*A.A.S.* **4**, I, 105-121 (2009)

# CHAIN-LADDER AS MAXIMUM LIKELIHOOD REVISITED

By D. Kuang, B. Nielsen and J. P. Nielsen

ABSTRACT

It has long been known that maximum likelihood estimation in a Poisson model reproduces the chain-ladder technique. We revisit this model. A new canonical parametrisation is proposed to circumvent the inherent identification problem in the parametrisation. The maximum likelihood estimators for the canonical parameter are simple, interpretable and easy to derive. The boundary problem where all observations in one particular development year or on particular underwriting year is zero is also analysed.

KEYWORDS

Boundary Problem; Canonical Parameter; Chain-Ladder; Identification Problem; Maximum Likelihood; Poisson Model

CONTACT ADDRESS

B. Nielsen, Nuffield College, Oxford OX1 5AD, UK.
E-mail: bent.nielsen@nuffield.ox.ac.uk

## 1. Introduction

The chain-ladder technique is the main tool when reserving in general insurance. Kremer (1985) and Mack (1991) have shown that the standard chain-ladder technique is reproduced through maximum likelihood estimation in a Poisson model; see also the discussion of Mack & Venter (2000), Verrall & England (2000). An early contribution to working out that stochastic models might be consistent with chain ladder estimates was made by Hachemeister & Stanard (1975). In this paper we study the Poisson model in further detail. We propose a new canonical parametrisation that gets around the over-parametrisation built into the previous formulations and analyse the model using exponential family techniques.

The choice of parametrisation might seem unimportant at first sight, any identifiable choice of parametrisation will eventually lead to the same model. While this argument is true for considerations within sample, it is not true once extrapolations are made beyond the observed development horizon. A related issue is that when inspecting the development of parameters visually, conclusions about the level of the parameters will depend on the choice of parametrisation. The aim of this paper is to clarify



https://doi.org/10.1017/S1748499500000610 Published online by Cambridge University Press

the underlying parametrisation of the chain-ladder model. The suggested parametrisation provides a coherent framework for predicting outstanding liabilities when we have not seen a complete run-off on our data yet. In this case, one has to extrapolate parameters from the development direction into the future and it becomes absolutely essential that the choice of parametrisation and its relationship to the extrapolation at hand is fully understood, see Kuang *et al.* (2008b) for a related case. In future work we intend to make further contributions in this direction.

The maximum likelihood estimator for the canonical parameter of the classical chain-ladder model turns out to have a simple, interpretable and explicit expression that is easy to write down and that can be linked to the classical empirical development factors in a straight forward way. Based on these findings, we are also able to identify exactly what the classical development factors are estimating in terms of the parameters of the underlying mathematical statistical model.

We also characterise boundary situations for the maximum likelihood problem. These arise when all observations in one particular development year or one particular underwriting year are zero. It is a well-known and non-trivial estimation problem of the theory of exponential families to find a proper solution to the estimation problem in such cases. We discuss in what sense the classic chain-ladder method can be viewed as a maximum likelihood estimator in boundary situations.

There are several advantages of addressing the over-parametrisation at the outset and translating the classical chain-ladder approach in terms of classical mathematical statistics and exponential families. First, we can take advantage of the general insights of the field of mathematical statistics. Secondly, the derivation of the maximum likelihood estimators is simple and largely avoiding the recursive arguments in many earlier proofs. Thirdly, the general approach of stating the underlying statistical model, addressing the identification issues surrounding the parameters and then estimating through the maximum likelihood approach might prove to be very useful in our attempts to get beyond the approach of classical chain-ladder without losing the well known advantages of this approach.

In Section 2 we identify what we call the canonical parameters of the chain-ladder method. In Section 3 we make a likelihood analysis of the Poisson model exploiting the canonical parametrisation. The boundary problem is discussed in Section 4. Proofs are given in an Appendix.

## 2. Parametrising the Chain-Ladder Technique

The chain-ladder technique has been described in numerous papers. We will follow the exposition of England & Verrall (2002), albeit with a slightly different notation. We first go through the classical chain-ladder approach

https://doi.org/10.1017/S1748499500000610 Published online by Cambridge University Press

and then we identify what we call the canonical parametrisation of the underlying mathematical statistical model. Among a wide class of different parametrisations we choose a parametrisation which is a maximal invariant, following a group theoretic terminology. When later formulating the statistical model this also turns out to be the natural parameter of a full exponential family. The end result is that the main ingredients in the parametrisation involve the growth of the parameters on a log scale as these growth rates always are fully identified. It should not be the levels of the log parameters. These are only identified up to an additive level parameter.

Consider a standard incremental insurance run-off triangle of dimension $k$. Each entry is denoted $Y_{ij}$ so that $i$ is the accident year index and $j$ is the development year index. Collectively, we have the data $Y = \{Y_{ij}, \forall (i,j) \in \mathcal{I}\}$, where $\mathcal{I}$ is the triangular index set

$$\mathcal{I} = \{(i,j) : i \text{ and } j \text{ belong to } (1,\dots,k) \text{ with } i+j-1 = 1,\dots,k\}. \tag{1}$$

From this the cumulative claims arise as $Z_{ij} = \sum_{\ell=1}^{j} Y_{i\ell}$.

The chain-ladder technique derives the development factors

$$F_j = \frac{\sum_{i=1}^{k+1-j}\sum_{\ell=1}^{j} Y_{i\ell}}{\sum_{i=1}^{k+1-j}\sum_{\ell=1}^{j-1} Y_{i\ell}} = \frac{\sum_{i=1}^{k+1-j} Z_{ij}}{\sum_{i=1}^{k+1-j} Z_{i,j-1}} \qquad \text{for } j = 2,\dots k. \tag{2}$$

These are then used to forecast future values of cumulative claims in the lower triangle where $i, j \le k$ through

$$\widetilde{Z}_{i,k-i+1+h} = Z_{i,k-i+1} \prod_{m=1}^{h} F_{k-i+1+m}. \tag{3}$$

In the stochastic formulation of the chain-ladder then $Y_{ij}$ is thought of as a stochastic variable with expectation

$$\mathsf{E}(Y_{ij}) = A_i B_j D \qquad \text{for } i,j = 1,\dots k,$$

giving parameters $A_1,\dots,A_k$, $B_1,\dots,B_k$ and $D$. The intrinsic over-parametrisation in this formulation is solved in various ways in the literature. Kremer (1985, p130), Mack (1991) and England & Verrall (2002) identify by $\sum_{j=1}^{k} B_j = D = 1$. When analysing the chain-ladder-type model based on a log normal distribution Kremer (1982, 1985, p138) and Verrall (1994) have $\sum_{i=1}^{k} A_i = \sum_{j=1}^{k} B_j = 1$. England & Verrall (2002) also use the identification $A_1 = B_1 = 1$.

