B.A.J. **4**, IV, 903-923 (1998)

A STOCHASTIC MODEL UNDERLYING THE CHAIN-LADDER TECHNIQUE

BY A. E. RENSHAW AND R. J. VERRALL

ABSTRACT

This paper presents a statistical model underlying the chain-ladder technique. This is related to other statistical approaches to the chain-ladder technique which have been presented previously. The statistical model is cast in the form of a generalised linear model, and a quasi-likelihood approach is used. It is shown that this enables the method to process negative incremental claims. It is suggested that the chain-ladder technique represents a very narrow view of the possible range of models.

KEYWORDS

Chain-Ladder Technique; Claims Reserving; Generalised Linear Models; Run-Off Triangles

CONTACT ADDRESS

Dr R. J. Verrall, Department of Actuarial Science & Statistics, The City University, Northampton Square, London EC1V 0HB, U.K. Tel: +44(0)171-477-8476; Fax: +44(0)171-477-8838; E-mail: r.j.verrall@city.ac.uk

1. INTRODUCTION

This paper is concerned with the estimation of outstanding claims. For ease of exposition, and without loss of generality, consider a triangle of data classified according to an index for accident year i, and an index for reporting delay j. Denote the claim amounts reported with accident year index i and delay index j by y_{ij}. Suppose that data have been collected for $i = 1, 2, \ldots, n$ and $j = 1, 2, \ldots, n$. Then the complete set of data is:

$$\{y_{ij} : i = 1, 2, \ldots, n; \quad j = 1, 2, \ldots, n - i + 1\}$$

which may be presented as a run-off triangle:

$$\begin{array}{lllll} y_{11} & y_{12} & \cdots & \cdots & y_{1n} \\ y_{21} & y_{22} & \cdots & y_{2,n-1} & \\ \vdots & & & & \\ y_{n1}. & & & & \end{array}$$

Denote the cumulative claim amounts with accident year index i reported up to, and including, delay index j by C_{ij}, so that:

$$C_{ij} = \sum_{k=1}^{j} y_{ik}. \tag{1.1}$$

Thus, we may either consider the increases in the claim amounts with each delay index, as above, or the cumulative claim amounts:

$$\begin{array}{lllll} C_{11} & C_{12} & \cdots & \cdots & C_{1n} \\ C_{21} & C_{22} & \cdots & C_{2,n-1} & \\ \vdots & & & & \\ C_{n1}. & & & & \end{array}$$

The aim is to obtain estimates of $\{C_{in} : i = 2, 3, \ldots, n\}$.

The chain-ladder technique was conceived as a deterministic method for predicting claim amounts. It is applied to cumulative claim amounts, and is designed to predict future incremental claim amounts in the empty cells of the well-defined triangular region to the immediate south-east of the run-off triangle. We call this region the target triangle. The run-off triangle may be truncated on its eastern fringes. It is also possible to generalise the geometrical configuration further through the removal of a symmetric triangular region in the north west corner of the run-off triangle and to allow for inflation. We also remark that, for the purposes of this paper, we are not expressly concerned with projections to the east of the run-off triangle.

Notwithstanding certain well-known shortcomings, the technique continues to occupy a prominent position with many practitioners as a claims reserving tool. There would appear to be some doubt as to the precise origin of the chain-ladder technique, but claims reserving techniques, in general, have generated a large body of research literature in the intervening years. One substantial strand of this literature is concerned with the development of stochastic claims reserving techniques which have clear advantages over deterministic techniques, such as provision for conducting diagnostic checks and the production of confidence intervals.

This paper identifies a statistical procedure which is exactly equivalent to the chain-ladder technique, in almost all circumstances. It should be noted that the method cannot be applied if the column sum of incremental claims for any development year is negative (this will be discussed in more detail in Section 2). We have always held the view that it is vital to subject any claims reserving procedure to a full statistical review, and we believe that the model presented in this paper provides an important framework to do this for the chain-ladder technique.

Broadly speaking, a stochastic claims reserving process involves three stages:

— *stage one*: the specification of a flexible parameterised model structure;

— *stage two*: a means of fitting the structure to the run-off data coupled with the means to conduct diagnostic checks on the fitted model; and
— *stage three*: a means of projecting the fitted structure into the target triangle.

Note that we assign the specification of a modelling distribution to stage two, since this leads to the construction of a likelihood or quasi-likelihood function which is maximised to obtain parameter estimates.

Thus, we are confronted with a wide choice of possibilities which is reflected in the number of published works on the subject. For the specific purpose of this paper, we find it both helpful and informative to relate the stochastic claims reserving process, whenever possible, to the generalised linear modelling technique, especially in relation to the first two stages described above.

A significant landmark in the development of stochastic versions of the chain-ladder technique was made by Kremer (1982), in which, in our opinion, he establishes the nature of the parameterised model structure which is inherent in the chain-ladder technique. This is equivalent to stage one of the stochastic claims reserving process, but it should be noted that Kremer only identified one of two possible ways of building the structure into the process as a whole. The structure is comprised of the linear predictor:

$$\eta_{ij} = \mu + \alpha_i + \beta_j \tag{1.2}$$

based on parameters corresponding to accident year i and delay j, which is connected to the logged incremental claim amounts. Section 3 contains more details of the theory of generalised linear models, including linear predictors. Implicit in Kremer's work is the suggestion that the log transformation should be applied to the incremental claim amounts. The other possibility is to apply the log transformation to the expected values. Thus, Kremer, in specifying stage two of the stochastic claims reserving process, elects to model the incremental data by imposing the log-normal distribution. Renshaw (1989), Verrall (1989, 1990, 1991a, 1991b), motivated by Kremer (1982), have investigated many different facets of a stochastic claims reserving process based on the log-normal assumption, taken in conjunction with the predictor (1.2).

Mack (1994) has criticised certain aspects of this work. This criticism is justified in so far as the model referred to above, derived by Kremer (1982), is not exactly equivalent to the chain-ladder technique. In Section 2 of this paper we will derive a generalised linear model which is exactly equivalent to the model which underpins the chain-ladder technique (noting the exception mentioned above). We believe that this equivalence is well known to a number of actuaries, but has not before been expressed in terms of generalised linear models. The contributions that this paper makes are to relate the chain-ladder technique directly to a generalised linear model, to show that it is not the most appropriate model for claims data, and to show how the model may be adapted in a straightforward way in order to be appropriate for claims data.

2. THE MODEL UNDERLYING THE CHAIN-LADDER TECHNIQUE

In this section we describe the connection between the chain-ladder technique and the following statistical model for incremental claims y_{ij} :

$$Y_{ij} \sim \text{Poisson with mean } m_{ij}, \text{ independently } \forall i,j \tag{2.1}$$

where:

$$\log m_{ij} = \mu + \alpha_i + \beta_j \tag{2.2}$$

and $\alpha_1 = \beta_1 = 0$.

