

This article was downloaded by: [DEFF]

On: 6 June 2010

Access details: Access Details: [subscription number 789685088]

Publisher Taylor & Francis

Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



Scandinavian Actuarial Journal

Publication details, including instructions for authors and subscription information:

<http://www.informaworld.com/smpp/title~content=t713690025>

Prediction of outstanding payments in a Poisson cluster model

Anders Hedegaard Jessen^a; Thomas Mikosch^a; Gennady Samorodnitsky^b

^a Department of Mathematics, University of Copenhagen, Copenhagen, Denmark ^b School of Operations Research and Information Engineering, Cornell University, Ithaca, NY, USA

First published on: 04 June 2010

To cite this Article Jessen, Anders Hedegaard, Mikosch, Thomas and Samorodnitsky, Gennady (2010) 'Prediction of outstanding payments in a Poisson cluster model', *Scandinavian Actuarial Journal*, First published on: 04 June 2010 (iFirst)

To link to this Article: DOI: 10.1080/03461238.2010.481080

URL: <http://dx.doi.org/10.1080/03461238.2010.481080>

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: <http://www.informaworld.com/terms-and-conditions-of-access.pdf>

This article may be used for research, teaching and private study purposes. Any substantial or systematic reproduction, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

Original Article

Prediction of outstanding payments in a Poisson cluster model

ANDERS HEDEGAARD JESSEN*, THOMAS MIKOSCH* and
GENNADY SAMORODNITSKY†

*Department of Mathematics, University of Copenhagen, Universitetsparken 5, DK-2100
Copenhagen, Denmark

†School of Operations Research and Information Engineering, Cornell University, 220 Rhodes
Hall, Ithaca, NY 14853, USA

(Accepted 27 March 2010)

We consider a simple Poisson cluster model for the payment numbers and the corresponding total payments for insurance claims arriving in a given year. Due to the Poisson structure one can give reasonably explicit expressions for the prediction of the payment numbers and total payments in future periods given the past observations of the payment numbers. One can also derive reasonably explicit expressions for the corresponding prediction errors. In the (a, b) class of Panjer's claim size distributions, these expressions can be evaluated by simple recursive algorithms. We study the conditions under which the predictions are asymptotically linear as the number of past payments becomes large. We also demonstrate that, in other regimes, the prediction may be far from linear. For example, a staircase-like pattern may arise as well. We illustrate how the theory works on real-life data, also in comparison with the chain ladder method.

Keywords: Poisson cluster model; Prediction; Claims reserving; Chain ladder method; Panjer distributions; (a, b) class; Recursive algorithm

1. Introduction

Let N_k be the number of payments for claims arriving in an insurance portfolio in the year 0 and being executed in the year $k \in \{0, 1, \dots\}$. Moreover, let S_k be the corresponding total amount of the payments executed in year k . If one has observed the counts N_k , $k = 0, \dots, j$, for some $j \geq 0$, a major problem for an insurance company is to determine a *reserve* for the years $j+1, j+2, \dots$. This amounts to *predicting* the pairs (N_{j+l+1}, S_{j+l+1}) for $l = 0, 1, \dots$. In this context natural estimators are given by the conditional expectations given the past values N_0, \dots, N_j , i.e.

$$\hat{N}_{j+l+1} = E(N_{j+l+1} | \mathcal{F}_j) \quad \text{and} \quad \hat{S}_{j+l+1} = E(S_{j+l+1} | \mathcal{F}_j), \quad (1.1)$$

$l = 0, 1, \dots$, where $\mathcal{F}_j = \sigma(N_0, \dots, N_j)$, $j = 0, 1, \dots$. Assuming $\text{var}(S_{j+l+1}) < \infty$, $S = \hat{S}_{j+l+1}$ is the a.s. unique minimizer of the mean square error $E((S_{j+l+1} - S)^2)$ in the class of square

Corresponding author. E-mail: mikosch@math.ku.dk

integrable random variables S which are measurable functions of N_0, \dots, N_j , and a similar remark applies to $\hat{N}_{j+\ell+1}$. Moreover, $s = \hat{S}_{j+\ell+1}$ minimizes the *conditional* mean square error $E((S_{j+\ell+1} - s)^2 | \mathcal{F}_j)$.

One of the popular procedures in this context was suggested by Mack (1993, 1994a, 1994b) and Mack *et al.* (2006). In its simplest version, Mack's procedure declares the predictors $\hat{S}_{j+\ell+1}$ and $\hat{N}_{j+\ell+1}$ to be linear functions of $S_0 + \dots + S_j$ or $N_j = N_0 + \dots + N_j$, respectively. For example,

$$\hat{N}_{j+1} = (f_j - 1) (N_0 + \dots + N_j), \quad j = 0, 1, \dots, \quad (1.2)$$

for constants $f_j \geq 1$.

Assume that one observes the *run-off triangle*

$$(N_{i,i+k}, S_{i,i+k}), \quad i = 1, \dots, n, \quad k = 0, \dots, n - i, \quad (1.3)$$

where $((N_{i,i+k}, S_{i,i+k}))_{k=0, \dots, n-i}$ are the payment numbers and total payments for claims arriving in year i and being executed in year $i+k$; one assumes that $((N_{i,i+k}, S_{i,i+k}))_{k=0, 1, \dots}$ are iid copies of $((N_k, S_k))_{k=0, 1, \dots}$. Here n is the last year for which payments were observed. Mack's assumptions (1.2) give raise to constructing natural estimators \hat{f}_j of f_j which are referred to as *chain ladder estimators*. Then, replacing the unknown parameters f_j on the right-hand side of Eq. (1.2) by their estimators \hat{f}_j , one obtains a predictor of N_{j+1} . In Section 5 we will introduce the chain ladder estimators and compare the performance of the corresponding predictors with those proposed in this paper.

Mack's procedure (1.2) does not determine the dynamics of a particular stochastic process $(N_j)_{j=0, 1, \dots}$. For example, one cannot simulate this process from Eq. (1.2). Furthermore, it is not known for which natural classes of stochastic processes $(N_j)_{j=0, 1, \dots}$ the relation (1.2) holds. In fact, the very linearity of the predictor (1.2) is a simplification which is hard to reconcile with natural stochastic models for the count process.

In this paper, we consider a simple stochastic process model for the counts N_j , $j = 0, 1, \dots$, and the corresponding payments S_j , $j = 0, 1, \dots$. We emphasize that the model is simple, and some of its assumptions may look idealized. However, this is an honest and natural stochastic model, and the forecasting procedure we derive comes with all the advantages of a model-based procedure. Not the least among these advantages is the possibility of constructing justifiable prediction intervals. This is in contrast with the Mack procedure.

Our model is given by the following conditions which we assume throughout this paper.

1.1. The model

Let M be the number of claims arriving in a given year with distribution:

$$q_m = P(M = m), \quad m = 0, 1, \dots$$

The m th claim causes a stream of K_m payment years from the insurer to the insured through the next years. We assume that the k th of these payments is executed in the year Y_{mk} . We further assume that (K_m) is an iid sequence of Poisson(μ) distributed random

variables and that $(Y_{mk})_{m,k=1,2,\dots}$ constitutes an iid family with common distribution:

$$p_j = P(Y_{11} = j), \quad j = 0, 1, \dots \tag{1.4}$$

Finally, assume that M , (K_m) , and (Y_{mk}) are independent. Write,

$$N_j = \sum_{m=1}^M \sum_{k=1}^{K_m} I_{\{Y_{mk}=j\}}, \quad j = 0, 1, \dots,$$

i.e. N_j is the number of payments for claims arriving in a given year and being executed in year j . Assume further that $(X_{mk})_{m,k=1,2,\dots}$ is an iid family of non-negative random variables independent of M , (K_m) , and (Y_{mk}) . We interpret X_{mk} as the k th payment for the m th claim. Then,

$$S_j = \sum_{m=1}^M \sum_{k=1}^{K_m} X_{mk} I_{\{Y_{mk}=j\}}, \quad j = 0, 1, \dots,$$

are the total payments for the claims arriving in year 0 and being executed in year j .

Both processes (N_j) and (S_j) can easily be simulated. It is our aim to show that the predictors (1.1) and their errors can be calculated explicitly and are easily derived by numerical methods for certain special cases of the distribution of M . The expressions for the predictor of N_{j+1} are highly non-linear functions of $N_0 + \dots + N_j$, in contrast to Mack's procedure (1.2). However, under some condition on the distribution of M these predictors are *asymptotically linear* functions of $N_0 + \dots + N_j$ if the latter quantity increases to infinity. In other situations, the predictors stay non-linear even in the limit.

We are fully aware of the fact that our model is based on rather simplistic assumptions but our aim is not to give a complicated model with too many unknown distributions and parameters. The possibility to simulate our model is perhaps its main advantage, and the assumptions of the model can be relaxed in rather different directions.

The assumption which deserves most criticism is that the distribution of the payments X_{mk} does not depend on the payment period $Y_{mk}=j$. For example, there is evidence in data that the payments become the smaller the larger j . Small changes in the proofs below show that the complete analysis remains valid when we assume that the distribution of the payment X_{mk} depends on the payment year $Y_{mk}=j$. However, in order to keep the analysis as simple as possible, we stick to the iid assumption on the X_{mk} 's.

