\sum_{j=1}^{s}\sum_{k=1}^{K-s} Y_{kj}\mu_{kj}^{1-p}/\phi_{kj}} = \frac{\beta_{s+1}^{2-p} w_{s+1}\sum_{k=1}^{K-s} \alpha_k^{2-p} v_k}{\left[\sum_{j=1}^{s}\beta_j^{2-p} w_j\right]\left[\sum_{k=1}^{K-s}\alpha_k^{2-p} v_k\right]}, \tag{A54}$$

from which (A46) follows. The proof of (A47) is essentially the same with the roles of rows and columns interchanged. ■

Lemma A.6. *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 non-Bayesian 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.3 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*

https://doi.org/10.1017/asb.2016.23 Published online by Cambridge University Press

to the same multiplicative system of weights. Then, the following relations hold for any $s = 1, 2, \ldots, J_k - 1$*:*

$$[\xi(\mathfrak{D})]^{-1} \frac{\sum_{j=1}^{s} \beta_j^* \beta_j^{1-p} w_j}{\beta_{s+1}^* \beta_{s+1}^{1-p} w_{J_k}} \leq \frac{\sum_{j=1}^{s} \beta_j^{2-p} w_j}{\beta_{s+1}^{2-p} w_{s+1}} \leq \xi(\mathfrak{D}) \frac{\sum_{j=1}^{s} \beta_j^* \beta_j^{1-p} w_j}{\beta_{s+1}^* \beta_{s+1}^{1-p} w_{s+1}}, \tag{A55}$$

$$[\xi(\mathfrak{D})]^{-1} \frac{\sum_{j=1}^{s} \beta_j^* \beta_j^{1-p} w_j}{\sum_{j=1}^{s+1} \beta_j^* \beta_j^{1-p} w_j} \leq \frac{\sum_{j=1}^{s} \beta_j^{2-p} w_j}{\sum_{j=1}^{s+1} \beta_j^{2-p} w_j} \leq \xi(\mathfrak{D}) \frac{\sum_{j=1}^{s} \beta_j^* \beta_j^{1-p} w_j}{\sum_{j=1}^{s+1} \beta_j^* \beta_j^{1-p} w_j}. \tag{A56}$$

Proof. Select an arbitrary $Y_{rs} \notin \mathcal{S}$, and a path $\gamma_{r,j_r:k_s,s}$ from Y_{rj_r} to $Y_{k_s s}$, as defined in Section 4.3. It will be assumed that $s > j_r$; a similar argument deals with the case $s < j_r$. The path $\gamma_{r,j_r:k_s,s}$ involves all the observations appearing in (4.7), and so $\pi_{r,j_r:k_s,s}$ is defined.

By definition of $\gamma_{r,j_r:k_s,s}$ in Section 2

$$Y_{rs} = \left(1 + \pi_{r,j_r:k_s,s}\right) Y_{rj_r} \prod_{\eta_{r,j_r:k_s,s}} \frac{Y_{r_{i+1}s_{i+1}}}{Y_{r_i s_i}}. \tag{A57}$$

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 has been shown in the proof of Lemma A.3 that, in this case, $Y_{r_i s_{i+1}}/Y_{r_i,s_i} = \beta_{s_i+1}^*/\beta_{s_i}^*$ or $\beta_{s_{i-1}}^*/\beta_{s_i}^*$, with s_i running from j_r to s. Thus, (A57) simplifies to the following:

$$Y_{rs} = \left(1 + \pi_{r,j_r:k_s,s}\right) Y_{rj_r} \frac{\beta_s^*}{\beta_{j_r}^*} = \left(1 + \pi_{r,j_r:k_s,s}\right) \alpha_r^* \beta_s^*. \tag{A58}$$

Alternatively, if $Y_{rs} \in \mathcal{S}$, then immediately $Y_{rs} = \alpha_r^* \beta_s^*$, and then (A58) holds if one adopts the convention $\pi_{r,j_r:k_s,s} = 0$ for Y_{rj_r}, $Y_{k_s s} \in \mathcal{S}$. With this convention, (A58) holds for all $Y_{rs} \in \mathfrak{D}$.