To get around the intrinsic over-parametrisation we follow the approach of Kuang *et al.* (2008a) and develop a canonical parametrisation which leans

https://doi.org/10.1017/S1748499500000610 Published online by Cambridge University Press

naturally to the maximum likelihood estimation and the development factors. This is done on a log linear scale so

$$\log\{\mathsf{E}(Y_{ij})\} = \mu_{ij} = \alpha_i + \beta_j + \delta, \tag{4}$$

where $\alpha_i = \log A_i$, $\beta_j = \log B_j$ and $\delta = \log D$. In this log formulation the parameter vector is therefore

$$\theta = (\alpha_1, \ldots, \alpha_k, \beta_1, \ldots, \beta_k, \delta) \in \mathbf{R}^{2k+1}.$$

The identification problem is that constants can be added and subtracted to the elements of $\theta$ without changing $\mu_{ij}$. This can be understood through the terminology of an invariance property. This leads us to consider the invariance property of the parameter $\mu_{ij}$ with respect to the group of transformations given by

$$g : \begin{pmatrix} \alpha_i \\ \beta_j \\ \delta \end{pmatrix} \mapsto \begin{pmatrix} \alpha_i + a \\ \beta_j + b \\ \delta - a - b \end{pmatrix}, \tag{5}$$

which holds simultaneously for all $i$ and $j$ and where $a$, $b$ are arbitrary constants. Every choice of parameter $\theta$ will lead us to a particular distribution. However, there are different such choices, $\theta_1$ and $\theta_2$ say, that lead to the same distribution. The problems of extrapolations and visual inspection of parameters enter when our analysis of $\theta_1$ lead us to a different conclusion than our analysis of $\theta_2$ even though both these two parameters represent the same underlying distribution. We are therefore looking for a canonical parameter which holds simultaneously for all $i$ and $j$ and where $a, b$ will reveal themselves as arbitrary constants unimportant for the analysis at hand. Cox & Hinkley (1974, §5.3) and Eaton (2007, §7.2) give an overview of the invariance terminology.

In the following we will work out an invariant parametrisation $\xi$. Look first at the parameter $\alpha_i$. This can be rewritten as $\alpha_i = \alpha_1 + (\alpha_i - \alpha_1)$. Using telescopic sums this turns into $\alpha_i = \alpha_1 + \sum_{\ell=2}^{i} \Delta\alpha_\ell$, where $\Delta\alpha_\ell = \alpha_\ell - \alpha_{\ell-1}$ is the growth rate from $\alpha_{\ell-1}$ to $\alpha_\ell$. Doing the same for the parameter $\beta_j$, noting $\mu_{11} = \alpha_1 + \beta_1 + \delta$, and inserting in the equation (4) gives

$$\mu_{ij} = \mu_{11} + \sum_{\ell=2}^{i} \Delta\alpha_\ell + \sum_{\ell=2}^{j} \Delta\beta_\ell. \tag{6}$$

The parameters $\mu_{ij}$ can therefore be expressed in terms of a parameter vector $\xi \in \mathbf{R}^{2k-1}$ where

$$\xi = (\mu_{11}, \Delta\alpha_2, \ldots, \Delta\alpha_k, \Delta\beta_2, \ldots, \Delta\beta_k)'. \tag{7}$$

https://doi.org/10.1017/S1748499500000610 Published online by Cambridge University Press

The interpretation of $\xi$ derives from (6) so that $\mu_{11}$ determines the level, while the differenced parameters determine non-constant effects. Note that $\xi$ is a vector of length $2k-1$. Two elements shorter than the length $2k+1$ of the vector $\theta$. Getting the parameter space down to its fundamental $2k-1$ elements have the advantage that all the parameters vary freely. Therefore the interpretation of the canonical parameters is more straightforward than the interpretation of the classical parameters of the chain-ladder model where the model is overparametrised and one can not interpret each parameter component independently of the other parameter components of which it depends.

The parameter $\xi$ is a function of $\theta$ and it is invariant to $g$ such that $\xi(\theta) = \xi\{g(\theta)\}$. Therefore, one single $\xi$ catches the entire linear space of possible values of $\theta$ leading to the same model. It is simpler to deal with one single parameter $\xi$ consisting of real numbers than to deal with a space of parameters $\theta$ leading to the same model. Below we will see, that it is also simpler to carry through the maximum likelihood analyses with the parametrisation $\xi$ as starting point.

Theorem 1 shows that $\xi$ gives a unique parametrisation of the parameters $\mu_{ij}$ and that $\xi$ therefore is a maximal invariant. This is our theoretical reason for calling it a canonical parameter of the chain ladder model. This contrasts to the classical chain-ladder identification schemes discussed above, which are not invariant to $g$ because they fix the level of the parameters.

