A HOMOTOPY CLASS OF SEMI-RECURSIVE CHAIN LADDER MODELS

Greg Taylor

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

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

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

September 2011

Abstract

The chain ladder algorithm is known to produce maximum likelihood estimates of the parameters of certain recursive and non-recursive models. These types of models represent two extremes of dependency within rows of a data array.

Whereas observations within a row of a non-recursive model are stochastically independent, each observation of a recursive model is, in expectation, directly proportional to the immediately preceding observation from the same row. The correlation structures of forecasts also differ as between recursive and non-recursive models.

The present paper constructs a family of models that forms a bridge between recursive and non-recursive models and so provides a continuum of intermediate cases in terms of dependency structure. The intermediate models are called semi-recursive.

The statistical inference properties of semi-recursive models are investigated. It is found (Section 5.4) that the chain ladder algorithm is also maximum likelihood for semi-recursive models.

Sufficient, and minimally sufficient, statistics are found for the semi-recursive model (Section 6). They are found to be the same as for non-recursive models. The minimally sufficient statistic is complete, leading to minimum variance unbiased estimation (Section 7).

Keywords: chain ladder, correlation, non-recursive model, recursive model, minimally sufficient statistic, minimum variance unbiased estimation, ODP cross-classified model, ODP Mack model, semi-recursive model, sufficient statistic,

1. Introduction

The actuarial literature identifies two families of chain ladder models categorised by Verrall (2000) as **recursive** and **non-recursive** models respectively. Although the model formulations are fundamentally different, both are found to yield the same maximum likelihood estimators of age-to-age factors and the same forecasts of loss reserve. The properties of these models are studied by Taylor (2011a).

Whereas observations within a row of a non-recursive model are stochastically independent, each observation of a recursive model is, in expectation, directly proportional to the immediately preceding observation from the same row. It would be useful to define a family of models that forms a bridge between these two extreme cases of dependency, i.e. where a relation between consecutive observations in a row exists but is less than linear (in expectation).

Further, distinct forecasts within a row of a run-off array are known to be correlated differently under recursive and non-recursive models (Taylor, 2011b). It would be useful to define a family of models displaying intermediate correlations.

The purpose of the present paper is to define just such a family and explore its statistical inference properties.

2. Framework and notation

2.1 Claims data

Consider a K x J rectangle of claims observations Y_{kj} with:

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

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

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

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

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

Also let

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

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

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

Define the **cumulative row sums**

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

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

$$H_k = \sum_{j=1}^{\min(J, K-k+1)} Y_{kj}$$

$$V_j = \sum_{k=1}^{K-j+1} Y_{kj} \tag{2.2}$$

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

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

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

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

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

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

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

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

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

For a random variable A_{kj} with $(k,j) \in \mathcal{D}_K, A_K$ will denote the entire array $\{A_{kj} : (k,j) \in \mathcal{D}_K\}$. The first column $(j=1)$ of A_K will be denoted by A_K^1.

2.2 Families of distributions

2.2.1 Exponential dispersion family

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

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

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

For Y so distributed,

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

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

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

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

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

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

2.2.2 Tweedie family

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

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

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

For this family,

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

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

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

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

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

2.2.3 Over-dispersed Poisson family

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

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

By (2.8) – (2.10),

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

By (2.14),

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

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

3. Chain ladder models

3.1 Heuristic chain ladder

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

Define the following factors:

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

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

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

with

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

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

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

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

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

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

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

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

3.2 Recursive models

A recursive model takes the general form

$$E\left[X_{k,j+1} \mid X_{kj}\right] = \text{function of } \mathcal{D}_{k+j-1} \text{ and some parameters} \tag{3.8}$$

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

3.2.1 Mack model

The Mack model (Mack, 1993) is defined by the following assumptions.

(M1) Accident periods are stochastically independent, i.e. $Y_{k_1 j_1}, Y_{k_2 j_2}$ are stochastically independent if $k_1 \neq k_2$.

(M2) For each k = 1, 2, …, K, the X_{kj} (j varying) form a Markov chain.

(M3) For each k = 1, 2, …, K and j = 1, 2, …, J – 1,
(a) $E\left[X_{k,j+1} \mid X_{kj}\right] = f_j X_{kj}$ for some parameters $f_j > 0$; and
(b) $Var\left[X_{k,j+1} \middle| X_{kj}\right] = \sigma_j^2 X_{kj}$ for some parameters $\sigma_j^2 > 0$.

3.2.2 ODP Mack model

Taylor (2011) defined the over–dispersed Poisson (ODP) Mack model as that satisfying assumptions (M1), (M2) and

$(ODPM3)$ For each k=1,2,…,K and j=1,2,…,J-1,

$$Y_{k,j+1} \mid X_{kj} \sim ODP\left(\left(f_j - 1\right) X_{kj}, \phi_{k,j+1}\right)$$

where now $f_j \geq 1$.

Assumption (ODPM3) implies (M3a). Moreover, in the special case $\phi_{k,j+1} = \phi_{j+1}$ independent of k, (ODPM3) also implies (M3b) with $\sigma_j^2 = \phi_{j+1}\left(f_j - 1\right)$.

It is evident that, for this model to be valid, it is necessary that all $Y_{k,j} \geq 0$. Note also that, under $(ODPM3)$, $X_{kj} = 0$ implies that $X_{k,j+m} = 0$ for all $m > 0$. This means that, for each k, either $Y_{k1} > 0$ or $X_{kj} = 0$ for all j.

A summary of these requirements in terms of the data array $\mathcal{D}_K$ is as follows.
(R1) $Y_{kj} \geq 0$ for all $Y_{kj} \in \mathcal{D}_K$
(R2) For each k = 1, 2, …, K, either:

(a) $Y_{k1} > 0$; or
(b) $Y_{kj} = 0$ for all $1 \le j \le \min(J, K-k+1)$

A data array satisfying these requirements will be called **ODPM-regular**.