We assume throughout that:

$$\sum_{i=1}^{n-j+1} y_{ij} \geq 0 \quad \text{for all } j. \tag{2.3}$$

The reason for this assumption will become clear after equation (2.5). Note that we do not assume that all the incremental claims are non-negative, but just that the column totals are non-negative. Note that the incremental claims are random variables, which is the reason for using upper case for y_{ij}. There are a number of points which should be made about this model. Firstly, the structure is the same as that used by Kremer (1982), except that it is embedded in a different modelling distribution; Kremer used a lognomal distribution. Secondly, it may seem strange to use a Poisson distribution for modelling claim amounts; it is strange, and this is reflected in the results. Thirdly, the specification of the Poisson modelling distribution does not mean that the model can only be applied to data which are positive integers; it is easy to write down a quasi-likelihood which has all the characteristics of a Poisson likelihood, without actually referring directly to the probability function for the Poisson random variable. This means that the model can be applied to negative incremental claims, etc., and the results are always the same as those by the chain-ladder technique (when $\sum_{i=1}^{n-j+1} y_{ij} \geq 0$ for all j). More details of this are contained in Sections 3 and 4. We also need to state that we have elected to present a different approach, in establishing the transition from the chain-ladder technique to the Poisson modelling distribution, to that which appears in Mack (1991).

We find it easiest to retain the assumption that the data have a Poisson distribution at the moment, although in all that follows in this section it is only the form of the likelihood which is important. Obviously, this model is very reasonable when the triangle consists of numbers of claims, rather than claim amounts. In order to show the equivalence with the chain-ladder technique, we first assume that the triangle consists of the numbers of claims, classified according to accident year i, and reporting delay j.

The maximum likelihood estimates for the model given by (2.1) and (2.2) may be obtained using GLIM (Francis, Green & Payne, 1993). The estimates of the total number of claims in each accident year may be obtained from the sums:

$$\hat{C}_{in} = C_{i,n-i+1} + \sum_{j=n-i+2}^{n} e^{\hat{\mu}+\hat{\alpha}_i+\hat{\beta}_j} \quad (i = 2, 3, \ldots, n). \tag{2.4}$$

where $\hat{\mu}$, $\hat{\alpha}_i$ and $\hat{\beta}_j$ are the maximum likelihood estimates of the parameters.

The same estimates of $\{C_{in} : i = 2, 3, \ldots, n\}$ are obtained using the following conditional likelihood. Denote the conditional probability that a claim with accident year index i, which we know has been reported, is reported with delay index j by $p_{(i)j}$. For accident year i, this is the probability that the delay index is j, given that it is less than, or equal to, $n-i+1$. The usual assumption of stationarity for the reporting process implies that the probability that a claim is reported in each delay year does not depend on the accident year. Then:

$$p_{(i)j} = \frac{p_j}{\sum_{k=1}^{n-i+1} p_k}. \tag{2.5}$$

where p_j is the (unconditional) probability that a claim is reported in delay year j and $\sum_{k=1}^{n} p_k = 1$.

This last assumption implies that we are assuming that all claims are reported by the end of delay year n. It is in line with the assumption made in Section 1, and by the chain-ladder technique generally, that we are not concerned with forecasting beyond C_{in}, i.e. beyond the latest delay year already observed. It also explains why the restriction (2.3) is necessary, since otherwise we could not use probabilities.

The conditional likelihood L_c for these data can be obtained using the multinomial distribution, and is given by:

$$L_c = \prod_{i=1}^{n} \left(\frac{C_{i,n-i+1}!}{\prod_{j=1}^{n-i+1} y_{ij}!} \prod_{j=1}^{n-i+1} p_{(i)j}^{y_{ij}} \right). \tag{2.6}$$

This is a *conditional* likelihood, since it conditions on the latest row totals, $C_{i,n-i+1}$. The fact that this is a *conditional* likelihood and is equivalent to the Poisson model is an important point, and it has often been noted that the chain-ladder technique conditions on the latest row totals. Maximising this likelihood gives the following estimates for the total number of claims:

$$\hat{C}_{in} = \frac{C_{i,n-i+1}}{1 - \sum_{j=n-i+2}^{n} \hat{p}_j} \quad (i = 2, 3, \ldots, n) \tag{2.7}$$

where $\hat{p}_j$ is the estimate of p_j obtained by maximising L_c.

It can be shown that the Poisson model gives the same estimates as the conditional likelihood; i.e. that equations (2.4) and (2.7) give the same results.

It is sometimes easier to consider the estimate for the accident year which has been reported up to delay index j instead of equation (2.7). The estimate of ultimate cumulative claims for accident year $n-j+1$ is:

$$\hat{C}_{n-j+1,n} = \frac{C_{n-j+1,j}}{1-\sum_{k=j+1}^{n} \hat{p}_k}. \tag{2.8}$$

This can be compared with the chain-ladder estimates:

$$\hat{C}_{n-j+1,n} = C_{n-j+1,j}\hat{\lambda}_{j+1}\hat{\lambda}_{j+2}\ldots\hat{\lambda}_n \tag{2.9}$$

where:

$$\hat{\lambda}_j = \frac{\sum_{i=1}^{n-j+1} C_{ij}}{\sum_{i=1}^{n-j+1} C_{i,j-1}}. \tag{2.10}$$

Rosenberg (1990) derived a simple technique for obtaining the estimates $\{\hat{p}_j: j=1,2,\ldots,n\}$, which is recursive. This can be stated simply as follows. Suppose that we have estimates of p_n, p_{n-1}, ..., p_{j+1}. Then the estimate of p_j is $\hat{p}_j$, where:

$$\hat{p}_j = \frac{y_{1j}+y_{2j}+\ldots+y_{n-j+1,j}}{C_{1n}+\dfrac{C_{2,n-1}}{1-\hat{p}_n}+\ldots+\dfrac{C_{n-j+1,j}}{1-\hat{p}_{j+1}-\ldots-\hat{p}_n}} \tag{2.11}$$

$$= \frac{y_{1j}+y_{2j}+\ldots+y_{n-j+1,j}}{C_{1n}+\hat{C}_{2,n}+\ldots+\hat{C}_{n-j+1,n}}. \tag{2.12}$$

Note that $\hat{p}_n = \dfrac{y_{1n}}{C_{1n}}$, which begins the recursion.

It is straightforward to show that the resulting estimates $\{\hat{p}_j: j=1,2,\ldots,n\}$ are the maximum likelihood estimates for the likelihood L_C.
This is proved in Theorem 1 of the Appendix.