**Theorem 1** *Let $\mu = \{\mu_{ij}; (i,j) \in \mathcal{I}\}$, where $\mu_{ij}$ satisfies (4) and $\mathcal{I}$ is given by (1). The parametrisation $\xi$ given by (7) satisfies*
*(i) $\xi$ is a function of $\theta$,*
*(ii) $\mu$ is a function of $\xi$, because of (6).*

*The parametrisation of $\mu$ by $\xi$ is exactly identified in that $\xi^\dagger \neq \xi^\ddagger$ implies $\mu(\xi^\dagger) \neq \mu(\xi^\ddagger)$.*

Parameters $\alpha_i, \beta_j, \delta$ can be constructed from (5) and (6) to satisfy arbitrary identification schemes. For instance, consider the identification of Kremer (1985) where $\sum_{j=1}^{k} B_j = D = 1$ corresponding to $\log \sum_{j=1}^{k} \exp(\beta_j) = \delta = 0$. For a given canonical parameter $\xi$ Kremer's parameter arises by choosing $\alpha_i = \sum_{\ell=2}^{i} \Delta\alpha_\ell + \mu_{11} + \psi$ and $\beta_j = \sum_{\ell=2}^{j} \Delta\beta_\ell - \psi$ as well as $\delta = 0$ where $\psi = \log\left\{1 + \sum_{j=2}^{k} \exp\left(\sum_{\ell=2}^{j} \Delta\beta_\ell\right)\right\}$. This is checked in two steps: Insert the expressions for $\alpha_i$, $\beta_j$ and $\delta$ in (4) to recognise (6); and recognise that $\psi$ is constructed so the $\beta_j$s satisfy the desired constraint.

## 3. The Poisson Model and its Analysis

As statistical model assume that $Y_{ij}$ are independent for $i,j$ varying in the triangular index set, $\mathcal{I}$, that $Y_{ij}$ is Poisson distributed so

https://doi.org/10.1017/S1748499500000610 Published online by Cambridge University Press

$$\mathsf{P}(Y_{ij} = y) = \frac{\exp(\mu_{ij} y)}{y!} \exp\{-\exp(\mu_{ij})\}$$

and that $\mu_{ij}$ is defined in terms of the canonical parameter $\xi \in \mathbf{R}^{2k-1}$ through (6), (7). In the current work, the aim of the model is the fundamental one of predicting the lower triangle entries. We do not consider extrapolation of development factors out of sample or other extensions of classical chain-ladder applications in this paper. The canonical parametrisation approach and its maximum likelihood analysis, however, appears to be a useful starting point when we in the future want to go beyond the classical applications of the classical chain-ladder model; see for instance Kuang *et al.* (2008a,b) for a discussion of the parametrisation of an extended chain-ladder model including calendar effects.

The simple chain-ladder prediction of the lower triangle entries will work well when those entries are also Poisson distributed. The Poisson assumption is, however, not essential. It is well-known that the chain-ladder technique applies beyond the case where observations are Poisson distributed, let alone non-negative integers. In such situations the Poisson likelihood function is of course still well-defined as a criterion function or quasi likelihood. The estimators and the predictions of the lower triangle make sense also in this case. However, when the underlying data are not Poisson distributed care should be taken when assessing the uncertainty of the estimates of the lower triangle. Some kind of bootstrap approach to assess uncertainty might be appropriate in this case, see England & Verrall (1999).

The log likelihood function of the statistical model based on the Poisson distribution is given by

$$\log \mathsf{L}(\xi;\, Y) = \sum_{i,j \in \mathcal{I}} \{\mu_{ij} Y_{ij} - \exp(\mu_{ij}) - \log(Y_{ij}!)\}. \tag{8}$$

To express the maximum likelihood estimators some further notation is needed. Denote row and column sums as

$$R_i = \sum_{j=1}^{k+1-i} Y_{ij}, \qquad C_j = \sum_{i=1}^{k+1-j} Y_{ij},$$

so $R_i = Z_{i,k+1-i}$ and define the rectangular sums

$$S_i = \sum_{\ell=1}^{i} \sum_{j=1}^{k-i} Y_{\ell j} = \sum_{\ell=1}^{i} R_\ell - \sum_{j=k-i+1}^{k} C_j \qquad \text{for } i = 1, \ldots, k-1.$$

Introduce column and row development factors

https://doi.org/10.1017/S1748499500000610 Published online by Cambridge University Press

$$F_j = \frac{\sum_{i=1}^{k+1-j}\sum_{\ell=1}^{j} Y_{i\ell}}{\sum_{i=1}^{k+1-j}\sum_{\ell=1}^{j-1} Y_{i\ell}} = \frac{S_{k+1-j} + C_j}{S_{k+1-j}} \qquad \text{for } j = 2, \ldots, k, \tag{9}$$

$$G_i = \frac{\sum_{j=1}^{k+1-i}\sum_{\ell=1}^{i} Y_{\ell j}}{\sum_{j=1}^{k+1-i}\sum_{\ell=1}^{i-1} Y_{\ell j}} = \frac{S_{i-1} + R_i}{S_{i-1}} \qquad \text{for } i = 2, \ldots, k, \tag{10}$$

where $F_j$ was previously defined in (2).

Combine the expression for the log likelihood in (8) with (6) to get

$$\log \mathsf{L}(\xi; Y) = \mu_{11} \sum_{i,j\in\mathcal{I}} Y_{ij} + \sum_{i=2}^{k} R_i \sum_{\ell=2}^{i} \Delta\alpha_\ell + \sum_{j=2}^{k} C_j \sum_{\ell=2}^{j} \Delta\beta_\ell - \sum_{i,j\in\mathcal{I}} \exp(\mu_{ij}) - \sum_{i,j\in\mathcal{I}} \log(Y_{ij}!). \tag{11}$$

This shows that the model is contained within the exponential family with minimal sufficient statistic $T = \left(\sum_{i,j\in\mathcal{I}} Y_{ij}, R_2, \ldots, R_k, C_2, \ldots, C_k\right)'$ and natural parameter $\mu_{11}, \alpha_2 - \alpha_1, \ldots, \alpha_k - \alpha_1, \beta_2 - \beta_1, \ldots, \beta_k - \beta_1$, which is a linear transformation of $\xi$. Since the natural parameter space, $\mathbf{R}^{2k-1}$, is unrestricted and open the exponential family is regular, see Barndorff-Nielsen (1978, p116). As the random variables are discrete boundary outcomes may arise in which the likelihood function does not have a maximum. We therefore need to study the convex support of the sufficient statistic $T$. This has the following property.