Assumption (ODPM3) may be expressed in the following form, suitable for GLM implementation of the OPD Mack model:

$$Y_{k,j+1} \mid X_{kj} \sim ODP\left(\exp\left[\ln X_{kj} + \ln\left(f_j - 1\right)\right], \phi / w_{k,j+1}\right) \tag{3.9}$$

where

$$w_{k,j+1} = \phi / \phi_{k,j+1} \tag{3.10}$$

In this form, the GLM of the $Y_{k,j+1}$ has log link, offsets $\ln X_{kj}$, parameters $\ln\left(f_j - 1\right)$, and weights $w_{k,j+1}$.

3.3 Non-recursive models

Taylor (2011) also defined the **ODP cross-classified model** as that satisfying the following assumptions:

(ODPCC1) The random variables $Y_{kj} \in \mathcal{D}_K^+$ are stochastically independent.

(ODPCC2) For each k = 1, 2,…, K and j = 1, 2, …, J,
(a) $Y_{kj} \sim ODP\left(\mu_{kj}, \phi_{kj}\right)$;
(b) $\mu_{kj} = \alpha_k \beta_j$ for some parameters $a_k, \beta_j > 0$; and
(c) $\sum_{j=1}^{J} \beta_j = 1$

Assumption (ODPCC2b) may be expressed in the following form, suitable for GLM implementation of the ODP cross-classified model:

$$Y_{kj} \sim ODP\left(\exp\left(\ln \alpha_k + \ln \beta_j\right), \phi / w_{kj}\right) \tag{3.11}$$

In this form, the GLM of the Y_{kj} has log link, parameters $\ln \alpha_k$ and $\ln \beta_j$, and weights w_{kj} satisfying

$$w_{kj} = \phi / \phi_{kj} \tag{3.12}$$

Assumption (ODPCC2b) removes one degree of redundancy from the parameter set, and would be reflected by the aliasing of one parameter in the GLM.

3.4 Semi-recursive models

First, a definition of homotopy is given. Let A and B be topological spaces and let $\xi : A \rightarrow B$ and $\eta : A \rightarrow B$ be continuous. A **homotopy** is a continuous function $H : A \times [0,1] \rightarrow B$ such that, for $a \in A$, $H(a,0) = \xi(a)$ and $H(a,1) = \eta(a)$. The collection of functions $\mathcal{H} = \{H(\cdot,t); t \in [0,1]\}$ will be referred to as the **homotopy class** associated with the homotopy just defined.

Consider a model that satisfies assumptions (M1), (M2) and the following:

(ODPSR3a) For each $k = 1,2,...,k$, and for some λ independent of k and j, subject to $0 \leq \lambda \leq 1$,

$$Y_{k1} \sim ODP\left((\alpha_k \beta_1)^{1-\lambda}, \phi_{k1}\right)$$

for some parameters $\alpha_k \geq 0, \beta_1 > 0$.

(ODPSR3b) For each $k = 1,2,...,K$, and $j = 1,2,...,J-1$,

$$Y_{k,j+1} \big| X_{kj} \sim ODP\left(\mathrm{X}_{kj}^{\lambda}\, \alpha_k^{1-\lambda} \psi_{j+1}, \phi_{k,j+1}\right)$$

for the same λ as in (ODPSR3a), and where $\psi_j, j = 2,3,...,J$ are parameters subject to $\psi_j \geq 0$.

By convention,

$$\begin{aligned} 0^{\lambda} &= 1, \text{ when } \lambda = 0 \\ &= 0, \text{ when } \lambda > 0 \end{aligned}$$

Such a model will be called **OPD semi-recursive**. It is valid only for a non-negative data arrays and, in the case $0 < \lambda \leq 1$, for ODPM-regular arrays. It will be assumed hence forth that all $\mathcal{D}_K$ satisfy these requirements.

Assumptions (ODPSR3a-b) may be expressed in the following form, suitable for GLM implementation of the ODP semi-recursive model:

$$Y_{k1} \sim ODP\left(\exp\left((1-\lambda)\ln \alpha_k + (1-\lambda)\ln \beta_1,\right), \phi / w_{k1}\right) \tag{3.13}$$

$$Y_{k,j+1} \big| X_{kj} \sim ODP\left(\exp\left(\lambda \ln \mathrm{X}_{kj} + (1-\lambda)\ln \alpha_k + \ln \psi_{j+1}\right), \phi / w_{k,j+1}\right) \tag{3.14}$$

with weights,

$$w_{kj} = \phi / \phi_{kj} \tag{3.15}$$

In this formulation, the terms $\lambda \ln \mathrm{X}_{kj}$, known quantities, are offsets, and the $(1-\lambda)\ln \alpha_k$ and $\ln \psi_j$ are unknown parameters requiring estimation.

ODP semi-recursive models are subject to the following representation lemma.

Lemma 3.1. The mean ODP parameter in (ODPSR3b) may be expressed in the form

$$\mathrm{X}_{kj}\alpha_k^{1-\lambda}\psi_{j+1} = \left[\mathrm{X}_{kj}\left(f_j - 1\right)\right]^{\lambda}\left(\alpha_k\beta_{j+1}\right)^{1-\lambda} \tag{3.16}$$

where, for $j = 1,2,3,...,J-1$, β_j is the unique non-negative solution of

$$\psi_j = \beta_j \Big/ \left(1 - \sum_{i=j}^{J}\beta_i\right)^{\lambda} \tag{3.17}$$

β_1 is given by the constraint

$$\sum_{i=1}^{J}\beta_i = 1 \tag{3.18}$$

and

$$f_j - 1 = \beta_{j+1} \Big/ \sum_{i=1}^{j}\beta_i, \quad j = 1,2,...,J-1 \tag{3.19}$$

The β_j are calculated recursively from (3.17) in the order $j = J,J\text{-}1,\ldots,2$.