We now show that Rosenberg's technique is equivalent to the chain-ladder technique. This will show that the chain-ladder technique is just a simple way to find the maximum likelihood estimates for the conditional likelihood L_C, and thus also for the Poisson model (2.1) and (2.2). It can be seen from equations (2.8) and (2.9) that:

$$\hat{\lambda}_{j+1}\hat{\lambda}_{j+2}\cdots\hat{\lambda}_n = \frac{1}{1-\hat{p}_{j+1}-\hat{p}_{j+2}-\ldots-\hat{p}_n} \tag{2.13}$$

and

$$\hat{\lambda}_j\hat{\lambda}_{j+1}\cdots\hat{\lambda}_n = \frac{1}{1-\hat{p}_j-\hat{p}_{j+1}-\ldots-\hat{p}_n}. \tag{2.14}$$

Thus:

$$\hat{\lambda}_j\hat{\lambda}_{j+1}\ldots\hat{\lambda}_n = \frac{1}{\dfrac{1}{\hat{\lambda}_{j+1}\hat{\lambda}_{j+2}\ldots\hat{\lambda}_n}-\hat{p}_j} \tag{2.15}$$

and

$$\hat{\lambda}_j = \frac{1}{1-\hat{p}_j\hat{\lambda}_{j+1}\hat{\lambda}_{j+2}\cdots\hat{\lambda}_n}. \tag{2.16}$$

We show, by backwards induction, that the estimate of $\hat{\lambda}_j$, obtained from equation (2.16) using $\hat{p}_j$ from equation (2.12), is the same as that given by equation (2.10). This is certainly true for $j = n$, since:

$$\begin{aligned}
\hat{\lambda}_n &= \frac{1}{1-\hat{p}_n} \quad \text{from (2.16)} \\
&= \frac{1}{1-\dfrac{y_{1n}}{C_{1n}}} \quad \text{from (2.12) with } j=n \\
&= \frac{C_{1n}}{C_{1n}-y_{1n}} \\
&= \frac{C_{1n}}{C_{1,n-1}}
\end{aligned}$$

which is the same as equation (2.10) when $j = n$.

Now suppose that the result holds for $j+1, j+2, \ldots, n$. We may thus write equation (2.11) as:

$$\hat{p}_j = \frac{y_{1j} + y_{2j} + \ldots + y_{n-j+1,j}}{C_{1n} + C_{2,n-1}\hat{\lambda}_n + \ldots + C_{n-j+1,j}\hat{\lambda}_{j+1} \ldots \hat{\lambda}_n}.$$

Substituting into equation (2.16):

$$\hat{\lambda}_j = \frac{1}{1 - \dfrac{y_{1j} + y_{2j} + \ldots + y_{n-j+1,j}}{C_{1n} + C_{2,n-1}\hat{\lambda}_n + \ldots + C_{n-j+1,j}\hat{\lambda}_{j+1} \ldots \hat{\lambda}_n}\hat{\lambda}_{j+1}\hat{\lambda}_{j+2} \ldots \hat{\lambda}_n}. \tag{2.17}$$

Theorem 2 from the Appendix, proves that:

$$C_{1n} + C_{2,n-1}\hat{\lambda}_n + \ldots + C_{n-j+1,j}\hat{\lambda}_{j+1} \ldots \hat{\lambda}_n = \hat{\lambda}_{j+1} \ldots \hat{\lambda}_n \sum_{i=1}^{n-j+1} C_{ij}.$$

Hence, equation (2.17) may be rewritten as:

$$\begin{aligned}
\hat{\lambda}_j &= \frac{1}{1 - \dfrac{y_{1j} + y_{2j} + \ldots + y_{n-j+1,j}}{\hat{\lambda}_{j+1} \ldots \hat{\lambda}_n \sum_{i=1}^{n-j+1} C_{ij}}\hat{\lambda}_{j+1}\hat{\lambda}_{j+2} \ldots \hat{\lambda}_n} \\
&= \frac{\sum_{i=1}^{n-j+1} C_{ij}}{\sum_{i=1}^{n-j+1} C_{ij} - (y_{1j} + y_{2j} + \ldots + y_{n-j+1,j})} \\
&= \frac{\sum_{i=1}^{n-j+1} C_{ij}}{\sum_{i=1}^{n-j+1} C_{i,j-1}}
\end{aligned}$$

which is the chain-ladder estimate. Hence, it has been proved by induction that Rosenberg's technique is equivalent to the chain-ladder technique.

We have thus shown in this section that the chain-ladder technique gives exactly the same estimates as the model (2.1) and (2.2), under assumption (2.3). We have a number of different ways of obtaining the same estimates of $\{C_{in} : i = 2, 3, \ldots, n\}$: the chain-ladder technique; Rosenberg's technique; and maximum likelihood estimation in the generalised linear model. We would be justified in calling this generalised linear model the chain-ladder linear model, under assumption (2.3). However, we choose to relax the assumption that the data have a Poisson distribution while retaining the same form for the model and the likelihood. Thus, we obtain the same results, but we avoid specifying the distribution of the data. This means that it is possible to apply the chain-ladder linear model to claim amounts (rather than just claim numbers). Further, the linear model can then handle negative incremental claims, under condition (2.3). This is covered more fully in Sections 3 and 4.

Note also that column parameters in the various models are related as well. The parameters for the chain-ladder technique $\{\lambda_j : j = 2, 3, \ldots, n\}$ are related to those of the conditional likelihood $\{p_j : j = 1, 2, \ldots, n\}$ recursively via:

$$\lambda_j \lambda_{j+1} \ldots \lambda_n = \frac{1}{1 - p_j - p_{j+1} - \ldots - p_n}.$$

It should be noted that there are other ways of expressing this relationship; see Kremer (1982) or Verrall (1991b). The column parameters of the generalised linear model are related to the parameters of the chain-ladder technique and the conditional likelihood via:

$$p_j = \frac{e^{\beta_j}}{\sum_{k=1}^{n} e^{\beta_k}} \quad \text{and} \quad \lambda_j = 1 + \frac{e^{\beta_j}}{\sum_{k=1}^{j-1} e^{\beta_k}}.$$

These relationships are derived in Verrall (1991b). Since we are using maximum likelihood estimation, the estimates of the parameters for any of the models can be obtained directly from the estimates for the linear model by substitution. $\hat{C}_{in}$ can be obtained from either (2.4), (2.8) or (2.9) — the same results are obtained in all cases.

The main point to be noted from this section is that we have identified a generalised linear model which gives exactly the same results as the chain-ladder technique, under assumption (2.3).