The paper is organized as follows. We start in Section 2 by giving the relevant formulas for the predictors of $S_{j+\ell+1}$, $\ell = 0, 1, \dots$. Since $S_j = N_j$ if $X_{mi} = 1$ for all m, i , the prediction of $N_{j+\ell+1}$, $\ell = 0, 1, \dots$, is a special case. We also determine the prediction errors. The predictors and conditional prediction errors involve certain derivatives of the Laplace–Stieltjes transform of M . In general, these derivatives are difficult to obtain. However, in the (a, b) class of Panjer distributions, including the Poisson, binomial, and negative binomial distributions, there exist simple recursive algorithms for calculating these derivatives; see Section 3. In Section 4 we study the asymptotic behavior of the predictors as the number of the previously observed payments grows. In particular, we give conditions under which the predictors $\hat{S}_{j+\ell+1}$ are asymptotically linear functions of $N_0 + \dots + N_j$. We also consider other situations, where different asymptotic patterns of the behavior of the predictors arise. An interesting feature is the staircase-like pattern discussed in that section. In Section 5 we apply our predictors to a non-life insurance data

Downloaded By: [DEFF] At: 13:42 6 June 2010

set. We compare the performance of these predictors with the corresponding ones based on chain ladder estimation.

2. The prediction problem

We intend to predict the future numbers of payments $N_{j+\ell+1}$ and the corresponding total claim amounts $S_{j+\ell+1}$, $\ell=0,1,\dots$, given the past payment numbers N_0,\dots,N_j . This means we will calculate the predictors $\hat{S}_{j+\ell+1}$ and $\hat{N}_{j+\ell+1}$ in Eq. (1.1) provided these quantities are well defined. Since we always assume the conditions of the model introduced in Section 1, the independence of (X_{mi}) and the rest of the random ingredients in the model imply that,

$$\hat{S}_{j+\ell+1} = EX_{11} E(N_{j+\ell+1}|\mathcal{F}_j) = EX_{11} \hat{N}_{j+\ell+1}, \quad (2.1)$$

$\ell=0,1,\dots$ Therefore the prediction problem for $S_{j+\ell+1}$ reduces to the one for $N_{j+\ell+1}$. Conversely, if $X_{mi}=1$ a.s. for all m, i then $N_j=S_j, j=0,1,\dots$ Therefore it suffices to study the prediction of $S_{j+\ell+1}$ given N_0,\dots,N_j .

We will derive expressions for the predictors (2.1) and determine their errors. We start with the one-step ahead prediction problem, i.e. $\ell=0$.

2.1. One-step ahead prediction

We introduce some notation to be used throughout the paper. We will need the *Laplace–Stieltjes transform* of M , i.e.

$$L(\gamma) = Ee^{-\gamma M} = \sum_{m=0}^{\infty} q_m e^{-\gamma m}, \quad \gamma \geq 0,$$

and its derivatives

$$L^{(\ell)}(\gamma) = (-1)^\ell E(M^\ell e^{-\gamma M}), \quad \gamma > 0, \quad \ell = 0, 1, \dots,$$

with the convention that $L^{(0)}=L$. Moreover, define

$$R_\ell(\gamma) = -\frac{L^{(\ell+1)}(\gamma)}{L^{(\ell)}(\gamma)} = \frac{E(M^{\ell+1} e^{-\gamma M})}{E(M^\ell e^{-\gamma M})}, \quad \gamma > 0, \quad \ell = 0, 1, \dots$$

Finally, write

$$\theta_j = \mu \sum_{d=0}^j p_d, \quad j = 0, 1, \dots;$$

recall that μ is the Poisson rate of the number of payments per claim, and (p_k) is the displacement probabilities in Eq. (1.4).

Next we formulate our main result on the prediction of S_{j+1} . Recall that the corresponding result for N_{j+1} follows by setting $X_{mi}=1$ a.s. for all m, i .

THEOREM 2.1 Assume that $EM < \infty$ and $EX_{11} < \infty$.

(1) The predictor \hat{S}_{j+1} of S_{j+1} given N_0, \dots, N_j has the form

$$E(S_{j+1}|N_0 = n_0, \dots, N_j = n_j) = \mu p_{j+1} EX_{11} R_{n_0 + \dots + n_j}(\theta_j), \quad (2.2)$$

$$n_0, \dots, n_j = 0, 1, \dots, j = 0, 1, \dots$$

(2) Assume, in addition, that $\text{var}(M) < \infty$ and $\text{var}(X_{11}) < \infty$. Then the unconditional prediction error for S_{j+1} , $j = 0, 1, \dots$, is given by

$$E((S_{j+1} - \hat{S}_{j+1})^2) = E(X_{11}^2) \mu p_{j+1} EM + (EX_{11} \mu p_{j+1})^2 E(M^2) - E(\hat{S}_{j+1}^2). \quad (2.3)$$

(3) Assume, in addition, that $\text{var}(M) < \infty$ and $\text{var}(X_{11}) < \infty$. Then the conditional prediction error for S_{j+1} given the past observations N_0, \dots, N_j , $j = 0, 1, \dots$, is

$$\text{var}(S_{j+1}|N_0 = n_0, \dots, N_j = n_j) = E(X_{11}^2) \mu p_{j+1} R_{n_0 + \dots + n_j}(\theta_j) + (EX_{11} \mu p_{j+1})^2 R_{n_0 + \dots + n_j}(\theta_j) \times [R_{n_0 + \dots + n_j + 1}(\theta_j) - R_{n_0 + \dots + n_j}(\theta_j)]. \quad (2.4)$$

REMARK 2.2 Writing $\ell_j = n_0 + \dots + n_j$, $j = 0, 1, \dots$, we observe by virtue of Eq. (2.2) that

$$E(S_{j+1}|N_0 = n_0, \dots, N_j = n_j) = E(S_{j+1}|N_0 = \ell_0, N_0 + N_1 = \ell_1, \dots, N_0 + \dots + N_j = \ell_j) = E(S_{j+1}|N_0 + \dots + N_j = \ell_j),$$

or, alternatively,

$$E(S_0 + \dots + S_{j+1}|N_0 + \dots + N_j = \ell_j) = EX_{11} \ell_j + E(S_{j+1}|N_0 + \dots + N_j = \ell_j). \quad (2.5)$$

By virtue of Eq. (2.2), the conditional expectation (2.5) is in general not a linear function of ℓ_j , in disagreement with Mack's procedure (1.2). In Section 4 we will give conditions on the distribution of M ensuring that Eq. (2.5) is asymptotically linear as $\ell_j \rightarrow \infty$.

REMARK 2.3 In Section 3 we will give a recursive algorithm for evaluating the quantities $L^{(\ell)}$ when the distribution of M belongs to the (a, b) class used for Panjer recursion.

Proof. (1) By the splitting property of the Poisson process, the double array $Z_{jl} = \sum_{k=1}^{K_j} I_{\{Y_k=j\}}$, $j = 0, 1, \dots$, $l = 1, 2, \dots$ consists of independent Poisson random variables. In particular, $Z_{jl} \sim \text{Poisson}(\mu p_j)$ and, conditionally on $M = m$, $N_j = \sum_{l=1}^m Z_{jl}$, $j = 0, 1, \dots$, constitutes a sequence of independent Poisson random variables. In particular, given $M = m$, $N_j \sim \text{Poisson}(m \mu p_j)$. Therefore,

$$E(N_{j+1}|N_0, \dots, N_j, M) = M \mu p_{j+1}, \quad j = 0, 1, \dots, \quad (2.6)$$

and

$$\hat{S}_{j+1} = EX_{11} \mu p_{j+1} E(M|\mathcal{F}_j), \quad j = 0, 1, \dots \quad (2.7)$$

Even more precisely, for any m , $j = 0, 1, \dots$ and $n_0, \dots, n_j = 0, 1, \dots$,

$$\begin{aligned}
P(N_0 = n_0, \dots, N_j = n_j, M = m) &= q_m P\left(\sum_{l=1}^m Z_{0l} = n_0\right) \cdots P\left(\sum_{l=1}^m Z_{jl} = n_j\right) \\
&= q_m \prod_{d=0}^j e^{-m\mu p_d} \frac{(m\mu p_d)^{n_d}}{n_d!} \\
&= q_m e^{-m\theta_j} m^{\sum_{d=0}^j n_d} \prod_{d=0}^j \frac{(\mu p_d)^{n_d}}{n_d!}. \tag{2.8}
\end{aligned}$$

We conclude that for $j \geq 0$,

$$P(M = m | N_0 = n_0, \dots, N_j = n_j) = \frac{q_m e^{-m\theta_j} m^{\sum_{d=0}^j n_d}}{\sum_{r=0}^{\infty} q_r e^{-r\theta_j} r^{\sum_{d=0}^j n_d}}, \quad m, n_0, \dots, n_j = 0, 1, \dots \tag{2.9}$$

In particular,

$$E(M | N_0 = n_0, \dots, N_j = n_j) = \frac{\sum_{m=0}^{\infty} q_m e^{-m\theta_j} m^{\sum_{d=0}^j n_d + 1}}{\sum_{r=0}^{\infty} q_r e^{-r\theta_j} r^{\sum_{d=0}^j n_d}} = R_{n_0 + \dots + n_j}(\theta_j). \tag{2.10}$$

We conclude, using Eq. (2.7), that Eq. (2.2) holds.