In the event that $\sum_{1-j}^{J}\beta_j = 1$ in (3.16),

$$\beta_1 = ... = \beta_j = 0 \tag{3.20}$$

Proof. The uniqueness of $\beta_1, \beta_2, ..., \beta_J$ is first proven for given $\psi_2, \psi_3, ..., \psi_{J.}$

Consider relation (3.17) and note that

$$\partial\psi_j / \partial\beta_j = \left\{\lambda\beta_j + \left[1 - \sum_{i=j}^{J}\beta_i\right]\right\} \Big/ \left[1 - \sum_{i=j}^{J}\beta_j\right]^{\lambda+1}$$
$$> 0 \tag{3.21}$$

in the case

$$\sum_{i=j+1}^{J}\beta_i < 1 \tag{3.22}$$

Hence (3.17) has at most one solution in β_j in the case that (3.22) holds. Note also that the right side of (3.17) varies from 0 to ∞ as β_j varies from 0 to $1 - \sum_{i=j+1}^{J}\beta_i$. Hence (3.17) has a non-negative solution in β_j, which is therefore unique.

Note also that the existence of a solution to (3.17) implies that

$$\sum_{i=j}^{J}\beta_i < 1 \tag{3.23}$$

Thus (3.22) implies (3.23), and so the required recursive calculation of the β_j can proceed over $j = J, J-1, ..., 2$.

Relation (3.23) holds for $j = 2$, so define

$$\beta_1 = 1 - \sum_{i=2}^{J} \beta_i \tag{3.24}$$

Then $\beta_1 > 0$, as required by ODPRS1 and (3.18) is satisfied.

Substitution of (3.17) and (3.19) in the right side of (3.16) now yields

$$X_{kj}^{\lambda} \alpha_k^{1-\lambda} \beta_{j+1} / \left[\sum_{i=1}^{j} \beta_i \right]^{\lambda} = X_{kj}^{\lambda} \alpha_k^{1-\lambda} \beta_{j+1} / \left[1 - \sum_{i=j+1}^{J} \beta_i \right]^{\lambda}$$

by (3.18). By (3.17), this is equal to the left side of (3.16), and so (3.16) holds. □

Note that the semi-recursive models form a homotopy class. Let A be the set each of whose members consists of a data array $\mathcal{D}_K^+$ and a parameter set $\left\{ f_j, j = 1, 2, ..., J-1, \alpha_k, \beta_j, \phi_{kj}, k = 1, 2, ..., K, j = 1, 2, ..., J \right\}$. Let a be a specific member of A and let $0 \le \lambda \le 1$. Define H to be the mapping that sends (a, λ) to the distributions of Y_{k1} and $Y_{k,j+1} \big| X_{kj}$ defined by (ODPSR3a-b). Convert H to a metric space by imposing the metric

$$\mu\left(H\left(a^{(1)}, \lambda^{(1)}\right), H\left(a^{(2)}, \lambda^{(2)}\right)\right) = \sum_{k=1}^{K} \left[\left(\alpha_k^{(1)} \beta_1^{(1)}\right)^{1-\lambda^{(1)}} - \left(\alpha_k^{(2)} \beta_1^{(2)}\right)^{1-\lambda^{(2)}} \right]^2$$

$$+ \sum_{k=1}^{K} \sum_{j=1}^{J-1} \left[\left(\alpha_{kj}^{(1)}\right)^{1-\lambda^{(1)}} \psi_{j+1}^{(1)} - \left(\alpha_{kj}^{(2)}\right)^{1-\lambda^{(2)}} \psi_{j+1}^{(2)} \right]^2$$

$$+ \sum_{k=1}^{K} \sum_{j=1}^{J} \left[\phi_{kj}^{(1)} - \phi_{kj}^{(2)} \right]^2$$

where the superscripts (1) and (2) on the right designate the parameters $\alpha_k, \beta_j, \psi_{j+1}, \phi_{kj}$ associated with the respective members $\left(a^{(1)}, \lambda^{(1)}\right)$ and $\left(a^{(2)}, \lambda^{(2)}\right)$ of $A \times [0,1]$.

The defined metric is continuous in λ, as required for homotopy. Moreover, it is evident from Lemma 3.1 that H(a,0) generates an OPD cross-classified model and H(a,1) an ODP Mack model. The homotopy class includes all the intermediate models between these two.

4. Correlation between observations

4.1 Semi-recursive models

Consider the model defined in Section 3.4, and specifically the conditional covariance for $Cov\left[X_{k_1,j_1+m}, X_{k_2,j_2+m+n} \middle| X_{k_1,j_1}, X_{k_2 j_2}\right]$ with $m>0, n\geq 0$. The following lemma is immediate from assumption (M1).

Lemma 4.1. In the semi-recursive model defined in Section 3.4

$$Cov\left[X_{k_1,j_1+m}, X_{k_2,j_2+m+n} \middle| X_{k_1,j_1}, X_{k_2 j_2}\right]=0 \text{ for } k_1 \neq k_2$$ □

In view of this result, attention will be focused on **within-row** covariances $Cov\left[X_{k,j+m}, X_{k,j+m+n} \middle| X_{kj}\right]$ and correlations $Corr\left[X_{k,j+m}, X_{k,j+m+n} \middle| X_{kj}\right]$. Let the latter be denoted $\rho^{\lambda}_{k,j+m,j+m+n|j}$, with $\rho^{0}_{k,j+m,j+m+n|j}$ and $\rho^{1}_{k,j+m,j+m+n|j}$ representing the boundary cases of the ODP cross-classified and ODP Mack models respectively.

These boundary correlations are evaluated by Taylor (2011b) with the following results:

$$\rho^{\lambda}_{k,j+m,j+m+n|j} = (1+B^{\lambda}_{j+m,j+m+n|j})^{-\frac{1}{2}} \text{ for } \lambda=0,1 \tag{4.1}$$

with

$$B^{0}_{j+m,j+m+n|j} = \sum_{i=j+m}^{j+m+n-1} \phi_{i+1}\beta_{i+1} \Big/ \sum_{i=j}^{j+m-1} \phi_{i+1}\beta_{i+1} \tag{4.2}$$

$$B^{1}_{j+m,j+m+n|j} = \frac{\sum_{i=j+m}^{j+m+n-1} f^2_{j+m+n-1}\cdots f^2_{i+1}\sigma^2_i f_{i-1}\cdots f_j}{\sum_{i=j}^{j+m-1} f^2_{j+m+n-1}\cdots f^2_{i+1}\sigma^2_i f_{i-1}\cdots f_j} \tag{4.3}$$

The properties of $\rho^{0}_{k,j+m,j+m+n|j}$ and $\rho^{1}_{k,j+m,j+m+n|j}$ and the relation between them are discussed by Taylor (2011b). It is established that, while there are distinct similarities between the two, there are also distinct differences. Certainly $\rho^{0}_{k,j+m,j+m+n|j}$ and $\rho^{1}_{k,j+m,j+m+n|j}$ are numerically different.

