Anders H. Jessen:

Claims reserving and other topics in non-life insurance mathematics

Preface

This thesis has been prepared in fulfillment of the requirements for the Ph. D. degree at the Department of Mathematical Sciences of the University of Copenhagen. The work has been carried out under the supervision of Prof. Thomas Mikosch and Prof. Jens Perch Nielsen in the period June 2006 to May 2009 at the Department of Mathematical Sciences of the University of Copenhagen. The project was funded in equal parts by the University of Copenhagen, the Danish Research Council (FNU) and Codan Insurance.

In this thesis each chapter is written as an academic paper. The chapters are self-contained and can be read independently. Each chapter is set up as . This structure resulted in some minor notational discrepancies among the different chapters. There will also be minor overlap in the contents.

Acknowledgements

A tremendous thank goes to Prof. Thomas Mikosch for tirelessly reading through my manuscripts; without our discussions and your important comments and suggestion this thesis would not exit. I have learned so much from you while tutoring on your course FM0 in 2002 and 2003, while writing my Master Thesis in 2005 and 2006 and through the last three years' working process. I would also like to thank Thomas Mikosch for facilitating my visit to Prof. Richard Davis at Colombia University in New York, for sending me to the Conference in Actuarial Science and Finance on Samos 2006 and for inviting me to participate in seminars at the Mathematisches Forschungsinstitut Oberwolfach 2007 and 2008. A great thank to Prof. Richard Davis and his wife Patty for letting me and my girlfriend stay at their apartment in New York, for their hospitality and help.

I would like to thank Prof. Jens Perch Nielsen for encouragement and support from the start of the working process. Also thanks to Jørgen Thau and Jens Perch Nielsen for the help to finance travels to Cass Business School in London, RS&A in Horsham, London and Delhi, the GIRO conference 2008 in Newport and Nuffield College in Oxford. In this context I would also like to express gratitude for the impressive hospitality shown to me by Paul Mumford and the actuaries in RS&A Group, Bent Nielsen at Nuffield College and

Richard Verrall at Cass Business School.

On a personal level I would like to thank my colleagues at the Department of Mathematics; Jesper Lund Pedersen, Mogens Steffensen and Jeffrey Collamore. I am also greatful to Martin Jacobsen, Ernst Hansen, Søren Tolver Jensen, Rolf Poulsen and Anders Rahbek for answering questions on diffusion processes, time series, asymptotic theory and R code. My colleagues, Jørgen, Martin, Fredrik, Robert, Jim, Kaspar, Ole, Lars and Tine, at the Research Department in Codan Insurance have all been a great inspiration to work with.

Furthermore, I want to thank my family and friends for general support and encouragement through the past three years. Special thoughts go to my parents Jonna and Henning and my girlfriend Anne-Lene.

Anders H. Jessen
Copenhagen, May 2009

Summary

This thesis studies two main subjects: claims reserving in non-life insurance and regularly varying functions.

The work on regular variation is a continuation of the work presented in my Master Thesis and the paper *Regularly varying functions* was chronologically the first part of this thesis to be written. It is included as the last paper in this thesis.

The primary part (four papers) of this thesis deals with claims reserving. The papers on claims reserving are, for the most part, kept simplistic for the theory to be easily applicable to data. The theory is centered around the methodology of run-off triangles but it is in some cases assumed that additional data is available. Apart from the first introductory section, *Chain ladder and its extension*, the papers on claims reserving do all include data studies based on data from Codan Insurance or Winterthur Insurance.

The chain ladder method is the most commonly used way of calculating a reserve in non-life insurance. The original idea of the chain ladder method is elementary and based on heuristic reasoning but the chain ladder method has since been formulated in a precise mathematical framework. The first paper in this thesis, *Chain ladder and its extension*, serves as an introduction to the chain ladder method considered in a multiplicative Poisson model (and multiplicative GLM) framework. In this framework a number of modifications and extensions have appeared in the literature. Some modifications simply belong to the folklore of the practicing actuary and these modifications do not necessarily have reference in the literature. The aim of the paper is to collect results and modifications related to the chain ladder method and formulate them in a mathematically precise way. Some results are proven in this paper. For results proven elsewhere we give references.

The second paper, *Diagonal effects in claims reserving*, studies one modification of the chain ladder method that includes diagonal effects — effects on payments triggered by, for example, changes in reporting patterns, law practice or economic inflation. For paid low-dimensional run-off triangles empirical analysis suggests that the estimation uncertainty becomes too large for the prediction to be meaningful. To overcome this problem we consider similar models with fewer parameters. The mean value structure of such a model and the according estimation of the parameters is known as the separation method. We extend the separation method to a GLM type framework

and propose a credibility model leading to a Benktander type prediction including diagonal effects. The Benktander method gives a simplistic reserve predictor based on a weighted average of the chain ladder method and the prior expected value of the total payments for an underwriting year.

The third paper, *Prediction of RBNS and IBNR claims using claim amounts and claim counts*, considers two run-off triangles: one including the number of reported claims and the other a paid run-off triangle. An insurance claim runs through a number of states before it is finally settled. First the claim was incurred but is unknown to the insurance company, then the claim is reported to the insurance company and finally the claim is settled with no, one or more payments to cover the damage. The waiting time from when a claim was incurred to when it is reported is referred to as IBNR-delay (Incurred But Not Reported) and the waiting time from when a claim is reported and until it is fully settled is called RBNS-delay (Reported But Not Settled). The inclusion of a run-off triangle of the number of reported claims add extra information to the reserve prediction compared to the chain ladder method as reported claims are not necessarily represented in the paid triangle. The model we propose explicitly models the IBNR- and RBNS-delays allowing for the prediction of both IBNR- and RBNS-reserves. The sum of these two add up to the total reserve. It is an important argument in this paper that data comes in two run-off triangles because this data format is the most common one in insurance companies. This assumption complicates the statistical analysis but this is a trade-off we are willing to make.

The fourth paper, *Prediction of outstanding payments in a Poisson cluster model*, deals with a similar but more general framework than the paper *Prediction of RBNS and IBNR claims using claim amounts and claim counts*. The model considered is also based on the number of reported claims and paid amounts but allows for a number of different distributional assumptions related to the total number of claims in a year (say). The number of payments triggered by each claim is assumed to be Poisson distributed. Therefore a Poisson cluster model is defined. The asymptotic behavior of the predictors turns out to be closely related to the chain ladder method which to some extent justifies the use of the chain ladder method in the broad framework of a Poisson cluster model. A comparison of the two methods based on a data study shows the resemblance between the models.

The final paper in this thesis, *Regularly varying functions*, considers random variables with regularly varying tail distributions; random variables with such distributions are referred to as regularly varying. Regularly variation of the tail is maintained under a large number of operations on regularly varying random variables. We have collected a number of results related to functions of regularly varying random variables and proved some in the paper. Regular variation is relevant when considering heavy-tailed distributions and in the contexts of extreme value theory and quantitative risk management.

Sammenfatning

I denne afhandling studeres to hoved emner: Skadesreservering i skadesforsikring og regulært varierende funktioner.

Arbejdet med regulær variation er en fortsættelse af arbejdet præsenteret i mit speciale og dette afsnit, *Regularly varying functions*, er inkluderet som den sidste del af denne afhandling.

Den hovedvægten i denne afhandling er lagt på skadesreservering. Papirerne om skadesreservering er generelt forsøgt skrevet i en let tilgængelig form, da vi håber, at dette vil lette anvendelserne i praksis. Teorien tager udgangspunkt i afløbstrekanter, men det antages i nogle tilfælde, at vi har yderligere information til rådighed for at kunne gennemføre analysen. Bort set fra det første afsnit, *Chain ladder and its extensions*, er teorien vedrørende reservering anvendt på data, som er lånt fra Codan forsikring eller Wintertour Insurance.

Chain ladder metoden er den mest anvendte metode til at beregne reserver i skadesforsikring. Den originale idé er ikke baseret på en statistisk model men på intuitiv argumentation. Efterfølgende er modeller blevet formuleret, som på en naturlig måde medfører chain ladder metoden som prædiktion eller estimation af reserven. Identifikationen af sådanne modeller giver således mulighed for at analysere den oprindelige chain ladder metode i en matematisk ramme. Det første afsnit, *Chain ladder and its extensions*, tjener som en introduktion til chain ladder metoden i rammen af en multiplikativ Poisson- og GLM-model. Da chain ladder metoden er vidt udbredt og samtidig meget simpel findes et utal af modifikationer og variationer af den oprindelige Poisson model. En stor del af disse modifikationer er velkendt blandt praktiserende aktuarer dog uden, at der altid findes konkrete henvisninger til resultaterne i litteraturen om skadesforsikring. Målet med dette afsnit er således at samle resultater fra litteraturen, der vedrører chain ladder i rammen af en multiplikativ Poisson model. I tilfælde, hvor vi ikke har kunnet finde et bevis, giver vi selv et.

Andet afsnit, *Diagonal effects in claims reserving*, omhandler en modifikation af chain ladder modellen, der inkluderer diagonale effekter. Studierne tager udgangspunkt i allerede kendte metoder, som er beskrevet i Afsnit 1, men som vi har forsøgt at optimere. Diagonale effekter findes ofte i afløbstrekanter pga. fx ændringer i anmeldelsesmønstre, økonomisk inflation eller ændring i lovpraksis. Vi har studeret lav-dimensionelle afløbstrekanter (under 13 års afløb) og fundet frem til, at det overvejende gælder, at tilføjelsen af diagonale parametre til chain ladder modellen øger estimationsusikkerheden så meget, at reserveprædiktionen bliver ubrugelig. For at omgås dette problem betragter vi en tilsvarende model, der er opbygget over færre parametre. Denne kan ses som en udvidelse af, hvad der kendes som *The separation method*. Vores set up tillader dog, at vi kan opbygge en kredibilitetsmodel omkring første momentstrukturen og af den vej konstruere en Benktander reserveprædiktion, hvor diagonale effekter er inkluderet. Benktander metoden i sin oprindelige form er baseret på et vægtet gennemsnit af en på forhånd forventning til forsikringsårets totale betalinger og chain ladder prædiktionen.

Tredje afsnit, *Prediction of RBNS and IBNR claims using claim amounts and claim counts*, tager udgangspunkt i to afløbstrekanter: En med antallet af anmeldte skader og en med betalinger. En forsikringsskade gennemløber forskellige stadier, før den afsluttes. Først er skaden indtruffet uden at være meldt til forsikringsselskabet, derefter bliver den anmeldt og til sidst afsluttet med ingen, en eller flere udbetalinger til dækning af skaden. Tidsrummet mellem skades indtræffelse og til dens anmeldelse kaldes IBNR-delay'et (Incurred But Not Reported) mens tiden mellem anmeldelse og endelig afslutning kaldes RBNS-delay (Reported But Not Settled). Viden om anmeldelsestidspunktet tillader os eksplicit at modellere både IBNR-delay og RBNS-delay og ad denne vej beregne IBNR og RBNS reserver særskilt. Dvs. reservering relativt til skader, der er anmeldte, og skader, som endnu ikke er anmeldt. Vi antager, at skader enten afsluttes med en betaling eller ingen. Der kan i modellen ikke være flere betalinger per skade. Det er vigtigt for argumenterne i papiret, at dataformatet er to afløbstrekanter, da dette dataformat kan forventes at være tilgængeligt i de fleste tilfælde i praksis.

Det fjerde papir, *Prediction of outstanding payments in a Poisson cluster model*, betragter et mere generelt set up end afsnit tre. Formler for prædiktorer udledes for en lang række fordelinger. Beregningerne bliver særligt pæne, når de indgående fordelinger er indenfor (a,b)-klassen. En ufravigelig antagelse er dog, at antallet af betalinger per skade er Poisson fordelt - heraf også titlen. Den asymptotiske opførsel af prædiktorerne i denne model viser sig at replicere en chain ladder struktur, hvilket i noget omfang retfærdiggør brugen af chain ladder metoden i Poisson cluster modeller. En sammenligning af de to forskellige prædiktionsmetoder ud fra et datastudie viser tilsvarende resultater.

Femte og sidste afsnit, *Regularly varying functions*, betragter fordelinger med regulært varierende halesandsynligheder. Stokastiske variable med regulært varierende halesandsynligheder betegnes regulært varierende stokastiske variable. Regulær variation af en stokastisk variabel viser sig at være invariant under forskellige operationer på variablerne, og i papiret har vi samlet en lang række resultater relateret til dette. En række resultater, for hvilke vi ikke kunne finde beviser i literaturen, er bevist i afsnittet. Regulær variation er relevant i relation til tunghalede fordelinger i ekstrem værditeori samt i risk managemet.

The papers

This thesis consists of the following papers:

1. **Chain ladder and its extensions**
   by Anders H. Jessen.

2. **Diagonal effects in claims reserving**
   by Anders H. Jessen and Niels Rietdorf. Submitted.

3. **Prediction of RBNS and IBNR claims using claim amounts and claim counts**
   by Anders H. Jessen, Jens P. Nielsen and Richard Verrall. Submitted.

4. **Prediction of outstanding payments in a Poisson cluster model**
   by Anders H. Jessen, Thomas Mikosch and Gennady Samorodnitsky. Submitted.

5. **Regularly varying functions**
   by Anders H. Jessen and Thomas Mikosch. Published in 2006 in Publications De L'Institut Mathematique, Nouvelle serie, tome 80 (94), 173-192.

Chain ladder and its extensions

Anders Hedegaard Jessen
Department of Mathematical Sciences, University of Copenhagen

May 27, 2009

Abstract
The chain ladder method is the most commonly used way of calculating a reserve in a non-life insurance company. Although the original idea of the chain ladder method is elementary and based on heuristic reasoning the chain ladder method has since been formulated in a precise mathematical framework. In this framework a number of modifications and extensions have appeared in the literature. Many of these extensions are used in practise without references to specific scientific papers, and they have as such become part of the folklore of the practising actuary. The aim of this paper is to collect results related to the chain ladder method which are commonly used by practitioners. We will use the chain ladder estimates whenever possible, and if this is not the case we will prefer to give results where closed form expressions can be derived.

1 Introduction

Practising actuaries often refer to the *chain ladder method* (CLM) as a way of estimating a reserve for a line of business in a non-life insurance company. The original method was based on heuristic reasoning and probably one out of many competing methods for determining a reserve. The CLM proved robust and intuitively appealing and, by now, the CLM is, to a large extent, standard. The simplistic nature of the CLM allows for easy ad hoc modifications of the calculation procedure. Thus the actuary may include *business knowledge* — practical knowledge based on personal experience about the nature of the business. This pragmatic way of handling claims reserving makes the CLM applicable to a large number of data sets.

The actuarial tradition however requires a more scientific approach to the calculation of reserves. Rather than heuristic reasoning and referring to empirical evidence it is advantageous to analyze the CLM and its modifications in a probabilistic framework, i.e., a stochastic model needs to be formulated in such a way that the CLM appears as a natural consequence of the model specifications. Models related to the CLM were formulated by Kremer (1982, 1985), Mack (1991, 1993, 1994), Hachemeister and Stanard (1975). We consider these models and discuss possible extensions and modifications of the original ideas in a structured way. For known results we give references. In some cases however, we elaborate or clarify the results. If we could not find suitable references in the literature we give proofs of the results.

In order for the results to be applicable in practise the calculations should not be numerically challenging or time consuming as this aspect would disqualify many ideas simply because a busy actuary has not got the time to implement them. We therefore make an effort to use the chain ladder estimates whenever possible. In some cases this approach will make the estimation suboptimal but this is a trade-off we are willing to accept.

We proceed as follows. In Section 2 we introduce some notation and explain the general idea of claims reserving and the CLM. In Section 3 we start by considering the setup of Mack (1991), where count data is the object of interest. In Section 4 we suggest how paid data can be understood in the framework of the CLM. Then we argue that most of the results from Section 3 carry over to Section 4. Finally, in Section 5 we consider the setup of Mack (1993) and relate it to the framework of Section 4. We will use the embedding in a time series to find extensions of the model and even make a connection between the CLM and a diffusion process.

2 Claims reserving

In this section we give a brief introduction to claims reserving and the CLM. We consider a non-life insurance company which sells policies (in one line of business) in a period of time, a year say. This year is referred to as *accident or underwriting year*. The claims regarding an underwriting year will not necessarily all be paid within this year. Due to legal issues, general consideration of the claim, the delay from the claim's occurrence time to the reporting time, etc., some claims are reported and paid in the following years. At some point in time there will however be no more payments regarding underwriting year one; we say that *year one has run off.*

Formally, let X_{ij} be the total claims regarding the underwriting period i which have been paid with j periods delay. The claim amounts X_{ij} with $i+j=n$ have thus been paid in the same calender period (year), namely in period $n \in \mathbb{N}$, but they regard different underwriting periods (period i). At the end of period $m \in \mathbb{N}$ one thus observes the total payments

$$X_{ij}, \qquad (i,j) \in \mathcal{A}_m, \tag{1}$$

where

$$\mathcal{A}_m = \{(i,j) \in \mathbb{N} \times \mathbb{N}_0 : 1 \leq i+j \leq m\}.$$

The triangular array (1) is referred to as a *run-off triangle.*

Assume that all periods of insurance are fully run off in $m_0 \in \mathbb{N}$ periods, i.e., $X_{ij} = 0$ for $j > m_0$. Then the reserve for the underwriting period i is defined as a predictor of the not yet observed amount $X_{i,m-i} + \cdots + X_{i,m_0}$ based on the observations (1). The *total reserve*, or just the *reserve*, is a prediction of the sum of the variables X_{ij}, $(i,j) \in \mathcal{B}_{m,m_0}$, where

$$\mathcal{B}_{m,m_0} = \{(i,j) \in \{1,...,m\} \times \{0,...,m_0\} : i+j \geq m+1\}.$$

We will often assume that $m = m_0$ in which case we write $\mathcal{B}_{m,m_0} = \mathcal{B}_m$.

The CLM works on the accumulated payments

$$Y_{ij} = X_{i0} + \cdots + X_{ij}, \qquad (i,j) \in \mathcal{A}_m \cup \mathcal{B}_m.$$

In this setup, prediction of the random variables Y_{ij}, $(i,j) \in \mathcal{B}_m$, is equivalent to prediction of X_{ij} as $X_{ij} = Y_{ij} - Y_{i,j-1}$, $(i,j) \in \mathcal{B}_m$ where $Y_{i,-1} = 0$. The CLM is based on the *chain ladder factors* which are given as

$$\widehat{f}_j = \frac{\sum_{i=1}^{m-j-1} Y_{i,j+1}}{\sum_{i=1}^{m-j-1} Y_{ij}}, \qquad 0 \leq j \leq m-2. \tag{2}$$

The chain ladder factors can be seen as weighted averages of the increments in the payments from delay j to $j+1$ as

$$\widehat{f}_j = \sum_{i=1}^{m-j-1} W_{ij} \frac{Y_{i,j+1}}{Y_{ij}}, \qquad W_{ij} = \frac{Y_{ij}}{\sum_{i=1}^{m-j-1} Y_{ij}}.$$

Notice that $\sum_{i=1}^{m-j-1} W_{ij} = 1$. Given this interpretation it is appealing to define the corresponding *chain ladder predictors*

$$\widehat{Y}_{ij} = Y_{i,m-i} \prod_{k=m-i}^{j-1} \widehat{f}_k \, , \qquad (i,j) \in \mathcal{B}_m. \tag{3}$$

Of course, the word 'predictor' has not yet been made precise in a mathematical sense since a stochastic model has not been formulated. Assuming that $m = m_0$, the reserve calculated by using the CLM is defined by

$$\widehat{R}_m = \sum_{i=2}^{m} \left[\widehat{Y}_{i,m-1} - Y_{i,m-i} \right] = \sum_{i=2}^{m} \sum_{j=m-i+1}^{m-1} \left[\widehat{Y}_{ij} - \widehat{Y}_{i,j-1} \right] = \sum_{(i,j)\in\mathcal{B}_m} \widehat{X}_{ij},$$

where $\widehat{X}_{ij} = \widehat{Y}_{ij} - \widehat{Y}_{i,j-1}$ for $(i,j) \in \mathcal{B}_m$.

It has now been formulated what is understood by the CLM. In the next section we introduce a stochastic model which reproduces the above formulas.

3 A Poisson model for claim counts

In this section we consider the multiplicative Poisson model suggested by Mack (1991), pp. 105-106. In the framework of Mack's model the CLM appears in a natural way in order to predict reserves.

Assume that N_{ij}, $(i,j) \in \mathcal{A}_m \cup \mathcal{B}_m$, are mutually independent and

$$N_{ij} \sim \text{Pois}\left(e^{\mu_{ij}}\right), \tag{4}$$

where

$$\mu_{ij} = \delta + \alpha_i + \beta_j, \qquad (i,j) \in \mathcal{A}_m \cup \mathcal{B}_m.$$

In a reserving context N_{ij} denotes the number of claims from the underwriting period i which is reported with j periods delay. The parameter vector of model (4) is given by

$$\theta_m = (\alpha_1, ..., \alpha_m, \beta_0, ..., \beta_{m-1}, \delta) \in \mathbb{R}^{2m+1}, \tag{5}$$

i.e., the dimension of the parameter space increases with m.

Before we go through the result of Mack (1991), some observations on the identifiability of the model are in place. The model defined in (4) is obviously not fully identified because

$$\delta + \alpha_i + \beta_j = (\delta + a + b) + (\alpha_i - a) + (\beta_j - b)$$

for all $(i, j) \in \mathcal{A}_m \cup \mathcal{B}_m$ and $a, b \in \mathbb{R}$. Various papers suggest different identification schemes such as for example $\alpha_1 = \beta_0 = 0$ in England and Verrall (1999) or $(\delta, \sum_{j=0}^{m-1} e^{\beta_j}) = (0, 1)$ in Mack (1991), p. 97. In this section we let $\alpha_1 = \beta_0 = 0$. We discuss the identification problem more in detail in Section 3.3.

3.1 Estimation and the connection to the CLM

In this section the connection between the model (4) and the CLM (3) is made precise; see Mack (1991), Appendix A.

Write

$$M_{ij} = N_{ij} + \cdots + N_{i0}, \qquad (i, j) \in \mathcal{A}_m \cup \mathcal{B}_m.$$

Since N_{ij} are mutually independent and Poisson distributed the likelihood equations based on N_{ij}, $(i, j) \in \mathcal{A}_m$, are given by the intuitively appealing equations

$$\begin{aligned} \sum_{i=1}^{m-j} N_{ij} &= \sum_{i=1}^{m-j} e^{\delta+\alpha_i+\beta_j}, \ 1 \le j \le m-1, && (6) \\ M_{i,m-i} = \sum_{j=0}^{m-i} N_{ij} &= \sum_{j=0}^{m-i} e^{\delta+\alpha_i+\beta_j}, \ 2 \le i \le m, && (7) \\ \sum_{i=1}^{m} M_{i,m-i} = \sum_{i=1}^{m}\sum_{j=0}^{m-i} N_{ij} &= \sum_{i=1}^{m}\sum_{j=0}^{m-i} e^{\delta+\alpha_i+\beta_j}. && (8) \end{aligned}$$

As pointed out in Schmidt and Wünsche (1998) equations (6)−(8) correspond to the principal of marginal totals. It follows directly from (6) − (8) that

$$M_{i,m-i} = \sum_{j=0}^{m-i} e^{\widehat{\delta}+\widehat{\alpha}_i+\widehat{\beta}_j} = e^{\widehat{\delta}+\widehat{\alpha}_i} \sum_{j=0}^{m-i} e^{\widehat{\beta}_j}, \qquad 2 \le i \le m. \tag{9}$$

The *chain ladder factors*, here considered as parameters, are defined by

$$f_j = \frac{\sum_{k=0}^{j+1} e^{\delta+\alpha_i+\beta_k}}{\sum_{k=0}^{j} e^{\delta+\alpha_i+\beta_k}} = \frac{\sum_{k=0}^{j+1} e^{\beta_k}}{\sum_{k=0}^{j} e^{\beta_k}}, \qquad 0 \le j \le m-2.$$

It now follows from Mack (1991), Appendix A, that the maximum likelihood estimates for f_j, $0 \le j \le m-2$, are given by

$$\widehat{f}_j = \frac{\sum_{k=0}^{j+1} e^{\widehat{\beta}_k}}{\sum_{k=0}^{j} e^{\widehat{\beta}_k}} = \frac{\sum_{i=1}^{m-j-1} M_{i,j+1}}{\sum_{i=1}^{m-j-1} M_{ij}}, \qquad 0 \le j \le m-2,$$

in agreement with (2). Together with (9) this formula yields

$$\begin{aligned}
\sum_{k=0}^{j} e^{\widehat{\delta}+\widehat{\alpha}_i+\widehat{\beta}_k} &= M_{i,m-i} \prod_{k=j}^{m-i} \left(\widehat{f}_k\right)^{-1}, \qquad (i,j) \in \mathcal{A}_m, \\
\sum_{k=0}^{j} e^{\widehat{\delta}+\widehat{\alpha}_i+\widehat{\beta}_k} &= M_{i,m-i} \prod_{k=m-i}^{j} \widehat{f}_k, \qquad (i,j) \in \mathcal{B}_m.
\end{aligned}$$

To calculate the individual parameter estimators we use the restriction $\alpha_1 = \beta_0 = 0$ together with the formula

$$e^{\widehat{\delta}+\widehat{\alpha}_i+\widehat{\beta}_j} = \sum_{k=0}^{j} e^{\widehat{\delta}+\widehat{\alpha}_i+\widehat{\beta}_k} - \sum_{k=0}^{j-1} e^{\widehat{\delta}+\widehat{\alpha}_i+\widehat{\beta}_k}, \qquad (i,j) \in \mathcal{A}_m \cup \mathcal{B}_m. \tag{10}$$

It follows directly from (10) that, with $(i,j) = (1,0)$, we obtain the maximum likelihood estimator of $e^{\widehat{\delta}}$. Then with $i = 1$ estimators of

$$e^{\widehat{\delta}+\widehat{\beta}_j}, \qquad 1 \le j \le m-1, \tag{11}$$

are obtained. Similarly with $j = 0$ we get estimators

$$e^{\widehat{\delta}+\widehat{\alpha}_i}, \qquad 2 \le i \le m. \tag{12}$$

Finally, if we divide all expressions in (11) and (12) by $e^{\widehat{\delta}}$, we arrive at the full set of the maximum likelihood estimators for the components of θ_m in (5). Thus the chain ladder factors are introduced to help calculating estimators of θ_m in (5).

We continue by studying the asymptotic behavior of the latter parameter estimators.

3.2 Asymptotic results

In this section we inspect the asymptotic behavior of the maximum likelihood estimators given in Section 3.1, in particular the chain ladder factors.

For simplicity we start by considering $(\widehat{f}_0, \widehat{f}_1)$. Let

$$t = \sum_{i=1}^{m-1} e^{\delta+\alpha_i+\beta_0} \qquad \text{and} \qquad s = \frac{\sum_{i=1}^{m-2} e^{\alpha_i}}{\sum_{i=1}^{m-1} e^{\alpha_i}}$$

such that

$$\begin{aligned} \sum_{i=1}^{m-1} M_{i0} &\sim \text{Pois}\,(t)\,, & \sum_{i=1}^{m-1} M_{i1} &\sim \text{Pois}\,(tf_0)\,, \\ \sum_{i=1}^{m-2} M_{i1} &\sim \text{Pois}\,(tsf_0)\,, & \sum_{i=1}^{m-2} M_{i2} &\sim \text{Pois}\,(tsf_0f_1)\,. \end{aligned}$$

By Mikosch (2006), p. 65, it follows that

$$t^{-1}\begin{bmatrix} \sum_{i=1}^{m-1} M_{i0} \\ \sum_{i=1}^{m-1} M_{i1} \\ \sum_{i=1}^{m-2} M_{i1} \\ \sum_{i=1}^{m-2} M_{i2} \end{bmatrix} \overset{asymp}{\sim} \mathcal{N}\left(\xi, t^{-1}\Sigma\right), \qquad t \to \infty,$$

where

$$\xi = \begin{bmatrix} 1 \\ f_0 \\ sf_0 \\ sf_0f_1 \end{bmatrix} \qquad \text{and} \qquad \Sigma = \begin{bmatrix} 1 & 1 & s & s \\ 1 & f_0 & f_0 s & f_0 s \\ s & f_0 s & f_0 s & f_0 s \\ s & f_0 s & f_0 s & f_0 f_1 s \end{bmatrix}.$$

Define the function $h : \mathbb{R}_+^4 \to \mathbb{R}_+^2$ by

$$h((x_1, ..., x_4)') = (x_1 x_2^{-1}, x_3 x_4^{-1})'.$$

Then the Delta method, see for example Casella and Berger (2002), used on h implies that

$$\begin{bmatrix} \widehat{f}_0 \\ \widehat{f}_1 \end{bmatrix} \overset{asymp}{\sim} \mathcal{N}\left(\begin{bmatrix} f_0 \\ f_1 \end{bmatrix}, t^{-1}\begin{bmatrix} f_0(f_0-1) & 0 \\ 0 & (sf_0)^{-1}f_1(f_1-1) \end{bmatrix}\right), \qquad t \to \infty.$$

This result can be generalized to m dimensions as formulated below.

Proposition 1
Let N_{ij}, $(i,j) \in \mathcal{A}_m$, follow the assumption (4) *with $\alpha_1 = \beta_0 = 0$. Then as $\delta \to \infty$,*

$$\begin{bmatrix} \widehat{f}_0 \\ \vdots \\ \widehat{f}_{m-2} \end{bmatrix} \overset{asymp}{\sim} \mathcal{N}\left(\begin{bmatrix} f_0 \\ \vdots \\ f_{m-2} \end{bmatrix}, \operatorname{diag}\begin{pmatrix} \left(\sum_{i=1}^{m-1} e^{\delta+\alpha_i+\beta_0}\right)^{-1} f_0(f_0-1) \\ \vdots \\ \left(\sum_{j=0}^{m-2} e^{\delta+\alpha_1+\beta_j}\right)^{-1} f_{m-2}(f_{m-2}-1) \end{pmatrix}\right).$$

Since $M_{i,m-i}$ is independent of $\widehat{f}_{m-i}, ..., \widehat{f}_{m-2}$ and

$$M_{i,m-i} \overset{asymp}{\sim} \mathcal{N}\left(\sum_{j=0}^{m-i} e^{\delta+\alpha_i+\beta_j}, \sum_{j=0}^{m-i} e^{\delta+\alpha_i+\beta_j}\right), \qquad \delta \to \infty,$$

it follows directly from Proposition 1 that

$$\sup_{x\in\mathbb{R}} \left| P\left(M_{i,m-i} \prod_{j=m-i}^{m-2} \widehat{f}_j \le x\right) - P\left(U \prod_{j=m-i}^{m-2} U_j \le x\right)\right| \to 0, \qquad \delta \to \infty,$$

where $U, U_{m-i}, ..., U_{m-2}$ are mutually independent,

$$U \sim \mathcal{N}\left(\sum_{j=0}^{m-i} e^{\delta+\alpha_i+\beta_j}, \sum_{j=0}^{m-i} e^{\delta+\alpha_i+\beta_j}\right)$$

and

$$U_j \sim \mathcal{N}\left(f_j, \left(\sum_{i=1}^{m-j-1} e^{\delta+\alpha_i+\beta_j}\right)^{-1} f_j(f_j-1)\right), \qquad m-i \le j \le m-2.$$

A similar result is derived in Mikosch (2006), Theorem 2.2.4, p. 60, which immediately yields that

$$\widehat{f}_j = \frac{\sum_{i=1}^{m-j-1} M_{i,j+1}}{\sum_{i=1}^{m-j-1} M_{ij}} = f_j \frac{(\sum_{i=1}^{m-j-1}\sum_{k=0}^{j+1} e^{\mu_{ik}})^{-1} \sum_{i=1}^{m-j-1} M_{i,j+1}}{(\sum_{i=1}^{m-j-1}\sum_{k=0}^{j} e^{\mu_{ik}})^{-1} \sum_{i=1}^{m-j-1} M_{ij}} \overset{a.s.}{\to} f_j$$

as $\delta \to \infty$.

Taylor (2003) shows that the chain ladder estimates are upward biased on the conditional probability measure where $\widehat{f}_j$ is finite. More precisely Taylor (2003) shows that when the chain ladder factors are well defined then

$$E\left(\widehat{f}_j \mid \widehat{f}_j < \infty\right) > f_j$$

for $0 \le j \le m-1$. With $M_{ij} = N_{i0} + \cdots + N_{ij}$ this implies that $\widehat{M}_{i,m-1}$ as defined in (3) is also biased upward in the sense that

$$E\left(\widehat{M}_{i,m-1} \mid \prod_{j=m-i}^{m-2} \widehat{f}_j < \infty\right) = E\left(M_{i,m-i} \prod_{j=m-i}^{m-2} \widehat{f}_j \mid \prod_{j=m-i}^{m-2} \widehat{f}_j < \infty\right) > EM_{i,m-1}$$

for $1 \le i \le m$; see Taylor (2003) for a proof.