(2) & (3) We start by calculating the prediction error of S_{j+1} given the values N_0, \dots, N_j . First observe that

$$\text{var}(S_{j+1} | N_0, \dots, N_j, M) = M E(X_{11}^2) \mu p_{j+1}.$$

Therefore, and in view of Eq. (2.6),

$$\begin{aligned}
E(S_{j+1}^2 | \mathcal{F}_j) &= E(\text{var}(S_{j+1} | N_0, \dots, N_j, M) + (E(S_{j+1} | N_0, \dots, N_j, M))^2 | \mathcal{F}_j) \\
&= \mu p_{j+1} E(X_{11}^2) E(M | \mathcal{F}_j) + (\mu p_{j+1} E X_{11})^2 E(M^2 | \mathcal{F}_j)
\end{aligned}$$

Finally, using Eq. (2.7), we see that the conditional prediction error can be written as,

$$\text{var}(S_{j+1} | \mathcal{F}_j) = E(X_{11}^2) E(N_{j+1} | \mathcal{F}_j) + (E X_{11} \mu p_{j+1})^2 \text{var}(M | \mathcal{F}_j). \tag{2.11}$$

Using Eq. (2.9), we can replace the conditional moments of M by the corresponding derivatives of L , leading to Eq. (2.4). Taking expectations in Eq. (2.11), we obtain the prediction error,

$$\begin{aligned}
E((S_{j+1} - \hat{S}_{j+1})^2) &= E[\text{var}(S_{j+1} | \mathcal{F}_j)] \\
&= E(X_{11}^2) E N_{j+1} + (E X_{11} \mu p_{j+1})^2 [E(M^2) - E((E(M | \mathcal{F}_j))^2)] \\
&= E(X_{11}^2) \mu p_{j+1} E M + (E X_{11} \mu p_{j+1})^2 E(M^2) - E(\hat{S}_{j+1}^2).
\end{aligned}$$

This finishes the proof. \square

REMARK 2.4 A simple upper bound of the unconditional prediction error (2.3) is given by

$$E[(S_{j+1} - \hat{S}_{j+1})^2] \leq E(X_{11}^2) \mu p_{j+1} EM + (EX_{11} \mu p_{j+1})^2 E(M^2).$$

Evaluation of $E(\hat{S}_{j+1}^2)$ in Eq. (2.3) is complicated. Following the lines of the proof above, one can derive a more explicit expression for this term:

$$\begin{aligned} E(\hat{S}_{j+1}^2) &= (EX_{11} \mu p_{j+1})^2 E[(E(M|\mathcal{F}_j))^2] = (EX_{11} \mu p_{j+1})^2 E[(R_{N_0+\dots+N_j}(\theta_j))^2] \\ &= (EX_{11} \mu p_{j+1})^2 \sum_{k=0}^{\infty} \sum_{m=0}^{\infty} P(N_0 + \dots + N_j = k, M = m) (R_k(\theta_j))^2. \end{aligned}$$

Applying Eq. (2.8), the right-hand double sum turns into

$$\begin{aligned} &\sum_{k=0}^{\infty} \left[\sum_{m=0}^{\infty} q_m e^{-m\theta_j} m^k \right] \sum_{n_0+\dots+n_j=k} \prod_{d=0}^j \frac{(\mu p_d)^{n_d}}{n_d!} (R_k(\theta_j))^2 \\ &= e^{\theta_j} \sum_{k=0}^{\infty} (-1)^k \frac{L^{(k+1)}(\theta_j)^2}{L^{(k)}(\theta_j)} P(\Theta_j = k) \\ &= e^{\theta_j} E \left((-1)^{\Theta_j} \frac{L^{(\Theta_j+1)}(\theta_j)^2}{L^{(\Theta_j)}(\theta_j)} \right), \end{aligned}$$

where Θ_j is $Poisson(\theta_j)$ distributed.

2.2. Multi-step ahead prediction

In this subsection we consider the prediction problem for $\ell + 1$ periods ahead. This means we are interested in the quantities $\hat{S}_{j+\ell+1}$, $\ell = 0, 1, \dots$, defined in Eq. (1.1), and the corresponding prediction errors.

THEOREM 2.5 Assume that $EM < \infty$ and $EX_{11} < \infty$.

(1) The predictor $\hat{S}_{j+\ell+1}$ of $S_{j+\ell+1}$ given N_0, \dots, N_j has the form

$$E(S_{j+\ell+1} | N_0 = n_0, \dots, N_j = n_j) = EX_{11} \mu p_{j+\ell+1} R_{n_0+\dots+n_j}(\theta_j), j, \ell = 0, 1, \dots$$

(2) Assume, in addition, that $\text{var}(M) < \infty$ and $\text{var}(X_{11}) < \infty$. Then the unconditional prediction error for $S_{j+\ell+1}$, $j, \ell = 0, 1, \dots$, is given by

$$E[(S_{j+\ell+1} - \hat{S}_{j+\ell+1})^2] = E(X_{11}^2) \mu p_{j+\ell+1} EM + (EX_{11} \mu p_{j+\ell+1})^2 E(M^2) - E(\hat{S}_{j+\ell+1}^2).$$

(3) Assume, in addition, that $\text{var}(M) < \infty$ and $\text{var}(X_{11}) < \infty$. Then the conditional prediction error for $S_{j+\ell+1}$, $j, \ell = 0, 1, \dots$, is given by

$$\begin{aligned} \text{var}(S_{j+\ell+1} | N_0 = n_0, \dots, N_j = n_j) &= E(X_{11}^2) \mu p_{j+\ell+1} R_{n_0+\dots+n_j}(\theta_j) + (EX_{11} \mu p_{j+\ell+1})^2 \\ &\quad \times R_{n_0+\dots+n_j}(\theta_j) [R_{n_0+\dots+n_j+1}(\theta_j) - R_{n_0+\dots+n_j}(\theta_j)]. \end{aligned}$$

Proof. (1) We start by observing that for $\ell \geq 0$,

$$\hat{S}_{j+\ell+1} = E[E(S_{j+\ell+1}|\mathcal{F}_{j+\ell})|\mathcal{F}_j].$$

Hence, using Eq. (2.7), we obtain

$$\hat{S}_{j+\ell+1} = EX_{11} \mu p_{j+\ell+1} E[E(M|\mathcal{F}_{j+\ell})|\mathcal{F}_j] = EX_{11} \mu p_{j+\ell+1} E(M|\mathcal{F}_j).$$

Now use relation (2.10).

(2) & (3) For the conditional prediction error for $S_{j+\ell+1}$, we observe that by Eq. (2.11)

$$\begin{aligned} \text{var}(S_{j+\ell+1}|\mathcal{F}_j) &= E(\text{var}(S_{j+\ell+1}|\mathcal{F}_{j+\ell})|\mathcal{F}_j) + \text{var}(E(S_{j+\ell+1}|\mathcal{F}_{j+\ell})|\mathcal{F}_j) \\ &= E(X_{11}^2) \hat{N}_{j+\ell+1} + (EX_{11} \mu p_{j+\ell+1})^2 E(\text{var}(M|\mathcal{F}_{j+\ell})|\mathcal{F}_j) + (EX_{11} \mu p_{j+\ell+1})^2 \\ &\quad \times \text{var}(E(M|\mathcal{F}_{j+\ell})|\mathcal{F}_j) = E(X_{11}^2) \hat{N}_{j+\ell+1} + (EX_{11} \mu p_{j+\ell+1})^2 \text{var}(M|\mathcal{F}_j). \end{aligned}$$

The conditional moments of M can, once again, be expressed using Eq. (2.9). Taking expectations, we obtain the unconditional prediction error

$$\begin{aligned} E[(S_{j+\ell+1} - \hat{S}_{j+\ell+1})^2] &= E[\text{var}(S_{j+\ell+1}|\mathcal{F}_j)] \\ &= E(X_{11}^2) \mu p_{j+\ell+1} EM + (EX_{11} \mu p_{j+\ell+1})^2 [E(M^2) - E((E(M|\mathcal{F}_j))^2)] \\ &= E(X_{11}^2) \mu p_{j+\ell+1} EM + (EX_{11} \mu p_{j+\ell+1})^2 E(M^2) - E(\hat{S}_{j+\ell+1}^2). \end{aligned}$$

□

REMARK 2.6 Notice that

$$\hat{S}_{j+\ell+1} = \frac{p_{j+\ell+1}}{p_{j+1}} \hat{S}_{j+1}, \quad \ell = 0, 1, \dots,$$

provided $p_{j+1} > 0$. Moreover, if $p_{j+\ell+1} = 0$ then $\hat{S}_{j+\ell+1} = 0$.

2.3. Conditionally independent payments

In this subsection we consider a slightly more general model. As before, we assume that the sequences $(X_{mk})_{k=1,2,\dots}$, $m=1,2,\dots$, are iid and independent of the rest of random variables defining the model. We further assume that each sequence $(X_{mk})_{k=1,2,\dots}$ consists of *conditionally* iid random variables or, equivalently, that $(X_{mk})_{k=1,2,\dots}$ is exchangeable. This situation is similar to models in credibility theory, where the claim sizes occurring in an individual policy are assumed conditionally iid; see Mikosch (2004), Chapters 5 and 6.

Since the random variables $(X_{mk})_{m,k=1,2,\dots}$ and $(Y_{mk})_{m,k=1,2,\dots}$ are independent the form of the one-step ahead predictor is again given by Eq. (2.1) but the prediction error changes.

PROPOSITION 2.7 Assume $\text{var}(M) < \infty$ and $\text{var}(X_{11}) < \infty$. Then the unconditional prediction error for S_{j+1} is given by

$$\begin{aligned} E((S_{j+1} - \hat{S}_{j+1})^2) &= EM [E(X_{11}^2) \mu p_{j+1} + \text{cov}(X_{11}, X_{12}) (\mu p_{j+1})^2] \\ &\quad + (EX_{11} \mu p_{j+1})^2 E(M^2) - E(\hat{S}_{j+1}^2), \quad j = 0, 1, \dots \end{aligned}$$

The conditional prediction error of S_{j+1} given N_0, \dots, N_j , $j=0,1, \dots$, has the form

$$\begin{aligned} \text{var}(S_{j+1}|N_0 = n_0, \dots, N_j = n_j) &= [E(X_{11}^2) \mu p_{j+1} + \text{cov}(X_{11}, X_{12}) (\mu p_{j+1})^2] R_{n_0 + \dots + n_j}(\theta_j) \\ &\quad + (EX_{11} \mu p_{j+1})^2 R_{n_0 + \dots + n_j}(\theta_j) \\ &\quad \times [R_{n_0 + \dots + n_j + 1}(\theta_j) - R_{n_0 + \dots + n_j}(\theta_j)]. \end{aligned}$$

A comparison of this result with Theorem 1 shows that the prediction error increases by the additional term with the factor $\text{cov}(X_{11}, X_{12})$. It is non-negative as an application of the conditional Jensen inequality shows.