**Theorem 2** *Consider the Poisson model over the upper triangle $\mathcal{I}$. The convex support of the sufficient statistic $T$ is closed. The following conditions are equivalent:*

*(i)* $T$ *is interior to its convex support;*
*(ii)* $S_1, \ldots, S_{k-1}, R_2, \ldots, R_k, C_2, \ldots, C_k$ *are positive;*
*(iii)* $F_2, \ldots, F_k, G_2, \ldots, G_k$ *are larger than one and finite;*
*(iv)* $R_1, \ldots, R_k$ *are positive and* $F_2, \ldots, F_k$ *are larger than one and finite;*
*(v)* $C_1, \ldots, C_k$ *are positive and* $G_2, \ldots, G_k$ *are larger than one and finite.*

It is illustrative to give examples of outcomes of reserving data not included in the interior of the convex support of $T$. If the first column is zero then $S_{k-1}$ is zero. If another column is zero. If all of the triangle except the last diagonal is zero then all $S_j$ are zero and the development factors not defined.

It is interesting to note that Theorem 2 also is informative about the situation where the data are not non-negative integers so the Poisson assumption is clearly violated. In that situation, as long as the data belong to

https://doi.org/10.1017/S1748499500000610 Published online by Cambridge University Press

the interior of the convex support of $T$ then the likelihood equations, which are now quasi likelihood equations, have a unique solution.

The maximum likelihood estimator for the canonical parameter can now be found.

**Theorem 3** *Consider the Poisson model over the upper triangle $\mathcal{I}$ with canonical parametrisation, $\xi$. The maximum likelihood estimator for $\xi$ is unique if and only if $T$ is interior to its convex support. It is given by*

$$\begin{aligned}
\Delta\widehat{\alpha}_i &= \Delta \log R_i + \log F_{k+2-i} \qquad \text{for } i = 2, \ldots, k,\\
\Delta\widehat{\beta}_j &= \Delta \log C_j + \log G_{k+2-j} \qquad \text{for } j = 2, \ldots, k,\\
\widehat{\mu}_{11} &= \log R_1 - \sum_{l=2}^{k} \log F_l = \log C_1 - \sum_{l=2}^{k} \log G_l.
\end{aligned}$$

*Moreover, the development factors, $F_j = \hat{\Phi}_j$ and $G_i = \hat{\Gamma}_i$ say, are maximum likelihood estimators for the parameters, for $i, j > 1$,*

$$\Phi_j = \frac{\sum_{\ell=1}^{j} \exp\left(\sum_{h=2}^{\ell} \Delta\beta_h\right)}{\sum_{\ell=1}^{j-1} \exp\left(\sum_{h=2}^{\ell} \Delta\beta_h\right)}, \qquad \Gamma_i = \frac{\sum_{\ell=1}^{i} \exp\left(\sum_{h=2}^{\ell} \Delta\alpha_h\right)}{\sum_{\ell=1}^{i-1} \exp\left(\sum_{h=2}^{\ell} \Delta\alpha_h\right)}. \tag{12}$$

The estimators for $\Delta\alpha_i$ and $\Delta\beta_j$ have rather simple expressions. These expressions can be interpreted as follows. If the claims $Y_{ij}$ had been observed for the full square $1 \le i, j \le k$ then the maximum likelihood estimators for $\Delta\alpha_i$ and $\Delta\beta_j$ would have been $\Delta \log R_i$ and $\Delta \log C_j$, which could be seen by a modification of the proof of Theorem 3. For the triangular data the estimators involve a correction in terms of the development factor. In the case of, for instance, $\Delta\beta_j$, the correction involves the row-wise development factor $G_{k+2-j}$, so the column-wise development factor $F_j$ is not directly involved.

With Theorems 2 and 3 at hand we can give a clear mathematical description and interpretation of what it is that the development factors really are estimating. For a long period of time these development factors have been considered as useful estimators, but they have not clearly been interpreted as estimators of parameters of a mathematical statistical model. But this is indeed what they are!

One can even consider the underlying parameters of the development factors to be an alternative invariant parametrisation. This parametrisation can be constructed from the development parameters $\Phi_j$ and $\Gamma_i$ as an alternative to the canonical parameter $\xi$. Insert the notation $\alpha_i = \log A_i$ and $\beta_j = \log B_j$ in (12) and note, for instance, $\exp\left(\sum_{\ell=2}^{j} \Delta\beta_\ell\right) = B_j/B_1$. This results in

https://doi.org/10.1017/S1748499500000610 Published online by Cambridge University Press

$$\Phi_j = \frac{\sum_{\ell=1}^{j} B_\ell/B_1}{\sum_{\ell=1}^{j-1} B_\ell/B_1}, \qquad \Gamma_i = \frac{\sum_{\ell=1}^{i} A_\ell/A_1}{\sum_{\ell=1}^{i-1} A_\ell/A_1} \qquad \text{for } i,j > 1. \tag{13}$$

Although the terms $B_1$ and $A_1$ cancel out they are kept in as the parameters $B_\ell$ and $A_\ell$ are not invariant to the group $g$ in (5), whereas $B_\ell/B_1$ and $A_\ell/A_1$ are invariant which serves to show that $\Phi_j$ and $\Gamma_i$ are also invariant to the group $g$.

Due to the expression (14) below there is a non-linear, one-one relation between the canonical parameter $\xi$ and the parameter

$$\zeta = (\mu_{11}, \Phi_2, \ldots, \Phi_k, \Gamma_2, \ldots, \Gamma_k)' \in \mathbf{R} \times (1,\infty)^{2k-2}.$$

Thus, the parameter $\zeta$ is also maximal invariant function of $\theta$ under $g$ and $\left(\sum_{i,j\in\mathcal{I}} Y_{ij}, G_2, \ldots, G_k, F_2, \ldots, F_k\right)$ is also a minimal sufficient statistic, albeit not a natural statistic of an exponential family because of its non-linear relation with $T$.

The statistic $R_1, \ldots, R_k, F_2, \ldots, F_k$ is another candidate for a minimal sufficient statistic, albeit not a natural statistic of an exponential family either. To argue that there is a one-one relation between this statistic and $T$ extract $C_j$ from the expression (9) for $F_j$ recursively for $j = k, \ldots, 2$.