3.3 Identification and prediction

In this section we follow the more general approach given in Kuang et. al (2008) where it is pointed out that a constant can be removed from α_i and β_j by taking differences such as $\Delta\alpha_i = \alpha_i - \alpha_{i-1}$. In particular by choosing $a = \alpha_1$ and $b = \beta_0$ we get

$$\mu_{ij} = \mu_{10} + \sum_{k=2}^{i} \Delta\alpha_k + \sum_{k=1}^{j} \Delta\beta_k,$$

where $\sum_{k=2}^{1} \Delta\alpha_k = \sum_{k=1}^{0} \Delta\beta_k = 0$ by convention. A canonical parametrization, θ_m^*, can thus be generated using differences and one initial point,

$$\theta_m^* = (\Delta\alpha_2, ..., \Delta\alpha_m, \Delta\beta_1, ..., \Delta\beta_{m-1}, \mu_{10}).$$

In group-theoretical terms this can be formulated as follows. Consider

$$\mu = \{\mu_{ij}, (i,j) \in \mathcal{A}_m \cup \mathcal{B}_m\}$$

a function of θ_m and define the group

$$g : \begin{pmatrix} \alpha_i, & 1 \le i \le m \\ \beta_j, & 0 \le j \le m-1 \\ \delta & \end{pmatrix} \mapsto \begin{pmatrix} \alpha_i - a, & 1 \le i \le m \\ \beta_j - b, & 0 \le j \le m-1 \\ \delta + a + b & \end{pmatrix}.$$

Then μ is invariant under g because $\mu(g(\theta_m)) = \mu(\theta_m)$. In Kuang (2008) it is proved that θ_m^* is the maximal invariant function of θ_m under g.

The identification scheme proposed in England and Verrall (1999), $\alpha_1 = \beta_0 = 0$, is obtained by letting

$$\delta = \mu_{10}, \qquad \alpha_i = \sum_{k=2}^{i} \Delta\alpha_k, \qquad \beta_j = \sum_{k=1}^{j} \Delta\beta_k.$$

We will use this convention throughout this paper unless mentioned otherwise. The identification scheme of Mack (1991), $(\delta, \sum_{j=0}^{m-1} e^{\beta_j}) = (0, 1)$, can be obtained from the canonical parameter by

$$\delta = 0, \qquad \alpha_i = C + \mu_{10} + \sum_{k=2}^{i} \Delta\alpha_k, \qquad \beta_j = -C + \sum_{k=1}^{j} \Delta\beta_k,$$

where

$$C = \log\left(\sum_{j=1}^{m-1} e^{\sum_{k=1}^{j} \Delta\beta_k} \right).$$

Notice that e^{β_j} is that proportion of the total number of claims which is reported with j periods delay, whereas e^{α_i} is the (expected) total claim amount for underwriting period i. This identification scheme is in agreement with the idea of the Bornheutter-Ferguson method which is considered in Sections 3.5 and 3.6 below.

Out of sample prediction

In the models which have been considered up to this point it has been assumed that $m_0 = m$ (using the notation of Section 2). That is, it has been assumed that $EN_{ij} = 0$ for $j \geq m$. In many cases of interest this assumption does not hold. If payment of claims related to underwriting year i takes more than m periods one should somehow estimate EN_{ij} for $m \leq j \leq m_0$. The estimator of EN_{ij} for $m \leq j \leq m_0$ should however not depend on the choice of identification scheme and the following result insures that this is not the case.

Proposition 2
An estimator of EN_{ij}, $m \le j \le m_0$, which is a function of

$$\theta_m^* = (\Delta\alpha_2, ..., \Delta\alpha_m, \Delta\beta_1, ..., \Delta\beta_{m-1}, \mu_{10})$$

does not depend on the choice of identification scheme.

The proof of Proposition 4 (in a slightly more general setting) is given in Kuang (2008).

A straightforward consequence of Proposition 4 is that any estimation of $f_m, ..., f_{m_0-1}$ based on $\widehat{f}_0, ..., \widehat{f}_{m-1}$ is invariant under the choice of identification scheme. A number of ways of predicting from the development factors $f_0, ..., f_{m-1}$ are given in Boor (2006). Another straightforward way of estimating EN_{ij}, $m \le j \le m_0$, is simply by estimating the differences $\Delta\beta_m, ..., \Delta\beta_{m_0-1}$ based on $\Delta\widehat{\beta}_1, ..., \Delta\widehat{\beta}_{m-1}$.

3.4 An extended model with overdispersion

Overdispersion is often empirically observed in data. That is

$$VN_{ij} = \gamma EN_{ij}$$

for some $\gamma > 1$. A model compatible with this property is given if N_{ij} has a negative binomial distribution. We assume this distribution in this section.

Let Θ_{ij}, $(i,j) \in \mathcal{A}_m \cup \mathcal{B}_m$, be mutually independent random variables with distribution

$$\Theta_{ij} \sim \Gamma\left(\frac{\gamma^{-1}}{1-\gamma^{-1}} e^{\delta+\alpha_i+\beta_j}, \frac{\gamma^{-1}}{1-\gamma^{-1}}\right).$$

Assume further that given Θ_{ij}, $(i,j) \in \mathcal{A}_m \cup \mathcal{B}_m$, the variables N_{ij}, $(i,j) \in \mathcal{A}_m \cup \mathcal{B}_m$, are conditionally independent and

$$N_{ij} \sim \mathrm{Pois}(\Theta_{ij}).$$

With

$$r_{ij} = \frac{\gamma^{-1}}{1-\gamma^{-1}} e^{\mu_{ij}}$$

it follows; see Mikosch (2006), p. 72, that

$$N_{ij} \sim \mathrm{NegBin}(\gamma^{-1}, r_{ij}).$$

Straightforward calculation yields that $EN_{ij} = EE(N_{ij}|\Theta_{ij}) = e^{\mu_{ij}}$ which is similar to the model (4), and

$$VN_{ij} = EV(N_{ij} \mid \Theta_{ij}) + VE(N_{ij} \mid \Theta_{ij}) = e^{\mu_{ij}} + \frac{(1-\gamma^{-1})e^{\mu_{ij}}}{\gamma^{-1}} = \gamma EN_{ij},$$

which is the desired overdispersion property. These conditions define a GLM setup as in McCullagh and Nelder (1989), Table 9.1. McCullagh and Nelder propose a Poisson quasi likelihood function to estimate the parameters α_i, β_j and δ. This procedure leads to the same estimators as in (10). McCullagh and Nelder (1989), p. 328, give a moment estimator of γ by

$$\widehat{\gamma} \;=\; \frac{2}{m(m-1)} \sum_{i=1}^{m} \sum_{j=0}^{m-i} \frac{(N_{ij} - e^{\widehat{\mu}_{ij}})^2}{e^{\widehat{\mu}_{ij}}}$$

which also suggests an estimator for r_{ij}, namely,

$$\widehat{r}_{ij} = \frac{\widehat{\gamma}^{-1}}{1-\widehat{\gamma}^{-1}} e^{\widehat{\mu}_{ij}}, \qquad (i,j) \in \mathcal{A}_m \cup \mathcal{B}_m.$$

3.5 The Bornheutter-Ferguson method

In this section we consider a simple modification of the chain ladder method called the Bornheutter-Ferguson method; see for example Schmidt and Zocher (2008) for a detailed description of this method and related issues. It will be convenient to use the identification scheme

$$\left(\delta, \sum_{j=0}^{m-1} e^{\beta_j}\right) = (0, 1),$$

in which case e^{β_j} is the proportion of the total number of claims which are reported with j periods delay, whereas e^{α_i} is the (expected) total claim for underwriting period i (the net premium for period i). The Bornheutter-Ferguson method simply exploits the idea that e^{α_i} is replaced by its prior expected value, the net premium, T_i. In this way the reserve estimate becomes

$$\widehat{R}_m = \sum_{i=2}^{m} T_i \sum_{j=m-i+1}^{m-1} e^{\widehat{\beta}_j}.$$

This simple idea proves to be an easy and practical way of modifying the chain ladder estimates if the estimates of e^{α_i} seem unprecise and far away from T_i. The Bornheutter-Ferguson method bases the reserve estimate purely on prior information, whereas the chain ladder method starts a recursion purely from the observed data, namely from $M_{i,m-i} = N_{i0} + \cdots + N_{i,m-i}$. An obvious way of combining these two outer positions is based on a Bayesian idea which used both prior information as well as the actual observed values. Such an approach is described in the next section.

3.6 A credibility model combining the chain ladder and Bornheutter-Ferguson methods

The idea of combining the CLM and the Bornheutter-Ferguson method has been addressed in a number of papers; see for example Benktander (1976), Hovinen (1981), Neuhaus (1992), Mack (2000) and Wüthrich (2008). We will however choose an alternative approach which is closely related to Verrall (2004) but formulated in the framework of this paper. In this framework a simple credibility estimate can be derived. The credibility model is a conditional chain ladder model defined as follows. Let

$$\Theta_i \sim \Gamma(\alpha T_i, \alpha)$$

be mutually independent for $i \geq 1$, and assume that given Θ_i,

$$N_{ij} \sim \text{Pois}(\Theta_i e^{\beta_j}), \qquad \sum_{j=0}^{m-1} e^{\beta_j} = 1. \tag{13}$$

Then $E\Theta_i = T_i$, $V\Theta_i = \alpha^{-1} T_i$ and

$$\log E(N_{ij} \mid \Theta_i) = \log \Theta_i + \beta_j.$$

This situation is similar to (4) provided the identification scheme $(\delta, \sum_{j=0}^{m-1} e^{\beta_j}) = (0, 1)$ is applied.

The idea is to estimate Θ_i by the Bayes estimator, $\widehat{\mu}_i$, defined as the minimizer of $E(\widehat{\mu}_i - \Theta_i)^2$ over all measurable, finite variance functions of N_{ij}, $(i,j) \in \mathcal{A}_m$. Following an approach similar to, for example Mikosch (2006), p. 197, we obtain the Bayes estimate

$$\widehat{\mu}_i = \omega_i T_i + (1-\omega_i)\frac{\sum_{j=0}^{m-i} N_{ij}}{\sum_{j=0}^{m-i} e^{\beta_j}}, \tag{14}$$

where $\omega_i = (\alpha^{-1}\sum_{j=0}^{m-i} e^{\beta_j} + 1)^{-1}$. Notice that

$$\begin{aligned}\frac{\sum_{j=0}^{m-i} N_{ij}}{\sum_{j=0}^{m-i} e^{\beta_j}} &= \left(\sum_{j=0}^{m-i} N_{ij}\right)\frac{\sum_{j=0}^{m-i+1} e^{\beta_j}}{\sum_{j=0}^{m-i} e^{\beta_j}}\frac{\sum_{j=0}^{m-i+2} e^{\beta_j}}{\sum_{j=0}^{m-i+1} e^{\beta_j}}\cdots\frac{\sum_{j=0}^{m-1} e^{\beta_j}}{\sum_{j=0}^{m-2} e^{\beta_j}} \\ &= \left(\sum_{j=0}^{m-i} N_{ij}\right)\prod_{j=m-i}^{m-1} f_j \end{aligned} \tag{15}$$

as $\sum_{j=0}^{m-1} e^{\beta_j} = 1$ by definition. This proves that, if the chain ladder estimates are used for f_j, $0 \le j \le m-1$, then the right hand side in (14) is in fact the chain ladder reserve estimate.

To estimate the parameter α we use the following result.

Proposition 3
Define

$$\widehat{\alpha^{-1}} = \frac{1}{m}\sum_{i=1}^{m}\frac{\left(\sum_{j=0}^{m-i} N_{ij}\right)^2 - T_i\left(\sum_{j=0}^{m-i} e^{\beta_j}\right) - T_i^2\left(\sum_{j=0}^{m-i} e^{\beta_j}\right)^2}{T_i\left(\sum_{j=0}^{m-i} e^{\beta_j}\right)^2}.$$

Then $E\widehat{\alpha^{-1}} = \alpha^{-1}$ *and if*

$$\frac{1}{m^2}\sum_{i=1}^{m} T_i^{k-2}\left(\sum_{j=0}^{m-i} e^{\beta_j}\right)^{k-4} \to 0, \qquad m \to \infty,$$

for $1 \le k \le 4$ *it also holds that* $\widehat{\alpha^{-1}} \xrightarrow{P} \alpha^{-1}$ *as* $m \to \infty$.

Proof: First observe that

$$E(N_{i.}^2) = T_i^2\left(\sum_{j=0}^{m-i} e^{\beta_j}\right)^2 + \frac{T_i}{\alpha}\left(\sum_{j=0}^{m-i} e^{\beta_j}\right)^2 + T_i\sum_{j=0}^{m-i} e^{\beta_j}$$

with $N_{i.} = N_{i0} + \cdots + N_{i,m-i}$. This relation directly imply that the estimator

$$\widehat{\alpha^{-1}} = \frac{1}{m}\sum_{i=1}^{m} \frac{N_{i.}^2 - T_i\left(\sum_{j=0}^{m-i} e^{\beta_j}\right) - T_i^2\left(\sum_{j=0}^{m-i} e^{\beta_j}\right)^2}{T_i\left(\sum_{j=0}^{m-i} e^{\beta_j}\right)^2}$$

is unbiased. Chebyshev's inequality then yield that

$$P\left(\left|\widehat{\alpha^{-1}} - \alpha^{-1}\right| > \epsilon\right) \leq \frac{V\widehat{\alpha^{-1}}}{\epsilon^2}$$

where

$$\begin{aligned}
V\widehat{\alpha^{-1}} &= \frac{1}{m^2}\sum_{i=1}^{m} \frac{EE(N_{i.}^4 \mid \Theta_i) - [EE(N_{i.}^2 \mid \Theta_i)]^2}{T_i^2\left(\sum_{j=0}^{m-i} e^{\beta_j}\right)^4} \\
&\leq \frac{1}{m^2}\sum_{i=1}^{m} \frac{EE\left(N_{i.}^4 \mid \Theta_i\right)}{T_i^2\left(\sum_{j=0}^{m-i} e^{\beta_j}\right)^4} \\
&\leq \frac{1}{m^2}\sum_{i=1}^{m}\sum_{k=1}^{4} \frac{c_k T_i^{k-2}}{\left(\sum_{j=0}^{m-i} e^{\beta_j}\right)^{4-k}}.
\end{aligned}$$

The constants c_k are calculated in the following way. Since the factorial moments $E(N_{i.}^{(k)} \mid \Theta_i) = \Theta_i^k$ it holds that

$$\begin{aligned}
E(N_{i.}^4 \mid \Theta_i) &= \left(\Theta_i \sum_{j=0}^{m-i} e^{\beta_j}\right)^4 + 6\left(\Theta_i \sum_{j=0}^{m-i} e^{\beta_j}\right)^3 \\
&\quad +7\left(\Theta_i \sum_{j=0}^{m-i} e^{\beta_j}\right)^2 + \left(\Theta_i \sum_{j=0}^{m-i} e^{\beta_j}\right). \qquad (16)
\end{aligned}$$

And as Θ_i is gamma distributed

$$E\Theta_i^k = \frac{\prod_{l=0}^{k-1}(\alpha T_i + l)}{\alpha^l}.$$

It directly follows that $T_i \mapsto EE(N_{i.}^4 \mid \Theta_i)$ is a polynomial of order 4 where the constants $c_k > 0$ are determined by plugging the moments of Θ_i in (16) .

$\square$

The original credibility estimator of Θ_i proposed by Benktander (1976) is given in the form (14) with $\omega_i = 1 - \sum_{j=0}^{m-i} e^{\beta_j}$. The so-called Benktander estimator of Θ_i is thus given by

$$\widehat{\mu}_i^B = T_i \left(1 - \sum_{j=0}^{m-i} e^{\beta_j}\right) + \sum_{j=0}^{m-i} N_{ij}. \qquad (17)$$

The according reserve related to (14) and (17) is given as

$$\widehat{R}_m = \sum_{i=2}^{m} \widehat{\mu}_i^B \sum_{j=m-i+1}^{m-1} e^{\widehat{\beta}_j}.$$

3.7 Diagonal effects

In this section we consider diagonal effects. Recall that N_{ij} denotes the number of reported claims regarding underwriting period i which have been reported with j periods delay. As motivation assume that we consider data from third part liability auto policies where a lawsuit to have whiplash injuries accepted as legitimate claims has been won in period $n \leq m$. It is then likely that a number of other whiplash claims will be reported in this calender period $i + j = n$ (and possibly the subsequent periods, $n+1, n+2, ...$, as well). The average number of claims reported along the nth diagonal will thus be relatively high. The model (4) however does not allow for diagonal effects.

For a number of reasons calender effects appear all kinds of data. In this section, (4) is extended to include diagonal effects such as described above. The following is based on Kuang, Nielsen and Nielsen (2008a, 2008b). To start with assume that N_{ij} are mutually independent such that $N_{ij} \sim \text{Pois}(e^{\mu_{ij}})$ where

$$\mu_{ij} = \delta + \alpha_i + \beta_j + \pi_{i+j}. \qquad (18)$$

Here π_{i+j} represents the diagonal effects and α_i and β_j have the same interpretations as in (4). The model (18) depends on the parameter

$$\theta_m = (\delta, \alpha_1, ..., \alpha_m, \beta_0, ..., \beta_{m-1}, \pi_1, ..., \pi_{2m-1})$$

and (18) is obviously not identified. In fact if $\mu = \{\mu_{ij}, (i,j) \in \mathcal{A}_m \cup \mathcal{B}_m\}$ is considered a function of θ_m then μ is invariant under g, i.e., $\mu(g(\theta_m)) = \mu(\theta_m)$, where

$$g : \begin{pmatrix} \alpha_i, \ 1 \leq i \leq m \\ \beta_j, \ 0 \leq j \leq m-1 \\ \pi_{i+j}, \ 1 \leq i+j \leq m \\ \delta \end{pmatrix} \mapsto \begin{pmatrix} \alpha_i - a - d(i-1), & 1 \leq i \leq m \\ \beta_j - b - dj, & 0 \leq j \leq m-1 \\ \pi_{i+j} - c + d(i+j-1), & 1 \leq i+j \leq 2m-1 \\ \delta + a + b + c & \end{pmatrix}$$

and a, b, c and d are arbitrary constants. In comparison to (4), in (18) one can add and subtract a linear trend in contrast to (4) where one only could subtract/add a constant. A linear trend can be removed by using double differences such as $\Delta^2\alpha_3 = \Delta(\alpha_3 - \alpha_2) = \alpha_3 - 2\alpha_2 + \alpha_1$. For example

$$\Delta^2(\alpha_i - a - d(i-1)) = \Delta(\Delta\alpha_i - d) = \Delta^2\alpha_i.$$

A canonical parametrization, θ_m^*, can thus be generated using double differences and three initial points,

$$\theta_m^* = (\Delta^2\alpha_3, ..., \Delta^2\alpha_m, \Delta^2\beta_2, ..., \Delta^2\beta_{m-1}, \Delta^2\pi_3, ..., \Delta^2\pi_{2m-1}, \mu_{10}, \mu_{11}, \mu_{20}).$$

Kuang (2008a, 2008b) shows that θ_m^* is the maximal invariant function of θ_m under g. In particular, if one puts

$$\begin{aligned} \alpha_i &= (i-1)(\mu_{20} - \mu_{10}) + \sum_{k=3}^{i}\sum_{l=3}^{k} \Delta^2\alpha_l, \\ \beta_j &= j(\mu_{11} - \mu_{10}) + \sum_{k=3}^{j}\sum_{l=3}^{k} \Delta^2\beta_l, \\ \pi_{i+j} &= \sum_{k=3}^{i+j}\sum_{l=3}^{k} \Delta^2\pi_l, \\ \delta &= \mu_{10}, \end{aligned}$$

then the identification scheme $\alpha_1 = \beta_0 = \pi_1 = \pi_2 = 0$ is obtained. With this choice of identification scheme Venter (2007) gives likelihood equations which can be recursively run through until the parameter estimates converge,

$$
\begin{aligned}
e^{\alpha_i} &= \frac{\sum_{j=0}^{m-i} X_{ij}}{\sum_{j=0}^{m-i} e^{\beta_j+\pi_{i+j}}}, \qquad 2 \le i \le m, \\
e^{\beta_j} &= \frac{\sum_{i=1}^{m-j} X_{ij}}{\sum_{i=1}^{m-i} e^{\alpha_i+\pi_{i+j}}}, \qquad 1 \le j \le m-1, \\
e^{\pi_k} &= \frac{\sum_{i=1}^{m-k} X_{i,m-k-i}}{\sum_{i=1}^{m-k} e^{\alpha_i+\beta_{m-k-i}}}, \qquad 3 \le k \le m.
\end{aligned}
$$

Venter (2007) mentions that usually around 50 loops are sufficient when using the usual chain chain ladder estimates for e^{α_i} and e^{β_j} and $e^{\pi_k} = 1$ as starting points for the recursion. The canonical parameter is then estimated by, for example,

$$
e^{\widehat{\Delta^2\alpha_i}} = e^{\widehat{\alpha}_i - 2\widehat{\alpha}_{i-1} + \widehat{\alpha}_{i-2}}
$$

for $3 \le i \le m$ and accordingly for $e^{\Delta^2\pi_{i+j}}$ and $e^{\Delta^2\beta_j}$.

In order to estimate EN_{ij}, $(i,j) \in \mathcal{B}_m$, it is necessary to predict π_{i+j} for $m+1 \le i+j \le 2m-1$. The most common approach is to assume that the likelihood estimates are observations from some time series from which we can predict future calender effects. When applying a method to do this one should take the following result into account.

Proposition 5
A predictor of π_{i+j}, $m+1 \le i+j \le 2m-1$, which is a function of θ_m^* does not depend on the choice of identification scheme.

□

A similar result holds for β_j, $j \ge m-1$.

This concludes our studies of the CLM in connection with count data. We continue by studying CLM used on a so-called paid run-off triangle in the next section.

4 Paid amounts

In most cases the real objects of interest are the payments triggered by the reported number of claims. In claims reserving one usually wants to estimate the ultimate amount which will be paid for each underwriting period along with the related cashflows. As described in Section 2 it is an empirically proven fact that the CLM works well on payments. In this section we will explain why the CLM is a reasonable way of estimating the parameters describing the payments triggered by N_{ij}, $(i, j) \in \mathcal{A}_m$. We will hence propose a model for the payments which inherits the important properties of the model (4). In fact most of the results in Section 3 carry over to the model for paid amounts.

4.1 A model for paid amounts

Assume that the number of reported claims, N_{ij}, triggers a vector of mutually independent numbers of payments

$$N_{ijl} \sim \text{Pois}(e^{\delta+\alpha_i+\beta_j+\rho_l}), \qquad l \geq 0.$$

For fixed i, j and l payments, $Z_{ijl}^{(k)}$, $k \geq 1$, are assumed iid with finite second moment. Moreover, $Z_{ijl}^{(k)}$, $k \geq 1$, and N_{ijl} are all assumed mutually independent for $(i, j) \in \mathcal{A}_m \cup \mathcal{B}_m$ and $0 \leq l \leq d$. Let the relative variance of each payment, defined by

$$\phi = \frac{E\left(Z_{ijl}^{(k)}\right)^2}{EZ_{ijl}^{(k)}},$$

be constant. Some bookkeeping tells us that the total payments regarding period i which are paid with j periods delay are given by

$$X_{ij} = \sum_{l=0}^{j} \sum_{k=1}^{N_{i,j-l,l}} Z_{i,j-l,l}^{(k)},$$

where $N_{ijl} = 0$ for $l < 0$ by convention. It now follows that

$$EX_{ij} = e^{\delta+\alpha_i} \sum_{l=0}^{j} e^{\beta_{j-l}+\rho_l} EZ_{i,j-l,l}^{(1)} = e^{\delta+\alpha_i+\zeta_j} \tag{19}$$

where

$$e^{\zeta_j} = \sum_{l=0}^{j} e^{\beta_{j-l}+\rho_l} E Z^{(1)}_{i,j-l,l}.$$

Further, as X_{ij}, $(i,j) \in \mathcal{A}_m \cup \mathcal{B}_m$, are mutually independent we get that

$$VX_{ij} = e^{\delta+\alpha_i} \sum_{l=0}^{j} e^{\beta_{j-l}+\rho_l} E\left(Z^{(1)}_{i,j-l,l}\right)^2 = \phi e^{\delta_0+\alpha_i+\zeta_j}$$

which yields the relation

$$VX_{ij} = \phi E X_{ij}, \qquad \forall (i,j) \in \mathcal{A}_m \cup \mathcal{B}_m. \tag{20}$$

The mean and variance structures (19) and (20) are closely related to the model (4) and it turns out to be sufficient to explain why the CLM works on paid data as described above.

For simplicity we will use the same notation as in the latter sections and let $\mu_{ij} = \delta + \alpha_i + \zeta_j$.

4.2 Estimation

In this section it is assumed that N_{ij}, $(i,j) \in \mathcal{A}_m$, are not observed. The estimations are based purely on the observations X_{ij}, $(i,j) \in \mathcal{A}_m$.

The only distributional characteristics specified for X_{ij}, $(i,j) \in \mathcal{A}_m \cup \mathcal{B}_m$ are mean and variance along with mutual independence. For data with such a specification McCullagh and Nelder (1989), p. 326, propose a Poisson quasi likelihood function. The idea consists of borrowing the likelihood function from a known distribution with similar mean and variance relation which belongs to an exponential family of distributions. As in Section 3.4, this approach yields the chain ladder estimates of e^{δ}, e^{α_i} and e^{ζ_j} based on X_{ij}, $(i,j) \in \mathcal{A}_m$, which are similar to (10) – (12).

It remains to estimate ϕ. In the same way as in Section 3.4, the following moment estimator is proposed:

$$\widehat{\phi} = \frac{2}{m(m-1)} \sum_{(i,j)\in\mathcal{A}_m} \frac{(X_{ij} - e^{\widehat{\mu}_{ij}})^2}{e^{\widehat{\mu}_{ij}}}.$$

We continue by considering the asymptotic behavior of the latter estimates which is similar to the asymptotics as given in Proposition 1.

4.3 Asymptotic results

The asymptotics of the model for paid amounts is closely related to the asymptotics of the counts data. In fact by Mikosch (2006), pp. 118-119, the payment X_{ij} has a representation

$$X_{ij} = \sum_{k=1}^{N^*_{ij}} Z^*_{ijk}$$

where N^*_{ij} is independent of Z^*_{ijk}, $k \geq 1$,

$$N^*_{ij} \sim \text{Pois}\left(\sum_{l=0}^{j} e^{\delta+\alpha_i+\zeta_{j-l}}\right)$$

and Z^*_{ijk}, $k \geq 1$, are iid with

$$Z^*_{ijk} \stackrel{d}{=} \sum_{l=0}^{j} I\{J=l\} Z^{(1)}_{ijl}, \qquad P(J=l) = \frac{e^{\rho_l}}{\sum_{l=0}^{j} e^{\rho_l}},$$

where J and $Z^{(k)}_{ijl}$ are mutually independent. Since X_{ij} can be interpretated as a compound Poisson process it follows from Mikosch (2006), p. 81, that as $\delta \to \infty$,

$$\left(\sum_{l=0}^{j} e^{\delta+\alpha_i+\zeta_{j-l}}\right)^{-1} X_{ij} \stackrel{asymp}{\sim} \mathcal{N}\left(EZ^*_{ij1}, \phi E Z^*_{ij1}\left(\sum_{l=0}^{j} e^{\delta+\alpha_i+\zeta_{j-l}}\right)^{-1}\right).$$

Similar arguments as in Section 3.2 yield the following result.

Proposition 4
Let X_{ij}, $(i,j) \in \mathcal{A}_m$, follow the assumption (19) *with $\alpha_1 = \zeta_0 = 0$. Then as $\delta \to \infty$,*

$$\begin{bmatrix} \widehat{f}_0 \\ \vdots \\ \widehat{f}_{m-2} \end{bmatrix} \stackrel{asymp}{\sim} \mathcal{N}\left(\begin{bmatrix} f_0 \\ \vdots \\ f_{m-2} \end{bmatrix}, \text{diag}\begin{pmatrix} \phi\left(\sum_{i=1}^{m-1} e^{\delta+\alpha_i+\zeta_0}\right)^{-1} f_0(f_0-1) \\ \vdots \\ \phi\left(\sum_{j=0}^{m-2} e^{\delta+\alpha_1+\zeta_j}\right)^{-1} f_{m-2}(f_{m-2}-1) \end{pmatrix}\right)$$

where $\widehat{f}_j$, $0 \le j \le m-2$, *are given in* (2).

In the same way as in Section 3.2, it follows directly from Proposition 4 that

$$\sup_{x \in \mathbb{R}} \left| P\left(Y_{i,m-i} \prod_{j=m-i}^{m-2} \widehat{f}_j \le x \right) - P\left(U \prod_{j=m-i}^{m-2} U_j \le x \right) \right| \to 0, \qquad \delta \to \infty,$$

where $U, U_{m-i}, ..., U_{m-2}$ are mutually independent,

$$U \sim \mathcal{N}\left(\sum_{j=0}^{m-i} e^{\delta+\alpha_i+\zeta_j}, \phi \sum_{j=0}^{m-i} e^{\delta+\alpha_i+\zeta_j} \right)$$

and

$$U_j \sim \mathcal{N}\left(f_j, \phi \left(\sum_{i=1}^{m-j-1} e^{\delta+\alpha_i+\zeta_j} \right)^{-1} f_j(f_j-1) \right), \qquad m-i \le j \le m-2.$$

Recall the definition of $\widehat{f}_j$ in (2). Then, by Mikosch (2006), p. 81, it follows as $\delta \to \infty$,

$$\widehat{f}_j = \frac{\sum_{i=1}^{m-j-1} Y_{i,j+1}}{\sum_{i=1}^{m-j-1} Y_{ij}} = f_j \frac{(\sum_{i=1}^{m-j-1} \sum_{k=0}^{j+1} e^{\mu_{ik}})^{-1} \sum_{i=1}^{m-j-1} Y_{i,j+1}}{(\sum_{i=1}^{m-j-1} \sum_{k=0}^{j} e^{\mu_{ik}})^{-1} \sum_{i=1}^{m-j-1} Y_{ij}} \stackrel{a.s.}{\to} f_j.$$

Taylor (2003) shows that the chain ladder estimates are upward biased on the conditional probability measure where $\widehat{f}_j$ is finite; see the remark on p. 8 above.

4.4 The bootstrap method

The bootstrap method as a general resampling method was introduced by Efron (1979). In the context of claims reserving the bootstrap method was utilized by England and Verrall (1999) and England (2002); see also Cairns (2000) for a discussion on how one should quantify the reserve uncertainty. In the claims reserving context, it is common to base the bootstrap resampling on the quantities

$$\widehat{\epsilon}_{ij} = \frac{X_{ij} - e^{\widehat{\mu}_{ij}}}{e^{\frac{1}{2}\widehat{\mu}_{ij}}}, \qquad (i,j) \in \mathcal{A}_m.$$

The basic assumption for the bootstrap method to be meaningful is that

$$\epsilon_{ij} = \frac{X_{ij} - e^{\mu_{ij}}}{e^{\frac{1}{2}\mu_{ij}}}, \qquad (i,j) \in \mathcal{A}_m,$$

are iid. It is then our hope that the iid property asymptotically carry over to $\widehat{\epsilon}_{ij}$, $(i,j) \in \mathcal{A}_m$, as $\delta \to \infty$. For simplicity we will refer to both $\widehat{\epsilon}_{ij}$ and ϵ_{ij} as (Pearson) residuals.