Proof. We start by calculating

$$\begin{aligned} \text{var}(S_{j+1}|N_0, \dots, N_j, M) &= M \text{var}\left(\sum_{k=1}^{K_1} X_{1k} I_{\{Y_{1k}=j+1\}} | \mathcal{F}_j\right) \\ &= ME\left(\left(\sum_{k=1}^{K_1} X_{1k} I_{\{Y_{1k}=j+1\}}\right)^2 | \mathcal{F}_j\right) - M (EX_{11} \mu p_{j+1})^2. \end{aligned}$$

We observe that

$$\begin{aligned} E\left(\left(\sum_{k=1}^{K_1} X_{1k} I_{\{Y_{1k}=j+1\}}\right)^2 | \mathcal{F}_j\right) &= E\left(\sum_{k=1}^{K_1} X_{1k}^2 I_{\{Y_{1k}=j+1\}} | \mathcal{F}_j\right) \\ &\quad + E\left(\sum_{k=1}^{K_1} \sum_{l=1, l \neq k}^{K_1} X_{1k} X_{1l} I_{\{Y_{1k}=j+1\}} I_{\{Y_{1l}=j+1\}} | \mathcal{F}_j\right) \\ &= [E(X_{11}^2) - E(X_{11} X_{12})] \mu p_{j+1} + E(X_{11} X_{12}) \\ &\quad \times E\left(\left(\sum_{k=1}^{K_1} I_{\{Y_{1k}=j+1\}}\right)^2 | \mathcal{F}_j\right) \\ &= E(X_{11}^2) \mu p_{j+1} + E(X_{11} X_{12}) (\mu p_{j+1})^2. \end{aligned}$$

Here we used the fact that, by the exchangeability, $E(X_{11} X_{12}) = E(X_{1k} X_{1l})$ for $k \neq l$. Overall, we obtain,

$$\text{var}(S_{j+1}|N_0, \dots, N_j, M) = M [E(X_{11}^2) \mu p_{j+1} + \text{cov}(X_{11}, X_{12}) (\mu p_{j+1})^2].$$

Therefore,

$$\begin{aligned} \text{var}(S_{j+1}| \mathcal{F}_j) &= E(M | \mathcal{F}_j) (E(X_{11}^2) \mu p_{j+1} + \text{cov}(X_{11}, X_{12}) (\mu p_{j+1})^2) \\ &\quad + (EX_{11} \mu p_{j+1})^2 \text{var}(M | \mathcal{F}_j). \end{aligned}$$

For the unconditional prediction error we have

$$\begin{aligned} E[(S_{j+1} - \hat{S}_{j+1})^2] &= EM [E(X_{11}^2) \mu p_{j+1} + \text{cov}(X_{11}, X_{12}) (\mu p_{j+1})^2] + E(M^2) (EX_{11} \mu p_{j+1})^2 \\ &\quad - (EX_{11} \mu p_{j+1})^2 E[(E(M | \mathcal{F}_j))^2] \\ &= EM [E(X_{11}^2) \mu p_{j+1} + \text{cov}(X_{11}, X_{12}) (\mu p_{j+1})^2] + E(M^2) \\ &\quad \times (EX_{11} \mu p_{j+1})^2 - E(\hat{S}_{j+1}^2). \end{aligned}$$

□

3. Prediction in the (a, b) class

In the previous sections we have learned that, for predicting the values $N_{j+\ell+1}$ and $S_{j+\ell+1}$, $\ell \geq 0$, given N_0, \dots, N_j , it is crucial to be able to evaluate the derivatives $(-1)^l L^{(l)}(\gamma) = E(M^l e^{-\gamma M})$. In this section we assume that the distribution of M belongs to the (a, b) class which is used in the Panjer recursive algorithm; see Mikosch (2004), Section 3.3. This class is given by the recursive relation

$$q_0 > 0, \quad q_m = (a + b/m) q_{m-1}, \quad m = 1, 2, \dots, \quad a, b \in \mathbb{R}. \quad (3.1)$$

Notice that necessarily $a < 1$ and $a + b > 0$; otherwise $(q_m)_{m=0,1,\dots}$ would not constitute a probability distribution.

This class contains exactly three non-degenerate distributions:

1. The Poisson(b) distribution with $a = 0$, $b > 0$.
2. The Bin(n , p) distribution with $a < 0$, $a = -p/(1-p)$, $b = -a(n+1)$, and $p \in (0, 1)$, $n \geq 1$.
3. The negative binomial distribution with parameter (p, v) :

$$q_m = \binom{v+m-1}{m} p^v (1-p)^m, \quad m = 0, 1, \dots, \quad p \in (0, 1), \quad v > 0,$$

where $0 < a = 1-p$, $b = (1-p)(v-1)$.

We will derive a recursion for the expressions $(-1)^l L^{(l)}(\gamma) = E(M^l e^{-\gamma M})$. Using the (a, b) structure and the binomial formula, we have for $l \geq 1$,

$$\begin{aligned} (-1)^l L^{(l)}(\gamma) &= \sum_{m=1}^{\infty} e^{-\gamma m} m^l (a + b/m) q_{m-1} \\ &= ae^{-\gamma} \sum_{m=0}^{\infty} e^{-\gamma m} (m+1)^l q_m + be^{-\gamma} \sum_{m=0}^{\infty} e^{-\gamma m} (m+1)^{l-1} q_m \\ &= ae^{-\gamma} \sum_{m=0}^{\infty} e^{-\gamma m} \sum_{r=0}^l \binom{l}{r} m^r q_m + be^{-\gamma} \sum_{m=0}^{\infty} e^{-\gamma m} \sum_{r=0}^{l-1} \binom{l-1}{r} m^r q_m \\ &= ae^{-\gamma} (-1)^l L^{(l)}(\gamma) + e^{-\gamma} \sum_{r=0}^{l-1} \left[a \binom{l}{r} + b \binom{l-1}{r} \right] (-1)^r L^{(r)}(\gamma). \end{aligned}$$

Hence,

$$(-1)^l L^{(l)}(\gamma) = \frac{e^{-\gamma}}{1 - ae^{-\gamma}} \sum_{r=0}^{l-1} \left[a \binom{l}{r} + b \binom{l-1}{r} \right] (-1)^r L^{(r)}(\gamma). \quad (3.2)$$

Notice that these formulas are meaningful because a is always smaller than 1. Now we consider the three different classes of (a, b) distributions.

PROPOSITION 3.1 *Assume that the distribution of M is in the (a, b) class.*

(1) In the Poisson (b) case, $a = 0, b > 0$, we have

$$L(\gamma) = e^{-b(1-e^{-\gamma})}, \quad \gamma \geq 0,$$

$$(-1)^l L^{(l)}(\gamma) = e^{-\gamma} b \sum_{r=0}^{l-1} \binom{l-1}{r} (-1)^r L^{(r)}(\gamma), \quad l \geq 1.$$

(2) In the Bin(n, p) case, $a = -p/(1-p), b = p(n+1)/(1-p), n = 1, 2, \dots$ and $p \in (0, 1)$, we have for $l \geq 1$,

$$L(\gamma) = (1 - p(1 - e^{-\gamma}))^n, \quad \gamma \geq 0,$$

$$(-1)^l L^{(l)}(\gamma) = \frac{p e^{-\gamma}}{1 - p(1 - e^{-\gamma})} \sum_{r=0}^{l-1} \binom{l-1}{r} \left[(n+1) \frac{l-r}{l} - 1 \right] (-1)^r L^{(r)}(\gamma).$$

(3) In the negative binomial case, $a = 1-p, b = (1-p)(v-1), p \in (0, 1)$ and $v > 0$, we have for $l \geq 1$,

$$L(\gamma) = \left(\frac{p}{1 - (1-p)e^{-\gamma}} \right)^v, \quad \gamma \geq 0,$$

$$(-1)^l L^{(l)}(\gamma) = \frac{(1-p)e^{-\gamma}}{1 - (1-p)e^{-\gamma}} \sum_{r=0}^{l-1} \binom{l-1}{r} \left[(v-1) \frac{l-r}{l} + 1 \right] (-1)^r L^{(r)}(\gamma).$$

REMARK 3.2 In the Poisson case, one can also get a different recursion for $L^{(k)}$. Introduce the polynomial H_n of degree n by the recursion

$$H_0(x) = 1, \quad H_n(x) = -x [H'_{n-1}(x) + H_{n-1}(x)], \quad n \geq 1, \quad x > 0.$$

Then calculation yields

$$L^{(k)}(\gamma) = H_k(be^{-\gamma}) L(\gamma), \quad k \geq 0, \quad \gamma > 0.$$

In particular, $R_k(\gamma)$ is a rational function of $be^{-\gamma}$ for each $k \geq 0$:

$$R_k(\gamma) = -\frac{H_{k+1}(be^{-\gamma})}{H_k(be^{-\gamma})} = be^{-\gamma} \left(1 + \frac{H'_k(be^{-\gamma})}{H_k(be^{-\gamma})} \right).$$

REMARK 3.3 Extensions of the (a, b) class were considered in Hess et al. (2002). They introduced distributions $(q_m)_{m=0,1,\dots}$ satisfying the (a, b) condition (3.1) with $q_0, \dots, q_k = 0$ for some $k \geq 0$ and $q_{k+l} > 0$. The calculations leading to the recursion (3.2) for $L^{(l)}$ remain valid in this case as well.