The development parameter $\Phi_j$ has previously been studied in the literature. A theorem of Verrall (1991, p493) gives an expression for $\Phi_j$ in terms of $\beta_j = \log(B_j)$ corresponding to (13). Conversely, $A_i$ and $B_j$ can be expressed in terms of $\Gamma_i$ and $\Phi_j$ as

$$\frac{A_i}{A_1} = (\Gamma_i - 1)\prod_{\ell=2}^{i-1}\Gamma_\ell, \qquad \frac{B_j}{B_1} = (\Phi_j - 1)\prod_{\ell=2}^{j-1}\Phi_\ell, \qquad \text{for } i,j > 1, \tag{14}$$

with the convention that empty products are unity. This is proved by inserting the expressions of (13) on the right hand side. An alternative expression for $B_j$ arises by noting that (13) implies $\sum_{\ell=1}^{k} B_\ell/B_1 = \prod_{\ell=2}^{k}\Phi_\ell$. Inserting this in (14) gives

$$\frac{B_j}{\sum_{\ell=1}^{k} B_\ell} = \frac{\Phi_j - 1}{\prod_{\ell=j}^{k}\Phi_\ell} \qquad \text{for } j = 2, \ldots, k.$$

Choosing the identification $\sum_{j=1}^{k} B_j = 1$ in this expression gives the result of Kremer (1985, p133).

A simple expression for the predictions of the model can be constructed. Use the expression for $\mu_{ij}$ in (6) and insert the maximum likelihood estimators for $\mu_{11}$ and $\sum_{\ell=2}^{i}\Delta\alpha_\ell$ reported in Theorem 3 to get, for $j = 1$,

https://doi.org/10.1017/S1748499500000610 Published online by Cambridge University Press

$$\tilde{Y}_{i1} = \exp(\hat{\mu}_{i1}) = \left(\frac{R_1}{\prod_{\ell=2}^{k} F_\ell}\right)\left(\frac{R_i}{R_1}\prod_{\ell=2}^{i} F_{k+2-\ell}\right) = \frac{R_i}{\prod_{\ell=2}^{k+1-i} F_\ell}. \tag{15}$$

In particular $\tilde{Y}_{k1} = R_k = Y_{k1}$. To get an expression for the predictor for $j > 1$ combine this with the expression for $\sum_{\ell=2}^{j} \Delta\beta_\ell = \log(B_j/B_1)$ in (14) to get

$$\tilde{Y}_{ij} = \exp(\hat{\mu}_{ij}) = R_i(F_j - 1)\frac{\prod_{\ell=2}^{j-1} F_\ell}{\prod_{\ell=2}^{k+1-i} F_\ell} \qquad \text{for } j > 1. \tag{16}$$

For the lower triangle, so $k - i + 1 < j \leq k$, this expression reduces to

$$\tilde{Y}_{ij} = R_i(F_j - 1)\prod_{\ell=k+2-i}^{j-1} F_\ell \quad \text{so} \quad \sum_{j=k+2-i}^{k} \tilde{Y}_{ij} = R_i \prod_{j=k+2-i}^{k} F_j. \tag{17}$$

This expression matches the reserve forecast in (3) since $R_i = Z_{i,k-i+1}$ and it was previously given by Mack (1991, p97). We also give a simplified expression for the prediction for the upper triangle, so $1 < j \leq k + 1 - i$. This could be used for checking the specification of the Poisson model for instance through residuals. For the upper triangle the prediction is, for $j > 1$,

$$\tilde{Y}_{ij} = R_i(F_j - 1)\prod_{\ell=j}^{k+1-i} \frac{1}{F_\ell}. \tag{18}$$

## 4. Boundary Outcomes

In applications the boundary outcomes frequently arise in which case the run-off triangle has one zero column or zero row. In those cases one of the development factors will be one (or infinite) so the sufficient statistic $T$ is on the boundary of its convex support according to Theorem 2. This implies that estimation cannot be done through Theorem 3. Two principled options are then available: either to look for different data or to look for a different statistical model. In the insurance context the latter will typically be appropriate. By formulating a different statistical model it is possible to show that the default option of simply using the chain-ladder technique regardless of the boundary issue actually can be viewed as a maximum likelihood approach.

Due to the symmetry of the model in rows and columns it suffices to consider the case where all incremental observations in a given development year, $j^\dagger > 1$ say, are zero, that is $C_{j^\dagger} = 0$. The default solution of using the

https://doi.org/10.1017/S1748499500000610 Published online by Cambridge University Press

chain-ladder technique evolves around the development factors. The development factors $F_j$ and $G_i$ are defined as in (9) and (10), noting that $C_{j^\dagger} = 0$ results in $F_{j^\dagger} = 1$. The traditional chain-ladder prediction in this case is the same as in (16). We will formulate a statistical model in which this is the maximum likelihood solution.

The solution proposed here could be extrapolated to cases of multiple zero columns and/or zero rows. An exception is when the first column or the first row is zero. In the case $C_1 = 0$ one could consider the triangle as commencing at column 2. This results in a full triangle of dimension $k-1$ which could be analysed the usual way. A consequence is, of course, that no chain-ladder-based predictions can be made for row $k$, since the development factor $F_2$ is not available.

Now, since $C_{j^\dagger} = 0$, it could be reasonable to assume $\mathsf{E}(Y_{ij^\dagger}) = 0$ for all $1 \le i \le k+1-j^\dagger$. This assumption has the particular consequence that when forecasting the reserve then all increments in developmental year $j^\dagger$ will be zero, that is $\tilde{Y}_{ij^\dagger} = 0$. The assumption should therefore be considered carefully against any contextual information in relation to the data. Let $\mathcal{I}_{/j^\dagger}$ denote the subset of $\mathcal{I}$ in which column $j^\dagger$ is excluded.