The naive bootstrap method corresponds to drawing independently from the empirical distribution function of $\widehat{\epsilon}_{ij}$ given by

$$F_m(x) = \frac{2}{m(m+1)} \sum_{(i,j)\in\mathcal{A}_m} I\{\widehat{\epsilon}_{ij} \le x\}, \qquad x \in \mathbb{R}.$$

If we pragmatically assume that that $\epsilon_{ij} \sim F_m$ then independent versions of the residuals

$$e^*_{ij}(k), \qquad (i,j) \in \mathcal{A}_m,$$

for $1 \le k \le B$ can be simulated. Independent versions of X_{ij} can then be generated by

$$X^*_{ij}(k) = e^{\frac{1}{2}\widehat{\mu}_{ij}}\epsilon^*_{ij}(k) + e^{\widehat{\mu}_{ij}}.$$

Generating a large number, B, of iid realizations, $X^*_{ij}(k)$, $1 \le k \le B$, enables one to estimate distributional properties of relevant functions of the data. If, in fact, the relation

$$X^*_{ij}(k) \overset{d}{=} X_{ij}, \qquad (i,j) \in \mathcal{A}_{ij},$$

were true then with

$$\widehat{f}^*_j(k) = \frac{\sum_{i=1}^{m-j-1}\sum_{l=0}^{j+1} X^*_{il}(k)}{\sum_{i=1}^{m-j-1}\sum_{l=0}^{j} X^*_{il}(k)}$$

the law of large numbers would yield

$$\frac{1}{B}\sum_{k=1}^{B} I\left\{\widehat{f}^*_j(k) \le x\right\} \xrightarrow{P} P\left(\widehat{f}_j \le x\right)$$

for $B \to \infty$. Similarly, it follows that

$$\frac{1}{B}\sum_{k=1}^{B} I\left\{\sum_{l=0}^{m-i} X^*_{il}(k) \le x\right\} \xrightarrow{P} P\left(\sum_{l=0}^{m-i} X_{il} \le x\right) = P\left(Y_{i,m-i} \le x\right).$$

Adjustments to obtain iid residuals

In some data sets, X_{ij}, $(i,j) \in \mathcal{A}_m$, the residuals are not homogeneous as a function of j (and possibly i). In this section we propose one possible solution to this problem. We assume that ϵ_{ij}, $(i,j) \in \mathcal{A}_m$, are independent such that

$$E\epsilon_{ij} = 0, \qquad E\epsilon_{ij}^2 = V\epsilon_{ij} = \phi_j,$$

where ϕ_j is some function of j. For simplicity, let $\phi_j = a + bj$ but ϕ_j can be defined in a number of other ways as well. The parameters $a, b \in \mathbb{R}$ are estimated by minimizing the sum

$$\sum_{i=1}^{m} \sum_{j=0}^{m-i} \left(\epsilon_{ij}^2 - \phi_j\right)^2.$$

In the case when $\phi_j = a + bj$ the minimum is obtained by

$$\begin{aligned}
\widehat{b} &= \frac{2\left(\sum_{i=1}^{m}\sum_{j=0}^{m-i} j\,\epsilon_{ij}^2\right)\kappa_m - m(m+1)\left(\sum_{i=1}^{m}\sum_{j=0}^{m-i} \epsilon_{ij}^2\right)}{2\kappa_m - m(m+1)\left(\sum_{i=1}^{m}\sum_{j=0}^{m-i} j^2\right)}, \\
\widehat{a} &= \frac{2}{m(m+1)}\left(\sum_{i=1}^{m}\sum_{j=0}^{m-i} \epsilon_{ij}^2 - \widehat{b}\kappa_m\right).
\end{aligned}$$

where

$$\kappa_m = \sum_{i=1}^{m}\sum_{j=0}^{m-i} j = \sum_{i=1}^{m} \frac{(m-i)^2 + m - i}{2} = \frac{(m-1)m}{4}\left(\frac{2m-1}{3} + 1\right)$$

New residuals can consequently be defined as

$$\widetilde{\epsilon}_{ij} = \phi_j^{-1/2}\epsilon_{ij},$$

where $E\widetilde{\epsilon}_{ij} = 0$ and $V\widetilde{\epsilon}_{ij} = 1$. To bootstrap the reserve uncertainty in this frame it is assumed that $\widetilde{\epsilon}_{ij}$ are iid such that they can be resampled. The formula

$$\widetilde{\epsilon}_{ij}\phi_j^{1/2} = \epsilon_{ij}$$

yields a sample of residuals as defined in Section 4.4. A similar approach as taken in Section 4.4 can thus be used.

As in Section 3 we go on to study the Bornheutter-Ferguson method in relation with paid data.

4.5 The Bornheutter-Ferguson method and a credibility model combining the chain ladder and Bornheutter-Ferguson methods

The Bornheutter-Ferguson method works on paid data in almost the same way as with claim numbers and most results from Sections 3.5 and 3.6 carry over to this section.

As in Sections 3.5 and 3.6 the identification scheme

$$\left(\delta, \sum_{j=0}^{m-1} e^{\beta_j}\right) = (0,1)$$

is used in this section. Again, let T_i be the prior expected losses related to underwriting year i. In the same way as in Section 3.5 the estimate of e^{α_i} is replaced by its prior estimate, T_i, to obtain the Bornheutter-Ferguson reserve

$$\widehat{R}_m = \sum_{i=2}^{m} T_i \sum_{j=m-i+1}^{m-1} e^{\widehat{\beta}_j}.$$

This reserve is based purely on the prior estimate T_i of the ultimate payments, $Y_{i,m-1}$. The CLM is purely based on data in the sense that

$$\widehat{Y}_{i,m-1} = \frac{Y_{i,m-i}}{\sum_{j=0}^{m-i} e^{\widehat{\beta}_j}}$$

as shown in (15). Thus it seems straightforward to propose a credibility model that combines these two positions.

As in Section 3.6 we assume that Θ_i are independent and let

$$E\Theta_i = T_i \qquad \text{and} \qquad V\Theta_i = \frac{T_i}{\alpha}.$$

Further assume that X_{ij}, $(i,j) \in \mathcal{A}_m \cup \mathcal{B}_m$, are conditionally independent given Θ_i, $1 \le i \le m$, and that

$$E(X_{ij} \mid \Theta_i) = e^{\beta_j}\Theta_i \qquad \text{and} \qquad V(X_{ij} \mid \Theta_i) = \phi e^{\beta_j}\Theta_i$$

for $(i,j) \in \mathcal{A}_m \cup \mathcal{B}_m$. Then the linear estimator, $\widehat{\mu}_i^L$, minimizing $E(\widehat{\mu}_i^L - \Theta_i)^2$ over all linear function of X_{ij}, $(i,j) \in \mathcal{A}_m$, is given by

$$\widehat{\mu}_i^L = \omega_i T_i + (1-\omega_i)\frac{Y_{i,m-i}}{\sum_{j=0}^{m-i} e^{\beta_j}}, \tag{21}$$

where $\omega_i = ((\phi\alpha)^{-1} \sum_{j=0}^{m-i} e^{\beta_j} + 1)^{-1}$. This is similar to Section 3.7 however with different weights, ω_i, where α^{-1} is replaced with $\alpha^{-1}\phi$. A proof of (21) follows directly from the Bühlmann-Straub model used on X_{ij}/e^{β_j}; see Mikosch pp. 213-214.

To estimate the parameters in this model first observe that

$$V\left(\sum_{j=0}^{m-i} X_{ij}\right) = \left(\phi + \frac{\sum_{j=0}^{m-i} e^{\beta_j}}{\alpha}\right) E\left(\sum_{j=0}^{m-i} X_{ij}\right).$$

Thus, if we put

$$b_i = \sum_{j=0}^{m-i} e^{\beta_j} \quad \text{and} \quad \widetilde{X}_i = \frac{\left(\sum_{j=0}^{m-i} X_{ij} - T_i \sum_{j=0}^{m-i} e^{\beta_j}\right)^2}{T_i \sum_{j=0}^{m-i} e^{\beta_j}}$$

then $E\widetilde{X}_i = \phi + \alpha^{-1} b_i$ which suggest a linear regression of $\widetilde{X}_i$ on b_i. A standard linear regression yields estimators of α^{-1} and ϕ given by

$$\widehat{\alpha^{-1}} = \frac{\left(\sum_{i=1}^m b_i\right)\left(\sum_{i=1}^m X_{i.}\right) - m \sum_{i=1}^m b_i X_{i.}}{\left(\sum_{i=1}^m b_i\right)^2 - m \sum_{i=1}^m b_i^2}, \qquad \widehat{\phi} = \frac{\sum_{i=1}^m X_{i.} - \widehat{\alpha^{-1}} \sum_{i=1}^m b_i}{m}.$$

The chain ladder estimates are used to estimate e^{β_j}, $0 \le j \le m-1$. Under similar assumption as in Section 3.3 it can be shown that the chain ladder estimates converge in probability to their true values.

4.6 Diagonal effects

The results in this section are related to count data, N_{ij}, $(i,j) \in \mathcal{A}_m$. The methods proposed for counts can easily be applied to paid data, X_{ij}, $(i,j) \in \mathcal{A}_m$, by considering the GLM model

$$EX_{ij} = e^{\mu_{ij}}, \qquad VX_{ij} = \phi E X_{ij},$$

for paid amounts. All results from Section 3.7 carry over directly.

5 Bibliography

[1] Benktander, G. (1976): An approach to credibility in calculating IBNR for casualty excess reinsurance. Actuar. Rev. 3, no. 2, 7.
[2] Boor, J. (2006): Estimating tail development factors - What to do when the triangle runs out. CAS Forum Winter 2006, 245-390.
[3] Bornhuetter, R. L. and Ferguson, R. E. (1972): The Actuary and IBNR. Proceedings of the Casualty Actuarial Society, vol. LIX, 181-195.
[4] Cairns, J. G. C. (2000): A discussion of parameter and model uncertainty in insurance. Insurance: Mathematics and Economics, vol. 27, 313-330.
[5] Casella, G. and Berger, R. L. (2002): Statistical Inference, 2nd Edt. Duxbury Press.
[6] Efron, B. (1979): Bootstrap methods: Another look at the jackknife. Ann. Statist., vol. 7, 1-26.
[7] England, P. D. and Verrall, R. J. (1999): Analytic and bootstrap estimates of prediction errors in claims reserving. Insurance: Mathematics and Economics, vol. 25, 281-293.
[8] England, P. D. (2002): Addendum to 'Analytic and bootstrap estimates of prediction errors in claims reserving'. Insurance: Mathematics and Economics, vol. 31, 461-466.
[9] Hachemeister, C. A. and Stanard, J. N. (1975): IBNR claims counts estimation with static lag function. 12th ASTIN Colloquium, IAA, Portimao, Portugal.
[10] Kremer, E. (1982): IBNR-claims and the two-way model of AVNOVA. Scandinavian Actuarial Journal, 47-55.
[11] Kremer, E. (1985): Einfuhrung in die Versicherungsmathematik. Vandenhoel und Ruprecht, Gottingen.
[12] Kuang, D., Nielsen, B. and Nielsen, J. P. (2008a): Forecasting with the age-period-cohort model and the extended chain ladder model. Biometrika, vol. 95, 979-986.
[13] Kuang, D., Nielsen, B. and Nielsen, J. P. (2008b): Identification of the age-period-cohort model and the extended chain ladder model. Biometrika, vol. 95, 987-991.
[14] Mack, T. (1991): A simple parametric model for rating automobile insurance or estimating IBNR claims reserves. Astin Bulletin, vol. 21, no. 1, 93-108.

[15] Mack, T. (1993): Distribution-free calculation of the standard error of chain ladder reserve estimates. Astin Bulletin, vol. 23, no. 2, 213-225.
[16] Mack, T. (1994): Measuring the variability of chain ladder reserve estimates. Casualty Actuarial Society Forum Spring 1994, vol. 1, 101-182.
[17] Mack, T. (2000): Credible claims reserves the Benktander method. Astin Bulletin, vol. 30, no. 2, 333-347.
[18] McCullagh, P. and Nelder, J. A. (1989): Generalized Linear Models, second edition. Chapman & Hall, London.
[19] Mikosch, T. (2004): Non-Life Insurance Mathematics. Springer-Verlag Berlin Heidelberg.
[20] Schmidt, K. D. and Wünsche, A. (1998): Chain ladder, marginal sum and maximum likelihood estimation. Blätter, vol. 23, no. 3, 267-277.
[21] Schmidt, K. D. and Zocher, M. (2008): The Bornhuetter-Ferguson principal. Variance no. 2, 85-110.
[22] Taylor, G. (2003): Chain ladder bias. Astin Bulletin, vol. 33, no. 2, 313-330.
[23] Taylor, G. (2000): Loss reserving, An Actuarial Perspective. KAP Boston.
[24] Venter, G. G. (2006): Discussion of the mean square error of prediction the chain ladder reserving method. Astin bulletin, vol. 36, 566-571.
[25] Venter, G. G. (2007): Refining reserve runoff Ranges. CAS E-forum:
http://casact.org/pubs/forum/07sforum/07s-venter2.pdf
[26] Verrall, R. J. (1990): Bayes and empirical Bayes estimates for the chains ladder. Astin Bulletin, vol. 20, 217-243.
[27] Verrall, R. J. (1991): Chain ladder and maximum likelihood. Journal of the institute of actuaries, vol. 118, 489-499.
[28] Verrall, R. J. and England, P. D. (2005): Incorporating expert opinion into a stochastic model for the chain-ladder technique. Insurance, mathematic and economics, vol. 37, 355-370 .

Diagonal Effects in Claims Reserving

Anders Hedegaard Jessen
Department of Mathematical Sciences, University of Copenhagen

Niels Rietdorf
Alpha Group

May 27, 2009

Abstract
In this paper we present two different approaches to how one can include diagonal effects in non-life claims reserving based on run-off triangles. Empirical analyses suggest that the approaches in Zehnwirth (2003) and Kuang, Nielsen and Nielsen (2008a, 2008b) do not work well with low-dimensional run-off triangles because estimation uncertainty is too large. To overcome this problem we consider similar models with a smaller number of parameters. These are closely related to the framework considered in Taylor (1977, 2000) and Verbeek (1972); the separation method. We explain that these models can be interpretated as extensions of the multiplicative Poisson models introduced by Hachemeister and Stanard (1975) and Mack (1991).

1 Introduction

Recently, some attention has been given to diagonal effects in run-off triangles $\{N_{ij}\}$ of claim counts. For example, the approach to estimation and prediction as taken in Zehnwirth (2003) was addressed in Kuang, Nielsen and Nielsen (2008a, 2008b), where it is also assumed that

$$N_{ij} \sim \mathrm{Pois}(\mu_{ij}) \tag{1}$$

with

$$\mu_{ij} = \alpha_i \beta_j \delta_{i+j}. \tag{2}$$

Kuang, Nielsen and Nielsen (2008a, 2008b) point out that the parameters (2) are not uniquely identified and, therefore, one has to be careful when it comes to the identification of the model. Indeed, the choice of identification scheme in Zehnwirth (2003) has an effect on the final prediction.

The model (1) is a generalization of the multiplicative Poisson models proposed in Hachemeister and Stanard (1975) and Mack (1991). There it is shown that (1) with $\delta_j = 1$, $j \geq 1$, implies the chain ladder method. As the chain ladder method is close to the industry standard it seems reasonable to take it as a starting point. We are, however, not interested in claim counts but paid amounts. In the spirit of Venter (2007) and England and Verrall (1999) we thus consider the GLM framework, where it is simply assumed for a paid run-off triangle $\{X_{ij}\}$ that

$$EX_{ij} = \mu_{ij} \qquad \text{and} \qquad VX_{ij} = \varphi EX_{ij}. \tag{3}$$

McCullagh and Nelder (1989) consider models of type (3). They suggest to use a Poisson quasi-likelihood equivalent to the likelihood in (1) to estimate the parameters in (3). This approach provides another link back to the chain ladder method (with $\delta_j = 1$, $j \geq 1$) for paid amounts because the reserve estimation remains identical to Hachemeister and Stanard (1975) and Mack (1991) for paid amounts.

Rietdorf (2008) points out that diagonal effects, δ_j, usually come from two sources: 1) economic inflation, i.e., claim payments follow a relevant price index which again follows the calender time and 2) legal issues, changes in the way claims are handled or similar give diagonals effects in the number of payments. 2) is referred to as claims inflation. Economic inflation is usually assumed to work in a multiplicative way suggesting that

$$EX_{ij} = \alpha_i \beta_j \delta_{i+j}, \qquad VX_{ij} = \varphi \alpha_i \beta_j \delta_{i+j}^2,$$

where claims inflation (see Kuang et. al (2008a, 2008b)) is assumed to satisfy

$$EX_{ij} = \alpha_i \beta_j \delta_{i+j}, \qquad VX_{ij} = \varphi \alpha_i \beta_j \delta_{i+j}.$$

Based solely on data from a paid run-off triangle one cannot distinguish whether changes along a diagonal are due to 1) or 2). In order to take this aspect into account Rietdorf (2008) considers models

$$EX_{ij} = \mu_{ij} \qquad \text{and} \qquad VX_{ij} = \varphi \delta_{i+j} EX_{ij} \tag{4}$$

by optimizing the extended Poisson quasi-likelihood functions suggested in Nelder and Pregibon (1987). In a similar way various modifications of (4), including structures like

$$VX_{ij} = \varphi\delta_{i+j}^{c} EX_{ij} \quad \text{and} \quad VX_{ij} = \varphi\delta_{i+j}^{c}(1+kj)EX_{ij} \tag{5}$$

for $c, k > 0$, are studied. The different modifications of (4) in (5) lead to different estimators of the diagonal effects causing dramatic diversifications in the ultimate reserve prediction.

The motivation for the work presented in this paper comes from the data studies of paid run-off triangles; see Rietdorf (2008). The conclusion based on Rietdorf (2008) is that it is intuitively difficult to understand and explain the estimated diagonal effects coming from (4). When diagonal effects are added to the GLM model

$$EX_{ij} = \alpha_i\beta_j, \qquad VX_{ij} = \varphi\alpha_i\beta_j,$$

one significantly overparameterizes this model which may cause unreliable parameter estimates in the framework of the data available in a paid run-off triangle. The estimation uncertainty leads to unreliable reserve prediction which in practice disqualifies these type models for actuarial purposes.

Therefore it is natural to search for models including diagonal effects with less parameters. Taylor (1977, 2000) and Verbeek (1972) consider the first order structure of a such model. This paper suggests two different variants of the models proposed in Taylor (1977, 2000) and Verbeek (1972) where a second order structures have been added; one variant leads to the same reserve estimates as Taylor and Verbeek, the other gives different reserve estimates through a credibility weighted model. In our framework the calculation of credibility weights is made possible because we can analyze the second order structure. As shown in Section 3 our results can be interpretated as a Benktander type method with diagonal effects included. The second order structure further allows for simulation studies of reserve predictors uncertainty; see Section 4.

We insist that model formulations remain as simplistic as (4) and estimation procedures are easily applicable to data. A data study shows that the estimated diagonal effects are intuitively appealing.

The paper is organized as follows. In Section 2 a Bornhuetter-Ferguson type method including diagonal effects is presented. In Section 3 we introduce a credibility model with diagonal effects which ultimately gives a prediction based on a weighted average of a chain ladder type estimator and the Bornhuetter-Ferguson method. Finally, in Section 4 data studies are given.

2 A Bornhuetter-Ferguson method including diagonal effects

In this section we will consider a model based on the Bornhuetter-Ferguson method; see Bornhuetter and Ferguson (1972). We will however use the framework of Schmidt and Zocher (2008) as they consider the row effects as known, leading to a reduction of the number of parameters. This approach is preferable because it is closely related to the chain ladder setup of Mack (1991) or Hachemeister and Standard (1975) as mentioned in the introduction.

Let X_{ij} be the entries in a paid incremental run-off triangle with dimension m:

$$\triangle_m = \{X_{ij} : (i,j) \in \mathcal{A}_m\}$$

where

$$\mathcal{A}_m = \{(i,j) \in \mathbb{N} \times \mathbb{N}_0 : 1 \leq i + j \leq m\}.$$

Write

$$\mathcal{B}_m = \{(i,j) \in \mathbb{N} \times \mathbb{N}_0 : i \leq m, j \leq m-1\}.$$

We are interested in the prediction of

$$X_{ij}, \qquad (i,j) \in \mathcal{B}_m \backslash \mathcal{A}_m,$$

which are unobserved random variables at time m. As a technical basis for prediction we consider a model for all random variables, X_{ij}, $(i,j) \in \mathcal{B}_m$, given by the following requirements.

i) X_{ij}, $(i,j) \in \mathcal{B}_m$, are mutually independent.

ii) For $(i,j) \in \mathcal{B}_m$

$$EX_{ij} = T_i \beta_j \delta_{i+j}, \qquad VX_{ij} = \varphi \delta^c_{i+j} EX_{ij}.$$

where the net premiums $T_i > 0$, $1 \leq i \leq m$, are assumed to be known, $c \in [0,1]$ and

$$\varphi, \beta_0, ..., \beta_{m-1} \quad \text{and} \quad \delta_1, ..., \delta_{2m-1}$$

are positive unknown constants. For the purpose of identification we also assume that

$$\sum_{j=0}^{m-1} \beta_j = 1.$$

The parameter $c \in [0,1]$ is considered known. In particular we will chose $c \in \{0, 0.5, 1\}$ for the following reasons:

• $c = 0$ corresponds to claims inflation; see 2) p. 2. This means that the diagonal effects are additive suggesting that changes in diagonals are triggered by an increased number or claims reported (due to, for example, changes in law practice).
• $c = 1$ corresponds to economic inflation; see 1) p. 2. If there have been no big changes in reporting patterns and the only diagonal effects in data are due to economic inflation (which usually acts in a multiplicative way).
• $c = 1/2$ is chosen if we are in a situation where both effects described above have an impact on data.

The specific choice of c can be based on intuition as well as plots of residuals; see the definition of $\widehat{e}_{ij}$ in (10).

Notice that if we had assumed T_i, $1 \leq i \leq m$, unknown and $\delta_j = 1$, then model **i)**, **ii)** would have been identical to the GLM chain ladder setup; see from for example England and Verrall (1999). Assuming T_i, $1 \leq i \leq m$, known and $\delta_j = 1$, **i)**, **ii)** is identical to the setup in Schmidt and Zocher (2008). If T_i, $1 \leq i \leq m$, are a priori expected values of discounted accumulated claims, then δ_j, $1 \leq j \leq m$, can be compared directly with externally given inflation rates. If T_i, $1 \leq i \leq m$, are inflation adjusted then δ_j, $1 \leq j \leq m$, may be thought of as inflation in this portfolio less the estimated a priori inflation.

We estimate the parameters in the model **i)**, **ii)** in a two-step procedure.

First, we use the separation method; see Taylor (1977, 2000) and Verbeek (1972). This method is based on the following observation where a transformation makes diagonals into row effects:

$$\widetilde{X}_{ij} = \frac{X_{i+j,j}}{T_{i+j}} \quad \text{then} \quad E\widetilde{X}_{ij} = \delta_i \beta_j \tag{6}$$

for $1 \le i \le m$ and $0 \le j \le i-1$. The well known multiplicative structure (6) suggests that the estimation can be based on the total marginals principal (see for example Schmidt and Wünsch (1998));

$$\widehat{\delta}_i \sum_{j=0}^{i-1} \widehat{\beta}_j = \sum_{j=0}^{i-1} \widetilde{X}_{ij}, \;\; 1 \le i \le m, \quad \text{and} \quad \widehat{\beta}_j \sum_{i=j+1}^{m} \widehat{\delta}_i = \sum_{i=j+1}^{m} \widetilde{X}_{ij}, \;\; 0 \le j \le m-1.$$

Schmidt and Wünsch points out that the total marginals principal yields a chain ladder method — here under the restriction $\sum_{j=0}^{m-1} \beta_j = 1$. Therefore the estimation procedure can be carried out using the following calculation. First put

$$\sum_{j=0}^{m-1} \widehat{\beta}_j \;\; = \;\; 1. \tag{7}$$

Then use the ratios

$$\frac{\sum_{j=0}^{k} \widehat{\beta}_j}{\sum_{j=0}^{k-1} \widehat{\beta}_j} \;\; = \;\; \frac{\sum_{j=0}^{k} \sum_{i=k+1}^{m} \widetilde{X}_{ij}}{\sum_{j=0}^{k-1} \sum_{i=k+1}^{m} \widetilde{X}_{ij}}, \qquad 1 \le k \le m-1, \tag{8}$$

to recursively calculate estimators $\widehat{\beta}_0, ..., \widehat{\beta}_{m-1}$ starting from (7). Finally, calculate

$$\widetilde{\delta}_i \;\; = \;\; \frac{\sum_{j=0}^{i-1} \widetilde{X}_{ij}}{\sum_{j=0}^{i-1} \widehat{\beta}_j}, \qquad 1 \le i \le m. \tag{9}$$

The second part of the estimation procedure is related to the variance structure. Define

$$\widehat{e}_{ij} = \frac{X_{ij} - T_i \widehat{\beta}_j \widehat{\delta}_{i+j-1}}{(T_i \widehat{\beta}_j \widehat{\delta}_{i+j-1}^{1+c})^{1/2}}, \qquad (i,j) \in \mathcal{A}_m. \tag{10}$$

The estimation of φ is carried out by using the method of moments:

$$\widehat{\varphi} = \frac{2}{m(m+1)} \sum_{i=1}^{m} \sum_{j=0}^{m-i} \widehat{e}_{ij}^2.$$

Another motivation for the form of the estimators (7) – (9) is that the estimators are in fact consistent in the following sense.

Proposition 1
Assume the model **i)**, **ii)**. If $T_i \to \infty$ for all $1 \le i \le m$ then

$$\widehat{\delta}_{j+1} \xrightarrow{P} \delta_{j+1} \text{ and } \widehat{\beta}_j \xrightarrow{P} \beta_j$$

for $0 \le j \le m-1$.

Proof:
From assumptions **i)** and **ii)** and Chebyshev's inequality it is straightforward that

$$\frac{X_{ij}}{T_i} \xrightarrow{P} \beta_j \delta_{i+j}$$

as $T_i \to \infty$. The recursive scheme (8) and (7) together with the continuous mapping theorem give the desired result.

$\square$

As X_{ij}, $(i,j) \in \mathcal{B}_m$, are assumed independent it is natural to predict X_{ij}, $(i,j) \in \mathcal{B}_m \backslash \mathcal{A}_m$, by its expected value:

$$\mu(X_{ij}) = EX_{ij} = T_i \beta_j \delta_{i+j}.$$

The parameters, δ_j, $0 \le j \le m-1$, are estimated directly from data using recursion (9). The diagonal parameters, δ_j, $m+1 \le j \le 2m-1$, are related to unobservable data, X_{ij}, $(i,j) \in \mathcal{B}_m \backslash \mathcal{A}_m$, and these can thus not be estimated directly using (9). Since the latter predictor, $\mu(X_{ij})$, is solely dependent on diagonal parameters δ_j, $m+1 \le j \le 2m-1$, we however need to somehow be able to estimate or predict these. The idea is to construct a predictor of δ_j, $m+1 \le j \le 2m-1$, based on the in-sample diagonals, $\delta_1, ..., \delta_m$. In the following, one way of generating a prediction is suggested and we also add a remark of a more theoretical character justifying this approach.

Example 2
Assume that $\widehat{\delta}_j$, $j \ge 1$, follow an $AR(1)$-process, i.e., that

$$\widehat{\delta}_j \Gamma + \epsilon_j = \widehat{\delta}_{j+1},$$

where ϵ_j, $1 \leq j \leq 2m-1$, is a mean zero white noise process. Based on the in-sample diagonal parameter estimators we estimate Γ by

$$\widehat{\Gamma} = \frac{\sum_{j=1}^{m-1} \widehat{\delta}_{j+1}\widehat{\delta}_j}{\sum_{j=1}^{m-1} \widehat{\delta}_j^2}.$$

The $AR(1)$ assumption suggests a predictor of $\widehat{\delta}_j$, $j \geq m+1$, given by

$$\widetilde{\delta}_j = \widehat{\delta}_m \widehat{\Gamma}^{j-m}.$$

Consequently, the prediction of X_{ij}, $(i,j) \in \mathcal{B}_m \backslash \mathcal{A}_m$, is given by

$$\widetilde{\mu}(X_{ij}) = T_i \widehat{\beta}_j \widetilde{\delta}_{i+j}.$$

$\square$

Remark 3

The approach to prediction taken in Example 2 indicates that δ_j, $j \geq 1$, should be considered as a time series rather than fixed parameters. Formally, this suggests the following alternative model assumptions:

i') assume that X_{ij}, $(i,j) \in \mathcal{A}_m$, are conditionally independent given δ_j, $j \geq 1$, and
ii') that

$$\begin{aligned} E(X_{ij} \mid \delta_{i+j-1}) &= T_i \beta_j \delta_{i+j} \\ V(X_{ij} \mid \delta_{i+j-1}) &= \varphi \delta_{i+j}^c E X_{ij} \end{aligned}$$

for $(i,j) \in \mathcal{B}_m$.

In this setup Proposition 1 remains valid. In fact by Proposition 1 it follows that

$$P(|X_{ij}/T_i - \beta_j \delta_{i+j}| > \epsilon \mid \delta_{i+j}) \to 0, \qquad T_i \to \infty,$$

and dominated convergence yields that

$$\frac{X_{ij}}{T_i} \xrightarrow{P} \beta_j \delta_{i+j}, \qquad T_i \to \infty.$$

In the same way as in the proof of Proposition 1 it now follows that if $T_i \to \infty$ for all $1 \le i \le m$ then

$$\widehat{\delta}_{j+1} \xrightarrow{P} \delta_{j+1}, \qquad \widehat{\beta}_j \xrightarrow{P} \beta_j, \qquad 0 \le j \le m-1,$$

where the latter estimators are given by (7).

Moreover, if we want to predict by minimizing the mean square error,

$$E[X_{ij} - \mu(X_{ij})]^2$$

over all measurable finite variance functions, $\mu(X_{ij})$, of $\triangle_m, \delta_1, ..., \delta_m$, then we obtain the predictor

$$\mu(X_{ij}) = T_i \beta_j E(\delta_{i+j} | \delta_1, ..., \delta_m).$$

Since $\delta_1, ..., \delta_m$ are not observable we have to replace them by the pseudo-observations $\widehat{\delta}_1, ..., \widehat{\delta}_m$, ultimately leading to the same prediction as suggested in Example 2.

$\square$

To estimate the uncertainty in model **i)**, **ii)** we can apply a bootstrap method similar to England and Verrall (1999). The bootstrap method should be based on the assumption that the residuals $\widehat{e}_{ij}$, $(i,j) \in \mathcal{A}_m$, are iid such that these can be resampled B times, $e^*_{ij}(k)$, $1 \le k \le B$, to generate iid versions of $\triangle_m$ by

$$X_{ij}(k) = T_i \widehat{\beta}_j \widehat{\delta}_{i+j} + e^*_{ij}(k) \left[\widehat{\varphi} T_i \widehat{\beta}_j \widehat{\delta}^{1+c}_{i+j} \right]^{1/2}, \qquad 1 \le k \le B. \tag{11}$$

In this section the row effects, T_i, $1 \le i \le m$, are considered fixed giving a lot of importance to the a priori estimates T_i, $1 \le i \le m$, of the underwriting years total payments. In the next section we consider a credibility model that leads to estimators of the row effects based on a weighted average of T_i and a chain ladder type estimator of the total payments.

3 A credibility model including diagonal effects

In this section we consider a credibility model in a similar framework as in Section 2 but with random rowwise effects. In a special case ($c = 1$) this approach yields predictors based on a weighted average of the chain ladder estimates (see for example Mack (1991,1993,1994)) and the Bornhuetter-Ferguson estimates given in Section 2. This type of results can also be found in Verrall (2004) and Wüthrich (2007) where diagonal effects are not included.

Using the same notation as in Section 2, we assume

i) Θ_i, $1 \leq i \leq m$, be mutually independent with $E\Theta_i = T_i$ and $V\Theta_i = \xi^{-1}T_i$,
ii) given Θ_i, $1 \leq i \leq m$, X_{ij}, $(i, j) \in \mathcal{B}_m$, are mutually independent and
iii) for $(i, j) \in \mathcal{B}_m$

$$E(X_{ij} \mid \Theta_i) = \Theta_i \beta_j \delta_{i+j}, \qquad V(X_{ij} \mid \Theta_i) = \varphi \delta_{i+j}^c E(X_{ij} \mid \Theta_i)$$

where the net premiums $T_i > 0$, $1 \leq i \leq m$, are considered known and

$$\varphi, \xi, \beta_0, ..., \beta_{m-1}, \delta_1, ..., \delta_{2m-1}$$

are positive unknown constants. For the model to be identified let

$$\sum_{j=0}^{m-1} \beta_j = 1.$$

This leads directly to the following result.

Proposition 4
Assume that **i)-iii)** holds. Then the best linear Bayes estimator of Θ_i is given by

$$\mu^{LB}(\Theta_i) = T_i w_i + (1 - w_i)\overline{X}_{i.}$$

where

$$\overline{X}_{i.} = \frac{\sum_{j=0}^{m-i} \delta_{i+j}^{-c} X_{ij}}{\sum_{j=0}^{m-i} \beta_j \delta_{i+j}^{1-c}} \qquad \text{and} \qquad w_i = \left(1 + (\varphi\xi)^{-1} \sum_{j=0}^{m-i} \beta_j \delta_{i+j}^{1-c}\right)^{-1}.$$

The best linear Bayes predictor of X_{ij} is moreover given by

$$\mu(X_{ij}) = \mu^{LB}(\Theta_i)\beta_j\delta_{i+j}.$$

Proof:
The proof of Proposition 4 follows directly from an application of the Bühlmann-Straub model (see for example Mikosch (2006) pp. 213-215) to

$$Z_{ij} = \frac{X_{ij}}{\beta_j \delta_{i+j}}, \qquad 0 \leq j \leq m - i.$$

The Bühlmann-Straub model assumes fixed and known parameters which is not the case in the above, but whether the parameters are fixed or not does not have an effect on the optimization problem considered. The result applies directly.