REMARK 3.4 The quantities $(-1)^l L^{(l)}(\gamma)$ grow rapidly as a function of l and therefore standard software delivers the value ∞ even for moderately large values l . This numerical problem can be avoided by writing Eq. (3.2) in terms of the ratios $R_r(\gamma)$ which are relevant for the prediction formulae considered in the previous sections:

$$R_l(\gamma) = \frac{e^{-\gamma}}{1 - ae^{-\gamma}} \sum_{r=0}^l \left[a \binom{l+1}{r} + b \binom{l}{r} \right] (R_r(\gamma) \cdots R_{l-1}(\gamma))^{-1}.$$

The latter recursion for R_l avoids the direct calculation of the large quantities $|L^{(l)}(\gamma)|$.

4. The asymptotic behavior of the prediction

4.1. The behavior of $R_k(\gamma)$ as $k \rightarrow \infty$

In this subsection we study the asymptotic behavior of the predictors $E(S_{j+1}|N_0 = n_0, \dots, N_j = n_j)$, $j = 0, 1, \dots$, when the number of payments $N_0 + \dots + N_j = n_0 + \dots + n_j = k \rightarrow \infty$. The same discussion will apply equally to the multi-step predictors $E(S_{j+\ell+1}|N_0 = n_0, \dots, N_j = n_j)$, $j, \ell = 0, 1, \dots$. In view of the results in Theorem 1 one needs to study the asymptotic behavior of the ratios $R_k(\gamma)$ as $k \rightarrow \infty$.

The interest in the asymptotic behavior of $R_k(\gamma)$ as $k \rightarrow \infty$ is triggered, in particular, by a comparison with Mack's procedure (1.2). The latter declares the predictor of $N_{j+\ell+1}$ given N_0, \dots, N_j to be a linear function of $k = N_0 + \dots + N_j$. In our setting, this predictor is a multiple of $R_k(\gamma)$ which has no reason to be linear. However, this observation does not exclude the case that the limit $k^{-1}R_k(\gamma)$ exists, is finite and positive. In such cases $R_k(\gamma)$ would be approximately linear for large k , as in Mack's procedure.

The following result yields a sufficient condition for asymptotic linearity of $R_k(\gamma)$.

LEMMA 4.1 Assume that $q_m > 0$ for $m \geq m_0$ and the limit

$$\lim_{m \rightarrow \infty} \frac{q_m}{q_{m-1}} = e^{-\tau} \in (0, 1] \tag{4.1}$$

exists. Then,

$$\lim_{k \rightarrow \infty} \frac{R_k(\gamma)}{k} = \frac{1}{\gamma + \tau}. \tag{4.2}$$

Proof. Let $\varepsilon \in (0, 1)$. We decompose $(-1)^k L^{(k)}(\gamma)$ for fixed γ :

$$\begin{aligned} (-1)^k L^{(k)}(\gamma) &= \left(\sum_{m < \frac{k(1-\varepsilon)}{\gamma+\tau}} + \sum_{m \in \left[\frac{k(1-\varepsilon)}{\gamma+\tau}, \frac{k(1+\varepsilon)}{\gamma+\tau} \right]} + \sum_{m > \frac{k(1+\varepsilon)}{\gamma+\tau}} \right) m^k q_m e^{-\gamma m} \\ &= I_1(\varepsilon) + I_2(\varepsilon) + I_3(\varepsilon). \end{aligned}$$

We start by studying $I_3(\varepsilon)$. For $m > k(1+\varepsilon)/(\gamma+\tau)$,

$$\frac{(m+1)^k e^{-(m+1)\gamma}}{m^k e^{-m\gamma}} = e^{-\gamma} (1+m^{-1})^k \leq e^{-\gamma} ((1+m^{-1})^m)^{\frac{\gamma+\tau}{1+\varepsilon}} \leq e^{-\gamma + (\gamma+\tau)/(1+\varepsilon)}.$$

Choose $\delta \in (0, \varepsilon/(1+\varepsilon))$. For k large enough, which implies that $m > k(1+\varepsilon)/(\gamma+\tau)$ is large enough, we have in view of Eq. (4.1),

$$\frac{q_{m+1}}{q_m} \leq e^{-\tau(1-\delta)}.$$

Combining these two bounds, we obtain

$$\frac{q_{m+1} (m+1)^k e^{-(m+1)\gamma}}{q_m m^k e^{-m\gamma}} \leq e^{-\gamma + (\gamma + \tau)/(1+\varepsilon)} e^{-\tau(1-\delta)} =: f(\varepsilon, \delta) < 1,$$

where we used the fact that $\delta \in (0, \varepsilon/(1+\varepsilon))$. Therefore, using the summation formula for geometric series, we obtain for large k ,

$$I_3(\varepsilon) \leq \frac{1}{1 - f(\varepsilon, \delta)} q_{\lceil k(1+\varepsilon)/(\gamma + \tau) \rceil} \left[\frac{k(1+\varepsilon)}{\gamma + \tau} \right]^k e^{-\lceil k(1+\varepsilon)/(\gamma + \tau) \rceil \gamma}. \tag{4.3}$$

Further, for large k the index set in $I_2(\varepsilon)$ contains the point $\lceil k/(\gamma + \tau) \rceil$. Therefore we obtain a lower bound, valid for large k :

$$I_2(\varepsilon) \geq q_{\lceil k/(\gamma + \tau) \rceil} \left[\frac{k}{\gamma + \tau} \right]^k e^{-\gamma \lceil k/(\gamma + \tau) \rceil}. \tag{4.4}$$

A combination of Eqs. (4.3) and (4.4) yields

$$\frac{I_3(\varepsilon)}{I_2(\varepsilon)} \leq \frac{1}{1 - f(\varepsilon, \delta)} \frac{q_{\lceil k(1+\varepsilon)/(\gamma + \tau) \rceil}}{q_{\lceil k/(\gamma + \tau) \rceil}} \left(\frac{\lceil k(1+\varepsilon)/(\gamma + \tau) \rceil}{\lceil k/(\gamma + \tau) \rceil} \right)^k \times e^{-\lceil k(1+\varepsilon)/(\gamma + \tau) \rceil - \lceil k/(\gamma + \tau) \rceil \gamma}.$$

By virtue of Eq. (4.1), for small $\alpha \in (0, 1)$ and large k ,

$$q_{\lceil k(1+\varepsilon)/(\gamma + \tau) \rceil} / q_{\lceil k/(\gamma + \tau) \rceil} \leq e^{-\tau(1-\alpha) (\lceil k(1+\varepsilon)/(\gamma + \tau) \rceil - \lceil k/(\gamma + \tau) \rceil)} \leq e^{-(\tau/(\gamma + \tau)) k (1-\alpha)\varepsilon}.$$

Furthermore, for large k and some positive constant c_1 ,

$$\begin{aligned} \left(\frac{\lceil k(1+\varepsilon)/(\gamma + \tau) \rceil}{\lceil k/(\gamma + \tau) \rceil} \right)^k &\leq \left(\frac{k(1+\varepsilon)/(\gamma + \tau) + 1}{k/(\gamma + \tau) - 1} \right)^k \\ &= (1+\varepsilon)^k \left(1 + \frac{1}{k(1+\varepsilon)/(\gamma + \tau)} \right)^k \left(1 - \frac{1}{k/(\gamma + \tau)} \right)^{-k} \\ &\leq c_1 (1+\varepsilon)^k. \end{aligned}$$

Finally,

$$e^{-\lceil k(1+\varepsilon)/(\gamma + \tau) \rceil - \lceil k/(\gamma + \tau) \rceil \gamma} \leq e^{-\gamma k \varepsilon / (\gamma + \tau)}.$$

Collecting the above bounds and choosing α such that

$$e^{\varepsilon(\tau(1-\alpha) + \gamma)/(\gamma + \tau)} > (1 + \alpha)(1 + \varepsilon),$$

we obtain, for some positive constant c_2 ,

$$\frac{I_3(\varepsilon)}{I_2(\varepsilon)} \leq c_2 (1 + \varepsilon)^k e^{-k\varepsilon(\tau(1-\alpha) + \gamma)/(\gamma + \tau)} \leq c_2 \left(\frac{1}{1 + \alpha} \right)^k.$$

Hence $I_3(\varepsilon) = o(I_2(\varepsilon))$ as $k \rightarrow \infty$.

Next we turn to the estimation of $I_1(\varepsilon)$. Once again, let $\delta \in (0,1)$ be small and choose m_0 so large that

$$\frac{q_{m+1}}{q_m} \geq e^{-\tau(1+\delta)} \quad \text{and} \quad \left(1 + \frac{1}{m}\right)^m \geq e^{1-\delta}, \quad m \geq m_0. \tag{4.5}$$

We further decompose $I_1(\varepsilon)$:

$$I_1(\varepsilon) = \left(\sum_{m < m_0} + \sum_{m \in \left[m_0, k \frac{1-\varepsilon}{\gamma+\tau} \right]} \right) m^k q_m e^{-\gamma m} = I_{11}(\varepsilon) + I_{12}(\varepsilon).$$

Trivially, $I_{11}(\varepsilon) = o(I_2(\varepsilon))$ as $k \rightarrow \infty$. For $m \in [k, k(1-\varepsilon)/(\gamma+\tau)]$ we have by Eq. (4.5)

$$(1 + m^{-1})^k \geq e^{(k/m)(1-\delta)} \geq e^{(\gamma+\tau)(1-\delta)/(1-\varepsilon)},$$

and, therefore,

$$\frac{q_{m+1} (m+1)^k e^{-(m+1)\gamma}}{q_m m^k e^{-m\gamma}} \geq e^{-\gamma} e^{-\tau(1+\delta)} e^{(\gamma+\tau)(1-\delta)/(1-\varepsilon)} = g(\varepsilon, \delta) > 1,$$

if we choose $\delta \in (0,1)$ so small that $(1-\delta)/(1-\varepsilon) > 1 + \delta$. Therefore for k large,

$$I_{12}(\varepsilon) \leq \frac{1}{1 - (g(\varepsilon, \delta))^{-1}} q_{\lfloor k(1-\varepsilon)/(\gamma+\tau) \rfloor} \left[\frac{k(1-\varepsilon)}{\gamma+\tau} \right]^k e^{-\lfloor k(1-\varepsilon)/(\gamma+\tau) \rfloor \gamma},$$

and an argument similar to the one above implies that $I_{12}(\varepsilon) = o(I_2(\varepsilon))$ as $k \rightarrow \infty$ for every $\varepsilon \in (0,1)$.