3. GENERALISED LINEAR MODELS AND CLAIMS RESERVING

This section gives an outline of the theory of generalised linear models relevant to claims reserving. A complete exposition of the statistical background can be found in McCullagh & Nelder (1989). A generalised linear model (GLM) is comprised of two components, one statistical, the other deterministic. The statistical component involves independent response variables:

$$\{Y_u : u \in S\}$$

together with the specification of their first two moments:

$$\mathrm{E}(Y_u) = m_u \qquad \mathrm{Var}(y_u) = \frac{\phi V(m_u)}{\omega_u}$$

where $\phi(>0)$ denotes a scale parameter, ω_u prior weights and $V(.)$ the variance function. In this case, the set S consists of $\{(i, j) : i = 1, 2, \ldots, n; j = 1, 2, \ldots, n-i+1\}$. The deterministic component consists of a linear predictor:

$$\eta_u = \sum_v x_{uv}\theta_v$$

involving a known covariate structure (x_{uv}) with unknown parameters θ_v, which is linked to the mean response through a monotonic differentiable link function g, where:

$$g(m_u) = \eta_u.$$

Parameter estimates for the θ_vs are needed to implement these models. Commonly these are determined by maximising the quasi log-likelihood expression:

$$q(\underline{y};\underline{m}) = \sum_u q_u = \sum_u \omega_u \int_{y_u}^{m_u} \frac{y_u - s}{\phi V(s)} ds. \tag{3.1}$$

We are expressly concerned in this paper with GLMs with power variance functions of the type:

$$V(m_u) = m_u^{\zeta}, \zeta \geq 0.$$

Specific cases to note are the normal or Gaussian distribution ($\zeta=0$, with $\omega_u=1 \ \forall \ u$, $\phi>0$), the Poisson distribution ($\zeta=1$, with $\omega_u=1 \ \forall \ u$, $\phi=1$) and the gamma distribution ($\zeta=2$, with $\omega_u=1 \ \forall \ u$, $\phi>0$). It is a trivial exercise to verify that the quasi log-likelihood expression (3.1) reduces to the appropriate log-likelihood expression for these three cases, with the implication that maximum likelihood estimators of the θ_vs are the same as quasi-likelihood estimators for these three cases. The properties of this class of power variance function GLMs are discussed in more detail in Renshaw (1993, 1994a, 1994b). In particular, we note that the case $\zeta \geq 2$ gives the class of power variance models most suitable for modelling claim amounts.

It should be noted that it is not necessary to name a distribution; all that is needed is a specification for ζ, ω_u and ϕ. Thus, the same estimates as would be obtained if the Poisson modelling distribution were specified can be obtained by setting $\zeta=1$, with $\omega_u=1 \ \forall \ u$, $\phi=1$.

Diagnostic checks of any fitted structure should include a graphical scrutiny of residuals as a basic requirement. There is a choice of residuals, based on the different statistics available to monitor goodness-of-fit. Denote the fitted values under the current model by $\hat{m}_u$. The (unscaled) deviance statistic is:

$$d(\underline{y};\underline{\hat{m}}) = \sum_u d_u = \sum_u 2\omega_u \int_{\hat{m}_u}^{y_u} \frac{y_u - s}{V(s)} ds \quad \{= -2\phi q(\underline{y};\underline{\hat{m}})\}$$

and gives rise to the deviance residuals:

$$r_u = sign(y_u - \hat{m}_u)\sqrt{d_u}$$

while the generalised Pearson statistic:

$$\chi^2(\underline{y};\underline{\hat{m}}) = \sum_u \omega_u \frac{(y_u - \hat{m}_u)^2}{V(\hat{m}_u)}$$

gives rise to (generalised) Pearson residuals:

$$r_u = \frac{(y_u - \hat{m}_u)}{\sqrt{\dfrac{V(\hat{m}_u)}{\omega_u}}}.$$

The stochastic claims reserving technique stemming directly from Kremer (1982) is underpinned by the model:

$$\log(y_{ij}) \sim \mathrm{N}(\mu + \alpha_i + \beta_j, \sigma^2) \quad \text{independently } \forall\, i, j.$$

Expressed as a GLM with power variance function, this is equivalent to modelling independent $\log(y_{ij})$ responses with $m_{ij} = \mathrm{E}[\log(y_{ij})]$, $\zeta = 0$, $\omega_{ij} = 1$ and $\phi = \sigma^2$ in combination with the identity link so that:

$$m_{ij} = \eta_{ij} = \mu + \alpha_i + \beta_j.$$

As an alternative to this formulation, we consider the class of GLMs with power variance function based on independent y_{ij} responses with mean $m_{ij} = \mathrm{E}(y_{ij})$ in combination with the log link and the same linear predictor, so that:

$$\log(m_{ij}) = \eta_{ij} = \mu + \alpha_i + \beta_j. \tag{3.2}$$

This alternative formulation models the incremental claim amounts y_{ij} directly, while preserving the basic structure associated with the chain-ladder technique established by Kremer (1982), by encapsulating it in the predictor-link relationship. It is also more general, since it offers the choice of values for $\zeta(\geq 0)$, $\phi(>0)$ and the prior weights ω_{ij}. In particular, to reproduce the chain-ladder technique results exactly, we specify $\zeta=1$, $\phi=1$, $\omega_{ij}=1$. Whether this is a sensible approach is discussed further in Sections 4, 5 and 6.

A further landmark in the development of this general theme is due to Mack (1991), in which the chain-ladder predictor-link structure (3.2) is modelled in combination with gamma distributed incremental claim amount ($\zeta=2$, $\phi>0$). For a comparison of aspects of the relative merit of using the gamma distribution as opposed to the log-normal distribution, the reader is referred to Firth (1988). We have noted that a further generalisation is possible, by resorting to the class of power variance function GLMs with $\zeta\geq 2$, $\phi>0$, discussed in Renshaw (1994a, 1994b). This includes the inverse-Gaussian distribution as a special case ($\zeta=3$).

4. A Stochastic Chain-Ladder Model

In this section we consider a well-defined square set of units $\{(i, j) : i=1, 2, \ldots, n; j=1, 2, \ldots, n\}$ formed by the union of a run-off triangle and its associated disjoint target triangle. Denote exposures, if present, by e_i. The exposure is a quantity relating to accident year, which can be used to introduce a standardising measure of the volume of business, such as premium income. Define the set of *augmented incremental claim amounts* Y_{ij}, to be:

$$Y_{ij} = \begin{cases} \text{incremental claim amount} & \text{if}(i,j) \in \text{run-off triangle} \\ 0 & \text{if}(i,j) \in \text{target triangle.} \end{cases}$$

We stress that this is purely a matter of computational convenience, and that all entries in the target triangle are weighted out of the ensuing calculations. Then a chain-ladder generalised linear model is defined as follows.