□

Remark 5
The special case $c = 1$ is particularly interesting. In this case

$$\overline{X}_{i.} = \frac{\sum_{j=0}^{m-i} \delta_{i+j}^{-1} X_{ij}}{1 - \sum_{m-i+1}^{m-1} \beta_j},$$

where one interpretation is that $\sum_{j=0}^{m-i} \delta_{i+j}^{-1} X_{ij}$ is a sum of discounted payments multiplied by a chain ladder type of factor. In fact if one calculates the chain ladder factors, $\widehat{f}_0, ..., \widehat{f}_{m-2}$, see Mack (1991), based on the discounted accumulated payments

$$C_{ij} = \sum_{j=0}^{m-i} \delta_{i+j}^{-1} X_{ij}, \qquad (i,j) \in \mathcal{A}_m,$$

then $(1 - \sum_{j=m-i+1}^{m-1} \widehat{\beta}_j)^{-1}$ and $\prod_{j=m-i}^{m-2} \widehat{f}_j$ are two estimators of the relative increase from $j = m-i+1$ to $j = m-1$. Since the C_{ij}'s define a GLM setup as in England et. al (1999), a Poisson quasi likelihood identical to the one in Mack (1991) and Hachemeister et. al (1975) can be used to estimate the parameters. This ultimately leads to prediction by the chain ladder method,

$$C_{ij} \prod_{j=m-i}^{m-2} \widehat{f}_j = \left[\sum_{j=0}^{m-i} \delta_{i+j}^{-1} X_{ij} \right] \prod_{j=m-i}^{m-2} \widehat{f}_j.$$

A look at Proposition 5 now explains how one can interpret the Bayes estimator, $\mu^{LB}(\Theta_i)$, as a weighted average between the chain ladder estimators and the Bornhuetter-Ferguson estimators.

□

In the same way as in Section 2 estimation is carried out in a two step procedure. In the first step the parameters δ_{j+1}, β_j, $0 \leq j \leq m-1$, are estimated by the recursive procedures (7) and (9). A motivation for reusing the estimation procedure from Section 2 is that Proposition 1 still holds under assumption **i)-iii)**.

The Bayes estimator given in Proposition 4 depends on parameters that are unknown. These unknown parameters must be estimated to produce the actual point estimators. Plugging in estimators for the parameters in Proposition 4 yields an empirical Linear Bayes procedure.

Proposition 6
Assume that **i)-iii)** holds. If $T_i \to \infty$ for all $1 \leq i \leq m$ then

$$\widehat{\delta}_{j+1} \xrightarrow{P} \delta_{j+1} \text{ and } \widehat{\beta}_j \xrightarrow{P} \beta_j$$

for $0 \leq j \leq m-1$.

Proof:
Put $Y_{ij} = X_{ij}/T_i$. By Chebyshev's inequality it holds that

$$\begin{aligned} P\left(|Y_{ij} - \beta_j \delta_{i+j}| > \epsilon \mid \Theta_i\right) &\leq \frac{E([X_{ij} - T_i \beta_j \delta_{i+j}]^2 \mid \Theta_i)}{T_i^2 \epsilon^2} \\ &\leq \frac{V(X_{ij} \mid \Theta_i) + (T_i \beta_j \delta_{i+j} - \Theta_i \beta_j \delta_{i+j})^2}{T_i^2 \epsilon^2}. \end{aligned}$$

Taking expectations on both sides, we obtain the inequality

$$P\left(|Y_{ij} - \beta_j \delta_{i+j}| > \epsilon\right) \leq \frac{\varphi \beta_j \delta_{i+j}^{1+c} T_i + (\beta_j \delta_{i+j})^2 \frac{T_i}{\xi}}{T_i^2 \epsilon^2},$$

where the right hand side goes to zero as $T_i \to \infty$. The recursive scheme (9) together with the continuous mapping theorem now yields the desired result.

$\square$

The second step in the estimation procedure is based on the definition of the row residuals

$$e_i = \frac{\sum_{j=0}^{m-i} X_{ij} - T_i \sum_{j=0}^{m-i} \beta_j \delta_{i+j}}{(T_i \sum_{j=0}^{m-i} \beta_j \delta_{i+j})^{1/2}}, \qquad 1 \leq i \leq m,$$

where

$$Ee_i^2 = \frac{1}{\xi}\left(\sum_{j=0}^{m-i} \beta_j \delta_{i+j}\right) + \varphi \frac{\sum_{j=0}^{m-i} \beta_j \delta_{i+j}^{1+c}}{\sum_{j=0}^{m-i} \beta_j \delta_{i+j}}.$$

By replacing the parameters β_j, δ_{j+1}, $0 \le j \le m-1$, by their estimators a least squares method based on the latter expectation can be used to estimate c, φ and ξ. In the case $c = 0$ the latter procedure can be reduced to an ordinary linear regression of e_i^2 against $\sum_{j=0}^{m-i} \beta_j \delta_{i+j}$, $1 \le i \le m$. If $c = 1$ then ξ and φ can also be estimated by a linear regression based on the relation

$$\frac{e_i^2}{\sum_{j=0}^{m-i} \beta_j \delta_{i+j}} \simeq \frac{1}{\xi} + \varphi \frac{\sum_{j=0}^{m-i} \beta_j \delta_{i+j}^2}{\left(\sum_{j=0}^{m-i} \beta_j \delta_{i+j}\right)^2}, \qquad 1 \le i \le m.$$

As there are only m row residuals, e_i, one or two outliers in the sample can influence the above regression a lot. It is therefore recommended to plot expected value against observed and in some case use subjective judgement to correct the estimation.

To simulate iid versions of $\triangle_m$ in the model **i)-iii)** we can also here apply a bootstrap method similar to England and Verrall (1999). The bootstrap method should be based on the assumption that the residuals

$$e_{ij}^* = \frac{X_{ij} - \mu^{LB}(\Theta_i)\beta_j \delta_{i+j}}{(\varphi \mu^{LB}(\Theta_i)\beta_j \delta_{i+j}^{1+c})^{1/2}}, \qquad (i,j) \in \mathcal{A}_m, \tag{12}$$

are iid such that these can be resampled.

In the next section the models proposed in Section 2 and 3 are applied to a data set.

4 A data study

In this section a data set from Codan Insurance in considered. Data contains 13 years run-off for a portfolio of a third-party liability for auto insurance. This run-off triangle is given below in Table 1.

$i\backslash j$	0	1	2	3	4	5	6	7	8	9	10	11	12
1	22564	17331	17377	7723	5058	2530	1443	1195	1889	106	33	139	14
2	22901	26734	8974	7089	3116	1911	3284	1591	879	21	575	476	
3	36152	26513	10973	6714	7155	2176	1656	1094	−89	8	115		
4	34733	29642	13593	11496	6256	6404	3900	2157	1133	25			
5	30709	28020	12465	8504	9929	5592	910	3413	1428				
6	33727	32190	13318	9211	8129	5225	2149	773					
7	30727	27677	9251	9221	6169	7492	2952						
8	32498	35446	18432	15110	13990	4986							
9	32228	42937	16231	12942	11078								
10	41947	41634	21056	15442									
11	37247	34135	19061										
12	32891	29719											
13	35993												

Table 1: Incremental run-off triangle.

In the models considered in Section 2 and 3 it is assumed that the expected rowwise effects are known. These are given below in Table 2.

i	1	2	3	4	5	6	7
T_i	85047	74409	86077	83082	83427	81557	79495
i	8	9	10	11	12	13	
T_i	101564	95482	107062	90091	85413	81995	

Table 2: Row effects, T_i, $1 \le i \le 13$.

To apply models **i)**, **ii)** in Section 2 and **i)** $-$ **iii)** in Section 3 to data Tables 1 and 2 is all the data we need. The estimation of the parameters β_j, δ_j, $0 \le j \le 12$, is identical in Section 2 and 3. The estimators are shown in Table 3 and Figure 1.

j	0	1	2	3	4	5	6	7	8	9	10	11	12
$\widehat{\beta}_j$	0.32	0.29	0.13	0.09	0.07	0.04	0.02	0.02	0.01	0.00	0.00	0.00	0.00
$\widehat{\delta}_{j+1}$	0.84	0.84	1.33	1.13	1.11	1.12	1.23	1.11	1.18	1.39	1.30	1.41	1.38

Table 3: The estimators $\widehat{\beta}_j, \widehat{\delta}_{j+1}$, $0 \le j \le 12$, as given in (7).

To calculate the final reserve estimate we need to predict the diagonal ef-

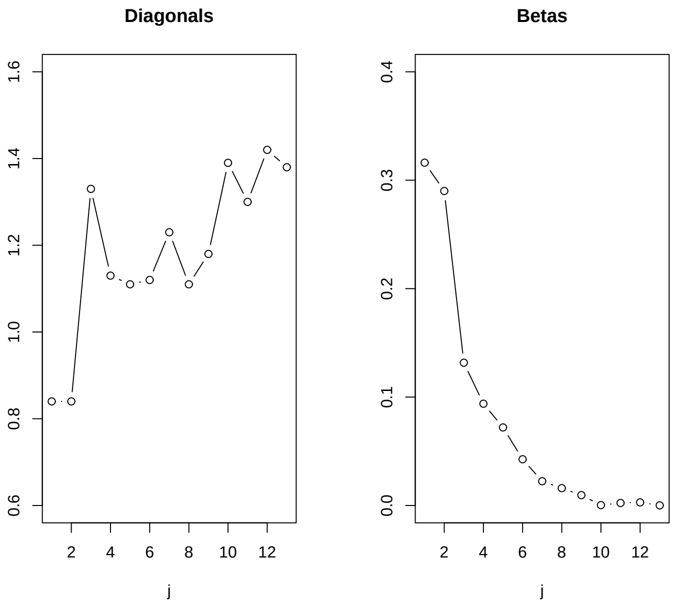

Figure 1: *Left: Diagonal effects,* δ_j*,* $1 \leq j \leq 13$*. Right: The payments pattern,* β_j*,* $0 \leq j \leq 12$*.*

fects, $\delta_{14}, ..., \delta_{25}$. We follow the procedure suggested in Example 2. The expected yearly increment, Γ, is estimated by

$$\widehat{\Gamma} = 1.0245$$

leading to prediction of $\delta_{14}, ..., \delta_{25}$ given by

$$\widetilde{\delta}_{13+j} = \widehat{\delta}_{13}\widehat{\Gamma}^j = 1.38 \cdot 1.0245^j, \qquad j \geq 1.$$

Moreover, if we assume that $\epsilon_j \sim \mathcal{N}(0, \sigma^2)$, $j \geq 1$, then

$$\widehat{\sigma}^2 = (m-1)^{-1} \sum_{j=1}^{m-1} (\widehat{\delta}_{j+1} - \widehat{\Gamma}\widehat{\delta}_j)^2 = 0.03037.$$

We go on to consider the model from Section 2 in detail.

4.1 Data study: A Bornhuetter-Ferguson method including diagonal effects

The first thing we have to determine is the parameter $c \in \{0, 1/2, 1\}$. One way of doing this is by considering the residuals, $\widehat{e}_{ij}$, $(i,j) \in \mathcal{A}_{13}$, as given in Section 2. For each $c = 0, 1/2, 1$ we have plotted points

$$(j, \widehat{e}_{ij}), \qquad (i,j) \in \mathcal{A}_m,$$

to check whether these residuals 'look iid'. If they do it can be taken as an indicator that the estimated model **i)**, **ii)** fit data. Judging from such residual plots $c = 0$ gives the best fit; see Figure 2 below for the case $c = 0$. With $c = 0$ the moment estimator of φ is $\widehat{\varphi} = 345.08$. Define the rowwise reserve

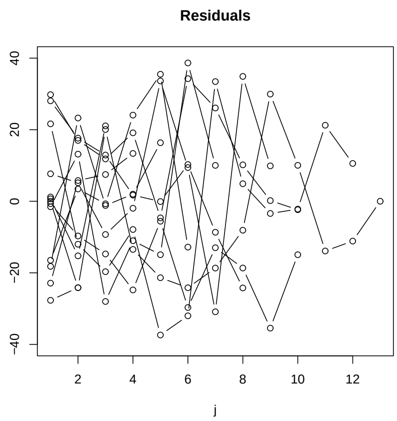

Figure 2: *Rowwise residuals:* $\{j, \widehat{e}_{ij}, 0 \le j \le 13 - i\}$ *for* $1 \le i \le 12$ *in the case* $c = 0$.

estimators by

$$\widehat{R}_i^{(1)} = \sum_{j=m-i+1}^{m-1} \widehat{\mu}(X_{ij}), \qquad 2 \le i \le 13.$$

Using this notation the reserve estimators are given below in Table 4.

i	2	3	4	5	6	7	8	9	10	11	12	13
$\widehat{R}_i$	12	364	631	690	1793	3589	7911	13377	26264	34599	49503	82308

Table 4: Reserve estimators $\widehat{R}_i^{(1)}$, $2 \le i \le 13$.

A simulation study

The definition of residuals allows us to quantify uncertainty related to the model **i)** $-$ **ii)** from Section 2 using a bootstrap method. We have considered the process error defined by the following simulation procedure. The same way as in (11) we generate iid realizations of X_{ij}, $(i,j) \in \mathcal{B}_m \backslash \mathcal{A}_m$, by

$$X_{ij}(k) = T_i \widehat{\beta}_j \widetilde{\delta}_{i+j}(k) + e^*_{ij}(k) \left[\widehat{\varphi} T_i \widehat{\beta}_j \widetilde{\delta}^{1+c}_{i+j}(k) \right]^{1/2}, \qquad 1 \le k \le B,$$

where $\widetilde{\delta}_j(k)$, $j \ge 1$, is simulated independently of $e^*_{ij}(k)$, $(i,j) \in \mathcal{B}_m \backslash \mathcal{A}_m$, using the recursion given in Example 2 with $\epsilon_j(k)$, $j \ge 1$, iid and normally distribution with mean zero and variance $\widehat{\sigma}^2$.

Table 5 below shows some relevant statistics of the 500000 simulation runs.

i	SD	$Qt(95\%)$	$ES(95\%)$
2	66	130	12
3	358	991	99
4	476	1446	194
5	505	1558	218
6	829	3216	355
7	1225	5697	554
8	2023	11399	928
9	2916	18384	1328
10	4932	34698	2231
11	6371	45349	2868
12	8896	64496	3941
13	13671	105310	5993
$Total$	34890	279392	14916

Table 5: Results related to the simulation study of the process error related to the reserve estimators $\widehat{R}^{(1)}_i$, $2 \le i \le 13$. *Column SD gives the standard deviation, column Qt(95%) gives the 95%-quantile and column ES(95%) gives expected shortfall at level 95%; see Mikosch (2006) p. 94 for a formal definition.*

4.2 Data study: A credibility model including diagonal effects

In this section we apply the model **i)-iii)** to data in Table 1. The above estimators of β_j, δ_{j+1}, $0 \leq j \leq 12$, are also used in this section. The estimation of ξ and ϕ is carried out by ordinary linear regression of the row residuals, e_i against $\sum_{j=0}^{m-i} \beta_j \delta_{i+j}$, $1 \leq i \leq 13$; see Figure 3.

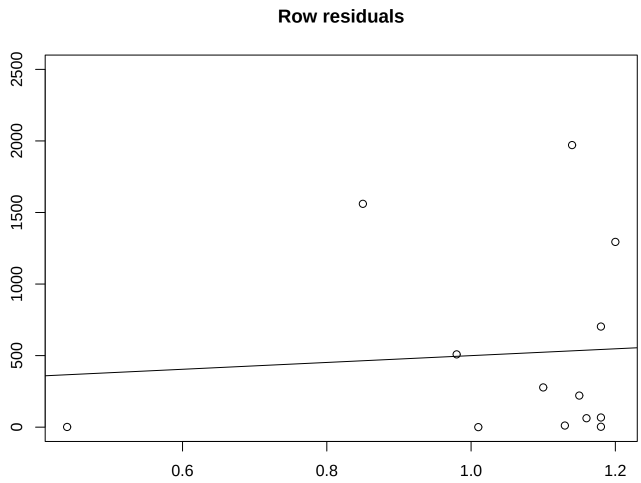

Figure 3: *Row residuals, $\widehat{e}_i^2$, $1 \leq i \leq 13$, plotted $\sum_{j=0}^{m-i} \beta_j \delta_{i+j}$, $1 \leq i \leq 13$. The solid line fitted to the row residuals by the least squares method.*

This procedure yields the estimates

$$\widehat{\xi} = 0.00383, \qquad \widehat{\varphi} = 238.50.$$

Using the convention that $\widehat{R}_i^{(2)} = \sum_{j=m-i+1}^{m-1} \widehat{\mu}(X_{ij})$ and the prediction of $\delta_{14}, ..., \delta_{25}$ given in Example 2 we can now calculate the credibility estimators of the rowwise reserves;

i	2	3	4	5	6	7	8	9	10	11	12	13
$\widehat{R}_i$	12	342	685	709	1883	3578	8015	13564	26124	34576	46251	82476

Table 6: Reserve estimators $\widehat{R}_i^{(2)}$, $2 \le i \le 13$.

see Figure 4 for the credibility weights w_i determining how much credibility should be put on the a priori expectation T_i.

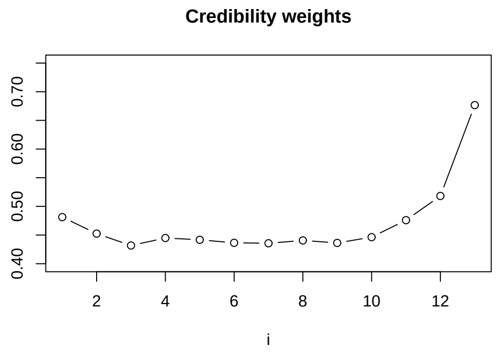

Figure 4: *Credibility weights, w_i, $1 \le i \le 13$, as given in Proposition 5.*

We finally compared the row reserve estimates $\widehat{R}_i^{(1)}$ and $\widehat{R}_i^{(2)}$ to see how much the credibility model effects the Bornhuetter-Ferguson way of prediction. The ratio of the reserve estimates is plotted in Figure 5.

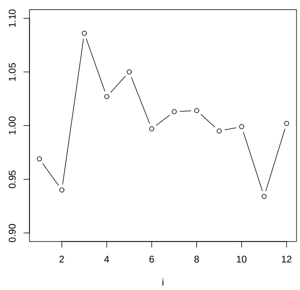

Figure 5: *Rowwise ratios, $\widehat{R}_i^{(2)}/\widehat{R}_i^{(1)}$, $2 \le i \le 13$, between the reserve estimates obtained for the model proposed in Section 3 and 2.*

A simulation study

We follow the same simulation procedure as described in Section 4.1 however based on the residual definition (12). As in Section 4.1 we have generated 500.000 independent simulations. Results are not surprising similar to Section 4.1 and given below in Table 6.

i	SD	$Qt(95\%)$	$ES(95\%)$
2	59	121	15
3	319	917	102
4	457	1461	203
5	471	1515	217
6	786	3225	360
7	1135	5506	528
8	1926	11289	904
9	2815	18340	1302
10	4750	34134	2154
11	6209	44936	2800
12	8221	60003	3652
13	13491	104857	5895
$Total$	34188	274513	14566

Table 7: Results related to the simulation study of the process error related to the reserve estimators $\widehat{R}_i^{(2)}$*,* $2 \leq i \leq 13$*. Column SD gives the standard deviation, column Qt(95%) gives the 95%-quantile and column ES(95%) gives expected shortfall at level 95%.*

Figure 6 shows the histogram of the 500.000 simulation runs. This also concludes our analysis of data.

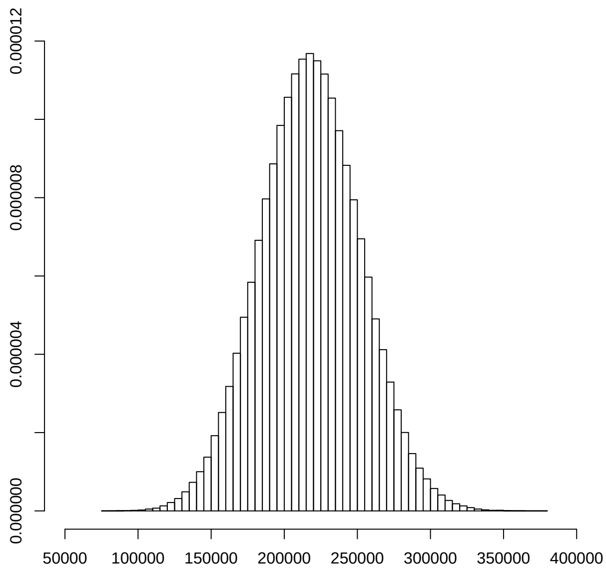

Figure 6: *Histogram for the total reserve.*

5 Bibliography

[1] Benktander, G. (1976): An approach to credibility in calculating IBNR for casualty excess reinsurance. Actuar. Rev. 3, no. 2, 7.
[2] Bornhuetter, R. L. and Ferguson, R. E. (1972): The Actuary and IBNR. Proceedings of the Casualty Actuarial Society, vol. LIX, 181-195.
[3] Buchwalder, M., Bühlmann, H., Merz, M. and Wüthrich, M. (2006): The Mean Square Error of Prediction in the Chain Ladder Reserving Method (Mack and Murphy Revisited). Astin Bulletin, vol 36, 521-543.
[4] England, P. D. and Verrall, R. J. (1999): Analytic and bootstrap estimates of prediction errors in claims reserving. Insurance: Mathematics and Economics, vol. 25, 281-293.
[5] Hachemeister, C. A. and Stanard, J. N. (1975): IBNR claims counts estimation with static lag function. 12th ASTIN Colloquium, IAA, Portimao, Portugal.

[6] Kuang, D., Nielsen, B. and Nielsen, J. P. (2008a): Forecasting with the age-period-cohort model and the extended chain ladder model. Biometrika, vol. 95, 979-986.

[7] Kuang, D., Nielsen, B. and Nielsen, J. P. (2008b): Identification of the age-period-cohort model and the extended chain ladder model. Biometrika, vol. 95, 987-991.

[8] Mack, T. (1991): A simple parametric model for rating automobile insurance or estimating IBNR claims reserves. Astin Bulletin, vol. 21, no. 1, 93-108.

[9] Mack, T. (1993): Distribution-free calculation of the standard error of chain ladder reserve estimates. Astin Bulletin, vol. 23, no. 2, 213-225.

[10] Mack, T. (1994): Measuring the variability of chain ladder reserve estimates. Casualty Actuarial Society Forum Spring 1994, vol. 1, 101-182.

[11] McCullagh, P. and Nelder, J. A. (1989): Generalized Linear Models, second edition. Chapman & Hall, London.

[12] Nelder, J. A. and Pregibon, D. (1987): An extended quasi-likelihood function. Biometrika, vol. 74, 221-232.

[13] Mikosch, T. (2004): Non-Life Insurance Mathematics. Springer-Verlag Berlin Heidelberg.

[14] Rietdorf, N. (2008): The chain ladder in a GLM setup with an extension that includes calender effects. Master Thesis, University of Copenhagen, Department of Mathematics Sciences.

[15] Schmidt, K. D. and Wünsche, A. (1998): Chain ladder, marginal sum and maximum likelihood estimation. Blätter, vol. 23, no. 3, 267-277.

[16] Schmidt, K. D. and Zocher, M. (2008): The Bornhuetter-Ferguson principal. Variance, vol. 2, 85-110.

[17] Taylor, G. (1977): Separation of inflation and other effects from the distribution of non-life insurance claim delays. Astin Bulletin, vol. 9, 219-230.

[18] Taylor, G. (2000): Loss reserving, An Actuarial Perspective. KAP Boston.

[19] Venter, G. G. (2007): Refining reserve runoff Ranges. CAS E-forum: http://casact.org/pubs/forum/07sforum/07s-venter2.pdf

[20] Verbeek, H. G. (1972): An approach to the analysis of claims experience in motor liability excess of loss reinsurance. Astin Bulletin, vol. 6, 195-202.

[21] Verrall, R. J. (2004): A Bayesian generalized linear model for the Bornhuetter-Ferguson method of claims reserving. North American Actuarial Journal, vol. 8, 67-89.

[22] Wüthrich, M. V. (2007): Using a Bayesian approach for claims reserv-

ing. Variance, vol. 1, 292-301.
[23] Zehnwirth, B. (2003): ICRFS-Plus 9.2 Manual. Stochastic loss reserving. Insureware Pty Ltd. St. Kilda, Australia.

Prediction of RBNS and IBNR claims using claim amounts and claim counts

Anders Hedegaard Jessen
Department of Mathematical Sciences, University of Copenhagen

Jens Perch Nielsen
Faculty of Actuarial Science and Insurance, Cass Business School, City University, London

Richard Verrall
Faculty of Actuarial Science and Insurance, Cass Business School, City University, London

May 27, 2009

Abstract
The paid run-off triangle is considered and it is assumed that also the numbers of reported claims (in a similar triangular array) are observable. In this paper only these two triangular arrays are used as data for the model we go on to set up. The data restrictions, to some extent, complicate the statistical analysis, but allow for the model proposed to be implemented on a large variety of data sets. On the basis of those data we suggest a stochastic model build on a compound Poisson framework. The model explicitly takes into account the delay from when a claim is incurred and to when it is reported (the IBNR delay) and the delay from when a claim is reported and to when it is fully paid (the RBNS delay). It is assumed that the single (unobserved) claims are iid and gamma distributed.

1 Introduction

We propose a bivariate stochastic model of loss reserving which is based on observable incremental reported claim numbers N_{ij} and paid amounts X_{ij} and which serves to predict RBNS and IBNR claims.

We start with a micro-model implying certain well defined properties of the reported aggregated claim numbers and aggregated paid amounts. We then study the maximum likelihood estimation of the parameters based on the run-off triangles $\{N_{ij}\}$ and $\{X_{ij}\}$. This is relatively straightforward since the multiplicative structure of the likelihood function allows for separate estimation of the entering parameters which we call $\{\alpha_i, \beta_j\}$ and of $p_0, ..., p_d, \delta, \nu$.

The model has three main ingredients, first the random variables of different accident years are assumed independent for every accident year i and the reported claim numbers

$$N_{i0}, ..., N_{i,m-1}$$

are assumed to have a multiplicative Poisson distribution with parameters $e^{\alpha_i}e^{\beta_0}$, ..., $e^{\alpha_i}e^{\beta_{m-1}}$.

Secondly, when we condition on

$$N_{ij} := \sum_{k=0}^{d} N_{ijk} + N_{ij}^{zero},$$

the numbers N_{ijk} with delay k and the number of zero claims N_{ij}^{zero} have a multinomial distribution with parameter N_{ij} and parameters $p_0, ..., p_d$ and p^{zero} (being independent of accident year i and development year j) with

$$\sum_{k=0}^{d} p_k + p^{zero} = 1.$$

Thirdly and last, when we condition on

$$N_{ij}^{paid} := \sum_{k=0}^{\min\{j,d\}} N_{i,j-k,k}$$

the paid claim amount X_{ij} are gamma distributed with parameters δN_{ij}^{paid} and ν (also being independent of accident year i and development year j).

In the final prediction step we use estimators of the outstanding members of the family $\{N_{ijk}\}$ to construct RBNS and IBNR claim numbers for every cell of the lower triangle. It is then assumed that the corresponding claim amounts have a gamma distribution as before, but with N_{ij}^{rbns} and N_{ij}^{ibnr}, respectively, in the place of N_{ij}^{paid}. Then we compute the Bayes predictor of X_{ij}^{ibnr} and X_{ij}^{rbns} given the triangle of reported claim numbers.

In this section we go through some of our arguments for why we have chosen to work with the above mentioned aggregated data and why we have chosen the above mentioned modeling. There are a number of stochastic models that can be used to estimate reserves in non-life insurance mathematics; see [12] for an extensive literature list. Most of these models have been designed to deal with data which have been aggregated in some way, as this is relatively easily done by the practising actuary. The aggregation of data leads to a loss of information that in some cases can give relatively poor estimation and prediction of the outstanding liabilities. This has been the subject of some recent papers on reserving: for example, [15] use a generalized linear model framework to model the characteristics of individual claims. Further back, a notable set of papers [7, 8] sets out a framework for the claims occurrence, reporting and payment process, at an individual claims level. These types of models are very detailed, often rather complex and use extensive data to estimate parameters. For the practising actuary however, they have certain limitations: in particular, they are difficult to implement because the use of data at an individual level is particularly computationally challenging. Further, very large and elaborate data sets are often hard to get in insurance companies, and it is often the case that a model will only get used in a practical situation if it can be applied to a wide variety of data sets across a wide variety of business lines.

We are therefore left with the dilemma of whether to use individual data, which is theoretically appealing, but computationally difficult, or whether to use aggregate data, which is much easier to deal with but from which some (possibly important) information has been lost. In this paper, we take a similar approach as [9, 19], in that we build the model from basic principles at the level of individual data and we improve the reserving accuracy by adding aggregated counts data readily available in most actuarial offices. Another interesting way of adding more information the classical chain ladder approach is introduced as the Munich chain ladder technique, see [10], where aggregated paid and aggregated incurred data are mixed in a joined model. Other interesting possibilities of adding extra data compared to the simple chain ladder method are [3, 9, 14, 16, 18]. The papers suggests a different ways of handling RBNS and IBNR claims. The extra information of aggregated counts data allow us to model payment patterns for RBNS claims. We believe that this provide better estimation and prediction of the outstanding liabilities. One of our strong arguments is that our model is based on an underlying realistic mathematical statistical micro model. The distributional assumptions are well defined as simple consequences of this underlying model.

The chain ladder method, which in many places is industry standard, was originally introduced without a stochastic model specified using heuristic reasoning to estimate the sum of Incurred But Not Reported (IBNR) claims and RBNS claims. In [2, 4, 5, 11] stochastic models have since been formulated that lead to the same estimates as the chain ladder method. In all three cases, these models take the data as given and do not attempt to build a model based on the commonly accepted compound Poisson framework, used elsewhere in risk theory. It could be argued that the over-dispersed Poisson model could be interpreted in this way (see, for example, [1]), but this was not the original approach taken. In this paper we take a starting point in a compound Poisson model. This is possible because we also observe and model the number of reported claims.

In this paper, we take as the starting point the compound Poisson model. A stringent model formulation for IBNR and RBNS claims is formulated and allows for prediction of IBNR and RBNS claims separately. This is possible because we have added aggregated counts to the data set. In Section 2 we define the notation and describe the data which we will assume is available. In Section 3, the theoretical development is given, working from assumptions at the level of individual data. Also in Section 3, the model which we will actually apply is given, as an approximation to the more detailed models for individual data. Section 4 considers estimation, Section 5 prediction, Section 6 examines some results based on the model and in Section 7 we collect the most important conclusions.

2 Data and notation

The data we work with is carefully chosen. The two main points of our choice of data is that the extra aggregated counts data is readily available for most practical actuaries and that this extra data gives a much better handling of RBNS and IBNR reserves. It also gives a well defined estimated cash flow that lends itself to many actuarial applications including capital considerations.

2.1 The data

We begin by defining the notation. Let us for now just say that a run-off triangle consists of the random variables $\triangle_m = \{X_{ij} : (i,j) \in \mathcal{A}_m\}$ where $\mathcal{A}_m = \{(i,j) \in \mathbb{N} \times \mathbb{N}_0 : 1 \leq i + j \leq m\}$.

$$\begin{array}{ccccc} X_{10} & X_{11} & \dots & X_{1,m-2} & X_{1,m-1} \\ X_{20} & X_{21} & \dots & X_{2,m-2} & \\ \vdots & \ddots & & & \\ X_{m0} & & & & \end{array}$$

The first suffix is a mark that denotes the origin period of the claim; the period when it incurred. The second suffix denotes the delay from the incurred period to when a payment (or a claims reserve) has been made.

There are different possibilities for what the random variables in $\triangle_m$ could represent. X_{ij}, $(i,j) \in \mathcal{A}_m$, could be the total claims incurred in period i and paid with j periods delay from when they incurred. This is then the triangle of paid claims. Another possibility defines X_{ij} as total claims incurred in period i with delay j . In the cases where the claims are reported but not paid a claim estimate is included, rather than the actual payment. In this case, the triangle represents the incurred claims. The use of either the paid or the incurred claims have different advantages and disadvantages. When including only paid data then $\triangle_m$ contains no human judgement; we deal with "real data". However, there is information about future payments, RBNS claims, which is then disregarded. On the other hand, the inclusion of claim estimates is debatable as one no longer considers "real data". There can be political or business related considerations which make the individual claim estimates unreliable. There is further some variability that is disregarded, since the claim estimates and the actual paid amount often differ. Finally claims estimates appear as paid at the wrong point in time which disrupts the cashflow modeling.