A Poisson model over $\mathcal{I}_{/j^\dagger}$ is defined through independent Poisson distributed variables with log expectation $\mu_{ij}$ defined by (4) as before. The modified canonical parameter is

$$\xi_{/j^\dagger} = (\mu_{11}, \Delta\alpha_2, \ldots, \Delta\alpha_k, \Delta\beta_2, \ldots, \Delta\beta_{j^\dagger-1}, \Delta_2\beta_{j^\dagger+1}, \Delta\beta_{j^\dagger+2}, \ldots, \Delta\beta_k),$$

using the definition $\Delta_2\beta_{j^\dagger+1} = \Delta\beta_{j^\dagger+1} + \Delta\beta_{j^\dagger} = \beta_{j^\dagger+1} - \beta_{j^\dagger-1}$. The likelihood function (11) is modified as

$$\begin{aligned}\log \mathsf{L}(\xi_{/j^\dagger};\, Y_{/j^\dagger}) = \mu_{11} \sum_{i,j\in\mathcal{I}_{/j^\dagger}} Y_{ij} + \sum_{i=2}^{k} R_i \sum_{\ell=2}^{i} \Delta\alpha_\ell + \sum_{j=2}^{j^\dagger-1} C_j \sum_{\ell=2}^{j} \Delta\beta_\ell \\ + \sum_{j=j^\dagger+1}^{k} C_j \left( \sum_{\ell=2}^{j^\dagger-1} \Delta\beta_\ell + \Delta_2\beta_{j^\dagger+1} + \sum_{\ell=j^\dagger+2}^{j} \Delta\beta_\ell \right) \\ - \sum_{i,j\in\mathcal{I}_{/j^\dagger}} \exp(\mu_{ij}) - \sum_{i,j\in\mathcal{I}_{/j^\dagger}} \log(Y_{ij}!). \end{aligned} \tag{19}$$

Just as before, this is an exponential family with minimal sufficient statistic $T_{/j^\dagger} = \left(\sum_{i,j\in\mathcal{I}_{/j^\dagger}} Y_{ij}, R_1, \ldots, R_k, C_1, \ldots, C_{j^\dagger-1}, C_{j^\dagger+1}, \ldots, C_k\right)$, and the natural parameter is a linear transformation of $\xi_{/j^\dagger}$. The convex support of $T_{/j^\dagger}$ can be characterised as before.

https://doi.org/10.1017/S1748499500000610 Published online by Cambridge University Press

**Theorem 4** *Consider the Poisson model over the upper triangle $\mathcal{I}$. The convex support of the sufficient statistic T is closed. T is interior to its convex support if and only if $F_2, \ldots, F_{j^\dagger-1}, F_{j^\dagger+1}, \ldots, F_k, G_2, \ldots, G_k$ are all larger than one and finite.*

The maximum likelihood estimators can now be found with the convention that $F_{j^\dagger} = 1$.

**Theorem 5** *Consider the Poisson model over $\mathcal{I}_{/j^\dagger}$ with canonical parameter, $\xi_{/j^\dagger}$. The maximum likelihood estimator for $\xi_{/j^\dagger}$ is unique if and only if $T_{/j^\dagger}$ is interior to its convex support. It is given by*

$$\begin{aligned}
\Delta\widehat{\alpha}_i &= \Delta \log R_i + \log F_{k+2-i}, \\
\Delta\widehat{\beta}_j &= \Delta \log C_j + \log G_{k+2-j} \qquad \text{for } j \neq j^\dagger, j^\dagger + 1, \\
\Delta_2\widehat{\beta}_{j^\dagger+1} &= \Delta_2 \log C_{j^\dagger+1} + \log G_{k+2-j^\dagger+1} + \log G_{k+2-j^\dagger}, \\
\widehat{\mu}_{11} &= \log R_1 - \sum_{l=2}^{k} \log F_l = \log C_1 - \sum_{l=2}^{k} \log G_l.
\end{aligned}$$

*The maximum likelihood estimator for* $\exp(\mu_{ij})$ *is given by (15), (16).*

It is interesting to note that $F_{j^\dagger} = 1$ implies that $\Delta\alpha_{k+2-j^\dagger}$ is estimated simply by $\Delta \log R_{k+2-j^\dagger}$. The interpretation is that since the rows $k+1-j^\dagger$ and $k+2-j^\dagger$ both have length $j^\dagger - 1$ due to the missing column there is no need to correct $\Delta \log R_{k+2-j^\dagger}$ with a development factor.

## Acknowledgements

The second author received financial support from ESRC grant RES-000-27-0179.

## References

Barndorff-Nielsen, O.E. (1978). *Information and exponential families.* Wiley, New York.

Cox, D.R. & Hinkley, D.V. (1974). *Theoretical statistics.* Chapman and Hall, London.

Eaton, M.L. (2007). *Multivariate statistics: a vector space approach.* Lecture Notes–Monograph Series 53. Beachwood, Institute of Mathematical Statistics, OH.

England, P.D. & Verrall, R.J. (1999). Analytic and bootstrap estimates of prediction errors in claims reserving. *Insurance: Mathematics and Economics*, **25**, 281–293.

England, P.D. & Verrall, R.J. (2002). Stochastic claims reserving in general insurance. *British Actuarial Journal*, **8**, 519-544.

Hachemeister, C.A. & Stanard, J.N. (1975). IBNR claims count estimation with static lag functions. Unpublished manuscript.

https://doi.org/10.1017/S1748499500000610 Published online by Cambridge University Press

KREMER, E. (1982). IBNR-claims and the two-way model of ANOVA. *Scandinavian Actuarial Journal*, 47-55.

KREMER, E. (1985). *Einführung in die Versicherungsmathematik*. Vandenhoek & Ruprecht, Göttingen.

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

KUANG, D., NIELSEN, B. & NIELSEN, J.P. (2008b). Forecasting with the age-period-cohort model and the extended chain-ladder model. *Biometrika*, **95**, 987-991.

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

MACK, T. & VENTER, G. (2000). A comparison of stochastic models that reproduce chain-ladder reserve estimates. *Insurance: Mathematics and Economics*, **26**, 101-107.

ROCKAFELLAR, R.T. (1970). *Convex analysis*. Princeton University Press, Princeton, NJ.

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

VERRALL, R.J. (1994). Statistical methods for the chain-ladder technique. *Casualty Actuarial Society Forum, Spring 1994*, 393-446.