Definition. A stochastic chain-ladder model consists of a GLM in which the augmented incremental claim amounts y_{ij} are modelled as independent responses with:

mean $m_{ij}=\text{E}(y_{ij})$ variance function $\text{V}(m_{ij})=m_{ij}$ scale parameter $\phi>0$

$$\text{prior weights } \omega_{ij} = \begin{cases} 1 \text{ if}(i,j) \in \text{run-off triangle} \\ 0 \text{ if}(i,j) \in \text{target triangle} \end{cases}$$

in combination with the predictor-link:

$$\log(m_{ij}) = \log(e_i) + \mu + \alpha_i + \beta_j.$$

The parameter estimates $\hat{\mu}, \hat{\alpha}_i, \hat{\beta}_j$ are determined by maximum quasi-likelihood estimation and the predicted claim amounts in the target triangle are provided by the matching values:

$$\hat{m}_{ij} = e_i \exp(\hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j).$$

There are some technical points to note. Firstly, the exposures e_i appear as the offset terms $\log(e_i)$ in the linear predictor, the default setting being $e_i = 1\forall$ units (i,j). Secondly, it is well known that predictors of this type are over-parameterised in a technical sense, so that it becomes necessary to set two constraints, typically $\alpha_1 = \beta_1 = 0$. Thirdly, note that it is not essential to pre-specify the value of the scale parameter ϕ, since the parameter estimates, and hence the predicted claim amounts, are invariant to the value of ϕ. In addition, the definition generalises to include the other possible geometric configurations described in Section 1. Implementation of the definition can be automated using the interactive facility of the GLIM software package (Francis, Green & Payne, 1993).

When there are any negative incremental claims, we would strongly advise using the (generalised) Pearson residuals, which, unlike the deviance residuals, are uniquely defined for negative as well as positive values of y_{ij}. This is because the formal evaluation of the integral expression (3.1), on which the values of the deviance residuals are based, gives rise to terms in log y_{ij} which are given zero default settings for negative y_{ij} values. This alone does not inhibit the optimisation of expression (3.1) in the quasi-likelihood parameter estimation procedure, provided also that the $\log m_{ij}$ terms are well defined, that is $m_{ij} > 0 \ \forall \ i, j$. Because of the specific nature of the parameterised structure of the linear predictor (1.2), this condition is met, provided that all the column (and row) sums in the incremental claims run-off triangle are positive. This is consistent with assumption (2.3).

The precise technique encapsulated by the definition has been known to the authors since 1990, when it was brought to our attention by our colleague Steven Haberman through his work into AIDS. Specifically, the technique has been applied to AIDS surveillance data comprising head counts for reporting delays. While it is perfectly natural to use the Poisson distribution in this context, it is quite a radical departure from accepted wisdom if precisely the same stochastic technique is to be applied to incremental claims amounts. These findings, therefore, represent yet one further hurdle which practitioners of the chain-ladder technique will need to reconcile.

5. ILLUSTRATION

In order to effect a comparison, we use the same data as Mack (1994), taken

Table 1. Predicted claim amounts; Poisson-type response model

										Total
5,012	3,257	2,638	898	1,734	2,642	1,828	599	54	172	
2,111.4	4,221.4	3,948.6	2,785.1	2,243.2	1,735.9	714.8	590.8	310.8	172.0	0.
106	4,179	1,111	5,270	3,116	1,817	−103	673	535	–	
1,889.9	3,778.5	3,534.4	2,492.9	2,007.8	1,553.8	639.8	528.8	278.2	154.0	154.
3,410	5,582	4,881	2,268	2,594	3,479	649	603	–	–	
2,699.9	5,398.0	5,049.2	3,561.4	2,868.4	2,219.7	914.0	755.4	397.4	219.9	617.
5,655	5,900	4,211	5,500	2,159	2,658	984	–	–	–	
3,217.8	6,433.4	6,017.7	4,244.5	3,418.6	2,645.5	1,089.4	900.3	473.7	262.1	1,636.
1,092	8,473	6,271	6,333	3,786	225	–	–	–	–	
3,242.8	6,483.6	6,064.6	4,277.6	3,445.3	2,666.1	1,097.9	907.4	477.4	264.2	2,746.
1,513	4,932	5,257	1,233	2,917	–	–	–	–	–	
2,186.2	4,370.9	4,088.5	2,883.8	2,322.6	1,797.4	740.1	611.7	321.8	178.1	3,649.
557	3,463	6,926	1,368	–	–	–	–	–	–	
1,989.8	3,978.3	3,721.2	2,624.7	2,114.0	1,635.9	673.6	556.8	292.9	162.1	5,435.
1,351	5,596	6,165	–	–	–	–	–	–	-	
2,692.7	5,383.6	5,035.7	3,551.9	2,860.8	2,213.8	911.6	753.4	396.4	219.4	10,907.
3,133	2,262	–	–	–	–	–	–	–	–	
1,798.7	3,596.3	3,363.9	2,372.7	1,911.0	1,478.8	609.0	503.3	264.8	146.5	10,650.
2,063	–	–	–	–	–	–	–	–	–	
2,063.0	4,124.7	3,858.2	2,721.3	2,191.8	1,696.1	698.4	577.2	303.7	168.1	16,339.
									Total	52,135.
						Root mean squared error of prediction				17,603.

from the *Historical Loss Development Study*, 1991 Edition, published by the Reinsurance Association of America.

The incremental claim amounts of the run-off triangle appear in Table 1, as the upper values in each row. This table also gives the predicted incremental claim amounts, for both the run-off triangle and the target triangle, as the lower values in each row. The last column shows the predicted reserves per accident year, determined by using the stochastic chain ladder of Section 4. The latter are in agreement with the values computed by Mack (1994). It is straightforward to see that these are the same as those obtained by applying the chain-ladder technique. It was stated in Section 1 that one advantage of using stochastic models is that confidence intervals can be produced. For the data in Table 1, the root mean square prediction error for the reserve, which measures the stochastic error in the reserve when used to estimate outstanding claims, is 17,603.

The generalised Pearson residual plots associated with the fit are reproduced in Figures 1, 2 and 3. These show that the pattern of the residuals is satisfactory. Broadly speaking, in any successful modelling exercise we would be looking for approximately 95% of (standardised) residuals to lie within the band 0 ± 2.