One might therefore wish to formulate a semi-recursive model with correlation structure intermediate between these two cases. It is evident from (3.16) that the homotopy class of semi-recursive models provides a continuum of correlation values between $\rho^{0}_{k,j+m,j+m+n|j}$ and $\rho^{1}_{k,j+m,j+m+n|j}$.

4.2 Evaluation of semi-recursive correlation structures

However, care may be required in the selection of a semi-recursive model as $\rho^{\lambda}_{k,j+m,j+m+n|j}$ does not appear to be related to $\rho^{0}_{k,j+m,j+m+n|j}$ and $\rho^{1}_{k,j+m,j+m+n|j}$ in any simple way. Consider the evaluation of $\rho^{\lambda}_{k,j+m,j+m+n|j}$. Let $c_{k,j+m,j+m+n|j}$ denote $Cov\left[X_{k,j+m}, X_{k,j+m+n} \middle| X_{kj}\right]$. Then

$$\begin{aligned} c_{k,j+m,j+m+n|j} &= E\left[\left\{X_{k,j+m}, X_{k,j+m+n} \middle| X_{kj}\right]\right\} \times \left\{X_{k,j+m+n} - \mathrm{E}\left[X_{k,j+m+n} \middle| X_{kj}\right]\right\} \middle| X_{kj}\right] \\ &= E\left[\left\{X_{k,j+m} - E\left[X_{k,j+m} \middle| X_{kj}\right]\right\} \times E\left[X_{k,j+m+n} - E\left[X_{k,j+m+n} \middle| X_{kj}\right] \middle| X_{k,j+m}\right] \middle| X_{kj}\right] \\ &= E\left[\left\{X_{k,j+m} - E\left[X_{k,j+m} \middle| X_{kj}\right]\right\} \times \left\{E\left[X_{k,j+m+n} \mid X_{k,j+m}\right] - E\left[X_{k,j+m+n} \middle| X_{kj}\right]\right\} \middle| X_{kj}\right] \end{aligned} \tag{4.4}$$

Difficulty arises in the evaluation of the terms $E\left[X_{k,j+m} \middle| X_{kj}\right]$. In the case of the recursive ODP Mack model, these are evaluated recursively, thus:

$$E\left[X_{k,j+m} \middle| X_{kj}\right] = E\left[E\left[X_{k,j+m} \mid X_{k,j+m-1}\right] \mid X_{kj}\right] = E\left[f_{j+m-1} X_{k,j+m-1} \mid X_{kj}\right] = etc.$$

where (ODPM3) has been used.

If, however, the same procedure is attempted for the semi- recursive model, then, by (3.16),

$$\begin{aligned} E\left[X_{k,j+m} \middle| X_{kj}\right] &= E\left[E\left[X_{k,j+m} \mid X_{k,j+m-1}\right] \mid X_{kj}\right] \\ &= E\left[X_{k,j+m-1} + \left[X_{kj}\left(f_j - 1\right)\right]^{\lambda} \left(\alpha_k \beta_{j+1}\right)^{1-\lambda} \mid X_{kj}\right] \\ &= E\left[X_{k,j+m-1} \mid X_{kj}\right] + \left(f_j - 1\right)^{\lambda} \left(\alpha_k \beta_{j+1}\right)^{1-\lambda} E\left[X^{\lambda}_{k,j+m-1} \mid X_{kj}\right] \end{aligned}$$

and difficulty arises in the evaluation of the last expectation.

4.3 Non-monotonicity of semi-recursive correlations

Care would also be necessary in the selection of semi-recursive correlation structures because, while it is known from Theorem 4.4 of Taylor (2011b) that $\rho^{1}_{k,j+m,j+m+n|j} \geq \rho^{0}_{k,j+m,j+m+n|j}$, it cannot be assumed that ρ changes monotonically between these extremes. Indeed, my colleague, Hugh Miller, provides the following counter-example.

Example. Consider a semi-recursive model in the representation of Lemma 3.1, with the following parameters: $\alpha = 100, \beta_1 = \beta_2 = \beta_3 = \beta_4 = 1, \phi_{kj} = 1$ (Poisson case). By (3.19), $f_1 = 2,\ f_2 = 1.5,\ f_3 = 1.33.$ Now consider the (unlikely) case of $X_{k1} = 1$ for some k (c.f. $E\left[X_{k1}\right] = 100$). Then simulation yields the values of $\rho^{\lambda}_{k,2,4|1}$ for various λ shown in Table 4.1.

Table 4.1 Values of $\rho^{\lambda}_{k,2,4|1}$ for varying λ $\left(X_{k1}=1\right)$

λ	P
0	0.578
0.1	0.569
0.2	0.593
1	0.769

The values of ρ for $\lambda = 0,1$ may be verified by the formulas (4.15) and (4.16) (λ = 0) and (4.6) and (4.7) (λ = 0) of Taylor (2011b) but note that ρ does not proceed monotonically between $\lambda = 0$ and 0.2.

If a less eccentric value of X_{k1} is chosen, say $X_{k1} = 50,$ the results are as in Table 4.2.

Table 4.2 Values of $\rho^{\lambda}_{k,2,4|1}$ for varying $\lambda\left(X_{k1}=50\right)$

λ	P
0	0.578
0.1	0.602
0.2	0.627
0.3	0.652
0.4	0.673
0.5	0.693
0.6	0.712
0.7	0.728
0.8	0.744
0.9	0.756
1	0.768

Evidence of slight sampling error is apparent from a comparison of Tables 4.1 and 4.2 at $\lambda = 1$. Nonetheless, monotonicity of $\rho^{\lambda}_{k,j+m,j+m+n|j}$ as a function of λ appears to have been achieved in this second example.