VERRALL, R.J. & ENGLAND, P.D. (2000). Comments on "A comparison of stochastic models that reproduce chain-ladder reserve estimates", by Mack and Venter. *Insurance: Mathematics and Economics*, **26**, 109-111.

https://doi.org/10.1017/S1748499500000610 Published online by Cambridge University Press

## APPENDIX

## PROOFS OF THEOREMS

**Proof of Theorem 1.** This is done by induction using formula (6), mimicking the proof of Kuang *et al.* (2008a, Theorem 1).

As a preliminary step, if $k = 1$ then $\xi = \mu_{11}$ and the statement is true.

To initialise the argument let $k = 2$. If $\mu_{11}^{\dagger} \neq \mu_{11}^{\ddagger}$ the statement holds. If $\mu_{11}^{\dagger} = \mu_{11}^{\ddagger}$, but $\Delta\alpha_2^{\dagger} \neq \Delta\alpha_2^{\ddagger}$ or $\Delta\beta_2^{\dagger} \neq \Delta\beta_2^{\ddagger}$ then $\mu_{21}^{\dagger} \neq \mu_{21}^{\ddagger}$ or $\mu_{12}^{\dagger} \neq \mu_{12}^{\ddagger}$ by formula (6).

For the induction step consider $\mu_{1,j}$. Assume that $\mu_{11}^{\dagger} = \mu_{11}^{\ddagger}$ and $\Delta\beta_{\ell}^{\dagger} = \Delta\beta_{\ell}^{\ddagger}$ for $\ell < j$ but $\Delta\beta_j^{\dagger} \neq \Delta\beta_j^{\ddagger}$. Then $\mu_{1j}^{\dagger} \neq \mu_{1j}^{\ddagger}$ by formula (6). If $\Delta\beta_{\ell}^{\dagger} = \Delta\beta_{\ell}^{\ddagger}$ for all $\ell \leq k$ apply the same argument for $\mu_{i,1}$.

**Proof of Theorem 2.** *Characterising convex support of T.* Let $Y = \{Y_{ij}$, for $i, j \in \mathcal{I}\}$ denote the original data triangle so $Y \in \mathbf{R}^n$, where $n = k(k+1)/2$. The convex support of $Y$ is the set $\mathcal{C}_Y = \{Y \in \mathbf{R}^n$ so $Y_{ij} \geq 0$ for all $i, j \in \mathcal{I}\}$. The set $\mathcal{C}_Y$ is a convex polyhedral cone (Rockafellar, 1970, p170). Since $T$ is a linear transformation of $Y$ then its convex support, $\mathcal{C}_T$ say, is a closed polyhedral convex set (Rockafellar, 1970, Theorems 19.1, 19.3). The relative interior of $\mathcal{C}_Y$ maps into the relative interior of $\mathcal{C}_T$ (Rockafellar, 1970, Theorem 6.6). Some boundary points of $\mathcal{C}_Y$ map into boundary points of $\mathcal{C}_T$ whereas others map into interior points of $\mathcal{C}_T$. The issue is to characterise which are which.

*Argue that* (*ii*) *implies* (*i*). Suppose $T'$ is the point implied by (*ii*). Since the relative interior of $\mathcal{C}_Y$ maps into the relative interior of $\mathcal{C}_T$ it suffices to construct a $Y' \in \mathrm{ri}(\mathcal{C}_Y)$ that maps into $T'$. First, for $j = 1$ then, by (15), (9),

$$Y'_{i1} = \frac{R_i}{\prod_{\ell=2}^{k+1-i} F_\ell} = R_i \prod_{\ell=2}^{k+1-i} \frac{S_{k+1-\ell}}{S_{k+1-\ell} + C_\ell},$$

which is positive by assumption (*ii*). Next, for $j > 1$ then, by (18), (9), the prediction for the upper triangle is

$$Y'_{ij} = R_i \frac{F_j - 1}{\prod_{\ell=j}^{k+1-i} F_\ell} = R_i \frac{C_j}{S_{k+1-j}} \prod_{\ell=j}^{k+1-i} \frac{S_{k+1-\ell}}{S_{k+1-\ell} + C_\ell},$$

which is also positive by assumption (*ii*).

*Argue that* (*i*) *implies* (*ii*). Suppose (*ii*) is not satisfied and let $T'$ be the point implied by (*ii*). It can be argued that $T'$ is a boundary point of $\mathcal{C}_T$ using Rockafellar (1970, Corollary 11.6.2). The idea is to construct a linear function $h(T)$ that achieves its maximum over $\mathcal{C}_T$ at $T'$.

https://doi.org/10.1017/S1748499500000610 Published online by Cambridge University Press

Suppose for instance that $R_i = 0$ for some $i$. Let $h(T) = -e'T$, where $e$ is the vector taking the value one for element $i$ and zero otherwise. Then $h(T) \leq 0$ for all points $T(Y)$ where $Y \in \mathcal{C}_Y$ with maximum $h(T') = 0$ at $T'$.

A similar argument can be made if $C_j = 0$ for some $j$.

Finally, suppose $S_i = 0$ for some $i$. Since $\sum_{i,j\in\mathcal{I}} Y_{ij} = S_i + \sum_{\ell=i+1}^{k} R_\ell + \sum_{\ell=k+1-i}^{k} C_\ell$ then $\sum_{i,j\in\mathcal{I}} Y_{ij} \geq \sum_{\ell=i+1}^{k} R_\ell + \sum_{\ell=k+1-i}^{k} C_\ell$ with equality if and only if $S_i = 0$. Thus, let $h(T) = -e'T$, where $e$ is a vector taking the value one in the first coordinate and minus one in the coordinates corresponding to $R_{i+1}, \ldots, R_k$ and $C_{k+1-i}, \ldots, C_k$. Apply the argument above.

*Argue that* (*ii*) *implies* (*iii*). The definitions of $F_j$ and $G_i$ in (9) and (10) imply that

$$F_j = 1 + C_j/S_{k+1-j}, \qquad G_i = 1 + R_i/S_{i-1}.$$

Thus, if (*ii*) holds so $C_j$, $S_{k+1-j}$, $R_i$ and $S_{i-1}$ are all positive then $F_j$ and $G_i$ are larger than one and finite.