For these reasons we adopt the first approach in this paper, and use the triangle of paid claims. Thus, X_{ij} is the total claims incurred in period i and paid with j periods delay. Paid amounts are real data and is in most companies easily accessible since they are the numbers taken directly from the book. We combine this paid triangle with incurred information. We also use a second triangle, in the exact same format as the paid triangle above. In order to incorporate our available information on the number of incurred $RBNS$ claims, we consider random variables $\aleph_m = \{N_{ij} : (i,j) \in \mathcal{A}_m\}$ where N_{ij} represents the number of claims incurred in period i and reported with j periods delay (in period $i+j$) for $(i,j) \in \mathcal{A}_m$. Note that these data consist of the incurred claims, and therefore use some of the information not used when just the aggregate paid claims are used. It would also have been possible to consider the number of payments, which would remove the claims which end up without a payment being made - the zero-claims - from the data. However, this can lead to a number of difficulties. For example, the number of payments is rarely easily accessible in insurance companies. The number of reported claims on the other hand is usually relatively easy to obtain.

We will therefore assume that we can observe $\{(N_{ij}, X_{ij}) : (i,j) \in \mathcal{A}_m\}$. In what follows the notation $(\aleph_m, \triangle_m) := \{(N_{ij}, X_{ij}) : (i,j) \in \mathcal{A}_m\}$ is used. This data straightforward to obtain in most cases. We note here that we are aware that the use of data to which we have limited ourselves to some extent complicates the statistical analysis. It would be better, from this point of view at least, to assume that we have data available at whatever level of detail we require. A disadvantage of this would be that the estimation of the models became much more computationally intense, and the models could not be used when the data requirements were not satisfied.

Thus, we have made a compromise about the data we use, but it will be seen in Section 5 that by just including the count data for the incurred claims, we can improve significantly on the chain ladder technique without completely giving up the well known chain ladder idea. The chain ladder method is in fact a special case of our model.

In the next section we bind together $\triangle_m$ and $\aleph_m$ using some unobservable random variables. The structure of the unobservable variables are intended to mimic the models [7,8], but we use a discrete time framework.

2.2 IBNR and RBNS claims

In this section we introduce our micro model including a number of (in practise often) unobservable random variables. Based on this micro model we are able to articulate a compound poisson interpretation of IBNR and RBNS claims (on individual level). In Section 2.1, the data available for the statistical analysis was outlined, in this section notation relating to some basic variables is defined which mimic the structure in the underlying data related to the payment of claims.

The detailed model of this paper with assumptions on individual claim development follow to some extent the theoretical papers [4, 7, 8]. This approach is operationalized through restrictions on available data. While assumptions are made on individual data level, only aggregated data is assumed to be available for the statistical estimation process.

Consider the kth claim of the N_{ij} claims incurred in period i and reported with j periods delay. Usually a claim is not paid immediately upon notification to the insurance company. The final claim amount is generally paid with some waiting time from notification, often due to general consideration of the case, legal issues, collection of further information concerning the case, etc. In other words, there is a delay from a claim being reported and until it is fully paid. The claims that have been reported but are not yet paid are the so called RBNS claims (or Incurred But Not Enough Reported (IBNER) claims). The related delay in payment is referred to as the RBNS delay.

In order to formulate mathematically how claims are paid, we introduce some stochastic variables that in some cases may be observable, but we assume in general that they are not, and the model does not rely on having observations for them. Denote by N_{ijk} the part of the N_{ij} claims which are (fully) paid with k periods delay, $k = 0, ..., d$. Here $k = 0$ corresponds to a claim being paid in the same period as it was reported whereas $k = d$ is the maximal possible RBNS delay in the model. d could be chosen using information from the underlying data or the judgement from a claims handler.

It sometimes occurs that reported claims are settled with no payment, for example, if there is consideration about who carries responsibility for a claim, fraud or similar. These are referred to as zero claims. Denote the number of zero claims that are incurred (or apparently incurred in fraud cases etc.) in period (i, j) and are settled at value zero at some point in time by N_{ij}^{zero}. We then must have $N_{ij0} + \cdots + N_{ijd} \leq N_{ij}$ and $N_{ij}^{zero} + N_{ij0} + \cdots + N_{ijd} = N_{ij}$ for $(i, j) \in \mathcal{A}_m$. The number of claims incurred in period i and (fully) paid with j periods delay after being reported is defined as

$$N_{ij0} + N_{i,j-1,1} + \cdots + N_{i,j-\min\{j,d\},d} = \sum_{k=0}^{\min\{j,d\}} N_{i,j-k,k} =: N_{ij}^{paid}$$

for $(i, j) \in \mathcal{A}_m$ where we put $N_{ijk} = 0$ for $j < 0$, $i, k \geq 1$, by convention. The non-negative random variable Y_{ijk} is the kth claim paid in period $i + j$ which was incurred in period i for $(i, j) \in \mathcal{A}_m$ and $k = 1, ..., N_{ij}^{paid}$.

Thus, we divide the lifetime of a claim into two: the IBNR delay and the RBNS (and IBNER) delay.

The notation has now been defined, and the next section is dedicated to formulating a model and discussing its possibilities and limitations.

3 The stochastic model

In this section basic assumptions on how claims are reported and paid are considered leading to a model for the paid amounts and reported numbers of claims.

The model for the reported number of claims is derived in Section 3.1, and models for claim amounts are derived in Sections 3.2. These models are built from principles closely related to the set-up in [7, 8].

3.1 The claim counts

In this section, a model for the number of reported claims, $\aleph_m$, is proposed. It is assumed that N_{ij} are independent random variables, which have a Poisson distribution with mean μ_{ij} where

$$\log \mu_{ij} = \alpha_i + \beta_j \tag{1}$$

and $\sum_{j=0}^{m-1} e^{\beta_j} = 1$. The maximum likelihood estimates of α_i, $1 \leq i \leq m$, and β_i, $0 \leq j \leq m-1$, are such that the estimation of the number of IBNR claims corresponds to the chain ladder method applied to $\aleph_m$: see for example [2, 5, 13, 18] for proofs of this well known result. As the chain ladder method is standard among practitioners and because the modeling of $\aleph_m$ is not the prime objective of this paper, we will simply assume the above model for $\aleph_m$.

3.2 Paid amounts: one payment per claim

The model of this paper assumes only one payment per claim. This model restriction simplifies theory, estimation and data questions considerable. For example, in real life there is often more than one payment per claim, however, modeling this through stochastic time series approaches is nontrivial. Also, data are often not available on the development of the payment patterns and definitions on payments may differ from one insurance company to the other, or even within the same insurance company. Assuming only one payment per claim should therefore be considered as a well chosen model simplification leading to more robustness, fewer parameters to estimate and a wider applicability of the method.

The proportion of claims settled at zero, $Q \in [0, 1)$, is assumed to be a known constant for $(i, j) \in \mathcal{A}_m$. Given N_{ij} let

$$(N_{ij0}, ..., N_{ijd}, N_{ij}^{zero}) \sim \text{Multi}(N_{ij}; p_0, ..., p_d, Q)$$

for $(i, j) \in \mathcal{A}_m$ and $p_0 + \cdots + p_d = 1 - Q$ where $p_i \in (0, 1)$, $0 \leq i \leq d$. In other words, it is assumed that the conditional density (w.r.t. the counting measure on $\mathbb{N}_0^{d+2}$) of $(N_{ij0}, ..., N_{ijd}, N_{ij}^{zero})$ is given by

$$f_{N_{ij0},...,N_{ijd},N_{ij}^{zero}|N_{ij}}(n_0, ..., n_{d+1}) = \binom{N_{ij}}{n_0, ..., n_{d+1}} p_0^{n_0} \cdots p_d^{n_d} Q^{n_{d+1}}$$

for $n_l \in \mathbb{N}_0$, $0 \le l \le d+1$ such that $n_0 + \cdots + n_{d+1} = N_{ij}$, see Section 2.2. The aggregated payments made with l periods delay are defined as

$$S_{ijl} = \sum_{k=N_{ij0}+\cdots+N_{ij,l-1}+1}^{N_{ij0}+\cdots+N_{ijl}} Y_{ij}^{(k)} \tag{2}$$

for $(i, j) \in \mathcal{A}_m$. Where $Y_{ij}^{(k)}$, $(i, j) \in \mathcal{A}_m$, $k \ge 1$, is as defined in Section 3.2. Here the second order structure is

$$\begin{aligned} E\left(S_{ijl} \mid N_{ij} = n_{ij}\right) &= n_{ij} E Y_{11}^{(1)} p_l, \\ \operatorname{Cov}\left(S_{ijl}; S_{ijl'} \mid N_{ij} = n_{ij}\right) &= n_{ij}\left[V\left(Y_{11}^{(1)}\right) p_l I\{l = l'\} - \left(E Y_{11}^{(1)}\right)^2 p_l p_{l'}\right] \end{aligned}$$

for $(i, j) \in \mathcal{A}_m$ and $l, l' = 0, ..., d$. In reality there are different numbers of payments related to different claims: sometimes there is one payment and sometimes there are more. From a pragmatic point of view this means that the parameter estimates in some cases may not have straightforward interpretations. While there are of course claims that are paid in one payment, we recognize that for many claims this is not the case. However, when there is more than one payment then often there is one main payment that is big compared to the others (which are adjustments, additional costs or similar). Thus the model (2) has the type of properties at the aggregate level that we believe are reasonable, and we can accept some theoretical shortcomings in the model. As was stated at the beginning of the paper, the logical process we have followed is to derive a model as far as possible, based on very basic, unobservable random variables, and then approximate the model as closely as possible to motivate a model for the data available. Accepting the convention that there is only one payment per claim, leads one to relatively simple distributional characteristics. The conditional distribution of S_{ijl} given N_{ijl} is a $\Gamma(\delta N_{ijl}, \nu)$-distribution and we thus have

$$X_{ij} := \sum_{l=0}^{d} \sum_{k=N_{ij0}+\cdots+N_{ij,l-1}+1}^{N_{ij0}+\cdots+N_{ijl}} Y_{ij}^{(k)} \overset{d}{=} \sum_{k=1}^{N_{ij0}+\cdots+N_{i,j-d,d}} Y_{11}^{(k)} \sim \Gamma(N_{ij}^{paid}\delta, \nu)$$

given N_{ij}^{paid} using the notation from Section 2.2.

In this section, we have discussed a model for Δ_m given $\aleph_m$ and argued that this model (1) and (2) is suitable for the data format available. The next sections discuss estimation and prediction for this model.

4 Estimation

In this section the likelihood functions for the model (1) and (2) proposed in Sections 3.1 and 3.2 given the data $(\aleph_m, \triangle_m)$ are calculated. The likelihood function for $(\aleph_m, \triangle_m)$ can be written as

$$\mathcal{L}_{\aleph_m,\triangle_m}(\{\alpha_i, 1 \le i \le m\}, \{\beta_j, 0 \le j \le m-1\}, \delta, \nu, \{p_l, 0 \le l \le d\}) \tag{3}$$

$$\begin{aligned}
&= \mathcal{L}_{\aleph_m}(\{\alpha_i, 1 \le i \le m\}, \{\beta_j, 0 \le j \le m-1\}) \times \mathcal{L}_{\triangle_m|\aleph_m}(\delta, \nu, p_0, ..., p_d) \\
&= \left(\prod_{i=1}^{m} \prod_{j=0}^{m-i} P(N_{ij} = n_{ij}) \right) \\
&\quad \times \left(\prod_{i=1}^{m} f_{X_{i0},...,X_{i,m-i}|N_{i0},...,N_{i,m-i}}(x_{i0}, ..., x_{i,m-i}|n_{i0}, ..., n_{i,m-i}) \right).
\end{aligned}$$

Since $\mathcal{L}_{\aleph_m}$ and $\mathcal{L}_{\triangle_m|\aleph_m}$ are not functions of the same parameters, it is sufficient to maximize $\mathcal{L}_{\aleph_m}$ and $\mathcal{L}_{\triangle_m|\aleph_m}$ separately to maximize $\mathcal{L}_{\triangle_m,\aleph_m}$ in (3).

The likelihood function of $\aleph_m$ can be optimized the following way using the chain ladder method. Define the accumulative run-off triangle by

$$C_{ij} = \sum_{k=0}^{j} N_{ik}, \qquad \text{for} \qquad (i,j) \in \mathcal{A}_m,$$

and let $\widehat{C}_{i,m-1}$ be the according chain ladder predictors of $C_{i,m-1}$, see for example [2,4]. Then the maximum likelihood estimates of e^{α_i} and e^{β_j} are given by

$$e^{\widehat{\alpha}_i} = \widehat{C}_{i,m-1}, \qquad 1 \le i \le m,$$

and

$$e^{\widehat{\beta}_j} = \frac{\widehat{C}_{mj} - \widehat{C}_{m,j-1}}{e^{\widehat{\alpha}_i}}, \qquad 0 \le j \le m-1.$$

See appendix A for a proof.

The log-likelihood function for $\triangle_m$ given $\aleph_m$, see Appendix A, is given by

$$\log(\mathcal{L}_{\triangle_m|\aleph_m}(\delta, \nu, p_0, ..., p_d)) = C' - \nu x_{..} + \sum_{i=1}^{m} \log \left(\sum_{j=0}^{m-i} Q^{n_{ij}} n_{ij}! \left[\sum_{n_{ij0}=0}^{n_{ij}} \left(\frac{p_0}{Q}\right)^{n_{ij0}} \frac{1}{n_{ij0}!} \right.\right.$$

$$\left.\left. \sum_{n_{ij1}=0}^{n_{ij}-n_{ij0}} \left(\frac{p_1}{Q}\right)^{n_{ij1}} \frac{1}{n_{ij1}!} \cdots \sum_{n_{ijd}=0}^{n_{ij}+n_{ijd}-n_{ij.}} \left(\frac{p_d}{Q}\right)^{n_{ijd}} \frac{1}{n_{ijd}!} \frac{1}{(n_{ij}-n_{ij.})!} \prod_{k=0}^{m-i} \frac{(\nu x_{ik})^{n_{ik}^{paid}\delta}}{\Gamma(n_{ik}^{paid}\delta)} \right] \right) \tag{4}$$

where $n_{ij}^{paid} = n_{ij0} + \cdots + n_{i,j-d,d}$ with $n_{ijl} := 0$ for $l \le -1$, $(i,k) \in \mathcal{A}_m$, and

$$x_{..} = \sum_{i=1}^{m} \sum_{j=0}^{m-i} x_{ij}.$$

If some of the n_{ij}'s are large numbers then numerical calculations of the sums in (4) requires a long computer time. To overcome this problem see Appendix B for an alternative estimation method using generalized linear models.

Now formulae for estimating the parameters in the model defined in (1) and (2) have been derived. The next section is dedicated to a discussion of what is meant by RBNS and IBNR claims on aggregate level and how these can be estimated.

5 Prediction

This section describes what RBNS and IBNR claims are (based on Section 2.2) and how they can be predicted. Firstly, we clarify the notions of RBNS and IBNR clams on aggregate level in the framework of a suitable model. The model for the triangular arrays $(\aleph_m, \triangle_m) = \{(N_{ij}, X_{ij}) : (i,j) \in \mathcal{A}_m\}$ defined in (1) and (2) can be extended in a natural way to the random variables N_{ij}, $(i,j) \in \mathcal{B}_m$. Here

$$\mathcal{B}_m = \{(i,j) \in \mathbb{N}_0^2 : 1 \leq i \leq m,\ 0 \leq j \leq m-1\}$$

and X_{ij}, $(i,j) \in \mathcal{C}_m$, where

$$\mathcal{C}_m = \{(i,j) \in \mathbb{N}_0^2 : 1 \leq i \leq m,\ 0 \leq j \leq m+d-1\}.$$

The random variables thus appear in this format

$$\begin{array}{ccc} N_{10} & \dots & N_{1,m-1} \\ N_{20} & \dots & N_{2,m-1} \\ \vdots & & \vdots \\ N_{m0} & \dots & N_{m,m-1} \end{array} \qquad \begin{array}{ccc} X_{10} & \dots & X_{1,m+d-1} \\ X_{20} & \dots & X_{2,m+d-1} \\ \vdots & & \vdots \\ X_{m0} & \dots & X_{m,m+d-1}. \end{array}$$

In compliance with (1), it is natural to assume that the variables N_{ij}, $m-i+1 \leq j \leq m-1$, are independent and

$$N_{ij} \sim \text{Pois}\,(\mu_{ij})\,, \qquad \log(\mu_{ij}) = \alpha_i + \beta_j,$$

for $1 \leq i \leq m$. Further, in analogy with (2), it is assumed that given N_{ij}

$$(N_{ij0}, ..., N_{ijd}, N_{ij}^{zero}) \sim \text{Multi}(N_{ij}; p_0, ..., p_d, Q)$$

for $(i,j) \in \mathcal{B}_m$. Define $N_{ij}^{rep} := 0$ for $j \geq m$ such that also $(N_{ij0}, ..., N_{ijd}, N_{ij}^{zero}) = (0, ..., 0)$ for $j \geq m$ and $1 \leq i \leq m$. In other words, it is assumed that no claims can be reported more than m periods after they incurred. As in Section 3.2, the total paid amounts, given N_{ij}^{paid}, $(i,j) \in \mathcal{C}_m$, are independent and

$$X_{ij} \sim \Gamma(N_{ij}^{paid}\delta, \nu)$$

where $N_{ij}^{paid} := N_{ij0} + \cdots + N_{i,j-d,d}$, $(i,j) \in \mathcal{C}_m$, again using the notation from Section 2.2.

The triangular arrays $(\aleph_m, \triangle_m)$ are subsets of N_{ij}^{rep}, $(i,j) \in \mathcal{B}_m$ and X_{ij}, $(i,j) \in \mathcal{C}_m$. The unobservable variables N_{ij}^{rep}, $(i,j) \in \mathcal{B}_m \backslash \mathcal{A}_m$ and X_{ij}, $(i,j) \in \mathcal{C}_m \backslash \mathcal{A}_m$ represent the number of claims that will be reported in the future and the future payments, respectively. It is natural to split the future payments into two parts: those related to claims that are already reported and as such appear in $\aleph_m$, and those that will be reported in the future, N_{ij}^{rep}, $(i,j) \in \mathcal{B}_m \backslash \mathcal{A}_m$.

Formally, these are related to the (RBNS) payments

$$X_{ij}^{rbns} \sim \Gamma(N_{ij}^{rbns}\delta, \nu)$$

which are independent given $N_{ij}^{rbns} := N_{i,m-i,i+j-m} + \cdots + N_{i,\max\{j-d,0\},\min\{d,j\}}$ for $m-i+1 \leq j \leq m-i+d$ and $1 \leq i \leq m$. For $j > m-i+d$ and $1 \leq i \leq m$, let $X_{ij}^{rbns} := 0$.

The claims that will be reported and paid in the future (IBNR claims) are independent given N_{ij}^{ibnr}, $(i,j) \in \mathcal{C}_m \backslash \mathcal{A}_m$, and

$$X_{ij}^{ibnr} \sim \Gamma(N_{ij}^{ibnr}\delta, \nu),$$

where $N_{ij}^{ibnr} := N_{ij0} + \cdots + N_{i,\max\{m-i+1;j-d\},\min\{j+i-m-1;d\}}$ for $m-i+1 \leq j \leq m+d-1$ and $2 \leq i \leq m$. Let $X_{1j}^{ibnr} := 0$ for $j \geq m$. Finally we assume that the random variables X_{ij}^{ibnr}, $m-i+1 \leq j \leq m-i+d$, and X_{ij}^{rbns}, $m-i+1 \leq j \leq m-i+d$ are mutually independent such that

$$X_{ij}^{ibnr} + X_{ij}^{rbns} = X_{ij}$$

for $(i,j) \in \mathcal{C}_m \backslash \mathcal{A}_m$.

5.1 Prediction of the total IBNR claims

This section considers the prediction of X_{ij}^{ibnr} for $m-i+1 \leq j \leq m-i+d$ and $2 \leq i \leq m$. In this case the predictor of total IBNR claims is an immediate functional of our parameters estimated by maximum likelihood. The predictor $\widehat{\mu}_{ij}^{rbns}$ of X_{ij}^{ibnr} is defined as the minimizer of

$$E\left(X_{ij}^{ibnr} - \mu_{ij}^{ibnr}\right)^2$$

over μ_{ij}^{ibnr} where μ_{ij}^{ibnr} is any finite variance measurable function of $(\aleph_m, \triangle_m)$, $m-i+1 \leq j \leq m-i+d$ and $2 \leq i \leq m$. As X_{ij}^{ibnr}, $m-i+1 \leq j \leq m-i+d$ and $2 \leq i \leq m$ are independent of $(\aleph_m, \triangle_m)$, the minimizer is simply given by

$$\widehat{\mu}_{ij}^{ibnr} = E\left(X_{ij}^{ibnr}\right) = \frac{e^{\alpha_i}\delta}{\nu} \sum_{k=0}^{\min\{d,j+i-m-1\}} p_k e^{\beta_{j-k}}. \tag{5}$$

5.2 Prediction of the total RBNS claims

The prediction of RBNS claims is nontrivial, even when all entering statistical parameters have been estimated from the maximum likelihood procedure. The reason is that the best distribution of RBNS claims conditional on available data is a complicated expression. While we do know the number of claims behind the RBNS claims, we do not know the exact number of claims that have been paid at this moment in time, and we do not know the exact amount paid. We only have aggregated data indicating trends, we do not have the exact underlying information available to us. In the following we derive the predictor of RBNS claims as the conditional mean based on observed information. RBNS claims are not independent of $(\aleph_m, \triangle_m)$ which complicates the prediction slightly. For simplicity we derive the predictor, μ_i^{rbns}, of the row-wise RBNS payments,

$$X_i^{rbns} := X_{i,m-i+1}^{rbns} + \cdots + X_{i,m+d-i}^{rbns}, \qquad 1 \leq i \leq m,$$

but it is also possible to derive predictors of X_{ij}^{rbns} in a similar fashion. The predictor, $\widehat{\mu}_i^{rbns}$, of X_i^{rbns} is again defined as the minimizer of

$$E\left(X_i^{rbns} - \mu_i^{rbns}\right)^2$$

over all finite variance functions of $(\aleph_m, \triangle_m)$. As $Y_{ij}^{(k)}$ are iid, $i, k \geq 1$ and $j \geq 0$, and independent of $\triangle_m$ it is straightforward that

$$\widehat{\mu}_i^{rbns} = \frac{\delta E(N_i^{rbns} \mid \aleph_m, \triangle_m)}{\nu}$$

for $1 \leq i \leq m$. One way of determining $E(N_i^{rbns} \mid \aleph_m, \triangle_m)$ is to derive a formula for $P(N_i^{rbns} = k \mid \aleph_m, \triangle_m)$, $0 \leq k \leq N_{i\cdot}$ where $N_{i\cdot} := N_{i0} + \cdots + N_{i,m-i}$, and $1 \leq i \leq m$. However, an application of Bayes formula shows that the probabilities

$$P\left(X_{ij} \in (x_j, x_j + h_j],\ \xi_i \leq j \leq m - i \mid N_m^{rbns} = k, \aleph_m\right)$$

appear in the expression for $E(N_i^{rbns} \mid \aleph_m, \triangle_m)$. And these probabilities are relatively complex to determine. In order for the prediction to be applicable in practise another predictor is thus proposed. To obtain relatively simple formulae we introduce the predictor, $\widetilde{\mu}_i^{rbns}$, of X_i^{rbns} as the minimizer of

$$E\left(X_i^{rbns} - \mu_i^{rbns}\right)^2$$

over all finite variance functions of $\aleph_m$ and $X_{i\cdot}$, $1 \leq i \leq m$. Conditioning only on how much has been paid in total, $X_{i\cdot}$, $1 \leq i \leq m$, leads to much less complex expressions for the predictors, $\widetilde{\mu}_i^{rbns}$, $1 \leq i \leq m$. A drawback is that the predictors, $\widetilde{\mu}_i^{rbns}$, are of course sub-optimal in the usual $\mathbb{L}^2$ sense.

Denote by $k \mapsto p_{n,q}(k)$ the point mass in $k \in \{0, ..., n\}$ of a binomial distribution with length n and success probability $1 - q \in (0, 1)$. Let $f_{\Gamma(a,b)}$ be the

Lebesgue density of a gamma distribution with parameter $(a,b) \in (0,\infty)^2$. For $0 \le k \le N_{i.}$ an application of Bayes formula yields that

$$P\left(N_i^{rbns} = k \mid \aleph_m, X_{i.} \in (x_i, x_i + h]\right) = \frac{P\left(N_i^{rbns} = k \mid \aleph_m\right) P\left(X_{i.} \in (x_i, x_i+h] \mid N_m^{rbns} = k, \aleph_m\right)}{P\left(X_{i.} \in (x_i, x_i+h] \mid \aleph_m\right)}$$

for $x_i, h > 0$ and $1 \le i \le m$. With $k_{m-i} := k - k_{\xi_i} - \cdots - k_{m-i-1}$ it follows directly from Section 3.2 that

$$P\left(N_i^{rbns} = k \mid \aleph_m\right) = \sum_{k_{\xi_i}=0}^{k} \sum_{k_{\xi_i+1}=0}^{k-k_{\xi_i}} \cdots \sum_{k_{m-i-1}=0}^{k-k_{\xi_i}-\cdots-k_{m-i-2}} \prod_{j=\xi_i}^{m-i} p_{N_{ij}, Q+\sum_{l=0}^{m-i-j} p_l}(k_j).$$

for $1 \le i \le m$. Particularly with $i = m$, then $P\left(N_i^{rbns} = k \mid \aleph_m\right) = p_{N_{m0}, Q+p_1}(k)$. Further, for $1 \le i \le m$ we have

$$\begin{aligned}
\frac{P\left(X_{i.} \in (x_i, x_i+h] \mid N_m^{rbns} = k, \triangle_m\right)}{P\left(X_{i.} \in (x_i, x_i+h] \mid \triangle_m\right)} &= \frac{P\left(X_{i.} \in (x_i, x_i+h] \mid N_m^{rbns} = k, N_{i.}\right)}{P\left(X_{i.} \in (x_i, x_i+h] \mid N_{i.}\right)} \\
&= (1+o(1)) \frac{h \sum_{l=1}^{N_{i.}-k} p_{N_{i.}-k, Q}(l) f_{\Gamma(l\delta,\nu)}(x_i)}{h \sum_{l=1}^{N_{i.}} p_{N_{i.}, Q}(l) f_{\Gamma(l\delta,\nu)}(x_i)}
\end{aligned}$$

as $h \to 0$ and it follows with $C_i' := (\sum_{l=1}^{N_{i.}} p_{N_{i.},Q}(l) f_{\Gamma(l\delta,\nu)}(x_i))^{-1}$ and $\xi_i := \max\{0, m+1+d-i\}$ that

$$\begin{aligned}
P\left(N_i^{rbns} = k \mid \aleph_m, X_{i.}\right) = & \\
= \; & C_i' \left(\sum_{k_{\xi_i}=0}^{k} \sum_{k_{\xi_i+1}=0}^{k-k_{\xi_i}} \cdots \sum_{k_{m-i-1}=0}^{k-k_{\xi_i}-\cdots-k_{m-i-2}} \prod_{j=\xi_i}^{m-i} p_{N_{ij}, Q+\sum_{l=0}^{m-i-j} p_l}(k_j) \right) \\
& \times \left(\sum_{l=0}^{N_{i.}-k} p_{N_{i.}-k, Q}(l) f_{\Gamma(l\delta,\nu)}(X_{i.}) \right).
\end{aligned}$$

The predictor, $\widetilde{\mu}_i^{rbns}$, of X_i^{rbns}, is therefore given by

$$\begin{aligned}
\widetilde{\mu}_i^{rbns} &= \frac{\delta E(N_i^{rbns} \mid \aleph_m, X_{i.})}{\nu} \qquad (6) \\
&= \frac{\delta}{\nu} \sum_{k=0}^{N_{i.}} k P\left(N_i^{rbns} = k \mid \aleph_m, X_{i.}\right).
\end{aligned}$$

Notice in particular that $\widetilde{\mu}_m^{rbns} = \widehat{\mu}_m^{rbns}$.

An alternative and simplistic predictor of the RBNS claims, X_{ij}^{rbns}, which does not take into account the models dependence structure, is given by

$$\breve{\mu}_{ij}^{rbns} := \frac{\delta}{\nu} \sum_{k=\xi_i}^{m-i} p_k N_{i,j-k}$$

for $(i, j) \in \mathcal{A}_m$. One could use $\breve{\mu}_{ij}$ as a benchmark prediction.

In this section the model (1) and (2) was extended in a natural way and in this work it has been specified what RBNS and IBNR claims are. Finally predictors minimizing the $\mathbb{L}^2$-norm have been derived. Prediction of the RBNS and IBNR claims is the last step in the theoretical analysis of model (1) and (2) based on the data $(\aleph_m, \triangle_m)$. In the next section an application to data of the results in Sections 4 and 5 is carried out.

6 Data study

In this section, the model (1) and (2) is applied to a dataset from Royal & Sun Alliance. The data relate to a portfolio of motor policies; in this example the auto third part liability (TPL) data is considered. The reason for choosing this data set is that we expect there to be reasonably long settlement delays (RBNS delays). This could be of particular interest as (1) and (2) explicitly models the RBNS delay. The data displayed in Table 1 is inflation corrected, so that

$$X_{ij} := \frac{Y_{ij}}{\delta_{i+j}}$$

where Y_{ij}, $(i, j) \in \mathcal{A}_{10}$, are the observed payments and δ_i is an inflation index, $1 \leq i \leq 10$. In a full analysis of a dataset such as this, the inflation index could be modeled independently, for example by a time series which should then be used in the prediction, $X_{ij}\delta_j$ for $j \geq 10-i+1$. For the purpose of this paper, we assume that the claims inflation has already been estimated, and we concentrate on modeling the inflation corrected payments, $\triangle_{10}$, which are shown in Table 1.

i\j	0	1	2	3	4	5	6	7	8	9
1	451288	339519	333371	144988	93243	45511	25217	20406	31482	1729
2	448627	512882	168467	130674	56044	33397	56071	26522	14346	
3	693574	497737	202272	120753	125046	37154	27608	17864		
4	652043	546406	244474	200896	106802	106753	63688			
5	566082	503970	217838	145181	165519	91313				
6	606606	562543	227374	153551	132743					
7	536976	472525	154205	150564						
8	554833	590880	300964							
9	537238	701111								
10	684944									

Table 1: The paid run-off triangle, X_{ij}, $(i, j) \in \mathcal{A}_{10}$, for the auto TPL data.

The incurred counts are shown in Table 2.

i\j	0	1	2	3	4	5	6	7	8	9
1	6238	831	49	7	1	1	2	1	2	3
2	7773	1381	23	4	1	3	1	1	3	
3	10306	1093	17	5	2	0	2	2		
4	9639	995	17	6	1	5	4			
5	9511	1386	39	4	6	5				
6	10023	1342	31	16	9					
7	9834	1424	59	24						
8	10899	1503	84							
9	11954	1704								
10	10989									

Table 2: The number of reported claims, N_{ij}, $(i,j) \in \mathcal{A}_{10}$, for the auto TPL.

Expert advice from a claims handler have been used to determine the fraction of reported zero-claims, $Q \in [0,1)$, and the maximal possible RBNS delay, $d \leq 10$. In this case $(Q,d) = (0.2, 7)$.

Estimation in the model (1) and (2) has been done using the chain ladder method as described in Section 4 and Appendix B respectively. Prediction of IBNR and RBNS claims has been conducted as proposed in (5) and (6).

6.1 Estimation in model (1) and (2)

The parameters in the model (1) are estimated by maximizing the likelihood function, $\mathcal{L}_{\aleph_m}(\gamma, \{\beta,0 \leq j \leq 9\}, \{\alpha_i, 1 \leq i \leq 10\})$, in (3).

Some of the values N_{i0}, $1 \leq i \leq 10$, exceed 10000. Therefore the maximization of (4) would require long computer time. Instead we use the quasi-likelihood function in (8) in Appendix B to estimate the parameters $(\delta, \nu, p_0, ..., p_7)$.

The optimizations are carried out in the statistical software **R**. The results are displayed in Table 4.

j	0	1	2	3	4	5	6	7	8	9
e^b	0.875	0.118	0.004	0.001	0.000	0.000	0.000	0.000	0.000	0.000
p	0.291	0.230	0.091	0.068	0.053	0.029	0.020	0.018		

Table 4: The row indexed by e^b are the maximum likelihood estimates of e^{β_j}, $0 \leq j \leq 9$ and the row indexed p are the maximum likelihood estimates of p_l, $0 \leq l \leq 7$.

Notice that the part of claim that is settled at value zero is set to $Q = 0.2$ such that

$$p_0 + \cdots + p_7 = 0.8.$$

The estimates of (ν, δ) are

$$(\widehat{\nu}, \widehat{\gamma}) = (0.0170, 8.395e - 5)$$

such that the distribution of the single (unobserved) claim, $Y_{ijl}^{(k)}$, is estimated to be a $\Gamma(0.0170, 8.395e - 5)$.

As was mentioned in the beginning of Section 6, the TPL claims are expected to have relatively long settlement delays (RBNS delays) as bodily injury claims often take a long time to settle. As may be seen from the left hand plot in Figure 1 there is empirical proof of this: long time after the majority of claims has been reported there are still significant payments. Also notice that p_0 may be relatively small. As claims happen, on average, in the middle of the year there is on average only half a year to receive the final payment in order to finish in the category of claims related to p_0. All other delay periods are full years.