Note also, by comparison of Tables 4.1 and 4.2 at λ = 0.1, 0.2, that $\rho^{\lambda}_{k,j+m,j+m+n|j}$ depends on the observed value of X_{kj} for 0 < λ < 1, whereas it is independent of this observation for λ = 0,1.

More detail on the relation between λ and $\rho^{\lambda}_{k,j+m,j+m+n|j}$ might be a fruitful area for future research.

5. Parameter estimation and forecasts

5.1 Recursive models

Consider MLE of parameters in the OPD Mack model defined in Section 3.2.2. The conditional log-likelihood of a single observation in $\mathcal{D}_k$ is (terms extraneous to MLE omitted)

$$\ell\left(Y_{kj} \middle| X_{k,j-1}\right) = \phi_{kj}^{-1}\left\{Y_{kj}\left[\ell n X_{k,j-1} + \ell n\left(f_{j-1}-1\right)\right] - X_{k,j-1}\left(f_{j-1}-1\right)\right\} \tag{5.1}$$

The conditional log- likelihood of the entire row $k\left(> K - J\right)$ of $\mathcal{D}_k$ is

$$\ell\left(Y_{k2}, Y_{k3}, \ldots Y_{k,K-k+1} \middle| X_{k1}\right) = \ell\left(Y_{k2} \middle| X_{k1}\right) +$$

$$\ell\left(Y_{k3}, \ldots, Y_{k,K-k+1} \middle| X_{k1}, Y_{k2}\right)$$

$$= \ell\left(Y_{k2} \middle| X_{k1}\right) + \ell\left(Y_{k3}, \ldots, Y_{k,K-k+1} \middle| X_{k2}\right)$$

by assumption (M2).

By extension of this argument

$$\ell\left(Y_{k2}, \ldots, Y_{k,K-k+1} \middle| X_{k1}\right) = \sum_{j=2}^{K-k+1} \ell\left(Y_{kj} \middle| X_{k,j-1}\right) \; for\; k > K - J \tag{5.2}$$

The reasoning for $k \le K - J$ is similar but with the upper limit of summation replaced by J. Then, by assumption (M1),

$$\ell\left(Y_K \middle| X_K^1\right) = \sum_{k=2}^{K-1} \sum_{j=2}^{\min(J,K-k+1)} \ell\left(Y_{kj} \middle| X_{k,j-1}\right) \tag{5.3}$$

Substitution of (5.1) into (5.3) and differentiation with respect to f_j for a particular value of j, yields

$$\partial\ell / \partial f_j = \sum_{k=1}^{K-j} \phi_{k,j+1}^{-1}\left\{Y_{j,k+1} \middle/ \left(f_j - 1\right) - X_{kj}\right\} \tag{5.4}$$

Setting this to zero and rearranging gives the following MLE of f_j:

$$\hat{f}_j = \frac{\sum_{k=1}^{K-j} \phi_{k,j+1}^{-1}\left(X_{kj} + Y_{k,j+1}\right)}{\sum_{k=1}^{K-j} \phi_{k,j+1}^{-1} X_{kj}} = \frac{\sum_{k=1}^{K-j} \phi_{k,j+1}^{-1} X_{k,j+1}}{\sum_{k=1}^{K-j} \phi_{k,j+1}^{-1} X_{kj}} \tag{5.5}$$

In the special case in which weights are column dependent only,

$$\phi_{kj} = \phi_j, \text{ independent of } k \tag{5.6}$$

the estimator (5.5) reduces to the usual chain ladder estimator

$$\hat{f}_j = \sum_{k=1}^{K-j} \mathrm{X}_{k,j+1} / \sum_{k=1}^{K-j} \mathrm{X}_{kj} \tag{5.7}$$

The forecast of a future (i.e. j+k > K+1) value of X_{kj} is

$$\hat{\mathrm{X}}_{kj}^{R} = \mathrm{X}_{k,\mathrm{K}-k+1} \hat{f}_{\mathrm{K}-k+1} \hat{f}_{\mathrm{K}-k+2} \cdots \hat{f}_{j-1},\ j = K-k+2, K-k+3, ..., J\ , \tag{5.8}$$

The estimation and forecast algorithm consisting of (5.7) and (5.8) constitute the **chain ladder algorithm** described in Section 3.1.

5.2 Non-recursive models

Consider MLE of parameters in the OPD cross-classified model defined in Section 3.3. The log-likelihood of a single observation in $\mathcal{D}_K$ is

$$\ell\left(Y_{kj}\right) = \phi_{kj}^{-1}\left\{Y_{kj}\left(\ell n\alpha_k + \ell n\beta_j\right) - \alpha_k\beta_j\right\} \tag{5.9}$$

The log-likelihood for the entire $\mathcal{D}_K$ is

$$\ell\left(Y_k\right) = \sum_{k=1}^{k} \sum_{j=1}^{\min(J,\mathrm{K}-k+1)} \ell\left(Y_{kj}\right) \tag{5.10}$$

Substitution of (5.9) into (5.10) and differentiation with respect to α_k for a particular value of k, yields

$$\partial\ell / \partial\alpha_k = \sum_{j=1}^{\min(J,\mathrm{K}-k+1)} \phi_{kj}^{-1}\left(Y_k / \alpha_k - \beta_j\right) \tag{5.11}$$

Differentiation with respect to β_j yields

$$\partial\ell / \partial\beta_j = \sum_{k=1}^{\mathrm{K}-j+1} \phi_{kj}^{-1}\left(Y_{kj} / \beta_j - \alpha_k\right) \tag{5.12}$$

Setting (5.11) and (5.12) to zero gives the following MLEs of α_k, β_j:

$$\hat{\alpha}_j = \sum_{j=1}^{\min(J,\mathrm{K}-k+1)} \phi_{kj}^{-1} Y_{kj} \Big/ \sum_{j=1}^{\min(J,\mathrm{K}-k+1)} \phi_{kj}^{-1}\hat{\beta}_k \tag{5.13}$$

$$\hat{\beta}_j = \sum_{k=1}^{\mathrm{K}-j+1} \phi_{kj}^{-1} Y_{kj} \Big/ \sum_{k=1}^{\mathrm{K}-j+1} \phi_{kj}^{-1}\hat{\alpha}_k \tag{5.14}$$