*Argue that* (*iii*) *implies* (*ii*). The above expressions for $F_j$ and $G_i$ imply

$$C_j = (F_j - 1)S_{k+1-j}, \qquad R_i = (G_i - 1)S_{i-1}.$$

Thus, if (*iii*) holds so $F_j$ and $G_i$ are finite then their denominators $S_{k+1-j}$ and $S_{i-1}$ are positive. Since $F_j$ and $G_i$ are also larger than one then $C_j$, $S_{k+1-j}$, $R_i$ and $S_{i-1}$ are all positive.

*Argue that* (*ii*) *implies* (*iv*). This holds since $R_1 = S_1 + C_k$ and $F_j = 1 + C_j/S_{k+1-j}$.

*Argue that* (*iv*) *implies* (*i*). Argue in the same ways as (*ii*) implies (*i*), noting that by (15) and (16) then $Y'_{i1} = R_i \prod_{\ell=2}^{k+1-i} F_\ell^{-1}$ and $Y'_{ij} = R_i(F_j - 1) \times \prod_{\ell=j}^{k+1-i} F_\ell^{-1}$ are both positive when (*iv*) is satisfied.

*Argue that* (*v*) *is an equivalent expression.* Same as (*iv*) as the chain ladder model is symmetric in rows and columns.

**Proof of Theorem 3.** First, for a full exponential family the maximum likelihood estimator is unique if and only if the natural statistic is interior to its convex support (Barndorff-Nielsen, 1978, Theorem 9.13).

Secondly, since the exponential family is regular the $2k-1$ likelihood equations are $T = \mathsf{E}T$ (Barndorff-Nielsen, 1978, Corollary 9.6). Since $\sum_{i,j\in\mathcal{I}} Y_{ij} = \sum_{i=1}^{k} R_i = \sum_{j=1}^{k} C_j$ this in turn implies the equations

$$R_i = \mathsf{E}R_i, \qquad C_j = \mathsf{E}C_j, \qquad \text{for } i, j = 1, \ldots k. \tag{20}$$

*Estimating the development factors.* The development factor $F_j$ can be written in terms of the minimal sufficient statistic as

https://doi.org/10.1017/S1748499500000610 Published online by Cambridge University Press

$$F_j = \frac{\sum_{i=1}^{k+1-j} R_i - \sum_{\ell=j+1}^{k} C_\ell}{\sum_{i=1}^{k+1-j} R_i - \sum_{\ell=j}^{k} C_\ell}.$$

Therefore, by the likelihood equations $F_j$ is maximum likelihood estimator for the parameter

$$\Phi_j = \frac{\sum_{i=1}^{k+1-j} \mathsf{E}R_i - \sum_{\ell=j+1}^{k} \mathsf{E}C_\ell}{\sum_{i=1}^{k+1-j} \mathsf{E}R_i - \sum_{\ell=j}^{k} \mathsf{E}C_\ell} = \frac{\sum_{i=1}^{k+1-j} \sum_{\ell=1}^{j} \mathsf{E}Y_{i\ell}}{\sum_{i=1}^{k+1-j} \sum_{\ell=1}^{j-1} \mathsf{E}Y_{i\ell}}.$$

Insert that $\mathsf{E}Y_{ij} = \exp(\mu_{ij})$ with $\mu_{ij} = \mu_{11} + \sum_{\ell=2}^{i} \Delta\alpha_\ell + \sum_{\ell=2}^{j} \Delta\beta_\ell$. Note that the expression factorises into $\mu_{11}$, $\Delta\alpha_i$ and $\Delta\beta_j$ terms, of which the $\mu_{11}$ and $\Delta\alpha_i$ terms cancel so

$$\Phi_j = \frac{\sum_{m=1}^{j} \exp\left(\sum_{\ell=2}^{m} \Delta\beta_\ell\right)}{\sum_{m=1}^{j-1} \exp\left(\sum_{\ell=2}^{m} \Delta\beta_\ell\right)}.$$

The expression for $G_i$ and $\Gamma_i$ arise similarly.

*Estimating the difference parameters.* The equations (20) also imply that $\Delta \log R_i = \Delta \log \mathsf{E}R_i$, where

$$\Delta \log \mathsf{E}R_i = \log\left\{\sum_{j=1}^{k+1-i} \exp\left(\mu_{11} + \sum_{\ell=2}^{i} \Delta\alpha_\ell + \sum_{\ell=2}^{j} \Delta\beta_\ell\right)\right\} - \log\left\{\sum_{j=1}^{k+2-i} \exp\left(\mu_{11} + \sum_{\ell=2}^{i-1} \Delta\alpha_\ell + \sum_{\ell=2}^{j} \Delta\beta_\ell\right)\right\},$$

which in turn reduces to

$$\Delta \log \mathsf{E}R_i = \Delta\alpha_i - \log \Phi_{k+2-i}.$$

Equating $\mathsf{E}R_i = R_i$ and $\Phi_{k+2-i} = F_{k+2-i}$ implies the desired expression for the maximum likelihood estimator for $\Delta\alpha_i$. The $\Delta\beta_j$ terms are dealt with similarly.

*Estimating the level.* The first expression for $\hat{\mu}_{11}$ arises from the likelihood equation for $R_k = Y_{k1}$. This is

$$\log(R_k) = \log(\mathsf{E}R_k) = \mu_{11} + \sum_{\ell=2}^{k} \Delta\alpha_\ell.$$

Insert, the estimators for $\Delta\alpha_\ell$ to get the desired expression.

https://doi.org/10.1017/S1748499500000610 Published online by Cambridge University Press

The second expression arises in a similar way from the likelihood equation $C_k = \mathsf{E}C_k$.

**Proof of Theorem 4.** Mimick the proof of Theorem 2.

**Proof of Theorem 5.** Mimick the proof of Theorem 3. Even though terms involving column $j^\dagger$ need to be handled with care, the result falls out as desired.

https://doi.org/10.1017/S1748499500000610 Published online by Cambridge University Press