By (A46)

$$\begin{aligned}
\frac{\sum_{j=1}^{s} \beta_j^{2-p} w_j}{\beta_{s+1}^{2-p} w_{s+1}} &= \frac{\sum_{j=1}^{s} \sum_{k=1}^{K-s} Y_{kj} \mu_{kj}^{1-p}/\phi_{kj}}{\sum_{k=1}^{K-s} Y_{k,s+1} \mu_{k,s+1}^{1-p}/\phi_{k,s+1}} \\
&= \frac{\sum_{j=1}^{s} \sum_{k=1}^{K-s} \left(1 + \pi_{k,j_k:k_j,j}\right) \alpha_k^* \beta_j^* \alpha_k^{1-p} \beta_j^{1-p} v_k w_j}{\sum_{k=1}^{K-s} \left(1 + \pi_{k,j_k:k_{s+1},\, s+1}\right) \alpha_k^* \beta_{s+1}^* \alpha_k^{1-p} \beta_{s+1}^{1-p} v_k w_{s+1}},
\end{aligned} \tag{A59}$$

[by (A58)]

$$\leq \xi(\mathfrak{D}) \frac{\sum_{j=1}^{s} \sum_{k=1}^{K-s} \alpha_k^* \beta_j^* \alpha_k^{1-p} \beta_j^{1-p} v_k w_j}{\sum_{k=1}^{K-s} \alpha_k^* \beta_{s+1}^* \alpha_k^{1-p} \beta_{s+1}^{1-p} v_k w_{s+1}},$$

[by definition of $\xi(\mathfrak{D})$]

$$= \xi(\mathfrak{D}) \frac{\left[\sum_{k=1}^{K-s} \alpha_k^* \alpha_k^{1-p} v_k\right] \left[\sum_{j=1}^{s} \beta_j^* \beta_j^{1-p} w_j\right]}{\left[\sum_{k=1}^{K-s} \alpha_k^* \alpha_k^{1-p} v_k\right] \left[\beta_{s+1}^* \beta_{s+1}^{1-p} w_{s+1}\right]} = \xi(\mathfrak{D}) \frac{\sum_{j=1}^{s} \beta_j^* \beta_j^{1-p} w_j}{\beta_{s+1}^* \beta_{s+1}^{1-p} w_{s+1}}. \tag{A60}$$

A parallel argument produces a corresponding inequality in the opposite direction so that (A55) holds.

In order to prove (A56), note that (A46) leads easily to the following relation, corresponding to the reciprocal of (A59) above:

$$\frac{\sum_{j=1}^{s+1}\beta_j^{2-p}w_j}{\sum_{j=1}^{s}\beta_j^{2-p}w_j}=1+\frac{\beta_{s+1}^{2-p}w_{s+1}}{\sum_{j=1}^{s}\beta_j^{2-p}w_j}=\frac{\sum_{j=1}^{s+1}\sum_{k=1}^{K-s}Y_{kj}\mu_{kj}^{1-p}/\phi_{kj}}{\sum_{j=1}^{s}\sum_{k=1}^{K-s}Y_{kj}\mu_{kj}^{1-p}/\phi_{kj}}. \tag{A61}$$

An argument may now be applied precisely parallel to that which led from (A59) to (A60), and (A56) follows. ■

Proof of Theorem 4.10.

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.3, a unique solution exists, and it is the proportional solution $Y_{kj}=\alpha_k\beta_j$. Thus, $\alpha_k=\alpha_k^*$, $\beta_j=\beta_j^*$.

As remarked just after (4.9), $\xi(\mathfrak{D})=1$ in this case. It follows that (4.11) is satisfied.

Case II. This is the case of Lemma A.6, in whose proof (A58) is established. Substitute this into (A48) to obtain

$$\sum_{s\in\mathcal{R}(r)}\left[\left(1+\pi_{r,j_r:k_s,s}\right)\alpha_r^*\beta_s^*-\alpha_r\beta_s\right]\mu_{rs}c'\left(\mu_{rs}\right)/\phi_{rs}=0. \tag{A62}$$

Now substitute (2.9) into (A62), recognise the special case of the weights, $\phi_{kj}^{-1}=v_kw_j$, and rearrange slightly, to obtain

$$\sum_{s\in\mathcal{R}(r)}\left[\left(1+\pi_{r,j_r:k_s,s}\right)\beta_s^*-\frac{\alpha_r}{\alpha_r^*}\beta_s\right]\beta_s^{1-p}w_s=0. \tag{A63}$$