Given $N_{1.}$, it holds that

$$(N_{11}, ..., N_{1m}) \sim \text{Multi}\left(N_{1.}, e^{\beta_0}, ..., e^{\beta_9}\right).$$

It hence seems natural to compare the estimates of the IBNR and RBNS delays by considering $p'_j := e^{\beta_j}$, $0 \le j \le 9$ and p_l, $0 \le l \le 7$. This comparison is made in the right hand plot in Figure 1. In this way the average IBNR delay related to p'_j, $0 \le j \le 9$ is 0.129 years whereas the average RBNS delay is 1.22. Hence the RBNS reserve is expected to be (about ten times) bigger than the IBNR reserve because the individual claims are assumed iid.

6.2 Model validation

In this section we verify whether the model (1) and (2) seems to be consistent with the dataset in Tables 1 and 2. As the random variables N_{ij}, X_{ij}, $(i, j) \in \mathcal{A}_m$, all have different distributions the following idea is used to validate the model.

A PP-plot is constructed in the following way. Denote the distribution function of a random variable, Z_j, by F_{Z_j}, $1 \le j \le N$. Then $U_j := F_{Z_j}(Z_j) \sim$Unif$(0, 1)$ provided F_{Z_j} is continuous and for the ordered sample $U_{(1)} \le U_{(2)} \le \cdots \le U_{(N)}$ it holds that $EU_{(j)} = j/(N+1)$ for $1 \le j \le N$. A PP-plot consists of the points $(j/(N+1), EU_{(j)})$ which approximate the line $\{(x, y) \in [0, 1]^2 : x = y\}$. In the left hand plot in Figure 2 the points $k/56$, $1 \le k \le 55$ are plotted against the ordered sample of $\widehat{F}_{N_{ij}}(N_{ij})$, $(i, j) \in \mathcal{A}_{10}$, where $\widehat{F}_{N_{ij}}$ is the estimated distribution function of N_{ij} (from Section 6.1). The right hand PP-plot corresponds to the variables $F_{X_{ij}|\triangle_m}(X_{ij})$, $(i, j) \in \mathcal{A}_{10}$. It seems that there is some systematic

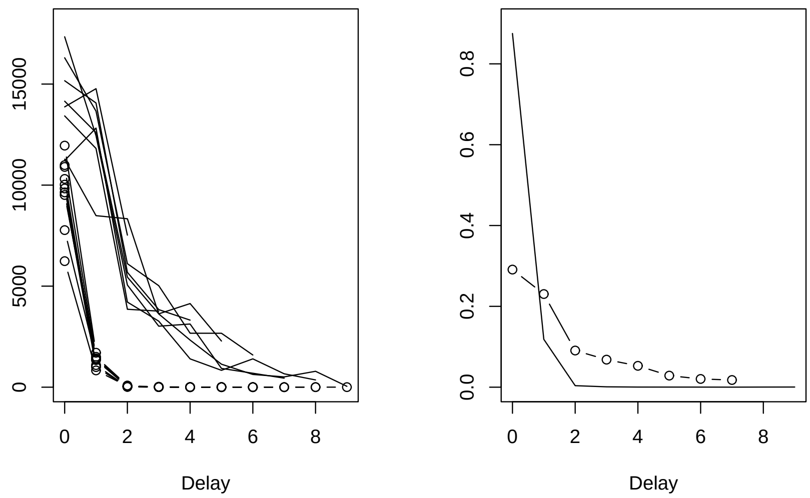

Figure 1: *Left: The dotted lines represent the points* $\{X_{ij}/40, 0 \leq j \leq 9\}$ *for* $1 \leq i \leq 10$. *The solid graphs are the number of reported claims,* $\{N_{ij}, 0 \leq j \leq 9\}$ *for* $1 \leq i \leq 9$. *Right: The estimates of* $(p_0, ..., p_7)$ *representing the RBNS delay are the dotted graphs and the solid graphs are the IBNR delay,* p'_j, $0 \leq j \leq 9$.

error in the left hand plot in Figure 2, but the right hand PP-plot looks convincing.

To visualize possible changes in the structure of the data through time we have also plotted $F_{N_{ij}}(N_{ij})$ and $F_{X_{ij}}(X_{ij})$ against the row and column indices respectively, see Figure 3. Notice in particular that the bottom right plot indicates that either p_l, $0 \leq l \leq d$, or the average claim, $EY_{1j}^{(1)}$, change as a function of the column, $0 \leq j \leq m-1$.

The plots indicate a reasonable fit of the model (1) and (2) to the data in Tables 1 and 2. In the next section we therefore go on to predict the IBNR and RBNS claims.

6.3 Estimation of IBNR and RBNS claims

In this section we predict the IBNR and RBNS claims as proposed in (5) and (6) respectively. We have chosen to consider only the rows $i = 4, ..., 10$ because the majority of the total claims reserve is related to these. Given the parameter estimates in Section 6.1 the estimated reserve are given in Table 5.

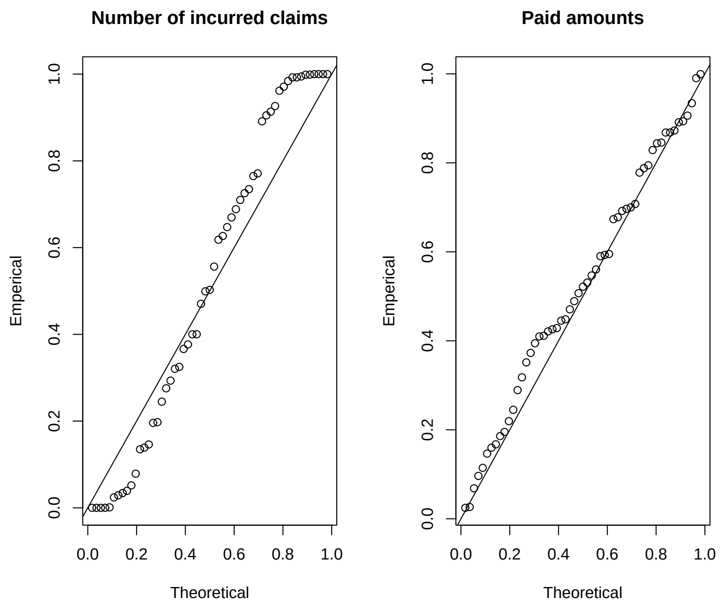

Figure 2: *Left: A PP-plot for* α_m. *Right: PP-plot for* $\triangle_m$ *given* $\aleph_m$.

i	10	9	8	7	6	5	4
IBNR	317326	17805	6670	3950	3215	2453	1883
RBNS	1137206	853576	511176	298823	168321	92933	57285
RESERVE	1454532	871381	517846	302773	171536	95386	59168

Table 5: The row wise reserve estimates split into IBNR and RBNS claims.

We saw in Section 6.1 that the IBNR delay is (on average) shorter than the RBNS delay, and hence the RBNS is expected to be larger then the IBNR reserve. The actual estimates divide the reserves such that the RBNS reserve takes up 91.6% of the total reserve and the IBNR only 8.4%: roughly $1:10$ as suggested in Section 6.1.

As proposed in Section 5.2 it is relevant to compare the more sophisticated estimator, $\widetilde{\mu}_i^{rbns}$, with the simple estimator $\breve{\mu}_{ij}^{rbns}$, $1 \leq i \leq 10$. This is done in Table 6 below.

i	10	9	8	7	6	5	4

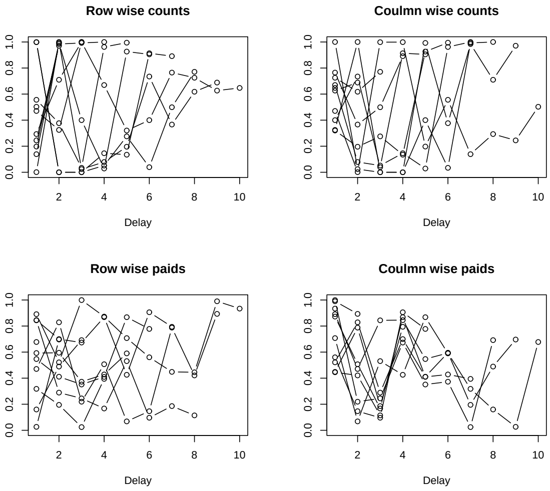

Figure 3: *Top left graphs: the points* $\{(j, F_{N_{ij}}(N_{ij}), 1 \le j \le m-i\}$, $1 \le i \le m$. *Top right graphs: the points* $\{(i, F_{N_{ij}}(N_{ij}), 1 \le i \le m-j\}$, $0 \le j \le m-1$. *Bottom left graphs: the points* $\{(j, F_{X_{ij}}(X_{ij}), 1 \le j \le m-i\}$, $1 \le i \le m$. *Top right graphs: the points* $\{(i, F_{X_{ij}}(X_{ij}), 1 \le i \le m-j\}$, $0 \le j \le m-1$.

```
----------------------------------------------------------
 mu_i(~)/(delta/nu)  |  5601  4204  2518  1472  829  458  282
 mu_i(v)/(delta/nu)  |  5594  4197  2509  1473  845  466  215
```

Table 6: The top row is the number of RBNS claims estimated using (6). *The bottom row is the number of RBNS claims estimated by* $\breve{\mu}_{i0}^{rbns} + \cdots + \breve{\mu}_{i,9-i}^{rbns}$, $4 \le i \le 10$, *where* $\breve{\mu}_{ij}^{rbns}$, $(i,j) \in \mathcal{A}_m$, *is as defined in Section 5.2.*

It was said in the introduction that the use of more data than, for example, the chain ladder allows for a more sophisticated analysis. However, as often done, we can use the chain ladder method as a benchmark to which one can relate the results in this section. In the next section we compare the results from the model (1) and (2) and the chain ladder reserve estimates. One could use bootstrapping to find the distributional characteristics for the chain ladder estimates, see [1]. The predictive distributions in model (1) and (2) are all explicitly given in Section 5.1 and 5.2.

6.4 A comparison with the chain ladder results

One of the most popular methods for calculating reserves in practise is the chain ladder method. Using the paid run-off triangle, $\triangle_m$, it is straightforward to estimate the row-wise reserves. The chain ladder reserves include both IBNR and a part of the RBNS claims. But it is not possible to split them explicitly. The reserve estimates are given in Table 7.

i	10	9	8	7	6	5	4
RESERVE	1459860	763919	475992	249349	173802	101158	53791

Table 7: The row-wise chain ladder reserve estimates.

The chain ladder reserve estimates are close to the result from the Section 6.3. The total chain ladder reserve for rows $i = 4, ..., 10$, is 3277871 compared to the 3472622 in Section 6.3. The difference could be explained by the fact that not all RBNS claims are estimated using the chain ladder method. It could also be due to model uncertainty.

In datasets with relatively few claims there is empirical evidence that the estimation in (1) and (2) gives an estimation error of the reserve estimates which is smaller than compared to the chain ladder.

7 Appendix A

In this section we assume the model (1) and (2).

7.1 The likelihood function for $\aleph_m$

In the papers [2,4] it is shown that

$$e^{\widehat{C}_{i,m-1}} = \sum_{j=0}^{m-1} e^{\widehat{\alpha}_i + \widehat{\beta}_j}, \qquad 1 \le i \le m$$

where $e^{\widehat{\alpha}_i}$ and $e^{\widehat{\beta}_j}$ denote the maximum likelihood estimators of e^{α_i} and e^{β_j}. Since the model is made identifiable by imposing the constraint

$$\sum_{j=0}^{m-1} e^{\beta_j} = 1$$

we directly obtain

$$e^{\widehat{C}_{i,m-1}} = e^{\widehat{\alpha}_i}, \qquad 1 \le i \le m.$$

It moreover follows directly that

$$\frac{e^{\widehat{C}_{mj}} - e^{\widehat{C}_{m,j-1}}}{e^{\widehat{\alpha}_m}} = \frac{1}{e^{\widehat{\alpha}_m}} \left(\sum_{k=0}^{j} e^{\widehat{\alpha}_i + \widehat{\beta}_k} - \sum_{k=0}^{j-1} e^{\widehat{\alpha}_i + \widehat{\beta}_k} \right) = e^{\widehat{\beta}_j}$$

for $0 \le j \le m-1$.

7.2 The likelihood function for $\triangle_m$ given $\aleph_m$

In addition to the conditions of Section 8.1 we assume that the paid amounts, $X_{ij} = x_{ij}$, $(i,j) \in \mathcal{A}_m$, are also known. We define $x_{i.} := x_{i0} + \cdots + x_{i,m-i}$, $n_{ij.} := n_{ij0} + \cdots + n_{ijd}$ and $n_{ij}^{paid} := n_{ij0} + n_{i,j-1,1} + \cdots + n_{i,j-d,d}$ where we put $n_{ijl} = 0$ for $l < 0$ by convention. The conditional likelihood function of $\triangle_m$ given $\aleph_m$ defined in (3) is given as

$$\mathcal{L}_{\triangle_m|\aleph_m}(\delta, \nu, p_0, ..., p_d) \quad =$$

$$\prod_{i=1}^{m} \sum_{j=0}^{m-i} \sum_{n_{ij0}=0}^{n_{ij}} \sum_{n_{ij1}=0}^{n_{ij}-n_{ij0}} \cdots \sum_{n_{ijd}=0}^{n_{ij}-n_{ij.}} \left[\binom{n_{ij}}{n_{ij0} \cdots n_{ijd} \cdot (n_{ij} - n_{ij.})} p_0^{n_{ij0}} \cdots p_d^{n_{ijd}} Q^{n_{ij}-n_{ij.}} \right.$$

$$\left. \prod_{k=0}^{m-i} \frac{\nu^{n_{ik}^{paid}\delta}}{\Gamma(n_{ik}^{paid}\delta)} x_{ik}^{n_{ik}^{paid}\delta - 1} e^{-\nu x_{ik}} \right].$$

Here

$$\prod_{k=0}^{m-i} \frac{\nu^{n_{ik}^{paid}\delta}}{\Gamma(n_{ik}^{paid}\delta)} x_{ik}^{n_{ik}^{paid}\delta-1} e^{-\nu x_{ik}} = e^{-\nu x_{i.}-\sum_{j=0}^{m-i}\log(x_{ij})} \prod_{k=0}^{m-i} \frac{\nu^{n_{ik}^{paid}\delta}}{\Gamma(n_{ik}^{paid}\delta)} x_{ik}^{n_{ik}^{paid}\delta}$$

and

$$\prod_{l=0}^{d} p_l^{n_{ijl}} Q^{n_{ij}-n_{ij.}} = Q^{n_{ij}} \prod_{l=0}^{d} \left(\frac{p_l}{Q}\right)^{n_{ijl}}$$

such that $\mathcal{L}_{\triangle_m|\aleph_m}$ can be written as

$$\prod_{i=1}^{m} e^{-\nu x_{i.}-\sum_{j=0}^{m-i}\log(x_{ij})} \sum_{j=0}^{m-i} Q^{n_{ij}} n_{ij}! \left[\sum_{n_{ij0}=0}^{n_{ij}} \sum_{n_{ij1}=0}^{n_{ij}-n_{ij0}} \cdots \sum_{n_{ijd}=0}^{n_{ij}-n_{ij.}} \left(\prod_{l=0}^{d} \left(\frac{p_l}{Q}\right)^{n_{ijl}} \frac{1}{n_{ijl}!} \right) \right.$$

$$\left. \frac{1}{(n_{ij}-n_{ij.})!} \prod_{k=0}^{m-i} \frac{(\nu x_{ik})^{n_{ik}^{paid}\delta}}{\Gamma(n_{ik}^{paid}\delta)} \right].$$

Taking logarithms, one finally obtains

$$\log(\mathcal{L}_{\triangle_m|\aleph_m}(\delta,\nu,p_0,...,p_d)) = C' + \sum_{i=1}^{m} \left[-\nu x_{i.} + \log\left(\sum_{j=0}^{m-i} Q^{n_{ij}} n_{ij}! \left[\sum_{n_{ij0}=0}^{n_{ij}} \left(\frac{p_0}{Q}\right)^{n_{ij0}} \frac{1}{n_{ij0}!} \right.\right.\right.$$

$$\left.\left.\left. \sum_{n_{ij1}=0}^{n_{ij}-n_{ij0}} \left(\frac{p_1}{Q}\right)^{n_{ij1}} \frac{1}{n_{ij1}!} \cdots \sum_{n_{ijd}=0}^{n_{ij}+n_{ijd}-n_{ij.}} \left(\frac{p_d}{Q}\right)^{n_{ijd}} \frac{1}{n_{ijd}!} \frac{1}{(n_{ij}-n_{ij.})!} \prod_{k=0}^{m-i} \frac{(\nu x_{ik})^{n_{ik}^{paid}\delta}}{\Gamma(n_{ik}^{paid}\delta)} \right] \right) \right]$$

where $C' = \sum_{i=1}^{m}\sum_{j=0}^{m-i}\log(x_{ij})$.

8 Appendix B

In this section we propose an approximative GLM approach to estimating the parameters $(\nu, \delta, p_0, ..., p_d)$. Section 9 in [6] is used to put up a Quasi-likelihood function. We however need the following convention: the relation $VY_{11}^{(1)}/EY_{11}^{(1)} \gg \max_{l=0,...,d}(1-p_l)$ holds, where $\gg$ means that the right hand side is negligible in comparison to the left hand side. This is often empirically observed. In Section 9.3 in [6] it is suggested that the random variables X_{ij} should be considered (conditionally given $\aleph_m$) independent as the covariance structure of $\triangle_m$ given $\aleph_m$ is completely determined by the parameters $(\nu, \delta, p_0, ..., p_d)$. Using [6], p. 326, the estimation is purely based on the relation between the conditional mean and variance of X_{ij}, $(i,j) \in \mathcal{A}_m$ given $\triangle_m$. For $(i,j) \in \mathcal{A}_m$ we see that

$$\begin{aligned} E(X_{ij}|\aleph_m) &= \sum_{l=0}^{d} N_{i,j-l}^{rep}\psi_l \\ V(X_{ij}|\aleph_m) &= \sum_{l=0}^{d} N_{i,j-l}^{rep}\psi_l \left[\frac{VY_{11}^{(1)}}{EY_{11}^{(1)}} + (1-p_l)\right] \simeq \varphi \sum_{l=0}^{d} N_{i,j-l}^{rep}\psi_l, \end{aligned} \tag{7}$$

where $\psi_l = p_l EY_{11}^{(1)}$, $0 \leq l \leq d$, and $\varphi = VY_{11}^{(1)}/EY_{11}^{(1)}$. The approximate formula (7) combined with Section 6.2.3 and Table 9.1 in [6] suggest the following conditional quasi-log-likelihood function for $\triangle_m$ given $\aleph_m$,

$$\sum_{(i,j)\in\mathcal{A}_m} \left[X_{ij} \log\left(\sum_{l=0}^{d} N_{i,j-l}^{rep}\psi_l\right) - \sum_{l=0}^{d} N_{i,j-l}^{rep}\psi_l\right].$$

Denote the maximizers of the latter quantity by $\widehat{\psi}_l$, $0 \leq l \leq d$. Following Section 9.2 in [6], the ratio, φ, can be estimated by

$$\widehat{\varphi} = \frac{2}{m(m+1)} \sum_{(i,j)\in\mathcal{A}_m} \frac{\left(X_{ij} - \sum_{l=0}^{d} N_{i,j-l}^{rep}\widehat{\psi}_l\right)^2}{\sum_{l=0}^{d} N_{i,j-l}^{rep}\widehat{\psi}_l}.$$

The actual parameter estimates are then obtained by the method of moments:

$$\widehat{\nu} = \frac{\sum_{k=0}^{d}\widehat{\psi}_k}{\widehat{\phi}(1-Q)}, \qquad \widehat{\delta} = \widehat{\nu}\frac{\sum_{k=0}^{d}\widehat{\psi}_k}{1-Q},$$

$$\widehat{p}_l = \frac{\widehat{\psi}_l(1-Q)}{\sum_{k=0}^{d}\widehat{\psi}_k}, \qquad 1 \leq l \leq d. \tag{8}$$

9 Bibliography

[1] England, P. D. and Verrall, R. J. (1999): Analytic and bootstrap estimates of prediction errors in claims reserving. Insurance: Mathematics and Economics, vol. 25, 281-293.

[2] Hachemeister, C. A. and Standard, J. N. (1975): IBNR claims count estimation with static lag functions. 12th Astin Colloquium, IAA, Portimao, Portugal.

[3] Liu, H. and Verrall, R. (2009): Predictive Distributions for Reserves which Separate True IBNR and IBNER Claims. To appear in Astin Bulletin.

[4] Mack, T. (1991): A simple parametric model for rating automobile insurance or estimating IBNR claims reserves. Astin Bulletin, vol. 21, no. 1, 93-108.

[5] McCullagh, P. and Nelder, J. A. (1989): Generalized Linear Models, second edition. Chapman & Hall, London.

[6] Neuhaus, W. (2004): On the estimation of outstanding claims. Australian Actuarial Journal, vol. 10, 485-518.

[7] Norberg, R., (1993): Prediction of outstanding liabilities in non-life insurance. Astin Bulletin, vol. 23, no. 1, 95-115.

[8] Norberg, R., (1999): Prediction of outstanding claims: Model variations and extensions. Astin Bulletin, vol. 29, no. 1, 5-25.

[9] Ntzoufras, I. and Dellaportas, P (2002): Bayesian modeling of outstanding liabilities incorporating claim count uncertainty. North American Actuarial Journal, vol. 6, no. 1, 113-137.

[10] Quarg, G. and Mack, T. (2004): Munich chain ladder. Blätter DGVM, vol. 26, 597-630.

[11] Renshaw, A. E. and Verrall, R. (1994): The stochastic model underlying the chain ladder technique. Actuarial research paper no. 63.

[12] Schmidt, K. D. (2007): A bibliography on loss reserving.
Available on: http://www.math.tu-dresden.de/sto/schmidt/dsvm/reserve.pdf

[13] Schmidt, K. D. and Wünsch, A (1998): Chain ladder, marginal sum, and maximum likelihood estimation. Blätter DGVM, vol. 23, 267-277.

[14] Schnieper, R. (1991): Separating true IBNR and IBNER claims. Astin Bulletin, vol. 21, 111-127.

[15] Taylor, G. C. and McGuire, G. (2004): Loss reserving with GLM - a case study. CAS discussion paper program, 327-391.

[16] Verrall, R. J. (1990): Bayes and empirical Bayes estimates for the chains ladder. Astin Bulletin, vol. 20, 217-243.

[17] Verrall, R. J. (1991): Chain ladder and maximum likelihood. Journal of the Institute of Actuaries, vol. 118, 489-499.

[18] Verrall, R. J. and England, P. D. (2005): Incorporating expert opinion into a stochastic model for the chain-ladder technique. Insurance, Mathematic and Economics, vol. 37, 355-370 .

[19] Wright, T. S. (1990): A stochastic method for claims reserving in general insurance. Journal of the Institute of Actuaries, vol. 117, 677-731.

PREDICTION OF OUTSTANDING PAYMENTS IN A POISSON CLUSTER MODEL

ANDERS HEDEGAARD JESSEN, THOMAS MIKOSCH,
AND GENNADY SAMORODNITSKY

Abstract. 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.

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

$$\widehat{N}_{j+\ell+1} = E(N_{j+\ell+1} \mid \mathcal{F}_j) \quad \text{and} \quad \widehat{S}_{j+\ell+1} = E(S_{j+\ell+1} \mid \mathcal{F}_j)\,, \tag{1.1}$$

$\ell = 0, 1, \ldots$, where $\mathcal{F}_j = \sigma(N_0, \ldots, N_j)$, $j = 0, 1, \ldots$. Assuming $\mathrm{var}(S_{j+\ell+1}) < \infty$, $S = \widehat{S}_{j+\ell+1}$ is the a.s. unique minimizer of the mean square error $E((S_{j+\ell+1} - S)^2)$ in the class of square integrable random variables S which

1991 *Mathematics Subject Classification.* Primary 60K10, Secondary 60G25, 60G55.

Key words and phrases. Poisson cluster model, prediction, claims reserving, chain ladder method, Panjer distributions, (a, b)-class, recursive algorithm.

Thomas Mikosch's research is partly supported by the Danish Research Council (FNU) Grant 272-06-0442. Gennady Samorodnitsky's research is partly supported by the ARO grant W911NF-07-1-0078 at Cornell University and a Villum Kann Rasmussen Visiting Professor Grant at the University of Copenhagen.

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

One of the popular procedures in this context was suggested by Mack; see Mack [2, 3, 4] and Mack et al. [6]. In its simplest version, Mack's procedure declares the predictors $\widehat{S}_{j+\ell+1}$ and $\widehat{N}_{j+\ell+1}$ to be linear functions of $S_0 + \cdots + S_j$ or $N_j = N_0 + \cdots + N_j$, respectively. For example,

$$\widehat{N}_{j+1} = (f_j - 1)\,(N_0 + \cdots + N_j)\,, \quad j = 0, 1, \ldots\,, \tag{1.2}$$

for constants $f_j \geq 1$.

Assume that one observes the *run-off triangle*

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

where $((N_{i,i+k}, S_{i,i+k}))_{k=0,\ldots,j-i}$ are the payment numbers and total payments for claims arriving in year i and being executed in year $i + k$; one assumes that $((N_{i,i+k}, S_{i,i+k}))_{k=0,1,\ldots}$ are iid copies of $((N_k, S_k))_{k=0,1,\ldots}$. Here n is the last year for which payments were observed. Mack's assumptions (1.2) give raise to constructing natural estimators $\widehat{f_j}$ of f_j which are referred to as *chain ladder estimators*. Then, replacing the unknown parameters f_j on the right-hand side of (1.2) by their estimators $\widehat{f_j}$, one obtains a predictor of N_{j+1}. In Section 5 we 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. For example, one cannot simulate a process $(N_j)_{j=0,1,\ldots}$ from (1.2). Moreover, the linearity of the estimator (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 $N_j, j = 0, 1, \ldots$, and the corresponding payments $S_j, j = 0, 1, \ldots$. The model is given by the following conditions which we assume throughout this paper.

The model. Let M be the number of claims arriving in a given year with distribution

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

The mth claim causes a stream of K_m payments from the insurer to the insured through the next years. We assume that the kth 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 $(Y_{mk})_{m,k=1,2,\ldots}$ constitutes an iid family with common distribution

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

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

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

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 $(X_{mk})_{m,k=1,2,\ldots}$ is an iid family of non-negative random variables independent of M, (K_m) and (Y_{mk}). We interpret X_{mk} as the kth payment for the mth claim. Then

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

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 $N_0 + \cdots + N_j$, in contrast to Mack's procedure (1.2). However, under some condition on the distribution of M these predictors are *asymptotically linear* functions of $N_0 + \cdots + N_j$ if the latter quantity increases to infinity. In other situations, the predictors stay non-linear even in the limit.

The paper is organized as follows. We start in Section 2 by giving the relevant formulas for the predictors of $S_{j+\ell+1}$, $\ell = 0, 1, \ldots$. Since $S_j = N_j$ if $X_{mi} = 1$ for all m, i, the prediction of $N_{j+\ell+1}$, $\ell = 0, 1, \ldots$, 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 $\widehat{S}_{j+\ell+1}$ are asymptotically linear functions of $N_0 + \cdots + N_j$. 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 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 $S_{j+\ell+1}$, $\ell = 0, 1, \ldots$, given the past payment numbers $N_0, \ldots, N_j$. This means we will calculate the predictors $\widehat{S}_{j+\ell+1}$ and

$\widehat{N}_{j+\ell+1}$ in (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 implies that

$$\widehat{S}_{j+\ell+1} = EX_{11}\, E(N_{j+\ell+1} \mid \mathcal{F}_j) = EX_{11}\, \widehat{N}_{j+\ell+1}\,, \tag{2.1}$$

$\ell = 0, 1, \ldots$. 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 $N_j = S_j$, $j = 0, 1, \ldots$. Therefore it suffices to study the prediction of $S_{j+\ell+1}$ given $N_0, \ldots, N_j$.

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.,

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

and its derivatives

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

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

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

Finally, write

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

recall that μ is the Poisson rate of the number of payments per claim, and (p_k) are the displacement probabilities in (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 $\widehat{S}_{j+1}$ of S_{j+1} given $N_0, \ldots, N_j$ has the form*

$$E(S_{j+1} \mid N_0 = n_0, \ldots, N_j = n_j) = \mu\, p_{j+1}\, EX_{11}\, R_{n_0+\cdots+n_j}(\theta_j)\,, \tag{2.2}$$

$n_0, \ldots, n_j = 0, 1, \ldots,$ $j = 0, 1, \ldots$.
(2) *Assume, in addition, that* $\mathrm{var}(M) < \infty$ *and* $\mathrm{var}(X_{11}) < \infty$. *Then the unconditional prediction error for S_{j+1}, $j = 0, 1, \ldots,$ is given by*

$$\begin{aligned} E\left((S_{j+1} - \widehat{S}_{j+1})^2\right) &= E(X_{11}^2)\, \mu\, p_{j+1}\, EM \\ &+ (EX_{11} \mu\, p_{j+1})^2\, E(M^2) - E(\widehat{S}_{j+1}^2)\,. \end{aligned} \tag{2.3}$$

.

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

$$\mathrm{var}(S_{j+1} \mid N_0 = n_0, \ldots, N_j = n_j) = E(X_{11}^2)\, \mu\, p_{j+1}\, R_{n_0+\cdots+n_j}(\theta_j) \tag{2.4}$$
$$+ (EX_{11}\, \mu\, p_{j+1})^2\, R_{n_0+\cdots+n_j}(\theta_j)\left[R_{n_0+\cdots+n_j+1}(\theta_j) - R_{n_0+\cdots+n_j}(\theta_j)\right] .$$

Remark 2.2. Writing $\ell_j = n_0 + \cdots + n_j$, $j = 0, 1, \ldots$, we observe by virtue of (2.2) that

$$\begin{aligned} & E(S_{j+1} \mid N_0 = n_0, \ldots, N_j = n_j) \\ = \; & E(S_{j+1} \mid N_0 = \ell_0, N_0 + N_1 = \ell_1, \ldots, N_0 + \cdots + N_j = \ell_j) \\ = \; & E(S_{j+1} \mid N_0 + \cdots + N_j = \ell_j)\,, \end{aligned}$$

or, alternatively,

$$\begin{aligned} & E(S_0 + \cdots + S_{j+1} \mid N_0 + \cdots + N_j = \ell_j) \\ & = EX_{11}\, \ell_j + E(S_{j+1} \mid N_0 + \cdots + N_j = \ell_j)\,. \end{aligned} \tag{2.5}$$

By virtue of (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 (2.5) is asymptotically linear as $\ell_j \to \infty$.

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, (N_j) constitutes, conditionally on M, a sequence of independent Poisson random variables. Therefore

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

and

$$\widehat{S}_{j+1} = EX_{11}\, \mu\, p_{j+1}\, E(M \mid \mathcal{F}_j)\,, \quad j = 0, 1, \ldots\,. \tag{2.6}$$

Even more precisely, let Z_{jl} denote the number of payments in the lth payment stream, $l = 1, 2, \ldots$, which are executed in year $j = 0, 1, \ldots$. Then (Z_{jl}) constitutes a double array of independent random variables with $Z_{jl} \sim \mathrm{Poisson}(\mu p_j)$. Therefore for any $m, j = 0, 1, \ldots$ and $n_0, \ldots, n_j = 0, 1, \ldots$,

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

$$= \quad q_m\, \mathrm{e}^{-m\,\theta_j}\; m^{\sum_{d=0}^{j} n_d} \prod_{d=0}^{j} \frac{(\mu\, p_d)^{n_d}}{n_d!}\,. \tag{2.7}$$

We conclude that for $j \geq 0$,

$$P(M = m \mid N_0 = n_0, \ldots, N_j = n_j) \quad = \quad \frac{q_m \mathrm{e}^{-m\,\theta_j}\; m^{\sum_{d=0}^{j} n_d}}{\sum_{r=0}^{\infty} q_r \mathrm{e}^{-r\,\theta_j}\; r^{\sum_{d=0}^{j} n_d}}\,,$$
$$m, n_0, \ldots, n_j = 0, 1, \ldots\,. \tag{2.8}$$

In particular,

$$E(M \mid N_0 = n_0, \ldots, N_j = n_j) \quad = \quad \frac{\sum_{m=0}^{\infty} q_m \mathrm{e}^{-m\,\theta_j}\; m^{\sum_{d=0}^{j} n_d + 1}}{\sum_{r=0}^{\infty} q_r \mathrm{e}^{-r\,\theta_j}\; r^{\sum_{d=0}^{j} n_d}}$$
$$= \quad R_{n_0+\cdots+n_j}(\theta_j)\,. \tag{2.9}$$

We conclude, using (2.6), that (2.2) holds.