We conclude that for $\varepsilon \in (0,1)$, as $k \rightarrow \infty$,

$$R_k(\gamma) \sim \frac{\sum_{m \in \left[\frac{k(1-\varepsilon)}{\gamma+\tau}, \frac{k(1+\varepsilon)}{\gamma+\tau} \right]} m^{k+1} q_m e^{-\gamma m}}{\sum_{m \in \left[\frac{k(1-\varepsilon)}{\gamma+\tau}, \frac{k(1+\varepsilon)}{\gamma+\tau} \right]} m^k q_m e^{-\gamma m}} \in \left[\frac{k(1-\varepsilon)}{\gamma+\tau}, \frac{k(1+\varepsilon)}{\gamma+\tau} \right].$$

Here $a_k \sim b_k$ for any sequences (a_k) and (b_k) of non-zero real numbers means that $\lim_{k \rightarrow \infty} a_k/b_k = 1$. Hence, relation (4.2) is immediate. This concludes the proof. \square

In view of Lemma 4.1 and Theorem 2.1 we conclude that, under suitable conditions on the distribution of M , the predictor for S_{j+1} given N_0, \dots, N_j is asymptotically linear.

COROLLARY 4.2 *Assume that $EX_{11} < \infty$ and that the distribution of M satisfies condition (4.1). Then, as $n_0 + \dots + n_j \rightarrow \infty$,*

$$E(S_{j+1} | N_0 = n_0, \dots, N_j = n_j) \sim EX_{11} \mu_{j+1} / (\tau + \theta_j) (n_0 + \dots + n_j).$$

In the rest of this subsection we study the behavior of $R_k(\gamma)$ for large k for the distributions in the (a, b) class introduced in Section 3.

4.1.1. The negative binomial distribution

The *negative binomial distribution* is the only member of the (a, b) class satisfying the condition (4.1) with $e^{-\tau} = 1-p$. Hence Corollary 4.2 applies. The asymptotically linear behavior of $R_k(\gamma)$ is nicely illustrated in the right graph of Figure 1.

4.1.2. The binomial distribution

Here we assume $M \sim \text{Bin}(n, p)$. In this case, it is clear that, as $k \rightarrow \infty$,

$$R_k(\gamma) = \frac{\sum_{m=1}^n (m/n)^k m e^{-\gamma m} q_m}{\sum_{m=1}^n (m/n)^k e^{-\gamma m} q_m} \rightarrow n.$$

The same result holds for any distribution $(q_m)_{m=0, \dots, n}$, $n \geq 1$, with $q_n > 0$. In this case, if $EX_{11} < \infty$, then, as $n_0 + \dots + n_j \rightarrow \infty$,

$$E(S_{j+1} | N_0 = n_0, \dots, N_j = n_j) \rightarrow n EX_{11} \mu p_{j+1}.$$

4.1.3. The Poisson distribution

If M is Poisson(b) distributed, then

$$E(M^k e^{-\gamma M}) = e^{-b(1-e^{-\gamma})} \sum_{m=1}^{\infty} m^k (b/e^{\gamma})^m \frac{1}{m!} e^{-be^{-\gamma}}, \quad k \geq 1.$$

Hence, the ratio $R_k(\gamma)$ is equal to the ratio of the moments $E(M')^{k+1}/E(M')^k$ for some Poisson random variable M' with a different mean, say, λ . In the sequel we study, therefore, the asymptotic behavior, as $k \rightarrow \infty$, of such ratios. For simplicity, we use the notation EM^{k+1}/EM^k instead of the proper $E(M')^{k+1}/E(M')^k$.

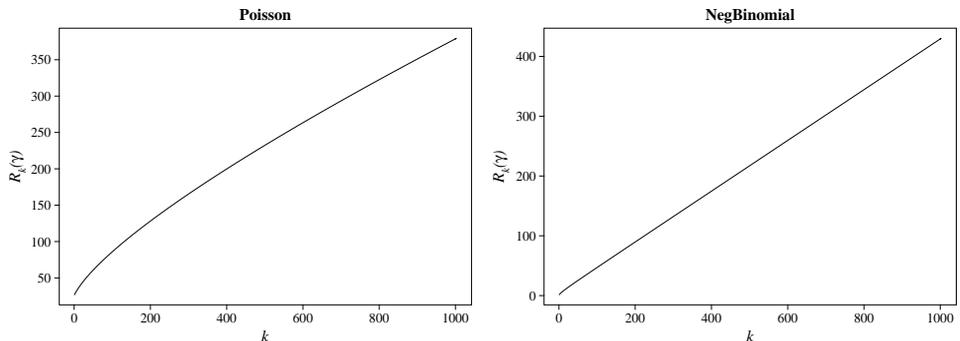


Figure 1. The ratio $R_k(2)$ for Poisson (200) distributed M (left) and for negative binomial M with parameters $p = 0.1$ and $v = 12.1$ (right).

The proof of the following lemma is similar to the proof of Lemma 4.1. We sketch the argument.

LEMMA 4.4 *Let M be Poisson (λ) distributed. Then, as $k \rightarrow \infty$,*

$$\frac{EM^{k+1}}{EM^k} \sim \frac{k}{\log k}. \quad (4.6)$$

Proof. For $\varepsilon \in (0,1)$ we decompose EM^k as

$$EM^k = \left(\sum_{m < \frac{k(1-\varepsilon)}{\log k} + \sum_{m \in \left[\frac{k(1-\varepsilon)}{\log k}, \frac{k(1+\varepsilon)}{\log k} \right]} + \sum_{m > \frac{k(1+\varepsilon)}{\log k}} \right) m^k e^{-\lambda} \frac{\lambda^m}{m!} = I_1(\varepsilon) + I_2(\varepsilon) + I_3(\varepsilon).$$

Beginning with $I_3(\varepsilon)$, it is straightforward to check, using Stirling's formula, that

$$\sup_{m > \frac{k(1+\varepsilon)}{\log k}} \frac{(m+1)^k \lambda^{m+1} / (m+1)!}{m^k \lambda^m / m!} \leq \frac{\lambda \log k}{1+\varepsilon} k^{-\varepsilon/(1+\varepsilon)} \rightarrow 0$$

as $k \rightarrow \infty$. Hence, as $k \rightarrow \infty$,

$$I_3(\varepsilon) = (1 + o(1))(m_*^{+\varepsilon})^k e^{-\lambda} \frac{\lambda^{m_*^{+\varepsilon}}}{(m_*^{+\varepsilon})!},$$

where

$$m_* = \lfloor k/\log k \rfloor, m_*^{+\varepsilon} = \lceil (1+\varepsilon)m_* \rceil.$$

It is clear that for large k ,

$$I_2(\varepsilon) \geq m_*^k e^{-\lambda} \frac{\lambda^{m_*}}{m_*!}.$$

Therefore, as $k \rightarrow \infty$,

$$\frac{I_3(\varepsilon)}{I_2(\varepsilon)} = e^{k(\log(1+\varepsilon) - \varepsilon + o(1))}.$$

Since $\log(1+\varepsilon) < \varepsilon$, we conclude that for every $\varepsilon > 0$,

$$\frac{I_3(\varepsilon)}{I_2(\varepsilon)} \rightarrow 0 \quad \text{as } k \rightarrow \infty.$$

Similarly, as $k \rightarrow \infty$,

$$I_1(\varepsilon) = (1 + o(1))(m_*^{-\varepsilon})^k e^{-\lambda} \frac{\lambda^{m_*^{-\varepsilon}}}{(m_*^{-\varepsilon})!},$$

where

$$m_*^{-\varepsilon} = \lfloor (1 - \varepsilon)m_* \rfloor$$

from which it is easy to check that, as $k \rightarrow \infty$,

$$\frac{I_1(\varepsilon)}{I_2(\varepsilon)} = e^{k(\log(1-\varepsilon) + \varepsilon + o(1))}.$$

Since $-\log(1-\varepsilon) > \varepsilon$ for $0 < \varepsilon < 1$, we conclude that for every $\varepsilon > 0$,

$$\frac{I_1(\varepsilon)}{I_2(\varepsilon)} \rightarrow 0 \quad \text{as } k \rightarrow \infty.$$

That is, for every $\varepsilon > 0$,

$$EM^k = (1 + o(1)) \sum_{m \in \left[\frac{k(1-\varepsilon)}{\log k}, \frac{k(1+\varepsilon)}{\log k} \right]} m^k e^{-\lambda} \frac{\lambda^m}{m!}$$

as $k \rightarrow \infty$ and, hence,

$$\begin{aligned} \frac{EM^{k+1}}{EM^k} &= (1 + o(1)) \frac{\sum_{m \in \left[\frac{k(1-\varepsilon)}{\log k}, \frac{k(1+\varepsilon)}{\log k} \right]} m^{k+1} e^{-\lambda} \frac{\lambda^m}{m!}}{\sum_{m \in \left[\frac{k(1-\varepsilon)}{\log k}, \frac{k(1+\varepsilon)}{\log k} \right]} m^k e^{-\lambda} \frac{\lambda^m}{m!}} \\ &\in \left[(1 + o(1)) \frac{k(1 - \varepsilon)}{\log k}, (1 + o(1)) \frac{k(1 + \varepsilon)}{\log k} \right]. \end{aligned}$$

Hence the statement of the lemma. □

Taking into account Lemma 4.4 and the remark preceding it, we conclude that, if M is Poisson(b) distributed, then, as $n_0 + \dots + n_j \rightarrow \infty$,

$$E(S_{j+1} | N_0 = n_0, \dots, N_j = n_j) \sim EX_{11} \mu_{j+1} \frac{n_0 + \dots + n_j}{\log(n_0 + \dots + n_j)}.$$

The asymptotic behavior of $R_k(\gamma)$ prescribed by Eq. (4.6) is nicely illustrated in the left graph of Figure 1.