It follows that

$$\frac{\alpha_r}{\alpha_r^*}=\frac{\sum_{j\in\mathcal{R}(r)}\left(1+\pi_{r,j_r:k_s,s}\right)\beta_j^*\beta_j^{1-p}w_j}{\sum_{j\in\mathcal{R}(r)}\beta_j^{2-p}w_j},$$

and then

$$\frac{\alpha_r\beta_s}{\alpha_r^*\beta_s^*}=\frac{\sum_{j\in\mathcal{R}(r)}\left(1+\pi_{r,j_r:k_s,s}\right)\beta_j^*\beta_j^{1-p}w_j}{\sum_{j\in\mathcal{R}(r)}\beta_j^{2-p}w_j}\times\frac{\beta_s}{\beta_s^*}\leq(1+\bar{\pi})\frac{\sum_{j\in\mathcal{R}(r)}\beta_j^*\beta_j^{1-p}w_j}{\sum_{j\in\mathcal{R}(r)}\beta_j^{2-p}w_j}\times\frac{\beta_s}{\beta_s^*}, \tag{A64}$$

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

For the trapezoidal array under consideration, the maximum value of j for which Y_{kj} exists in row k is defined in Lemma A.5 as J_k. It will be convenient to re-express (A64) in the form

$$\frac{\alpha_k\beta_s}{\alpha_k^*\beta_s^*}\leq(1+\bar{\pi})\left[1+\frac{\sum_{j=1}^{J_k-1}\beta_j^*\beta_j^{1-p}w_j}{\beta_{J_k}^*\beta_{J_k}^{1-p}w_{J_k}}\right]\Big/\left[1+\frac{\sum_{j=1}^{J_k-1}\beta_j^{2-p}w_j}{\beta_{J_k}^{2-p}w_{J_k}}\right]\times\frac{\beta_{J_k}^*}{\beta_s^*}\Big/\frac{\beta_{J_k}}{\beta_s}. \tag{A65}$$

https://doi.org/10.1017/asb.2016.23 Published online by Cambridge University Press

But Lemma A.6 yields the following for the special case $s = J_k - 1$:

$$\xi^{-1} \le \frac{\sum_{j=1}^{J_k-1} \beta_j^* \beta_j^{1-p} w_j}{\beta_{J_k}^* \beta_{J_k}^{1-p} w_{J_k}} \Big/ \frac{\sum_{j=1}^{J_k-1} \beta_j^{2-p} w_j}{\beta_{J_k}^{2-p} w_{J_k}} \le \xi, \tag{A66}$$

where, for brevity, the explicit dependence of ξ on $\mathfrak{D}$ has been temporarily omitted.

It follows that the ratio of square bracketed terms in (A65) is subject to the same bounds ξ^{-1} and ξ (*a fortiori*), and so (A65), together with the simple extension to a two-sided inequality, yields

$$\left(1 + \underline{\pi}\right) \xi^{-1} \frac{\beta_{J_k}^*}{\beta_s^*} \Big/ \frac{\beta_{J_k}}{\beta_s} \le \frac{\alpha_k \beta_s}{\alpha_k^* \beta_s^*} \le \left(1 + \bar{\pi}\right) \xi \frac{\beta_{J_k}^*}{\beta_s^*} \Big/ \frac{\beta_{J_k}}{\beta_s}, \tag{A67}$$

where, again for brevity, the explicit dependence of $\underline{\pi}, \bar{\pi}$ on $\mathfrak{D}$ has been temporarily omitted.

Inequalities can be placed on the ratio that appears on left and right of this result, as follows. There are three cases to be considered: $s < J_k$, $s = J_k$ and $s > J_k$, respectively.