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

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

Taking into account this relation and (2.1), we see that the conditional prediction error can be written as

$$\mathrm{var}(S_{j+1} \mid \mathcal{F}_j) = E(X_{11}^2)\, E(N_{j+1} \mid \mathcal{F}_j) + (EX_{11}\mu\, p_{j+1})^2\, \mathrm{var}(M \mid \mathcal{F}_j)\,. \tag{2.10}$$

Using (2.8), we can replace the conditional moments of M by the corresponding derivatives of L, leading to (2.4). Taking expectations in (2.10), we obtain the prediction error

$$E\left((S_{j+1} - \widehat{S}_{j+1})^2\right) = E\left[\mathrm{var}(S_{j+1} \mid \mathcal{F}_j)\right]$$
$$= \quad E(X_{11}^2) EN_{j+1} + (EX_{11}\mu\, p_{j+1})^2 \left[E(M^2) - E[(E(M \mid \mathcal{F}_j))^2]\right]$$
$$= \quad E(X_{11}^2)\, \mu\, p_{j+1}\, EM + (EX_{11}\mu\, p_{j+1})^2\, E(M^2) - E(\widehat{S}_{j+1}^2)\,.$$

This finishes the proof. □

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

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

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

$$E(\widehat{S}_{j+1}^2) = (EX_{11}\, \mu\, p_{j+1})^2\, E\left[(E(M \mid \mathcal{F}_j))^2\right]$$
$$= \quad (EX_{11}\mu\, p_{j+1})^2\, E\left[(R_{N_0+\cdots+N_j}(\theta_j))^2\right]$$

$$= \quad (EX_{11}\mu\, p_{j+1})^2 \sum_{k=0}^{\infty}\sum_{m=0}^{\infty} P(N_0+\cdots+N_j=k\,, M=m)\,(R_k(\theta_j))^2\,.$$

Applying (2.7), the right-hand double sum turns into

$$\sum_{k=0}^{\infty}\left[\sum_{m=0}^{\infty} q_m\,\mathrm{e}^{-m\,\theta_j}\,m^k\right]\sum_{n_0+\cdots+n_j=k}\prod_{d=0}^{j}\frac{(\mu\,p_d)^{n_d}}{n_d!}(R_k(\theta_j))^2$$

$$= \quad \mathrm{e}^{\theta_j}\sum_{k=0}^{\infty}(-1)^k\,\frac{\left(L^{(k+1)}\,(\theta_j)\right)^2}{L^{(k)}\,(\theta_j)}\,P(\Theta_j=k)$$

$$= \quad \mathrm{e}^{\theta_j}E\left((-1)^{\Theta}_j\,\frac{\left(L^{(\Theta_j+1)}\,(\theta_j)\right)^2}{L^{(\Theta_j)}\,(\theta_j)}\right)\,,$$

where Θ_j is Poisson(θ_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 $\widehat{S}_{j+\ell+1}$, $\ell=0,1,\ldots$, defined in (1.1), and the corresponding prediction errors.

Theorem 2.5. *Assume that* $EM<\infty$ *and* $EX_{11}<\infty$.
(1) *The predictor* $\widehat{S}_{j+\ell+1}$ *of* $S_{j+\ell+1}$ *given* $N_0,\ldots,N_j$ *has the form*

$$E(S_{j+\ell+1}\mid N_0=n_0,\ldots,N_j=n_j) \quad = \quad EX_{11}\,\mu\,p_{j+\ell+1}\,R_{n_0+\cdots+n_j}(\theta_j)\,,$$
$$j,\ell=0,1,\ldots\,.$$

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

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

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

$$\mathrm{var}(S_{j+\ell+1}\mid N_0=n_0,\ldots,N_j=n_j)$$
$$= \quad E(X_{11}^2)\,\mu\,p_{j+\ell+1}\,R_{n_0+\cdots+n_j}(\theta_j)$$
$$+(EX_{11}\mu p_{j+\ell+1})^2\,R_{n_0+\cdots+n_j}(\theta_j)\,\left[R_{n_0+\cdots+n_j+1}(\theta_j)-R_{n_0+\cdots+n_j}(\theta_j))\right]\,.$$

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

$$\widehat{S}_{j+\ell+1} \quad = \quad E\left[E(S_{j+\ell+1}\mid\mathcal{F}_{j+\ell})\mid\mathcal{F}_j\right]\,.$$

Hence, using (2.6), we obtain

$$\widehat{S}_{j+\ell+1} \quad = \quad EX_{11}\,\mu\,p_{j+\ell+1}E\left[E(M\mid\mathcal{F}_{j+\ell})\mid\mathcal{F}_j\right]$$

$$= \quad EX_{11}\,\mu\,p_{j+\ell+1}E(M \mid \mathcal{F}_j)\,.$$

Now use relation (2.9).

(2) & (3) For the conditional prediction error for $S_{j+\ell+1}$, we observe that by (2.10)

$$\begin{aligned}
&\operatorname{var}(S_{j+\ell+1} \mid \mathcal{F}_j)\\
= \;& E(\operatorname{var}(S_{j+\ell+1} \mid \mathcal{F}_{j+\ell}) \mid \mathcal{F}_j) + \operatorname{var}(E(S_{j+\ell+1} \mid \mathcal{F}_{j+\ell}) \mid \mathcal{F}_j)\\
= \;& E(X_{11}^2)\,\widehat{N}_{j+\ell+1} + (EX_{11}\mu\,p_{j+\ell+1})^2\,E(\operatorname{var}(M \mid \mathcal{F}_{j+\ell}) \mid \mathcal{F}_j)\\
& +(EX_{11}\mu\,p_{j+\ell+1})^2\operatorname{var}(E(M \mid \mathcal{F}_{j+\ell}) \mid \mathcal{F}_j)\\
= \;& E(X_{11}^2)\,\widehat{N}_{j+\ell+1} + (EX_{11}\mu\,p_{j+\ell+1})^2\operatorname{var}(M \mid \mathcal{F}_j)\,.
\end{aligned}$$

The conditional moments of M can, once again, be expressed using (2.8). Taking expectations, we obtain the unconditional prediction error

$$\begin{aligned}
&E\left[(S_{j+\ell+1} - \widehat{S}_{j+\ell+1})^2\right] = E\left[\operatorname{var}(S_{j+\ell+1} \mid \mathcal{F}_j)\right]\\
= \;& E(X_{11}^2)\,\mu\,p_{j+\ell+1}\,EM + (EX_{11}\mu\,p_{j+\ell+1})^2\left[E(M^2) - E\left((E(M \mid \mathcal{F}_j))^2\right)\right]\\
= \;& E(X_{11}^2)\,\mu\,p_{j+\ell+1}\,EM + (EX_{11}\mu\,p_{j+\ell+1})^2\,E(M^2) - E(\widehat{S}_{j+\ell+1}^2)\,.
\end{aligned}$$

□

Remark 2.6. Notice that

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

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

2.3. **Conditionally independent payments.** In this subsection we consider a slightly more general model. As before, we assume that the sequences $(X_{mk})_{k=1,2,\ldots}$, $m = 1, 2, \ldots$, are iid and independent of the rest of random variables defining the model. We further assume that each sequence $(X_{mk})_{k=1,2,\ldots}$ consists of *conditionally* iid random variables or, equivalently, that $(X_{mk})_{k=1,2,\ldots}$ 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 [7], Chapters 5 and 6.

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

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

$$\begin{aligned}
E((S_{j+1} - \widehat{S}_{j+1})^2) \quad = \quad & EM\,\left[E(X_{11}^2)\,\mu\,p_{j+1} + \operatorname{cov}(X_{11}, X_{12})\,(\mu\,p_{j+1})^2\right]\\
& +(EX_{11}\mu\,p_{j+1})^2E(M^2) - E(\widehat{S}_{j+1}^2)\,, \quad j = 0, 1, \ldots\,.
\end{aligned}$$

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

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

A comparison of this result with Theorem 2.1 shows that the prediction error increases by the additional term with the factor $\operatorname{cov}(X_{11},X_{12})$. It is non-negative as an application of the conditional Jensen inequality shows.

Proof. We start by calculating

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

We observe that

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

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

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

Therefore,

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

For the unconditional prediction error we have

$$\begin{aligned}
&E[(S_{j+1}-\widehat{S}_{j+1})^2]=EM\,\left[E(X_{11}^2)\,\mu\,p_{j+1}+\operatorname{cov}(X_{11},X_{12})\,(\mu\,p_{j+1})^2\right]\\
&+E(M^2)\,(EX_{11}\mu\,p_{j+1})^2-(EX_{11}\mu\,p_{j+1})^2E[(E(M\mid\mathcal{F}_j))^2]\\
&= EM\,\left[E(X_{11}^2)\,\mu\,p_{j+1}+\operatorname{cov}(X_{11},X_{12})\,(\mu\,p_{j+1})^2\right]
\end{aligned}$$

$$+E(M^2)\,(EX_{11}\mu\,p_{j+1})^2 - E(\widehat{S}^2_{j+1})\,.$$

□

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 $(-1)^l L^{(l)}(\gamma) = E(M^l \mathrm{e}^{-\gamma M})$. 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 [7], Section 3.3. This class is given by the recursive relation

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

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

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

Hence

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

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*

$$\begin{aligned} L(\gamma) &= \mathrm{e}^{-b\,(1-\mathrm{e}^{-\gamma})}\,, \quad \gamma\ge 0\,,\\ (-1)^l L^{(l)}(\gamma) &= \mathrm{e}^{-\gamma}\, b \sum_{r=0}^{l-1} \binom{l-1}{r} (-1)^r L^{(r)}(\gamma)\,, \quad l\ge 1\,. \end{aligned}$$

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

$$\begin{aligned} L(\gamma) &= (1-p(1-\mathrm{e}^{-\gamma}))^n\,, \quad \gamma\ge 0\,,\\ (-1)^l L^{(l)}(\gamma) &= \frac{p\,\mathrm{e}^{-\gamma}}{1-p(1-\mathrm{e}^{-\gamma})} \sum_{r=0}^{l-1} \binom{l}{r} \left[(n+1)\,\frac{l-r}{l}-1\right] (-1)^r L^{(r)}(\gamma)\,. \end{aligned}$$

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

$$\begin{aligned} L(\gamma) &= \left(\frac{p}{1-(1-p)\mathrm{e}^{-\gamma}}\right)^v\,, \quad \gamma\ge 0\,,\\ (-1)^l L^{(l)}(\gamma) &= \frac{(1-p)\mathrm{e}^{-\gamma}}{1-(1-p)\,\mathrm{e}^{-\gamma}} \sum_{r=0}^{l-1} \binom{l}{r} \left[(v-1)\,\frac{l-r}{l}+1\right] (-1)^r L^{(r)}(\gamma)\,. \end{aligned}$$

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

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

Then calculation yields

$$L^{(k)}(\gamma) = H_k(b\mathrm{e}^{-\gamma})\,L(\gamma)\,, \quad k\ge 0\,, \quad \gamma>0\,.$$

In particular, $R_k(\gamma)$ is a rational function of $b\mathrm{e}^{-\gamma}$ for each $k\ge 0$:

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

Remark 3.3. Extensions of the (a,b)-class were considered in Hess et al. [1]. They introduced distributions $(q_m)_{m=0,1,\ldots}$ satisfying the (a,b)-condition (3.1) with $q_0,\ldots,q_k=0$ for some $k\ge 0$ and $q_{k+1}>0$. 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 (3.2) in terms of the ratios $R_r(\gamma)$ which are relevant for the prediction formulae considered in the previous sections:

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

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 $k \to \infty$.**

In this subsection we study the asymptotic behavior of the predictors $E(S_{j+1} \mid N_0 = n_0, \ldots, N_j = n_j)$, $j = 0, 1, \ldots$, when the number of payments $N_0 + \cdots + N_j = n_0 + \cdots + n_j = k \to \infty$. The same discussion will apply equally to the multi-step predictors $E(S_{j+\ell+1} \mid N_0 = n_0, \ldots, N_j = n_j)$, $j, \ell = 0, 1, \ldots$. In view of the results in Theorem 2.1 one needs to study the asymptotic behavior of the ratios $R_k(\gamma)$ as $k \to \infty$.

The interest in the asymptotic behavior of $R_k(\gamma)$ as $k \to \infty$ is triggered, in particular, by a comparison with Mack's procedure (1.2). The latter *declares* the predictor of $N_{j+\ell+1}$ given $N_0, \ldots, N_j$ to be a linear function of $k = N_0 + \cdots + N_j$. 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*

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

exists. Then

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

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

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

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

$$\frac{(m+1)^k\, \mathrm{e}^{-(m+1)\gamma}}{m^k\, \mathrm{e}^{-m\,\gamma}} = \mathrm{e}^{-\gamma}\, \left(1 + m^{-1}\right)^k$$

$$\leq \mathrm{e}^{-\gamma}\, \left(\left(1 + m^{-1}\right)^m\right)^{\frac{\gamma+\tau}{1+\epsilon}} \leq \mathrm{e}^{-\gamma + (\gamma+\tau)/(1+\epsilon)}\,.$$

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

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

Combining these two bounds, we obtain

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

where we used the fact that $\delta \in (0, \epsilon/(1+\epsilon))$. Therefore for large k,

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

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

$$I_2(\epsilon) \ge q_{[k/(\gamma+\tau)]} \left[\frac{k}{\gamma+\tau}\right]^k \mathrm{e}^{-\gamma\,[k/(\gamma+\tau)]}\,. \tag{4.4}$$

A combination of (4.3) aud (4.4) yields

$$\begin{aligned}\frac{I_3(\epsilon)}{I_2(\epsilon)} \le \frac{1}{1-f(\epsilon,\delta)} & \frac{q_{\lceil k(1+\epsilon)/(\gamma+\tau)\rceil}}{q_{[k/(\gamma+\tau)]}} \Big(\frac{\lceil k(1+\epsilon)/(\gamma+\tau)\rceil}{[k/(\gamma+\tau)]}\Big)^k \\ & \mathrm{e}^{-[\lceil k(1+\epsilon)/(\gamma+\tau)\rceil - [k/(\gamma+\tau)]]\,\gamma}\,.\end{aligned}$$

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

$$\begin{aligned}\frac{q_{\lceil k(1+\epsilon)/(\gamma+\tau)\rceil}}{q_{[k/(\gamma+\tau)]}} &\le \mathrm{e}^{-\tau(1-\alpha)\,(\lceil k(1+\epsilon)/(\gamma+\tau)\rceil - [k/(\gamma+\tau)])} \\ &\le \mathrm{e}^{-\left(\tau/(\gamma+\tau)\right)\,k\,(1-\alpha)\,\epsilon}\,.\end{aligned}$$

Furthermore, for large k and some positive constant c_1,

$$\begin{aligned}\Big(\frac{\lceil k(1+\epsilon)/(\gamma+\tau)\rceil}{[k/(\gamma+\tau)]}\Big)^k &\le \Big(\frac{k(1+\epsilon)/(\gamma+\tau)+1}{k/(\gamma+\tau)-1}\Big)^k \\ = \quad & (1+\epsilon)^k\Big(1+\frac{1}{k(1+\epsilon)/(\gamma+\tau)}\Big)^k\Big(1-\frac{1}{k/(\gamma+\tau)}\Big)^{-k} \\ \le \quad & c_1\,(1+\epsilon)^k\,.\end{aligned}$$

Finally,

$$\mathrm{e}^{-[\lceil k(1+\epsilon)/(\gamma+\tau)\rceil - [k/(\gamma+\tau)]]\,\gamma} \le \mathrm{e}^{-\gamma\,k\,\epsilon/(\gamma+\tau)}\,.$$

Collecting the above bounds and choosing α such that

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

we obtain, for some positive constant c_2,

$$\frac{I_3(\epsilon)}{I_2(\epsilon)} \le c_2(1+\epsilon)^k\,\mathrm{e}^{-k\,\epsilon\,(\tau(1-\alpha)+\gamma)/(\gamma+\tau)} \le c_2\left(\frac{1}{1+\alpha}\right)^k\,.$$

Hence $I_3(\epsilon) = o(I_2(\epsilon))$ as $k \to \infty$.

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

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

We further decompose $I_1(\epsilon)$:

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

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

$$\left(1+m^{-1}\right)^k \geq \mathrm{e}^{(k/m)(1-\delta)} \geq \mathrm{e}^{(\gamma+\tau)(1-\delta)/(1-\epsilon)} ,$$

and, therefore,

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

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

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

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

We conclude that for $\epsilon \in (0,1)$, as $k \to \infty$

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

and, hence, relation (4.2) is immediate. This concludes the proof. □

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 $N_0,\ldots,N_j$ is asymptotically linear.

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

$$E(S_{j+1} \mid N_0 = n_0,\ldots,N_j = n_j) \sim \frac{EX_{11}\,\mu\, p_{j+1}}{\tau+\theta_j}\,(n_0+\cdots+n_j) .$$

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 $\mathrm{e}^{-\tau} = 1-p$. Hence Corollary 4.2 applies. The asymptotically linear behavior of $R_k(\gamma)$ is nicely illustrated in the right graph of Figure 4.3.

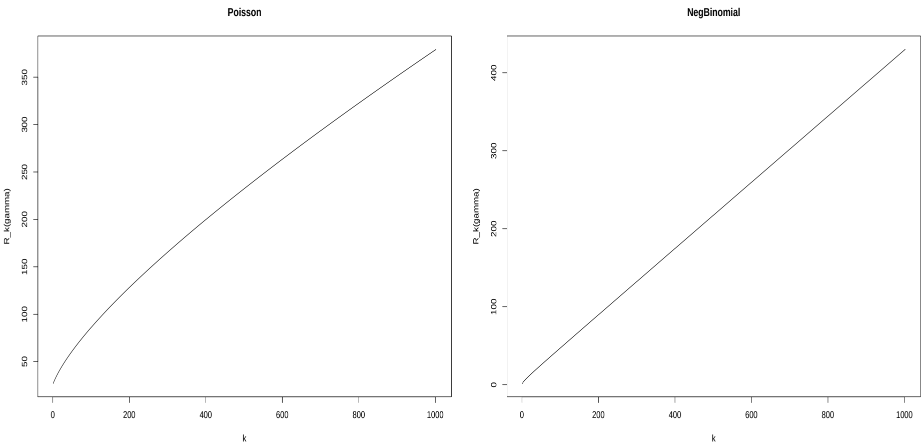

Figure 4.3. *The ratio* $R_k(2)$ *for* Poisson(200) *distributed* M (left) *and for negative binomial* M *with parameters* $p = 0.1$, $v = 12.1$ (right).

4.1.2. *The binomial distribution.* In this case, it is clear that, as $k \to \infty$,

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

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

$$E(S_{j+1} \mid N_0 = n_0, \ldots, N_j = n_j) \to n \, EX_{11} \, \mu \, p_{j+1} \,.$$

4.1.3. *The Poisson distribution.* If M is Poisson(b) distributed, then

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

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 $k \to \infty$, of such ratios. For simplicity, we use the notation EM^{k+1}/EM^k instead of the proper $E(M')^{k+1}/E(M')^k$.

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* $k \to \infty$,

$$\frac{EM^{k+1}}{EM^k} \sim \frac{k}{\log k} \,. \tag{4.6}$$

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

$$EM^k \quad = \quad \Big(\sum_{m < \frac{k\,(1-\epsilon)}{\log k}} + \sum_{m \in \left[\frac{k\,(1-\epsilon)}{\log k}, \frac{k\,(1+\epsilon)}{\log k}\right]} + \sum_{m > \frac{k\,(1+\epsilon)}{\log k}} \Big) m^k \, \mathrm{e}^{-\lambda} \, \frac{\lambda^m}{m!}$$

$$= \quad I_1(\epsilon) + I_2(\epsilon) + I_3(\epsilon)\,.$$

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

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

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

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

where

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

It is clear that for large k,

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

Therefore, as $k \to \infty$,

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

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

$$\frac{I_3(\epsilon)}{I_2(\epsilon)} \to 0 \;\; \text{as } k \to \infty.$$

Similarly, as $k \to \infty$,

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

where

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

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

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

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

$$\frac{I_1(\epsilon)}{I_2(\epsilon)} \to 0 \;\; \text{as } k \to \infty.$$

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

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

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

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

Hence the statement of the lemma. □

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

$$E(S_{j+1}\mid N_0=n_0,\ldots,N_j=n_j)\sim EX_{11}\,\mu\,p_{j+1}\frac{n_0+\cdots+n_j}{\log\big(n_0+\cdots+n_j\big)}\,.$$

The asymptotic behavior of $R_k(\gamma)$ prescribed by (4.6) is nicely illustrated in the left graph of Figure 4.3.

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 4.5. Below we give a limit theorem explaining this phenomenon.

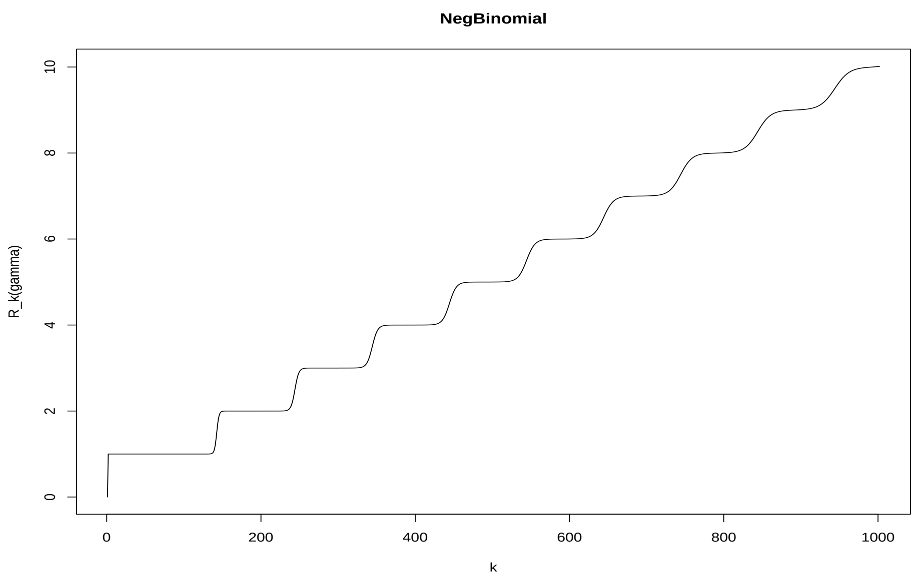

Figure 4.5. *The ratio $R_k(100)$ for negative binomial M with parameters $p=0.1$ and $v=12.1$.*

Assume that

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

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

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

which reaches its minimum at $s = t$. Therefore the sequence $(h_t(m))_{m=1,2,\ldots}$ reaches its minimum either at $m = m_-(t) = \lfloor t \rfloor$ if $m_-(t) \ge 1$ 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 $k \ge 1$,

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

By virtue of (4.7),

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

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,

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

Therefore for γ sufficiently large,

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

It now follows from (4.8)–(4.10) that

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

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

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

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

$$a_j = \inf\{t \ge 0 : m_*(t) = j\} = \frac{1}{\log j - \log(j-1)}.$$

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

Proposition 4.6. *For $j = 1, 2, \ldots$ let $t \in (a_j, a_{j+1})$ be such that $q_{m_*}(t) \ne 0$. If k and γ grow according to (4.7), then*

$$R_k(\gamma) \to j\,.$$

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 4.5.

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 contains 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, 7,302 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 (1.3). These data are needed for the construction of the *chain ladder estimators* of the factors f_j in (1.2):

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

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 5.1 contains the monthly claim arrival numbers M_j, $j = 1, \ldots, 12$, showing a clear seasonality in the data.

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
$100 \times p_j$	2.41	11.36	10.37	7.95	6.49	5.26	4.30	4.04	3.00	2.94	2.60	2.09

Table 5.1. *The monthly claim numbers M_j and the estimated probabilities p_j, $j = 1, \ldots, 12$.*

Table 5.2 contains the payment numbers in run-off triangle form. The ith row contains the payment numbers $N_{ii}, \ldots, N_{i,12}$ for claims arriving in the ith month and whose payments are executed in month $i + k \in \{i, \ldots, 12\}$. These data are supposed to be known (observed). For our data set, we also know the "future" monthly payment numbers $N_{i,12+k}$, $k \geq 1$, which we want to predict. They are presented in Table 5.3.

In our model, we assume that the monthly claim numbers M_j, $j = 1, \ldots, 12$, are Poisson distributed. Since there is a clear seasonality in the data we do not assume the M_j's identically distributed and simply take the

Month	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

Table 5.2. *The run-off triangle with the number of payments arising from 12 months of claim arrivals. The ith row contains the observed payment numbers* $N_{ii}, \dots, N_{i,12}$.

Month	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

Table 5.3. *The "future" payment numbers corresponding to the observations in Table 5.2. The ith row contains the monthly payment numbers* $N_{i,13}, \dots, N_{i,12+i-1}$. *These numbers have to be predicted.*

M_j's themselves as surrogates of their expectations EM_j. The Poisson assumption on the 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. We estimated both the average Poisson number of payments per claim, $\mu = 3.37$, and the distribution (p_j) from the empirical distribution of the data. The observed values M_j and estimated probabilities p_j, $j = 1, \dots, 12$, are given in Table 5.1.

In Table 5.4 we show the results of our prediction procedure. Since we know the values to be predicted (Table 5.3) we calculated the relative prediction error. The ith row in Table 5.4 shows the relative error in our prediction procedure for the values $N_{i,12+k}$, $k = 1, \dots, i-1$, based on the values $N_{ii}, \dots, N_{i,12}$. This prediction triangle has been chosen because it is standard to calculate it in this way for the chain ladder method. The relative errors fluctuate wildly both in the negative and in the positive directions. Reasons for these deviations are problems in choosing the right distributions for (q_m), (p_j), but also the statistical uncertainty when calibrating the model from a single data set. In addition, one may not expect miracles from mean

Month	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

Table 5.4. *Relative prediction error in % for our method. These values are calculated by usig the observations given in Tables 5.2 and 5.3.*

Month	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

Table 5.5. *Relative prediction errors in % for the chain ladder method. These values are calculated by usig the observations given in Tables 5.2 and 5.3.*

square predictions in the presence of distributions which are far away from the Gaussian.

In the left column of Figure 5.6 we show the performance of our prediction procedure in two different situations. In the top graph we are in the situation of the 11th line of Tables 5.2 and 5.3. 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 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 7th line in Tables 5.2 and 5.3. We predict the payment numbers arising from claims in the 7th month. Six payment have already been observed from which the payment numbers for months 13-18 get predicted.

Our predictions are compared with those prescribed by the chain ladder method under Mack's condition (1.2). Using the previous notation and the chain ladder estimator given in (5.1), the chain ladder predictors $\widetilde{N}_{i,12+k}$ of

$N_{i,12+k}$ are then given by

$$\widetilde{N}_{i,12+k} = (\widehat{f}_{12-i+k-1} - 1)\widehat{f}_{12-i+k-2} \cdots \widehat{f}_{12-i}\,(N_{ii} + \cdots + N_{i,12})\,,\ k = 1, 2, \ldots .$$

The form of these predictors is suggested by multiple use of Mack's condition (1.2):

$$\begin{aligned} & E(N_{i,12+k} \mid N_{ii}, \ldots, N_{i,12}) \\ &\quad = E\big(E(N_{i,12+k} \mid N_{ii}, \ldots, N_{i,12+k-1}) \mid N_{ii}, \ldots, N_{i,12}\big) \\ &\quad = (f_{12-i+k-1} - 1)\, E(N_{ii} + \cdots + N_{i,12+k-1} \mid N_{ii}, \ldots, N_{i,12}) \\ &\quad = (f_{12-i+k-1} - 1) f_{12-i+k-2} E(N_{ii} + \cdots + N_{i,12+k-2} \mid N_{ii}, \ldots, N_{i,12}) \\ &\quad = (f_{12-i+k-1} - 1) f_{12-i+k-2} \cdots f_{12-i}(N_{ii} + \cdots + N_{i,12})\,. \end{aligned}$$

The quantity $\widetilde{N}_{i,12+k}$ is then obtained by replacing the f-factors by their $\widehat{f}$-estimators.

In Table 5.5 the relative prediction errors of the chain ladder predictors $\widetilde{N}_{i,12+k}$, $k = 1, \ldots, 12 - i + 1$, are calculated from Tables 5.2 and 5.3. Again, the relative predicition errors fluctuate wildly in the positive and negative directions. A comparison with Table 5.4 shows that none of the two methods seems to outperform the other one, with the exception of the last row in Table 5.5 which is far from the true payment values. It is even impossible to say in which regions of the Tables 5.4 or 5.5 one or the other method has some advantages.

In the right column of Figure 5.6 we show the chain ladder prediction for the same situations as in the left column. Again, a direct comparison seems difficult although the error bands in the chain ladder case seem to be larger than for our method. In the right column, the bands around the predictions indicate ± 1.96 times the square of the conditional prediction error given at the bottom of p. 363 in Mack [5].

If we assume M_j Poisson distributed, Lemma 4.4 shows that the main assumption of the chain ladder approach, i.e., linearity of the conditional expectation in (1.2), is not satisfied even in an asymptotic sense. Given the fact that only 12 months of claim arrivals and the corresponding payment streams were available, the hypothesis about the Poisson distribution of M_j cannot be verified. Despite all these shortcomings, our study of real-life data clearly shows that our prediction method is not worse than the chain ladder prediction as regards closeness to the real-life payment numbers and magnitude 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 distributions (q_m), (p_j) and the Poisson parameter μ. These quantities are easy to estimate *if* data from different periods are available.

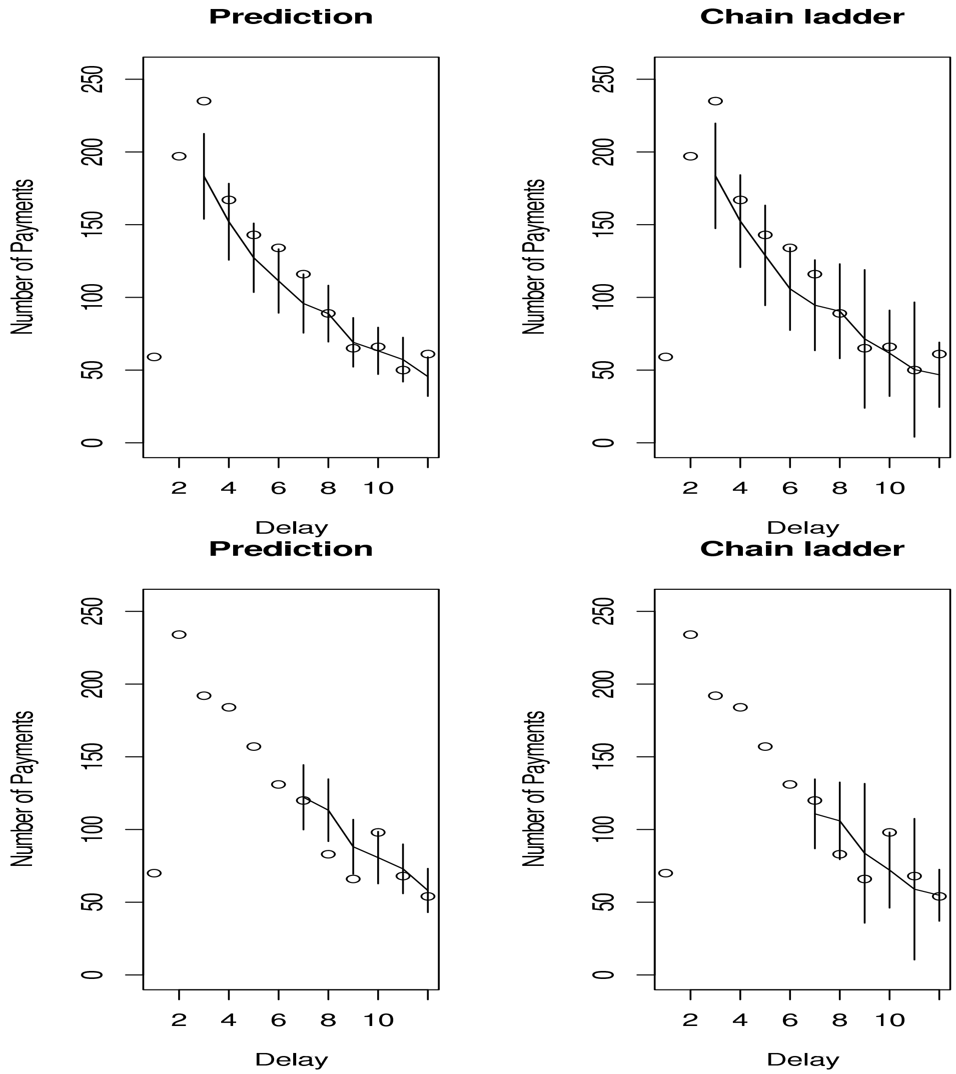

Figure 5.6. *Predictions* (solid ines) *of monthly payment numbers by using the theory in Section 2.2* (left graphs) *and the chain ladder approach* (right graphs)*. The dots represent the payment numbers of the real-life data. The bands around the predictions represent* ± 1.96 *times the square of the conditional prediction errors. The top graphs correspond to the* 11*th row in Tables 5.2 and 5.3, the bottom ones to the* 7*th row.*

The theoretical and empirical results of this paper show that it is worthwhile considering a stochastic process for modeling the dynamics in a non-life insurance portfolio. Even if one does not have enough information about all ingredients of the model, our stochastic model allows one to get exact predictions of future payments. Moreover, this model can easily be simulated.