4.2. The behavior of $R_k(\gamma)$ for large γ

When evaluating the ratio $R_k(\gamma)$ numerically, one observes a rather unusual phenomenon for large values of γ : $R_k(\gamma)$ oscillates rather strongly for moderately large values of k , whereas this effect gradually disappears when k becomes even larger. For small values of γ this behavior cannot be observed. A computer graph of $R_k(\gamma)$ which exhibits this staircase-like behavior is illustrated in Figure 2. Below we give a limit theorem explaining this phenomenon.

Assume that

$$\gamma \rightarrow \infty, \quad k = k(\gamma) \quad \text{and} \quad k/\gamma \rightarrow t \in (0, \infty). \tag{4.7}$$

For $t > 0$ we consider the strictly convex function

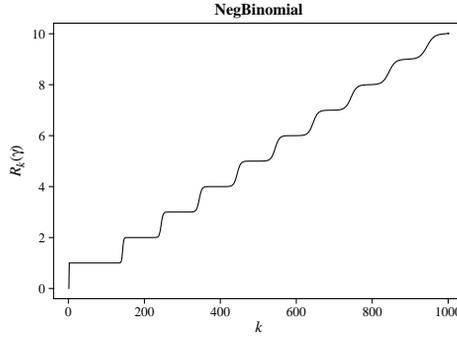


Figure 2. The ratio $R_k(100)$ for negative binomial M with parameters $p=0.1$ and $v=12.1$.

$$h_t(s) = s - t \log s, \quad s > 0,$$

which reaches its minimum at $s = t$. Therefore the sequence $(h_t(m))_{m=1,2,\dots}$ reaches its minimum either at $m = m_-(t) = \lfloor t \rfloor$ if $m_-(t) \geq 1$ or at $m = m_+(t) = \lceil t \rceil$, and it is possible that the values of h_t at $m_-(t)$ and $m_+(t)$ coincide. Let us assume that there exists a unique value $m = m_*(t)$ at which h_t is minimized. We have for $k \geq 1$,

$$E(M^k e^{-\gamma M}) = q_{m_*(t)} (m_*(t))^k e^{-\gamma m_*(t)} + \sum_{m \neq m_*(t)} m^k e^{-\gamma m} q_m.$$

By virtue of Eq. (4.7),

$$\begin{aligned} q_{m_*(t)} (m_*(t))^k e^{-\gamma m_*(t)} &= q_{m_*(t)} e^{-\gamma h_{k/\gamma}(m_*(t))} \\ &\rightarrow q_{m_*(t)} e^{-\gamma h_t(m_*(t))}. \end{aligned} \tag{4.8}$$

The assumption that the minimum of $h_t(m)$ is unique for a given t implies that the minimum of $h_u(m)$ is also unique for u in a small neighborhood of t and achieved at $m_*(t)$. Hence, for u sufficiently close to t ,

$$\begin{aligned} \min_{m \neq m_*(t)} h_u(m) &= \min(h_u(m_*(t) - 1), h_u(m_*(t) + 1)) \\ &\rightarrow \min(h_t(m_* - 1), h_t(m_* + 1)), u \rightarrow t. \end{aligned} \tag{4.9}$$

Therefore for γ sufficiently large,

$$\begin{aligned} \sum_{m \neq m_*(t)} m^k e^{-\gamma m} q_m &= \sum_{m \neq m_*(t)} e^{-\gamma h_{k/\gamma}(m)} q_m \\ &\leq e^{-\gamma \min(h_{k/\gamma}(m_*(t)-1), h_{k/\gamma}(m_*(t)+1))}. \end{aligned} \tag{4.10}$$

It now follows from Eqs. (4.8)–(4.10) that

$$E(M^k e^{-\gamma M}) \sim q_{m_*(t)} (m_*(t))^k e^{-m_*(t) \gamma},$$

provided $q_{m_*(t)} \neq 0$, and, therefore, under assumption (4.7),

$$R_k(\gamma) \rightarrow m_*(t). \tag{4.11}$$

For $j=1,2,\dots$, let

$$a_j = \inf\{t \geq 0 : m_x(t) = j\} = \frac{1}{\log j - \log(j-1)}.$$

Then $a_j < a_{j+1}$, and by relation (4.11) we obtain the following result.

PROPOSITION 4.6 For $j=1,2,\dots$ let $t \in (a_j, a_{j+1})$ be such that $q_{m_x}(t) \neq 0$. If k and γ grow according to Eq. (4.7), then

$$R_k(\gamma) \rightarrow j.$$

That is, if k and λ are large, and $k/\lambda \in (a_j, a_{j+1})$ for j that is not very large, then $R_k(\gamma)$ will be close to j . Once again, this staircase-like behavior is neatly visible in Figure 2.

5. Prediction in a non-life insurance data set

In this section we consider a non-life insurance data set which was kindly provided to us by Alois Gisler. The business line is not known to us. Our aim is to study the performance of our predictors on this data set, given suitable assumptions on the distributions (q_m) and (p_j) and the Poisson parameter μ . Moreover, we will compare our predictors with those prescribed by the chain ladder method under Mack’s conditions; see Section 1. We will focus on the prediction of the numbers of payments.

Our data contain claims that arrive in one year (1985) and the individual payment processes for each claim, including arrival date in 1985 and all dates and amounts of executed payments. Overall, 7302 claims were incurred which triggered 24,606 payments through more than 10 years. Since we want to compare our method with the chain ladder prediction in Mack’s framework, one of the problems we are facing is as follows. The chain ladder method requires a run-off triangle of data from different years; see Eq. (1.3). These data are needed for the construction of the *chain ladder estimators* of the factors f_j in Eq. (1.2):

$$\hat{f}_j = \frac{\sum_{i=1}^{n-j-1} \sum_{r=0}^{j+1} N_{i,i+r}}{\sum_{i=1}^{n-j-1} \sum_{r=0}^j N_{i,i+r}}, \tag{5.1}$$

where n is the number of years for which the run-off triangle is available. But we have only one year of claim arrivals at our disposal.

We solve this problem by switching from years to months. Then we have $n = 12$ months of claim arrivals and the corresponding individual payment processes accounted for by months. Table 1 contains the monthly claim arrival numbers $M_j, j=1,\dots,12$, showing a clear seasonality in the data.

Table 2 contains the payment numbers in run-off triangle form. The i th row contains the payment numbers N_{i1}, \dots, N_{i12} for claims arriving in the i th month and whose

Table 1. The monthly claim numbers M_j .

j	1	2	3	4	5	6	7	8	9	10	11	12
M_j	414	391	493	552	641	765	748	762	741	635	628	532

Downloaded By: [DEFF] At: 13:42 6 June 2010

Table 2. The run-off triangle with the number of payments arising from 12 months of claim arrivals. The i th row contains the observed payment numbers $N_{ii}, \dots, N_{i,12}$.

Months	1	2	3	4	5	6	7	8	9	10	11	12
1	40	147	141	120	99	84	67	63	42	44	47	35
2		40	144	132	100	85	71	53	63	52	50	25
3			39	179	159	125	118	80	85	85	85	48
4				36	209	169	158	117	91	103	98	54
5					60	217	224	159	136	125	110	86
6						75	280	229	186	191	145	121
7							70	234	192	184	157	131
8								63	242	250	197	142
9									62	246	227	195
10										64	200	175
11											59	197
12												28

payments are executed in month $i+k \in \{i, \dots, 12\}$. These numbers form the input to the chain ladder method under Mack's procedure. In this case, the data set contains also the actual 'future' monthly payment numbers $N_{i,12+k}, k \geq 1$. We will try to predict these using both our method and the chain ladder method. We will construct both forecasts for each month of a whole year *after* the month $i, i = 1, \dots, 12$. The constraint on the prediction range is caused by the structure of the chain ladder method, but not of our procedure. Clearly, we do not predict the 'observed' numbers in Table 2. The actual 'future' numbers $N_{i,13}, \dots, N_{i,i+11}$ are in our data set, and they are presented in Table 3.

In our prediction procedure, we assume that the monthly claim numbers $M_j, j = 1, \dots, 12$, are Poisson distributed. Since there is a clear seasonality in the data we do not assume the M_j 's identically distributed. Instead, we simply take each M_j as the substitute of its expectation EM_j . Recall that this is the only parameter of the assumed Poisson distribution needed to compute the functions R and, hence, to run our procedure; see Theorem 2.5. Clearly, the Poisson assumption on the distribution of M_j 's is rather ad hoc, but we have only one data set, which makes it impossible to estimate the distributions of the M_j 's. In general, such estimation has to be done on historical data.

Table 3. The actual 'future' payment numbers corresponding to the observations in Table 2. The i th row contains the monthly payment numbers $N_{i,13}, \dots, N_{i,12}$.