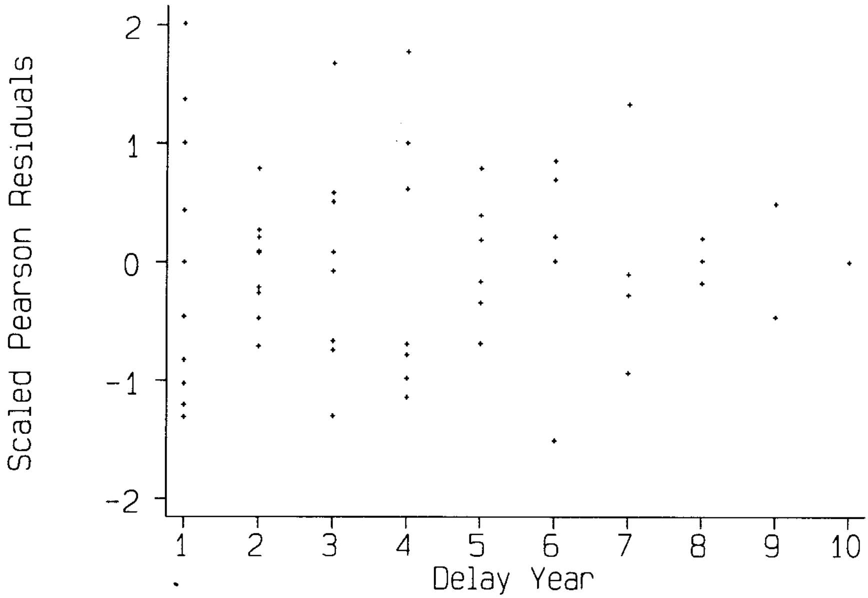

Figure 1. Pearson residuals vs delay; Poisson-type response model

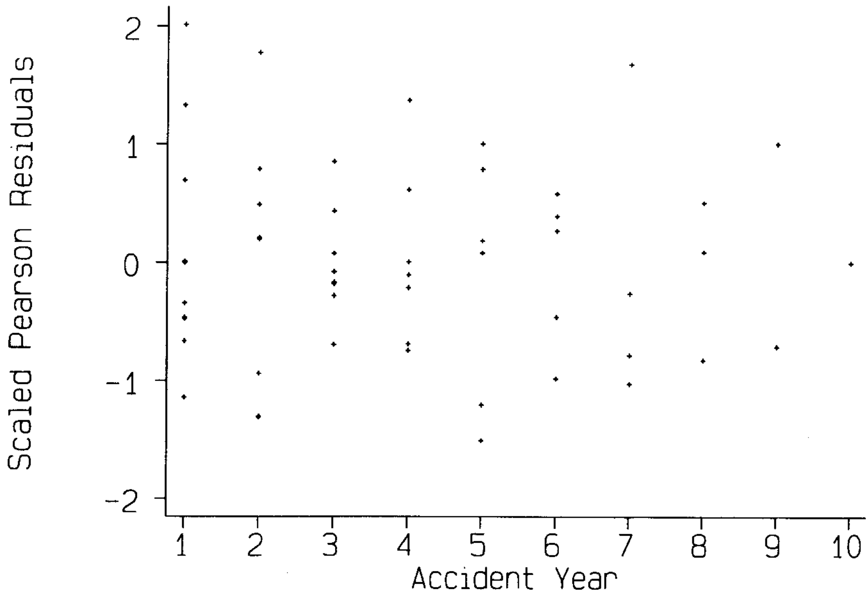

Figure 2. Pearson residuals vs accident year; Poisson-type response model

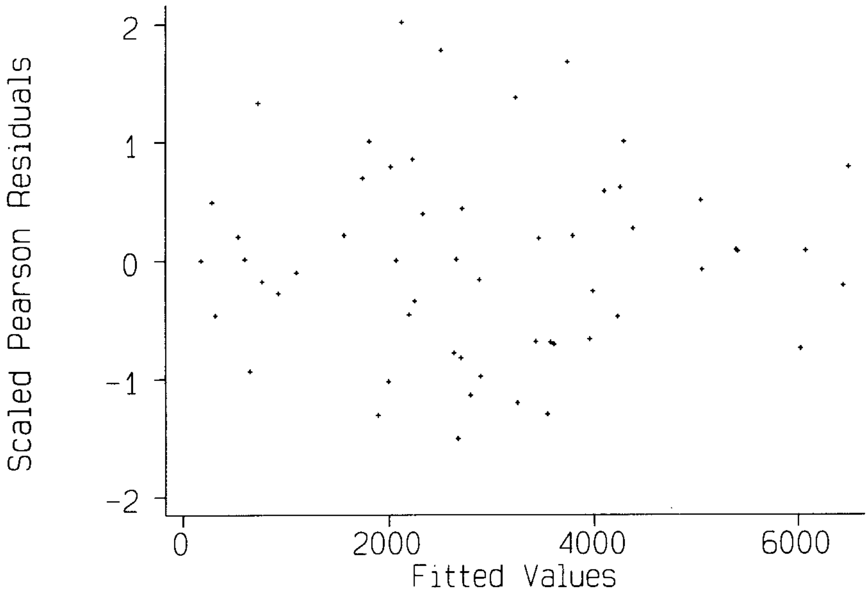

Figure 3. Pearson residuals vs fitted values; Poisson-type response model

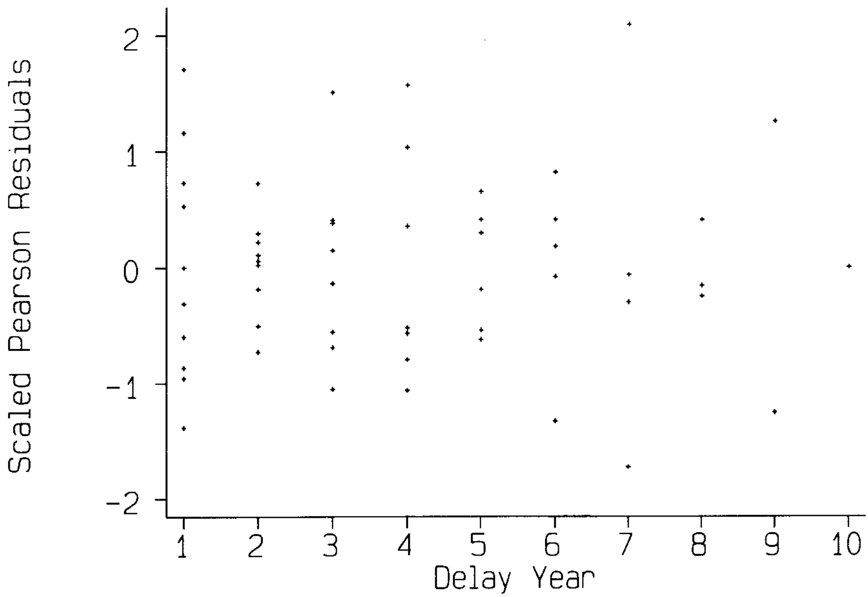

Figure 4. Pearson residuals vs delay; gamma-type response model

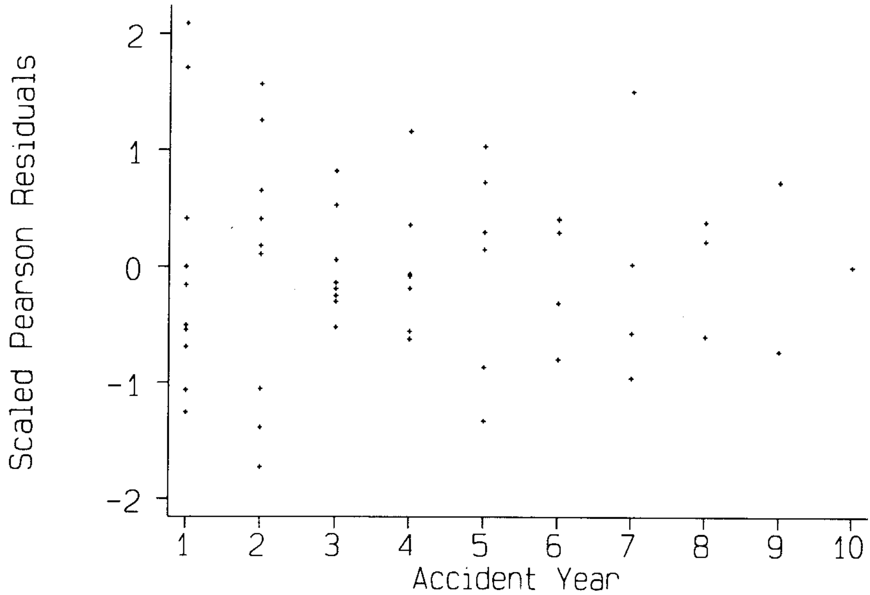

Figure 5. Pearson residuals vs accident year; gamma-type response model

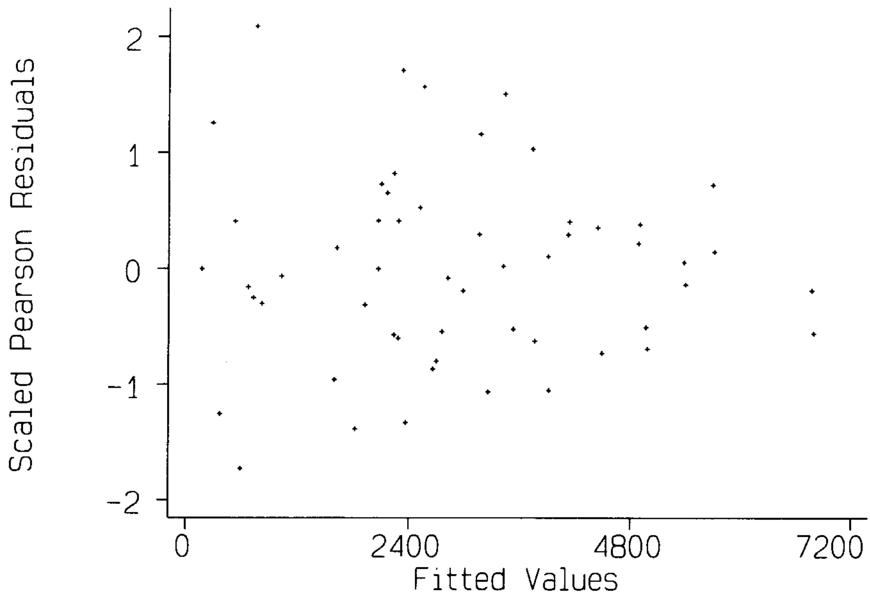

Figure 6. Pearson residuals vs fitted values; gamma-type response model

Table 1. Predicted claim amounts; Poisson-type response model

										Total
5,012	3,257	2,638	898	1,734	2,642	1,828	599	54	172	
2,111.4	4,221.4	3,948.6	2,785.1	2,243.2	1,735.9	714.8	590.8	310.8	172.0	0.
106	4,179	1,111	5,270	3,116	1,817	−103	673	535	–	
1,889.9	3,778.5	3,534.4	2,492.9	2,007.8	1,553.8	639.8	528.8	278.2	154.0	154.
3,410	5,582	4,881	2,268	2,594	3,479	649	603	–	–	
2,699.9	5,398.0	5,049.2	3,561.4	2,868.4	2,219.7	914.0	755.4	397.4	219.9	617.
5,655	5,900	4,211	5,500	2,159	2,658	984	–	–	–	
3,217.8	6,433.4	6,017.7	4,244.5	3,418.6	2,645.5	1,089.4	900.3	473.7	262.1	1,636.
1,092	8,473	6,271	6,333	3,786	225	–	–	–	–	
3,242.8	6,483.6	6,064.6	4,277.6	3,445.3	2,666.1	1,097.9	907.4	477.4	264.2	2,746.
1,513	4,932	5,257	1,233	2,917	–	–	–	–	–	
2,186.2	4,370.9	4,088.5	2,883.8	2,322.6	1,797.4	740.1	611.7	321.8	178.1	3,649.
557	3,463	6,926	1,368	–	–	–	–	–	–	
1,989.8	3,978.3	3,721.2	2,624.7	2,114.0	1,635.9	673.6	556.8	292.9	162.1	5,435.
1,351	5,596	6,165	–	–	–	–	–	–	–	
2,692.7	5,383.6	5,035.7	3,551.9	2,860.8	2,213.8	911.6	753.4	396.4	219.4	10,907.