In the special case in which weights are column dependent only, i.e. (5.6) holds, relations (5.13) and (5.14) reduce to the following:

$$\hat{\alpha}_k = \sum_{j=1}^{\min(J,\mathrm{K}-k+1)} \phi_j^{-1} Y_{kj} \Big/ \sum_{j=1}^{\min(J,\mathrm{K}-k+1)} \phi_j^{-1}\hat{\beta}_j \tag{5.13a}$$

$$\hat{\beta}_j = \sum_{k=1}^{\mathrm{K}-j+1} Y_{kj} \Big/ \sum_{k=1}^{\mathrm{K}-j+1} \hat{\alpha}_k \tag{5.14a}$$

In the alternative special case in which weights are row dependent only,

$$\phi_{kj} = \phi_k \text{, independent of j} \tag{5.15}$$

relations (5.13) and (5.14) reduce to the following:

$$\hat{\alpha}_k = \sum_{j=1}^{\min(J,K-k+1)} Y_{kj} \Big/ \sum_{j=1}^{\min(J,K-k+1)} \hat{\beta}_j \tag{5.13b}$$

$$\hat{\beta}_j = \sum_{k=1}^{K-j+1} \phi_k^{-1} Y_{kj} \Big/ \sum_{k=1}^{K-j+1} \phi_k^{-1} \hat{\alpha}_k \tag{5.14b}$$

In the even more specialised case in which weights are uniform across all cells, the relations simplify further, as follows

$$\hat{\alpha}_k = \sum_{j=1}^{\min(J,K-k+1)} Y_{kj} \Big/ \sum_{j=1}^{\min(J,K-k+1)} \hat{\beta}_j \tag{5.13c}$$

$$\hat{\beta}_j = \sum_{k=1}^{K-j+1} Y_{kj} \Big/ \sum_{k=1}^{K-j+1} \hat{\alpha}_k \tag{5.14c}$$

The last case includes the case $\phi_{kj} = 1$, i.e. the ODP distribution reduces to Poisson. This is a case where MLEs have been studied in detail by Hachemeister & Stanard (1975), Renshaw & Verrall (1998) and Taylor (2000), among others, where it is shown that (5.13c) and (5.14c) are equivalent to the chain ladder estimates (5.7) when the $\hat{f}_j$ and $\hat{\beta}_j$ are related by $\hat{f}_j - 1 = \hat{\beta}_{j+1} \Big/ \sum_{i=1}^{j} \hat{\beta}_i$. It is shown by England & Verrall (2002) that this result continues to hold in the more general case $\phi_{kj} = \phi$.

The forecast of a future value of Y_{kj} is

$$\hat{Y}_{kj}^{NR} = \hat{\alpha}_k \hat{\beta}_j,\ j = K-k+2, K-k+3, \ldots, J \tag{5.16}$$

5.3 Relation between recursive and non-recursive cases

5.3.1 Poisson distribution

Taylor (2000, Chapter 2) studies the ODP cross-classified model subject to $\phi_{kj} = 1$ (i.e. Poisson distribution in each cell of $\mathcal{D}_K$), with MLEs given by (5.13c) and (5.14c). It is shown there (equation (2.47)) that

$$\sum_{k=1}^{K-j} X_{k,j+1} \Big/ \sum_{k=1}^{K-j} X_{kj} = \sum_{i=1}^{j+1} \hat{\beta}_i \Big/ \sum_{i=1}^{j} \hat{\beta}_i \tag{5.17}$$

Comparison of this result with (5.7) shows that

$$\hat{f}_j = \sum_{i=1}^{j+1} \hat{\beta}_i \Big/ \sum_{i=1}^{j} \hat{\beta}_i \tag{5.18}$$

establishing the relation between the MLEs of the recursive and non-recursive models.

Verrall (2000, p.93) shows that $X_{k,K-k+1}$ is the MLE of z_k for $K - J < k \le K$ where z_k is defined by

$$\alpha_k \beta_j = z_k \beta_j \Big/ \sum_{i=1}^{K-k+1} \beta_i \qquad (5.19)$$

It follows that

$$\hat{\alpha}_k = X_{k,K-k+1} / \sum_{i=1}^{K-k+1} \hat{\beta}_i \qquad (5.20)$$

Substitution of (5.18) and (5.20) into (5.8) yields

$$\hat{X}_{kj}^R = \hat{\alpha}_k \sum_{i=1}^{j} \beta_i$$

in which case the forecast of Y_{kj} in the recursive model is

$$\hat{Y}_{kj}^R = \hat{X}_{kj}^R - \hat{X}_{k,j-1}^R = \hat{\alpha}_k \hat{\beta}_j = \hat{Y}_{kj}^{NR} \qquad (5.21)$$

by (5.16).

Thus, the recursive (Poisson Mack) and non- recursive (Poisson cross-classified) models produce the same forecasts when parameters are estimated by MLEs in the case $\phi_{kj} = 1$.

It then follows from Section 5.1 that those forecasts are obtainable from the chain ladder algorithm.

5.3.2 OPD distribution

Now consider the more general case in which $\phi_{kj} = \phi$, independent of k and j but not necessarily equal to unity. In both OPD Mack and ODP cross- classified models

$Y_{kj} \sim ODP(\mu_{kj}, \phi)$ for some mean μ_{kj}.

The meaning of this is

$$Y_{kj} \big/ \phi \sim Poiss(\mu_{kj} / \phi) \qquad (5.22)$$

Application of (5.22) to (ODPM3) yields

$$Y_{k,j+1} \big/ \phi \sim Poiss(X_{kj} f_j / \phi) \qquad (5.23)$$

For any fixed value of j, this last relation indicates that the MLE of f_j is obtained by application of the Poisson theory of Section 5.3.1 (i.e. with $\phi_{kj} = 1$) but with each Y_{kj} replaced by $Y_{kj} \big/ \phi$. This leaves the estimator (5.7) unchanged.