Case II(a): $s = J_k$ In this case (A67) reduces trivially to the following:

$$\left(1 + \underline{\pi}\right) \xi^{-1} \le \frac{\alpha_k \beta_s}{\alpha_k^* \beta_s^*} \le \left(1 + \bar{\pi}\right) \xi. \tag{A68}$$

Case II(b): $s < J_k$ Commence with the observation that

$$\begin{aligned}\frac{\beta_{J_k}^{2-p} w_{J_k}}{\beta_s^{2-p} w_s} &= \left(\frac{\sum_{j=1}^{s} \beta_j^{2-p} w_j}{\beta_s^{2-p} w_s}\right)\left(\frac{\sum_{j=1}^{s+1} \beta_j^{2-p} w_j}{\sum_{j=1}^{s} \beta_j^{2-p} w_j}\right) \cdots \left(\frac{\sum_{j=1}^{J_k-1} \beta_j^{2-p} w_j}{\sum_{j=1}^{J_k-2} \beta_j^{2-p} w_j}\right)\left(\frac{\beta_{J_k}^{2-p} w_{J_k}}{\sum_{j=1}^{J_k-1} \beta_j^{2-p} w_j}\right) \\ &= \left(1 + \frac{\sum_{j=1}^{s-1} \beta_j^{2-p} w_j}{\beta_s^{2-p} w_s}\right)\left(\frac{\sum_{j=1}^{s+1} \beta_j^{2-p} w_j}{\sum_{j=1}^{s} \beta_j^{2-p} w_j}\right) \cdots \left(\frac{\sum_{j=1}^{J_k-1} \beta_j^{2-p} w_j}{\sum_{j=1}^{J_k-2} \beta_j^{2-p} w_j}\right)\left(\frac{\beta_{J_k}^{2-p} w_{J_k}}{\sum_{j=1}^{J_k-1} \beta_j^{2-p} w_j}\right).\end{aligned} \tag{A69}$$

All of the ratios on the right are of the form (subject to reciprocation) of the subject quantities in inequalities (A55) and (A56). It, therefore, follows from these inequalities that

$$\begin{aligned}\frac{\beta_{J_k}^{2-p} w_{J_k}}{\beta_s^{2-p} w_s} \le \xi^{J_k-s} &\left(1 + \xi \frac{\sum_{j=1}^{s-1} \beta_j^* \beta_j^{1-p} w_j}{\beta_s^* \beta_s^{1-p} w_s}\right)\left(\frac{\sum_{j=1}^{s+1} \beta_j^* \beta_j^{1-p} w_j}{\sum_{j=1}^{s} \beta_j^* \beta_j^{1-p} w_j}\right) \cdots\cdots \left(\frac{\sum_{j=1}^{J_k-1} \beta_j^* \beta_j^{1-p} w_j}{\sum_{j=1}^{J_k-2} \beta_j^* \beta_j^{1-p} w_j}\right) \\ \left(\frac{\beta_{J_k}^* \beta_{J_k}^{1-p} w_{J_k}}{\sum_{j=1}^{J_k-1} \beta_j^* \beta_j^{1-p} w_j}\right) &\le \xi^{J_k-s+1} \left(1 + \frac{\sum_{j=1}^{s-1} \beta_j^* \beta_j^{1-p} w_j}{\beta_s^* \beta_s^{1-p} w_s}\right)\left(\frac{\sum_{j=1}^{s+1} \beta_j^* \beta_j^{1-p} w_j}{\sum_{j=1}^{s} \beta_j^* \beta_j^{1-p} w_j}\right) \cdots\cdots \left(\frac{\sum_{j=1}^{J_k-1} \beta_j^* \beta_j^{1-p} w_j}{\sum_{j=1}^{J_k-2} \beta_j^* \beta_j^{1-p} w_j}\right) \\ \left(\frac{\beta_{J_k}^* \beta_{J_k}^{1-p} w_{J_k}}{\sum_{j=1}^{J_k-1} \beta_j^* \beta_j^{1-p} w_j}\right) &= \xi^{J_k-s+1} \frac{\beta_{J_k}^* \beta_{J_k}^{1-p} w_{J_k}}{\beta_s^* \beta_s^{1-p} w_s}.\end{aligned} \tag{A70}$$

Then

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

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

$$\xi^{-(J_k-s+1)} \le \frac{\beta_{J_k}}{\beta_s} \Big/ \frac{\beta_{J_k}^*}{\beta_s^*} \le \xi^{J_k-s+1}. \tag{A71}$$