3,133	2,262	–	–	–	–	–	–	–	–	
1,798.7	3,596.3	3,363.9	2,372.7	1,911.0	1,478.8	609.0	503.3	264.8	146.5	10,650.
2,063	–	–	–	–	–	–	–	–	–	
2,063.0	4,124.7	3,858.2	2,721.3	2,191.8	1,696.1	698.4	577.2	303.7	168.1	16,339.
									Total	52,135.
									Root mean squared error of prediction	17,603.

By way of contrast, Table 2 gives the corresponding individual predicted incremental claim amounts and their associated totals, based on the gamma response model. Similarly, the generalised Pearson residual plots associated with this fit are reproduced in Figures 4, 5 and 6.

6. CONCLUSIONS

This paper has studied a stochastic model for the chain-ladder technique. It has been shown that this can be viewed in a number of ways. One of the possible varieties concerns whether a quasi-likelihood approach is taken. This allows the strictness of the Poisson modelling assumption to be relaxed in order to process data which are not exclusively positive integers. Having identified a stochastic chain-ladder model, it is possible to view the chain-ladder technique within a wider modelling framework. There are great benefits to be gained from this. It is essential that the data presented are scrutinised fully, and an important contribution to this is the use of residual plots. Other models, within the same general framework as the chain-ladder model, use more appropriate assumptions for claims data than the chain-ladder model. It is suggested that the chain-ladder technique is appropriate for analysing numbers of claims, but unsuitable for claim amounts on this basis.

REFERENCES

FIRTH, D. (1988). Multiplicative errors: log-normal or gamma? *J.R.S.S. Series B*, **50**, 266-268.