On the other hand, application of (5.22) to (ODPCC2) yields

$$Y_{kj} \big/ \phi \sim Poiss\left(\alpha_k \beta_j \big/ \phi\right) \tag{5.24}$$

This last relation indicates that the MLEs of the α_k and β_j / ϕ are again obtained by application of the Poisson theory of Section 5.3.1 but with Y_{kj} replaced by Y_{kj} / ϕ. Equations (5.13c) and (5.14c) are unchanged by these substitutions, indicating that the MLE, and forecasts of the OPD cross-classified model are the same as in the Poisson case. This fact was noted by England & Verrall (2002, p.449)

This leads to the following result.

Lemma 5.1. For a given data array $\mathcal{D}_K$, the OPD Mack and OPD cross-classified models with dispersion parameters uniform across $\mathcal{D}_K^+$ ($\phi_{kj} = \phi$), generate the same forecasts of $\mathcal{D}_K^c$ on the basis of ML. The ML parameter estimates for the two models are related through (5.18) and (5.20).

Proof. Section 5.3.1 gives the proof for the case $\phi = 1$. The present sub-section shows that all of the forecasts and parameter estimates discussed in Section 5.3.1 are unaffected by a change in ϕ to a value not equal to unity. □

5.4 Semi-recursive models

Consider MLE of parameters in the semi-recursive model defined in Section 3.4. The log-likelihood of the data array is

$$\ell\left(Y_K\right) = \ell\left(X_K^1\right) + \ell\left(Y_K \middle| X_K^1\right)$$

$$= \ell\left(\mathrm{X}_K^1\right) + \sum_{k=1}^{K} \sum_{j=1}^{\min(J-1,K-k)} \ell\left(Y_{k,j+1} \middle| \mathrm{X}_{kj}\right) \tag{5.25}$$

by the same argument as led to (5.3).

The partial log-likelihood for $\mathrm{X}_K^1 (= Y_K^1)$ can be obtained from assumption (ODPSR3a) and that for $Y_{k,j+1} \middle| \mathrm{X}_{kj}$ from Lemma 3.1. These give

$$\ell\left(Y_{k1}\right) = (1-\lambda)\,\ell^{NR}\left(Y_{k1}\right) \tag{5.26}$$

$$\ell\left(Y_{k,j+1} \middle| \mathrm{X}_{kj}\right) = \lambda\, \ell^R\left(Y_{k,j+1} \middle| \mathrm{X}_{kj}\right) + (1-\lambda)\, \ell^{NR}\left(Y_{k,j+1}\right) \tag{5.27}$$

where ℓ^R denotes a log-likelihood within the recursive model of Section 5.1 and ℓ^{NR} within the non-recursive model of Section 5.2.

Substitution of (5.26) and (5.27) into (5.25) gives

$$\ell\left(Y_K\right) = \lambda \ell^R\left(Y_K \mid \mathrm{X}_K^1\right) + (1-\lambda)\,\ell^{NR}\left(Y_K\right) \tag{5.28}$$

by (5.3) and (5.10).

It is shown in Section 5.1 that the chain ladder estimates (5.7) of the parameters f_j set the log-likelihood component ℓ^R to zero in the case of column dependent dispersion parameters (5.6).

Likewise, it is shown in Section 5.2 that the chain ladder estimates (5.7) of the parameters f_j set the log-likelihood component ℓ^{NR} to zero in the case of uniform dispersion parameters $\phi_{kj} = \phi$ when the $\hat{f}_j$ and $\hat{\beta}_j$ are related as in (5.18).

It follows that the chain ladder estimates (5.7) of the parameters f_j set the log-likelihood (5.28) to zero under the same conditions. These results may be summarised as follows.

Theorem 5.2. Suppose that the data array $\mathcal{D}_K$ is subject to a semi-recursive model as represented in Lemma 3.1 with $\phi_{kj} = \phi$, independent of k and j. Then
(a) the MLEs of its parameters f_j are obtained by treating the data array $\mathcal{D}_K$ as if subject to the (recursive) ODP Mack model.
(b) the MLEs of parameters α_k, β_j are obtained by treating the data array as if subject to the (non-recursive) ODP cross-classified model.
(c) these parameter estimates are related by (5.18) and (5.20) and the ODP Mack, ODP cross-classified, and semi-recursive forecasts of any particular future Y_{kj} are all identical. The forecasts are obtainable by application of the chain ladder algorithm. □

The theorem shows that the chain ladder algorithm provides ML parameter estimates and forecasts for the entire homotopy class of semi-recursive models defined Section 3.4.

6. Sufficient statistics

The following results are special cases of more general results appearing in Taylor (2011).

Lemma 6.1. (a) For an OPD Mack model, $\sum^{C_j} Y_{k,j+1} \phi^{-1}_{k,j+1}$ is a sufficient statistic for f_j.

(b) For an OPD cross-classified model, $\sum^{R(k)} Y_{kj} \phi^{-1}_{kj}$ is a sufficient statistic for α_k and $\sum^{C(j)} Y_{kj} \phi^{-1}_{kj}$ is a sufficient statistic for β_j.

(c) In case (b), the sufficient statistic for the full parameter set $\{\alpha_k, \beta_j\}$ consists of the K row sums and J column sums. This sufficient statistic is not minimal. A minimal sufficient statistic is obtained by deletion of an arbitrary single component.

Proof. (a) See Theorem 5.1 of Taylor (2011a).
(b) See Theorem 5.2 of Taylor (2011a).
(c) See Theorem 5.3 of Taylor (2011a). □

Remark. The minimal statistic defined in part (c) of the theorem is not unique. For full detail on the construction of alternative minimal sufficient statistics, see Theorem 5.3 of Taylor (2011a).

Theorem 6.2. For the semi-recursive model defined in Section 3.4,

(a) The vector s = $\left\{\sum^{\mathcal{R}(k)} Y_{kj}\phi_{kj}^{-1}, k=1,...,K, \sum^{C(j)} Y_{kj}\phi_{kj}^{-1}, j=1,...,J,\right\}$ is a sufficient statistic for the parameter set $\{f_1,..., f_{J-1}, \alpha_1..., \alpha_K, \beta_1,...,\beta_J\}$.
(b) A minimal sufficient statistic can be obtained by the deletion of an arbitrary single component of s. This statistic is complete.