Substitution of (A71) into (A67) yields

$$\left(1 + \underline{\pi}\right) \xi^{-(J_k-s+2)} \le \frac{\alpha_k \beta_s}{\alpha_k^* \beta_s^*} \le \left(1 + \bar{\pi}\right) \xi^{J_k-s+2}. \tag{A72}$$

https://doi.org/10.1017/asb.2016.23 Published online by Cambridge University Press

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 (A69):

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

and then continues in parallel with Case II(b), leading ultimately to the following result in place of (A72):

$$[1 + \underline{\pi}(\mathfrak{D})][\xi(\mathfrak{D})]^{-(|J_k - s|+2)} \le \frac{\alpha_k \beta_s}{\alpha_k^* \beta_s^*} \le [1 + \bar{\pi}(\mathfrak{D})][\xi(\mathfrak{D})]^{|J_k - s|+2} \text{ for all } k, s. \tag{A73}$$

Proof of Corollary 4.11

Use (A58) to express $Y_{kj}/\bar{\mu}_{kj}$ in the form

$$\frac{Y_{kj}}{\bar{\mu}_{kj}} = \frac{\left(1 + \pi_{r,j_r:k_s,s}\right)\alpha_k^* \beta_j^*}{\bar{\mu}_{kj}}. \tag{A74}$$

Now

$$1 - \underline{\pi} \le 1 + \pi_{r,j_r:k_s,s} \le 1 + \bar{\pi} \qquad [\text{by definition of } \underline{\pi}, \bar{\pi}], \tag{A75}$$

so that

$$\left(1 - \underline{\pi}\right)\frac{\alpha_k^* \beta_j^*}{\bar{\mu}_{kj}} \le \frac{Y_{kj}}{\bar{\mu}_{kj}} \le (1 + \bar{\pi})\frac{\alpha_k^* \beta_j^*}{\bar{\mu}_{kj}}. \tag{A76}$$

By Theorem 4.10

$$\alpha_k^* \beta_j^* \left(1 - \underline{\pi}\right)\xi^{-(|J_k - j|+2)} \le \alpha_k \beta_j \le \alpha_k^* \beta_j^* (1 + \bar{\pi})\xi^{|J_k - j|+2}, \tag{A77}$$

so set

$$\bar{\mu}_{kj} = \alpha_k^* \beta_j^* (1 + \bar{\pi})\xi^{|J_k - j|+2}, \ \underline{\mu}_{kj} = \alpha_k^* \beta_j^* \left(1 + \underline{\pi}\right)\xi^{-(|J_k - j|+2)}, \tag{A78}$$

to convert (A77) to the condition $\underline{\mu}_{kj} < \mu_{kj} < \bar{\mu}_{kj}$ required in Theorem 4.4.

Then substitution of (A78) into (A76) yields

$$\frac{1 - \underline{\pi}}{1 + \bar{\pi}}\xi^{-(|J_k - j|+2)} \le \frac{Y_{kj}}{\bar{\mu}_{kj}} \le \xi^{-(|J_k - j|+2)}.$$

By definition of ξ in (4.9), the first ratio on the left may be replaced by ξ^{-1} and, since $1 \le J_k, j \le J$,

$$\xi^{-(J+2)} \le \frac{Y_{kj}}{\bar{\mu}_{kj}} \le \xi^{-2}. \tag{A79}$$

If condition (4.12) holds, then combination of it with (A79) yields condition (4.6).

Now under the conditions of Corollary 4.11, those of Corollary 4.6 also hold, in which case (4.6) is a sufficient condition for a unique MLE of the model parameters. It then follows that, in the present case, (4.12) is a sufficient condition, and Corollary 4.11 is proved. ■

Proof of Theorem 4.13

The proofs of both Corollaries 4.11 and 4.12 proceed by establishing the necessary convexity property of the log-likelihood ℓ^{cond}. As was seen in the proof of Lemma 4.3, ℓ^{cond} is strictly convex except in any direction in which all μ_{kj} are invariant, where ℓ^{cond} is also invariant.

It was also shown in the proof of Lemma 4.1 that ℓ^{prior} is strictly convex. Therefore, $\ell^{\text{post}} = \ell^{\text{cond}} + \ell^{\text{prior}}$ is strictly convex. ■

https://doi.org/10.1017/asb.2016.23 Published online by Cambridge University Press