Months	13	14	15	16	17	18	19	20	21	22	23
1											
2	33										
3	52	42									
4	51	45	32								
5	68	53	56	42							
6	107	79	68	75	55						
7	120	83	66	98	68	54					
8	150	105	112	107	88	75	59				
9	148	140	115	131	107	85	80	56			
10	174	118	122	103	82	77	76	57	45		
11	235	167	143	134	116	89	65	66	50	61	
12	177	161	138	139	120	101	113	63	65	86	56

Downloaded By: [DEFF] At: 13:42 6 June 2010

Table 4. The estimated probabilities $p_j, j = 1, \dots, 12$.

j	1	2	3	4	5	6	7	8	9	10	11	12
$100 \times p_j$	2.41	11.36	10.37	7.95	6.49	5.26	4.30	4.04	3.00	2.94	2.60	2.09

On the other hand, the other ingredients needed for our procedure, the average Poisson number of payments per claim μ and the ‘displacement probabilities’ p_j in Eq. (1.4), can be easily estimated from our single data set. The estimate of the former is $\mu = 3.37$, and the estimated values of (p_j) are given in Table 4. Recall that we only need the values of p_j with $j \leq 12$, which are presented in Table 4. In particular, the entries in the table do not add up to 1.

In Table 5 we show the relative prediction error for our prediction procedure. The i th row in this table shows the relative error in our prediction procedure of the values $N_{i,12+k}, k = 1, \dots, i-1$, based on the values $N_{ii}, \dots, N_{i,12}$.

In the left column of Figure 3 we show the performance of our prediction procedure in two different situations. In the top graph we are in the situation of the 11th row of Tables 2 and 3. We predict the payments numbers for claims arriving in the 11th month. Two payment numbers have already been observed from which the payment numbers for months 13–22 get predicted (solid lines). The dots indicate the observed payment numbers and the bands around the predictions represent ± 1.96 times the square of the conditional prediction error given in Theorem 2.5. In the bottom graph we are in the situation of the seventh row in Tables 2 and 3. We predict the payment numbers arising from claims arriving in the seventh month. Six payment numbers have already been observed, and the payment numbers for months 13–18 are predicted.

Our predictions are compared with those prescribed by the chain ladder method under Mack’s procedure (1.2). The multi-step prediction version of the latter is

$$\tilde{N}_{i,12+k} = (f_{12-i+k-1} - 1)f_{12-i+k-2} \cdots f_{12-i} (N_{ii} + \cdots + N_{i,12}), k = 1, 2, \dots,$$

and then one replaces the f factors by their chain ladder \hat{f} estimators given in Eq. (5.1).

Table 5. Relative prediction error in percentage (%) by our method of the values in Table 3 using the observations given in Table 2.

Months	13	14	15	16	17	18	19	20	21	22	23
1											
2	-11.6										
3	-0.8	0.7									
4	-29.8	-33.0	-48.9								
5	-32.9	-56.1	-33.6	-41.9							
6	-2.4	-7.9	-14.8	5.9	-2.2						
7	-1.8	-36.5	-33.5	17.7	-7.3	-7.6					
8	6.0	-15.6	-0.5	18.2	8.9	3.3	2.2				
9	0.4	7.9	3.5	21.5	25.2	13.8	17.2	5.8			
10	14.1	-6.1	10.3	8.5	-6.5	11.7	18.1	1.3	0.5		
11	17.1	3.2	5.4	11.7	12.2	-6.1	-13.0	-1.9	-21.6	20.6	
12	-4.0	-6.1	-2.7	14.7	13.6	11.6	26.7	-2.2	9.3	38.0	24.2

Downloaded By: [DEFF] At: 13:42 6 June 2010

Table 6. Relative prediction error in percentage (%) by the chain ladder method of the values in Table 3 using the observations given in Table 2.

Months	13	14	15	16	17	18	19	20	21	22	23
1											
2	3.4										
3	17.9	2.5									
4	-13.3	-6.5	-28.4								
5	-13.8	-22.0	0.7	-18.7							
6	-11.2	-17.0	-17.1	11.8	-11.7						
7	8.3	-21.7	-21.1	35.9	15.3	-1.4					
8	14.3	-10.5	-0.2	20.8	15.2	20.1	1.7				
9	-6.9	7.3	-1.4	17.4	21.5	11.9	28.8	-2.9			
10	14.3	-8.3	15.4	9.0	-9.3	7.9	23.6	13.3	-3.7		
11	28.0	9.5	11.0	26.6	22.6	-1.6	-9.0	7.2	-0.7	30.4	
12	67.5	68.0	73.3	106.5	117.1	104.4	139.1	68.9	10.21	227.0	129.2

In Table 6 we show the relative prediction error for the chain ladder predictor under Mack's procedure. Once again, the i th row in this table shows the relative error in predicting the values $N_{i,12+k}$, $k=1, \dots, i-1$, based on the values $N_{ii}, \dots, N_{i,12}$.

Similarly, in the right column of Figure 3 we show the chain ladder prediction for the same situations as with our predictor in the left column. The error bands for the chain ladder method seem to be larger than for our method. Note that the bands around the predictions in the right column indicate ± 1.96 times the square of the conditional prediction error given in Mack (1999, p. 363). These numbers are not model based and, consequently, their meaning is unclear.

It should be noted that, in practice, our procedure is likely to be used under far more favorable conditions than in the application presented in this section. A larger historical data set will allow for better estimation of the parameters required to run our forecasting procedure. Most importantly, in the present situation we were unable to estimate the distribution of the number M of the arriving claims, and had to postulate the Poisson distribution. Further, usage of months instead of years introduced seasonality in the data, which our procedure does not take into account. Even under these unfavorable conditions, the present study shows that our predictor performs not worse than the chain ladder prediction as regards closeness to the actual payment numbers and magnitude (as well as reliability) of the conditional prediction errors.

For the three distributions of the (a, b) class our prediction method is easily implemented by using standard software. When taking into account Remark 3.4, one gets quick numerical answers to the prediction problem. Notice that the distributions in the (a, b) class are the most frequently used ones in applications. Our model requires knowledge of the probabilities (q_m) , (p_j) , and the Poisson parameter μ . Once again, these quantities are easy to estimate if historical data are available.

We conclude that our prediction approach, which is based on a well-defined stochastic model of claim arrivals and payment time distribution, works reasonably well in a 'difficult' situation. Our model can be easily simulated from. Recall that the chain ladder

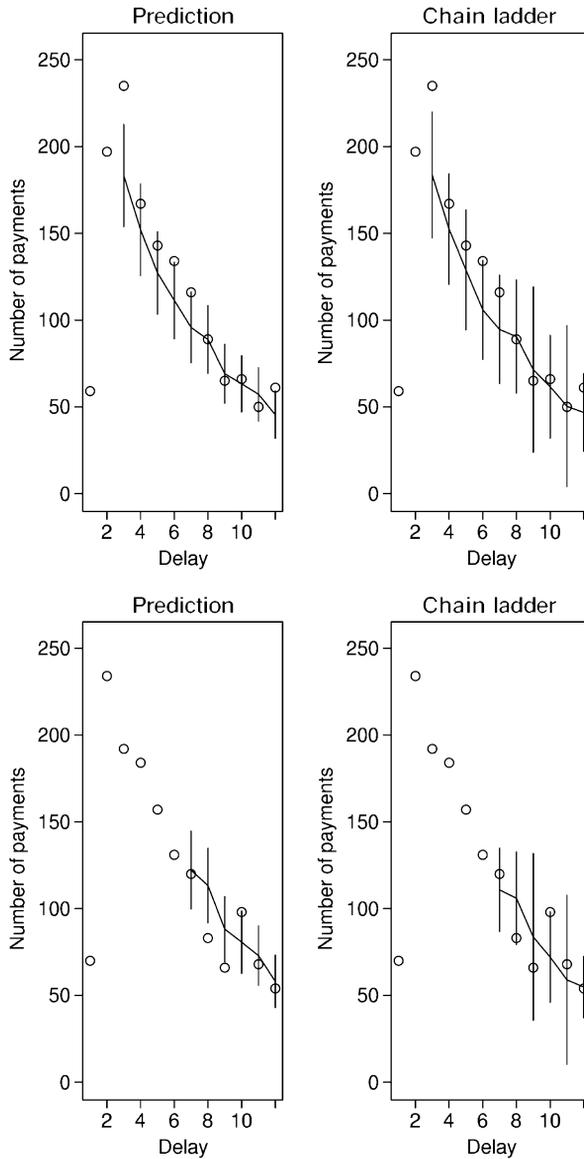


Figure 3. Predictions (solid lines) of monthly payment numbers by using the theory in Section 2.2 (left graphs) and the chain ladder approach (right graphs). The dots represent the payment numbers of the real-life data. The bands around the predictions represent ± 1.96 times the square of the conditional prediction errors. The top graphs correspond to the 11th row in Tables 2 and 3, the bottom ones to the seventh row.

estimator with Mack's procedure is not model based. Therefore, we consider it to be promising to use the suggested forecasting procedure under the dynamics in a non-life insurance portfolio.

Acknowledgements

We would like to thank the referee for carefully reading the paper. His/her suggestions allowed us to improve the presentation.

References

- Hess, K. T., Liewald, A. & Schmidt, K. D. (2002). An extension of Panjer's recursion. *ASTIN Bulletin* **32**, 283–297.
- Mack, T. (1993). Distribution-free calculation of the standard error of chain ladder reserve estimates. *ASTIN Bulletin* **23**, 213–225.
- Mack, T. (1994a). Which stochastic model is underlying the chain ladder method? *Insurance: Mathematics & Economics* **15**, 133–138.
- Mack, T. (1994b). Measuring the variability of chain ladder reserve estimates. In *Proceedings of the Casualty Actuarial Society Forum, Spring 1994*, Vol. 1, pp. 101–182.
- Mack, T. (1999). The standard error of chain ladder reserve estimates: recursive calculation and inclusion of a tail factor. *ASTIN Bulletin* **29**, 361–366.
- Mack, T., Quarg, G. & Braun, C. (2006). The mean square error of prediction in the chain ladder reserving method. A comment. *ASTIN Bulletin* **36**, 543–553.
- Mikosch, T. (2004). *Non-life insurance mathematics. An introduction with stochastic processes*. Heidelberg: Springer.