FRANCIS, B., GREEN, M. & PAYNE, C. (1993). *The GLIM system.* Release 4 Manual, Clarendon Press, Oxford.

KREMER, E. (1982). IBNR-claims and the two-way model of ANOVA. *Scand. Act. J.* **1**, 47-55.

MCCULLAGH, P. & NELDER, J.A. (1989). *Generalized linear models.* Second edition, New York: Chapman and Hall.

MACK, T. (1991). A simple parameteric model for rating automobile insurance or estimating IBNR claims reserves. *ASTIN Bulletin*, **22**, 93-109.

MACK, T. (1994). Which stochastic model is underlying the chain ladder method? *Insurance: Mathematics and Economics*, **15**, 133-138.

RENSHAW, A. E. (1989). Chain ladder and interactive modelling (claims reserving and GLIM). *J.I.A.* **116**, 559-587.

RENSHAW, A. E. (1993). On the application of exponential dispersion models in premium rating. *ASTIN Bulletin*, **23**, 145-148.

RENSHAW, A. E. (1994a). Modelling the claims process in the presence of covariates. *ASTIN Bulletin*, **24**, 265-285.

RENSHAW, A. E. (1994b). A note on some practical aspects of modelling the claims process in the presence of covariates. *Proceedings of the XXV ASTIN Colloquium*, Cannes.

ROSENBERG, P. S. (1990). A simple correction of AIDS surveillance data for reporting delays. *J. AIDS*, **3**, 49-54.

VERRALL, R. J. (1989). A state space representation of the chain ladder linear model. *J.I.A.* **116**, 589-610.

VERRALL, R. J. (1990). Bayes and empirical Bayes estimation for the chain ladder model. *ASTIN Bulletin*, **20**, No. 2, 217-243.

VERRALL, R. J. (1991a). On the unbiased estimation of reserves from loglinear models. *Insurance: Mathematics and Economics*, **10**, No. 1, 75-80.

VERRALL, R. J. (1991b). Chain ladder and maximum likelihood. *J.I.A.* **118**, 489-499.

APPENDIX

Theorem 1. The estimates obtained by maximising the conditional likelihood L_C can be obtained recursively by equation (2.11).

Proof

We require the estimates which maximise:

$$L_C = \prod_{i=1}^{n} \left(\frac{C_{i,n-i+1}!}{\prod_{j=1}^{n-i+1} y_{ij}!} \prod_{j=1}^{n-i+1} p_{(i)j}^{y_{ij}} \right).$$

It is equivalent to maximise log L_C, and to do this we need only consider the terms involving $p_{(i)j}$. Thus, we maximise:

$$\begin{aligned} l_C &= \sum_{i=1}^{n} \sum_{j=1}^{n-i+1} y_{ij} \log p_{(i)j} \\ &= \sum_{i=1}^{n} \sum_{j=1}^{n-i+1} y_{ij} (\log p_j - \log \sum_{k=1}^{n-i+1} p_k) \quad \text{using (2.5).} \end{aligned}$$

Note that p_n only occurs in the terms for which $i=1$. Thus:

$$\frac{\partial l_C}{\partial p_n} = \frac{y_{1n}}{p_n} - \sum_{j=1}^{n} \frac{y_{1j}}{\sum_{k=1}^{n} p_k}$$

and $\frac{\partial l_C}{\partial p_n} = 0$ implies that $\hat{p}_n = \frac{y_{1n}}{\sum_{j=1}^{n} y_{1j}}$ since $\sum_{k=1}^{n} p_k = 1$.

Thus, $\hat{p}_n = \frac{y_{1n}}{C_{1n}}$.

Now assume that the theorem is true for $j+1, j+2, \ldots, n$. Differentiating with respect to p_j gives:

$$\begin{aligned} \frac{\partial l_C}{\partial p_j} &= \sum_{i=1}^{n-j+1} \left\{ \frac{y_{ij}}{p_j} - \sum_{j=1}^{n-i+1} \frac{y_{ij}}{\sum_{k=1}^{n-i+1} p_k} \right\} \\ &= \frac{\sum_{k=1}^{n-i+1} y_{ij}}{p_j} - \sum_{i=1}^{n-j+1} \left(\frac{C_{i,n-i+1}}{\sum_{k=1}^{n-i+1} p_k} \right) \end{aligned}$$

and $\dfrac{\partial l_C}{\partial p_j} = 0$ implies that:

$$\hat{p}_j = \frac{\sum_{i=1}^{n-j+1} y_{ij}}{\sum_{i=1}^{n-j+1}\left(\dfrac{C_{i,n-i+1}}{\sum\limits_{k=1}^{n-i+1} p_k}\right)}.$$

Since $\sum_{k=1}^{n-i+1} p_k = 1 - \sum_{k=n-i+2}^{n} p_k$ if $i>1$ and $\sum_{k=1}^{n} p_k = 1$, this is equivalent to (2.10). Thus, the theorem has been proved by induction.

Theorem 2. $C_{1n} + C_{2,n-1}\hat{\lambda}_n + \ldots + C_{n-j+1,j}\hat{\lambda}_{j+1}\ldots\hat{\lambda}_n = \hat{\lambda}_{j+1}\ldots\hat{\lambda}_n \sum_{i=1}^{n-j+1} C_{ij}$.

Proof

$$\begin{aligned} C_{1n} + C_{2,n-1}\hat{\lambda}_n &= C_{1n} + C_{2,n-1}\frac{C_{1n}}{C_{1,n-1}} \\ &= \frac{C_{1n}}{C_{1,n-1}}(C_{1,n-1} + C_{2,n-1}) \\ &= \hat{\lambda}_n(C_{1,n-1} + C_{2,n-1}). \end{aligned}$$

Therefore:

$$\begin{aligned} C_{1n} + C_{2,n-1}\hat{\lambda}_n + C_{3,n-2}\hat{\lambda}_n\hat{\lambda}_{n-1} &= \hat{\lambda}_n[C_{1,n-1} + C_{2,n-1} + C_{3,n-2}\hat{\lambda}_{n-1}] \\ &= \hat{\lambda}_n[(C_{1,n-2} + C_{2,n-2})\hat{\lambda}_{n-1} + C_{3,n-2}\hat{\lambda}_{n-1}] \end{aligned}$$

since $\quad \hat{\lambda}_{n-1} = \dfrac{C_{1,n-1} + C_{2,n-1}}{C_{1,n-2} + C_{2,n-2}}$.

Hence $C_{1n} + C_{2,n-1}\hat{\lambda}_n + C_{3,n-2}\hat{\lambda}_n\hat{\lambda}_{n-1} = \hat{\lambda}_{n-1}\hat{\lambda}_n[C_{1,n-2} + C_{2,n-2} + C_{3,n-2}]$ and so on. By continuing for all terms in a similar way, the theorem is proved.