Proof. (a) Recall the form (5.28) for the log-likelihood $\ell(Y_K)$. Theorem 5.1 of Taylor (2011a) shows that ℓ^R satisfies Fisher-Neyman factorisation with respect to the parameter set $\{f_1,..., f_{J-1}\}$ and the statistic s.

Similarly, ℓ^{NR} with respect to the set $\{\alpha_1,...,\alpha_K,\beta_1,...\beta_J\}$. Thus $\ell(Y_K)$ satisfies Fisher-Neyman factorisation with respect to entire parameter set $\{f_1,..., f_{J-1}, \alpha_1..., \alpha_K, \beta_1,...,\beta_J\}$ and it follows that s is a sufficient statistic for that parameter set.
(b) Let the components of s be denoted $s_1,..., s_K, s_{K+1,},..., s_{K+J}$. By the relations at the end of Section 2,

$$s_1 + ... + s_K = s_{K+1} + ... + s_{K+J} \tag{6.1}$$

whence any component of s can be expressed in terms of the other components. This means that $s_{\min}$, obtained from s by the deletion of an arbitrary component, contains the same information as s and is therefore sufficient for $\{f_1,..., f_{J-1}, \alpha_1..., \alpha_K, \beta_1,...,\beta_J\}$.

Now note that this last parameter set can be reduced in dimension. By (3.18), each f_j may be expressed in terms of $\beta_1,...,\beta_{j+1}$ and so $\{f_1,..., f_{J-1}\}$ may be expressed in terms of $\{\beta_1,...,\beta_J\}$. Further, by (3.17), this last set may be reduced to $\{\beta_1,...,\beta_{J-1}\}$. Thus the parameter set for the semi-recursive model may be taken as $\{\alpha_1,...,\alpha_K,\beta_1,...\beta_{J-1}\}$, of dimension $K+J-1$.

Now $s_{\min}$ is of the same dimension and, for a regression model such as the semi-recursive model, with error terms distributed as a member of the EDF, this equality of dimensions is a necessary and sufficient condition for a sufficient statistic to be complete (Cox & Hinkley, 1974, p.31).

Finally, since $s_{\min}$ is a complete sufficient statistic, it is immediately minimally sufficient (Lehmann & Casella, 1998).

□

7. Minimum variance estimation

The Mack model is known to generate unbiased MLEs, of loss reserve (Mack, 1993). The ODP Mack model with column dependent dispersion parameters contains the same expectations and leads to the same MLEs (5.7) and (5.8). The same is not true, however, of the ODP cross-classified model.

Since the semi-recursive model is a mixture of these two, one can expect that its MLEs will, in general, be biased. However, any bias can be corrected as follows.

Let $Z : \mathcal{D}_K^c \to \mathcal{R}$ be some predictand and let $\hat{Z} : \mathcal{D}_K \to \mathcal{R}$ be a predictor of $Z \mid \mathcal{D}_K$. Define

$$\tilde{Z} = \hat{Z}\ \mathrm{E}\left[Z \middle| \mathcal{D}_K\right] \Big/ \mathrm{E}\left[\hat{Z} \middle| \mathcal{D}_K\right] \tag{7.1}$$

Then

$$\mathrm{E}\left[\tilde{Z} \middle| \mathcal{D}_K\right] = \mathrm{E}\left[Z \middle| \mathcal{D}_K\right] \tag{7.2}$$

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

Theorem 7.1. Let $\mathcal{D}_k^+$ be subject to the semi-recursive model of Section 3.4, with $\phi_{kj} = \phi$, const. Then the bias corrected chain ladder estimates $\tilde{X}_{kj}, \tilde{R}_k$ and $\tilde{R}$ (derived from $\hat{X}_{kj}; \hat{R}_k$ and $\hat{R}$ defined by (3.5) – (3.7) respectively) are minimum variance unbiased estimators (MVUEs) of $\mathrm{E}\left[X_{kj} \middle| \mathcal{D}_K\right]$, $\mathrm{E}\left[R_k \mid \mathcal{D}_K\right]$ and $\mathrm{E}\left[R \middle| \mathcal{D}_K\right]$.

Proof. Lemma 4.3 of Taylor (2011a) shows that the estimators $\hat{X}_{kj}, \hat{R}_k$ and $\hat{R}$ are MLE for the ODP cross-classified model with $\phi_{kj} = \phi$. The result also appears in England & Verrall (2002).

These estimators are expressible in terms of the statistic s, defined in Theorem 6.2. It is apparent from the proof of that theorem that s is expressible in terms of $s_{\min}$ which, by the same theorem, is a complete sufficient (in fact, minimal sufficient) statistic for the semi-recursive model parameter set $\left\{ f_1, \ldots, f_{J-1}, \alpha_1 \ldots, \alpha_K, \beta_1, \ldots, \beta_J \right\}$.

Thus, $\tilde{X}_{kj}, \tilde{R}_k$ and $\hat{R}$ are unbiased estimators that are functions of a complete sufficient statistic. By the Lehmann-Scheffe’ theorem, they are MVUEs. □

The application of Theorem 7.1 is limited by the fact that the bias correction factors in $\tilde{X}_{kj}$, etc. would rarely be known practice. On the other hand, however, the biases contained in chain ladder estimates are tolerated in practice and, in this context, the theorem shows that the chain ladder provides a minimum variance estimate of whatever it is estimating.

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

References

Cox DR & Hinckley DV (1974). **Theoretical Statistics**. Chapman and Hall, London UK.

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

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

Lehmann EL & Casella G (1998). **Theory of point estimation (2nd edition).** Springer.

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

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

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

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

Taylor (2011b). Chain ladder correlations. Research paper No. 220 at http://www.economics.unimelb.edu.au/actwww/wps2011.shtmlhttp://www.economics.unimelb.edu.au/actwww/wps2011.shtml.

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