References

[1] Hess, K.T., Liewald, A. and Schmidt, K.D. (2002) An extension of Panjer's recursion. *ASTIN Bulletin* **32**, 283–297.

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

[3] Mack, T. (1994) Which stochastic model is underlying the chain ladder method? *Insurance: Mathematics & Economics* **15**, 133–138.

[4] Mack, T. (1994) Measuring the variability of chain ladder reserve estimates. In: *Proceedings of the Casualty Actuarial Society Forum, Spring 1994,* Vol. **1**, pp. 101–182.

[5] Mack, T. (1999) The standard error of chain ladder reserve estimates: recursive calculation and inclusion of a tail factor. *ASTIN Bulletin* **29**, 361-366.

[6] Mack, T., Quarg, G. and Braun, C. (2006) The mean square error of prediction in the chain ladder reserving method. A comment. *ASTIN Bulletin* **36**, 543–553.

[7] Mikosch, T. (2004) *Non-Life Insurance Mathematics. An Introduction with Stochastic Processes.* Springer, Heidelberg.

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

E-mail address: ahj@math.ku.dk

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

E-mail address: mikosch@math.ku.dk

School of Operations Research and Information Engineering, Cornell University, 220 Rhodes Hall, Ithaca, NY 14853, U.S.A.

E-mail address: gennady@orie.cornell.edu

PUBLICATIONS DE L'INSTITUT MATHÉMATIQUE
Nouvelle série, tome 80(94) (2006), 171–192 DOI:10.2298/PIM0694171H

REGULARLY VARYING FUNCTIONS

Anders Hedegaard Jessen and Thomas Mikosch

In memoriam Tatjana Ostrogorski.

Abstract. We consider some elementary functions of the components of a regularly varying random vector such as linear combinations, products, minima, maxima, order statistics, powers. We give conditions under which these functions are again regularly varying, possibly with a different index.

1. Introduction

Regular variation is one of the basic concepts which appears in a natural way in different contexts of applied probability theory. Feller's [**21**] monograph has certainly contributed to the propagation of regular variation in the context of limit theory for sums of iid random variables. Resnick [**50, 51, 52**] popularized the notion of multivariate regular variation for multivariate extreme value theory. Bingham et al. [**3**] is an encyclopedia where one finds many analytical results related to one-dimensional regular variation. Kesten [**28**] and Goldie [**22**] studied regular variation of the stationary solution to a stochastic recurrence equation. These results find natural applications in financial time series analysis, see Basrak et al. [**2**] and Mikosch [**39**]. Recently, regular variation has become one of the key notions for modelling the behavior of large telecommunications networks, see e.g. Leland et al. [**35**], Heath et al. [**23**], Mikosch et al. [**40**].

It is the aim of this paper to review some known results on basic functions acting on regularly varying random variables and random vectors such as sums, products, linear combinations, maxima and minima, and powers. These results are often useful in applications related to time series analysis, risk management, insurance and telecommunications. Most of the results belong to the folklore but they are often wide spread over the literature and not always easily accessible. We will give references whenever we are aware of a proved result and give proofs if this is not the case.

2000 *Mathematics Subject Classification*: Primary 60G70; Secondary 62P05.

Mikosch's research is also partially supported by the Danish Research Council (SNF) Grant 21-04-0400. Jessen's research is partly supported by a grant from CODAN Insurance.

We focus on functions of finitely many regularly varying random variables. With a few exceptions (the tail of the marginal distribution of a linear process, functionals with a random index) we will *not* consider results where an increasing or an infinite number of such random variables or vectors is involved. We exclude distributional limit results e.g. for partial sums and maxima of iid and strictly stationary sequences, tail probabilities of subadditive functionals acting on a regularly varying random walk (e.g. ruin probabilities) and heavy-tailed large deviation results, tails of solutions to stochastic recurrence equations.

We start by introducing the notion of a multivariate regularly varying vector in Section 2. Then we will consider sum-type functionals of regularly varying vectors in Section 3. Functionals of product-type are considered in Section 4. In Section 5 we finally study order statistics and powers.

2. Regularly varying random vectors

In what follows, we will often need the notion of a regularly varying random vector and its properties; we refer to Resnick [**50**] and [**51**, Section 5.4.2]. This notion was further developed by Tatjana Ostrogorski in a series of papers, see [**42, 43, 44, 45, 46, 47**].

DEFINITION 2.1. An $\mathbb{R}^d$-valued random vector $\mathbf{X}$ and its distribution are said to be *regularly varying with limiting non-null Radon measure* μ on the Borel σ-field $\mathcal{B}(\overline{\mathbb{R}}_0^d)$ of $\overline{\mathbb{R}}_0^d = \overline{\mathbb{R}}^d \setminus \{\mathbf{0}\}$ if

$$\frac{P(x^{-1}\mathbf{X} \in \cdot)}{P(|\mathbf{X}| > x)} \xrightarrow{v} \mu\,, \quad x \to \infty\,. \tag{2.1}$$

Here $|\cdot|$ is any norm in $\mathbb{R}^d$ and $\xrightarrow{v}$ refers to vague convergence on $\mathcal{B}(\overline{\mathbb{R}}_0^d)$.

Since μ necessarily has the property $\mu(t\,A) = t^{-\alpha}\mu(A)$, $t > 0$, for some $\alpha > 0$ and all Borel sets A in $\overline{\mathbb{R}}_0^d$, we say that $\mathbf{X}$ is *regularly varying with index* α *and limiting measure* μ, for short $\mathbf{X} \in \mathrm{RV}(\alpha, \mu)$. If the limit measure μ is irrelevant we also write $\mathbf{X} \in \mathrm{RV}(\alpha)$. Relation (2.1) is often used in different equivalent disguises. It is equivalent to the *sequential definition of regular variation*: there exist $c_n \to \infty$ such that $n\,P(c_n^{-1}\,\mathbf{X} \in \cdot) \xrightarrow{v} \mu$. One can always choose (c_n) increasing and such that $n\,P(|\mathbf{X}| > c_n) \sim 1$. Another aspect of regular variation can be seen if one switches in (2.1) to a polar coordinate representation. Writing $\widetilde{\mathbf{x}} = \mathbf{x}/|\mathbf{x}|$ for any $\mathbf{x} \neq \mathbf{0}$ and $\mathbb{S}^{d-1} = \{\mathbf{x} \in \mathbb{R}^d : |\mathbf{x}| = 1\}$ for the unit sphere in $\mathbb{R}^d$, relation (2.1) is equivalent to

$$\frac{P(|\mathbf{X}| > x\,t\,, \widetilde{\mathbf{X}} \in \cdot)}{P(|\mathbf{X}| > x)} \xrightarrow{w} t^{-\alpha}\,P(\boldsymbol{\Theta} \in \cdot) \quad \text{for all } t > 0, \tag{2.2}$$

where $\boldsymbol{\Theta}$ is a random vector assuming values in $\mathbb{S}^{d-1}$ and $\xrightarrow{w}$ refers to weak convergence on the Borel σ-field of $\mathbb{S}^{d-1}$.

Plugging the set $\mathbb{S}^{d-1}$ into (2.2), it is straightforward that the norm $|\mathbf{X}|$ is regularly varying with index α.

The special case $d = 1$ refers to a regularly varying random variable X with index $\alpha \geqslant 0$:

$$P(X > x) \sim p\, x^{-\alpha}\, L(x) \quad \text{and} \quad P(X \leqslant -x) \sim q\, x^{-\alpha}\, L(x), \;\; p + q = 1, \tag{2.3}$$

where L is a *slowly varying function*, i.e., $L(cx)/L(x) \to 1$ as $x \to \infty$ for every $c > 0$. Condition (2.3) is also referred to as a *tail balance condition*. The cases $p = 0$ or $q = 0$ are not excluded. Here and in what follows we write $f(x) \sim g(x)$ as $x \to \infty$ if $f(x)/g(x) \to 1$ or, if $g(x) = 0$, we interpret this asymptotic relation as $f(x) = o(1)$.

3. Sum-type functions

3.1. Partial sums of random variables. Consider regularly varying random variables $X_1, X_2, \ldots$, possibly with different indices. We write

$$S_n = X_1 + \cdots + X_n, \quad n \geqslant 1,$$

for the partial sums. In what follows, we write $\overline{G} = 1 - G$ for the right tail of a distribution function G on $\mathbb{R}$.

LEMMA 3.1. *Assume $|X_1|$ is regularly varying with index $\alpha \geqslant 0$ and distribution function F. Assume $X_1, \ldots, X_n$ are random variables satisfying*

$$\lim_{x\to\infty} \frac{P(X_i > x)}{\overline{F}(x)} = c_i^+ \quad \textit{and} \quad \lim_{x\to\infty} \frac{P(X_i \leqslant -x)}{\overline{F}(x)} = c_i^-, \quad i = 1, \ldots, n\,, \tag{3.1}$$

for some non-negative numbers $c_i^\pm$ and

$$\begin{aligned} \lim_{x\to\infty} \frac{P(X_i > x\,, X_j > x)}{\overline{F}(x)} &= \lim_{x\to\infty} \frac{P(X_i \leqslant -x\,, X_j > x)}{\overline{F}(x)} \\ &= \lim_{x\to\infty} \frac{P(X_i \leqslant -x\,, X_j \leqslant -x)}{\overline{F}(x)} = 0\,, \quad i \neq j\,. \end{aligned} \tag{3.2}$$

Then

$$\lim_{x\to\infty} \frac{P(S_n > x)}{\overline{F}(x)} = c_1^+ + \cdots + c_n^+ \quad \textit{and} \quad \lim_{x\to\infty} \frac{P(S_n \leqslant -x)}{\overline{F}(x)} = c_1^- + \cdots + c_n^-\,.$$

In particular, if the X_i's are independent non-negative regularly varying random variables then

$$P(S_n > x) \sim P(X_1 > x) + \cdots + P(X_n > x)\,. \tag{3.3}$$

The proof of (3.3) can be found in Feller [**21**, p. 278], cf. Embrechts et al. [**18**, Lemma 1.3.1]. The general case of possibly dependent non-negative X_i's was proved in Davis and Resnick [**14**, Lemma 2.1]; the extension to general X_i's follows along the lines of the proof in [**14**]. Generalizations to the multivariate case are given in Section 3.6 below.

The conditions in Lemma 3.1 are sharp in the sense that they cannot be substantially improved. A condition like (3.1) with not all $c_i^\pm$'s vanishing is needed in order to ensure that at least one summand X_i is regularly varying. Condition (3.2) is a so-called *asymptotic independence condition*. It cannot be avoided as the trivial example $X_2 = -X_1$ for a regularly varying X_1 shows. Then (3.1) holds but (3.2) does not and $S_2 = 0$ a.s.

A partial converse follows from Embrechts et al. [**17**].

LEMMA 3.2. *Assume $S_n = X_1 + \cdots + X_n$ is regularly varying with index $\alpha \geqslant 0$ and X_i are iid non-negative. Then the X_i's are regularly varying with index α and*

$$P(S_n > x) \sim n\, P(X_1 > x)\,, \quad n \geqslant 1\,. \tag{3.4}$$

Relation (3.4) can be taken as the definition of a subexponential distribution. The class of those distributions is larger than the class of regularly varying distributions, see Embrechts et al. [**18**, Sections 1.3, 1.4 and Appendix A3]. Lemma 3.2 remains valid for subexponential distributions in the sense that subexponentiality of S_n implies subexponentiality of X_1. This property is referred to as *convolution root closure* of subexponential distributions.

PROOF. Since S_n is regularly varying it is subexponential. Then the regular variation of X_i follows from the convolution root closure of subexponential distributions, see Proposition A3.18 in Embrechts et al. [**18**]. Relation (3.4) is a consequence of (3.3). □

An alternative proof is presented in the proof of Proposition 4.8 in Faÿ et al. [**20**]. It strongly depends on the regular variation of the X_i's: Karamata's Tauberian theorem (see Feller [**21**, XIII, Section 5]) is used.

In general, one cannot conclude from regular variation of $X+Y$ for independent X and Y that X and Y are regularly varying. For example, if $X+Y$ has a Cauchy distribution, in particular $X+Y \in \mathrm{RV}(1)$, then X can be chosen Poisson, see Theorem 6.3.1 on p. 71 in Lukacs [**37**]. It follows from Lemma 3.12 below that $Y \in \mathrm{RV}(1)$.

3.2. Weighted sums of iid regularly varying random variables. We assume that (Z_i) is an iid sequence of regularly varying random variables with index $\alpha \geqslant 0$ and tail balance condition (2.3) (applied to $X = Z_i$). Then it follows from Lemma 3.1 that for any real constants ψ_i

$$P(\psi_1 Z_1 + \cdots + \psi_m Z_m > x) \sim P(\psi_1 Z_1 > x) + \cdots + P(\psi_m Z_1 > x)\,.$$

Then evaluating $P(\psi_i Z_1 > x) = P(\psi_i^+ Z_i^+ > x) + P(\psi_i^- Z_i^- > x)$, where $x^\pm = 0 \vee (\pm x)$ we conclude the following result which can be found in various books, e.g. Embrechts et al. [**18**, Lemma A3.26].

LEMMA 3.3. *Let (Z_i) be an iid sequence of regularly varying random variables satisfying the tail balance condition* (2.3). *Then for any real constants ψ_i and $m \geqslant 1$,*

$$P(\psi_1 Z_1 + \cdots + \psi_m Z_m > x) \sim P(|Z_1| > x) \sum_{i=1}^m \left[p\,(\psi_i^+)^\alpha + q\,(\psi_i^-)^\alpha\right]. \tag{3.5}$$

The converse of Lemma 3.3 is in general incorrect, i.e., regular variation of $\psi_1 Z_1 + \cdots + \psi_m Z_m$ with index $\alpha > 0$ for an iid sequence (Z_i) does in general not imply regular variation of Z_1, an exception being the case $m = 2$ with $\psi_i > 0$, $Z_i \geqslant 0$ a.s., $i = 1, 2$, cf. Jacobsen et al. [**27**].

3.3. Infinite series of weighted iid regularly varying random variables. The question about the tail behavior of an infinite series

$$X = \sum_{i=0}^{\infty} \psi_j Z_j \tag{3.6}$$

for an iid sequence (Z_i) of regularly varying random variables with index $\alpha > 0$ and real weights occurs for example in the context of extreme value theory for linear processes, including ARMA and FARIMA processes, see Davis and Resnick [**11, 12, 13**], Klüppelberg and Mikosch [**29, 30, 31**], cf. Brockwell and Davis [**5**, Section 13.3], Resnick [**51**, Section 4.5], Embrechts et al. [**18**, Section 5.5 and Chapter 7].

The problem about the regular variation of X is only reasonably posed if the infinite series (3.6) converges a.s. Necessary and sufficient conditions are given by Kolmogorov's 3-series theorem, cf. Petrov [**48, 49**]. For example, if $\alpha > 2$ (then $\mathrm{var}(Z_i) < \infty$), the conditions $E(Z_1) = 0$ and $\sum_i \psi_i^2 < \infty$ are necessary and sufficient for the a.s. convergence of X.

The following conditions from Mikosch and Samorodnitsky [**41**] are best possible in the sense that the conditions on (ψ_i) coincide with or are close to the conditions in the 3-series theorem. Similar results, partly under stronger conditions, can be found in Lemma 4.24 of Resnick [**51**] for $\alpha \leqslant 1$ (attributed to Cline [**7, 8**]), Theorem 2.2 in Kokoszka and Taqqu [**32**] for $\alpha \in (1, 2)$.

LEMMA 3.4. *Let (Z_i) be an iid sequence of regularly varying random variables with index $\alpha > 0$ which satisfy the tail balance condition* (2.3). *Let (ψ_i) be a sequence of real weights. Assume that one of the following conditions holds:*

(1) $\alpha > 2$, $E(Z_1) = 0$ *and* $\sum_{i=0}^{\infty} \psi_i^2 < \infty$.
(2) $\alpha \in (1, 2]$, $E(Z_1) = 0$ *and* $\sum_{i=0}^{\infty} |\psi_i|^{\alpha-\varepsilon} < \infty$ *for some* $\varepsilon > 0$.
(3) $\alpha \in (0, 1]$ *and* $\sum_{i=0}^{\infty} |\psi_i|^{\alpha-\varepsilon} < \infty$ *for some* $\varepsilon > 0$.

Then

$$P(X > x) \sim P(|Z_1| > x) \sum_{i=0}^{\infty} \left[p\,(\psi_i^+)^\alpha + q\,(\psi_i^-)^\alpha \right] . \tag{3.7}$$

The conditions on (ψ_j) in the case $\alpha \in (0, 2]$ can be slightly relaxed if one knows more about the slowly varying L. In this case the following result from Mikosch and Samorodnitsky [**41**] holds.

LEMMA 3.5. *Let (Z_i) be an iid sequence of regularly varying random variables with index $\alpha \in (0, 2]$ which satisfy the tail balance condition* (2.3). *Assume that $\sum_{i=1}^{\infty} |\psi_i|^\alpha < \infty$, that the infinite series* (3.6) *converges a.s. and that one of the following conditions holds:*

(1) *There exist constants $c, x_0 > 0$ such that $L(x_2) \leqslant c\,L(x_1)$ for all $x_0 < x_1 < x_2$.*

(2) *There exist constants* $c, x_0 > 0$ *such that* $L(x_1\, x_2) \leqslant c\, L(x_1)\, L(x_2)$ *for all* $x_1, x_2 \geqslant x_0 > 0$

Then (3.7) *holds.*

Condition (2) holds for Pareto-like tails $P(Z_1 > x) \sim c\, x^{-\alpha}$, in particular for α-stable random variables Z_i and for student distributed Z_i's with α degrees of freedom. It is also satisfied for $L(x) = (\log_k x)^\beta$, any real β, where $\log_k$ is the kth time iterated logarithm.

Classical time series analysis deals with the strictly stationary linear processes

$$X_n = \sum_{i=0}^{\infty} \psi_i\, Z_{n-i}\,, \quad n \in \mathbb{Z}\,,$$

where (Z_i) is an iid white noise sequence, cf. Brockwell and Davis [**5**]. In the case of regularly varying Z_i's with $\alpha > 2$, $\operatorname{var}(Z_1)$ and $\operatorname{var}(X_1)$ are finite and therefore it makes sense to define the autocovariance function $\gamma_X(h) = \operatorname{cov}(X_0, X_h) = \operatorname{var}(Z_1) \sum_i \psi_i \psi_{i+|h|}$, $h \in \mathbb{Z}$. The condition $\sum_i \psi_i^2 < \infty$ (which is necessary for the a.s. convergence of X_n) does not only capture *short range dependent* sequences (such as ARMA processes for which $\gamma_X(h)$ decays exponentially fast to zero) but also *long range dependent* sequences (X_n) in the sense that $\sum_h |\gamma_X(h)| = \infty$. Thus Lemma 3.4 also covers long range dependent sequences. The latter class includes fractional ARIMA processes; cf. Brockwell and Davis [**5**, Section 13.2], and Samorodnitsky and Taqqu [**56**].

Notice that (3.7) is the direct analog of (3.5) for the truncated series. The proof of (3.7) is based on (3.5) and the fact that the remainder term $\sum_{i=m+1}^{\infty} \psi_i\, Z_i$ is negligible compared to $P(|Z_1| > x)$ when first letting $x \to \infty$ and then $m \to \infty$. More generally, the following result holds:

LEMMA 3.6. *Let* A *be a random variable and let* Z *be positive regularly varying random variable with index* $\alpha \geqslant 0$. *Assume that for every* $m \geqslant 0$ *there exist finite positive constants* $c_m > 0$, *random variables* A_m *and* B_m *such that the representation* $A \stackrel{d}{=} A_m + B_m$ *holds and the following three conditions are satisfied:*

$$P(A_m > x) \sim c_m\, P(Z > x)\,, \quad \text{as } x \to \infty\,,$$
$$c_m \to c_0\,, \quad \text{as } m \to \infty\,,$$
$$\lim_{m\to\infty} \limsup_{x\to\infty} \frac{P(B_m > x)}{P(Z > x)} = 0 \quad \textit{and } A_m, B_m \textit{ are independent for every } m \geqslant 1 \text{ or}$$
$$\lim_{m\to\infty} \limsup_{x\to\infty} \frac{P(|B_m| > x)}{P(Z > x)} = 0\,.$$

Then $P(A > x) \sim c_0\, P(Z > x)$.

PROOF. For every $m \geqslant 1$ and $\varepsilon \in (0,1)$.

$$P(A > x) \leqslant P(A_m > x(1-\varepsilon)) + P(B_m > \varepsilon\, x)\,.$$

Hence

$$\begin{aligned}
\limsup_{x\to\infty}\frac{P(A>x)}{P(Z>x)} &\leqslant \limsup_{x\to\infty}\frac{P(A_m>x(1-\varepsilon))}{P(Z>x)}+\limsup_{x\to\infty}\frac{P(B_m>\varepsilon\,x)}{P(Z>x)}\\
&= c_m\,(1-\varepsilon)^{-\alpha}+\varepsilon^{-\alpha}\,\limsup_{x\to\infty}\frac{P(B_m>\varepsilon\,x)}{P(Z>\varepsilon\,x)}\\
&\to c_0\,(1-\varepsilon)^{-\alpha}\quad\text{as } m\to\infty\\
&\to c_0\quad\text{as } \varepsilon\downarrow 0.
\end{aligned}$$

Similarly, for independent A_m and B_m,

$$\begin{aligned}
\liminf_{x\to\infty}\frac{P(A>x)}{P(Z>x)} &\geqslant \liminf_{x\to\infty}\frac{P(A_m>x\,(1+\varepsilon)\,,B_m\geqslant-\varepsilon\,x)}{P(Z>x)}\\
&= \liminf_{x\to\infty}\frac{P(A_m>x\,(1+\varepsilon))}{P(Z>x)}\,P(B_m\geqslant-\varepsilon\,x)\\
&= c_m\,(1+\varepsilon)^{-\alpha}\to c_0\,,\quad\text{as } m\to\infty,\ \varepsilon\downarrow 0.
\end{aligned}$$

If A_m and B_m are not necessarily independent a similar bound yields

$$\begin{aligned}
\liminf_{x\to\infty}\frac{P(A>x)}{P(Z>x)} &\geqslant \liminf_{x\to\infty}\frac{P(A_m>x\,(1+\varepsilon)\,,|B_m|\leqslant\varepsilon\,x)}{P(Z>x)}\\
&\geqslant \liminf_{x\to\infty}\frac{P(A_m>x\,(1+\varepsilon))}{P(Z>x)}-\limsup_{x\to\infty}\frac{P(|B_m|>\varepsilon\,x)}{P(Z>x)}\\
&= c_m\,(1+\varepsilon)^{-\alpha}\to c_0\,,\quad\text{as } m\to\infty,\ \varepsilon\downarrow 0.
\end{aligned}$$

Combining the upper and lower bounds, we arrive at the desired result. □

We also mention that Resnick and Willekens [**53**] study the tails of the infinite series $\sum_i \mathbf{A}_i\,\mathbf{Z}_i$, where $(\mathbf{A}_i)$ is an iid sequence of random matrices, independent of the iid sequence $(\mathbf{Z}_i)$ of regularly varying vectors $\mathbf{Z}_i$.

3.4. Random sums. We consider an iid sequence (X_i) of non-negative random variables, independent of the integer-valued non-negative random variable K. Depending on the distributional tails of K and X_1, one gets rather different tail behavior for the random sum $S_K=\sum_{i=1}^K X_i$. The following results are taken from Faÿ et al. [**20**].

LEMMA 3.7. (1) *Assume X_1 is regularly varying with index $\alpha>0$, $EK<\infty$ and $P(K>x)=o(P(X_1>x))$. Then, as $x\to\infty$,*

$$P(S_K>x)\sim EK\,P(X_1>x)\,. \tag{3.8}$$

(2) *Assume K is regularly varying with index $\beta\geqslant 0$. If $\beta=1$, assume that $EK<\infty$. Moreover, let (X_i) be an iid sequence such that $E(X_1)<\infty$ and $P(X_1>x)=o(P(K>x))$. Then, as $x\to\infty$,*

$$P(S_K>x)\sim P(K>(E(X_1))^{-1}\,x)\sim(E(X_1))^{\beta}\,P(K>x)\,. \tag{3.9}$$

(3) *Assume S_K is regularly varying with index $\alpha>0$ and $E(K^{1\vee(\alpha+\delta)})<\infty$ for some positive δ. Then X_1 is regularly varying with index α and $P(S_K>x)\sim EK\,P(X_1>x)$.*

(4) *Assume S_K is regularly varying with index $\alpha > 0$. Suppose that $E(X_1) < \infty$ and $P(X_1 > x) = o(P(S_K > x))$ as $x \to \infty$. In the case $\alpha = 1$ and $E(S_K) = \infty$, assume that $x\,P(X_1 > x) = o(P(S_K > x))$ as $x \to \infty$. Then K is regularly varying with index α and*

$$P(S_K > x) \sim (E(X_1))^\alpha \, P(K > x)\,.$$

(5) *Assume $P(K > x) \sim c\,P(X_1 > x)$ for some $c > 0$, that X_1 is regularly varying with index $\alpha \geqslant 1$ and $E(X_1) < \infty$. Then*

$$P(S_K > x) \sim (EK + c\,(E(X_1))^\alpha)\,P(X_1 > x)\,.$$

Relations (3) and (4) are the partial converses of the corresponding relations (1) and (2). The law of large numbers stands behind the form of relation (3.9), whereas relation (3.8) is expected from the results in Section 3.1.

Relations of type (3.8) appear in a natural way in risk and queuing theory when the summands X_i are subexponential and K has a moment generating function in some neighborhood of the origin, see for example the proof of the Cramér-Lundberg ruin bound in Section 1.4 of Embrechts et al. [**18**].

For $\alpha \in (0,2)$ some of the results in Lemma 3.7 can already be found in Resnick [**50**] and even in the earlier papers by Stam [**57**], Embrechts and Omey [**19**]. The restriction to $\alpha < 2$ is due to the fact that some of the proofs depend on the equivalence between regular variation and membership in the domain of attraction of infinite variance stable distributions. Resnick [**50**] also extends some of his results to the case when K is a stopping time.

In the following example the assumptions of Lemma 3.7 are not necessarily satisfied. Assume (X_i) is a sequence of iid positive α-stable random variables for some $\alpha < 1$. Then $S_K \stackrel{d}{=} K^{1/\alpha} X_1$ and $P(X_1 > x) \sim c\,x^{-\alpha}$ for some $c > 0$; cf. Feller [**21**] or Samorodnitsky and Taqqu [**56**]. If $EK < \infty$ then Breiman's result (see Lemma 4.2 below) yields $P(S_K > x) \sim EK P(X > x)$ in agreement with (3.8). If $EK = \infty$ we have to consider different possibilities. If K is regularly varying with index 1, then $K^{1/\alpha} \in \mathrm{RV}(\alpha)$. Then we are in the situation of Lemma 4.2 below and S_K is regularly varying with index α. If we assume that $K \in \mathrm{RV}(\beta)$ for some $\beta < 1$, then $K^{1/\alpha} \in \mathrm{RV}(\beta\alpha)$ and the results of Lemma 4.2 ensure that $P(S_K > x) \sim E(X^{\alpha\beta})P(K^{1/\alpha} > x)$.

The latter result can be extended by using a Tauberian argument.

LEMMA 3.8. *Assume that $K, X_1 > 0$ are regularly varying with indices $\beta \in [0,1)$ and $\alpha \in [0,1)$, respectively. Then*

$$P(S_K > x) \sim P(K > [P(X > x)]^{-1}) \sim P(M_K > x)\,,$$

where $M_n = \max_{i=1,\dots,n} X_i$.

PROOF. By Karamata's Tauberian theorem (see Feller [**21**, XIII, Section 5]) $1 - E(\mathrm{e}^{-sK}) \sim s^\beta L_K(1/s)$ as $s \downarrow 0$ provided that $P(K > x) = x^{-\beta} L_K(x)$ for some slowly varying function L. In the same way, $1 - E(\mathrm{e}^{-tX_1}) \sim t^\alpha L_X(1/t)$ as $t \downarrow 0$.

Then

$$\begin{aligned} 1 - E(\mathrm{e}^{-t\,S_K}) &= 1 - E\left(\exp\left\{K\,\log\left(E\left(\mathrm{e}^{-t\,X_1}\right)\right)\right\}\right) \\ &\sim \left[-\log\left(E\left(\mathrm{e}^{-t\,X_1}\right)\right)\right]^{\beta}\, L_K(1/[-\log\left(E\left(\mathrm{e}^{-t\,X_1}\right)\right)]) \\ &\sim \left[1 - E\left(\mathrm{e}^{-t\,X_1}\right)\right]^{\beta} L_K\left(1/\left[1 - E\left(\mathrm{e}^{-t\,X_1}\right)\right]\right) \\ &\sim [t^{\alpha} L_X(1/t)]^{\beta} L_K([t^{\alpha} L_X(1/t)]^{-1}) \\ &= t^{\alpha\beta} L(1/t)\,, \end{aligned}$$

where $L(x) = L_X^{\beta}(x)\, L_K(x^{\alpha}/L_X(x))$ is a slowly varying function. Again by Karamata's Tauberian theorem, $P(S_K > x) \sim x^{-\alpha\,\beta}\, L(x)$. Notice that the right-hand side is equivalent to the tail $P(K > [P(X_1 > x)]^{-1}) \sim P(M_K > x)$. The latter equivalence follows from (5.1) below. □

3.5. Linear combinations of a regularly varying random vector. Assume $\mathbf{X} \in \mathrm{RV}(\alpha, \mu)$ and let $\mathbf{c} \in \mathbb{R}^d$, $\mathbf{c} \neq 0$, be a constant. The set $A_{\mathbf{c}} = \{\mathbf{x} : \mathbf{c}'\mathbf{x} > 1\}$ is bounded away from zero and $\mu(\partial A_{\mathbf{c}}) = 0$. Indeed, it follows from the scaling property of μ that $\mu(\{\mathbf{x} : \mathbf{c}'\mathbf{x} = y\}) = y^{-\alpha}\,\mu(\{\mathbf{x} : \mathbf{c}'\mathbf{x} = 1\})$, $y > 0$. If $\mu(\{\mathbf{x} : \mathbf{c}'\mathbf{x} = 1\}) > 0$ this would contradict the finiteness of $\mu(A_{\mathbf{c}})$.

Therefore, from (2.1),

$$\frac{P(x^{-1}\mathbf{X} \in A_{\mathbf{c}})}{P(|\mathbf{X}| > x)} = \frac{P(\mathbf{c}'\mathbf{X} > x)}{P(|\mathbf{X}| > x)} \to \mu(A_{\mathbf{c}})\,.$$

We conclude the following, see also Resnick [**52**], Section 7.3.

LEMMA 3.9. *For* $\mathbf{c} \in \mathbb{R}$, $\mathbf{c} \neq \mathbf{0}$, $\mathbf{c}'\mathbf{X}$ *is regularly varying with index* α *if* $\mu(A_{\mathbf{c}}) \neq 0$. *In general,*

$$\frac{P(\mathbf{c}'\mathbf{X} > x)}{P(|\mathbf{X}| > x)} \to \mu(\{\mathbf{x} : \mathbf{c}'\mathbf{x} > 1\})\,,$$

where the right-hand side possibly vanishes. In particular, if $\mu(\{\mathbf{x} : \mathbf{c}_i'\mathbf{x} > 1\}) > 0$ *for the basis vector* $\mathbf{c}_i = (0, \ldots, 0, 1, 0, \ldots, 0)'$ *with* 1 *in the ith component then* $(X_i)^+$ *is regularly varying with index* α.

A natural question arises: given that

$$\frac{P(\mathbf{c}'\mathbf{X} > x)}{L(x)x^{-\alpha}} = C(\mathbf{c}) \quad \text{for all } \mathbf{c} \neq \mathbf{0} \text{ and } C(\mathbf{c}) \neq 0 \text{ for at least one } \mathbf{c} \tag{3.10}$$

holds for some function C, is then $\mathbf{X}$ regularly varying in the sense of Definition 2.1? This would yield a Cramér–Wold device analog for regularly varying random vectors.

The answer to this question is not obvious. Here are some partial answers. The first three statements can be found in Basrak et al. [**1**], the last statements are due to Hult and Lindskog [**26**]. Statement (5) was already mentioned (without proof) in Kesten [**28**].

LEMMA 3.10. (1) (3.10) *implies that* $\mathbf{X}$ *is regularly varying with a unique spectral measure if* α *is not an integer.*

(2) (3.10) *when restricted to* $\mathbf{c} \in [0,\infty)^d \setminus \{\mathbf{0}\}$ *implies that* $\mathbf{X}$ *is regularly varying with a unique spectral measure if* $\mathbf{X}$ *has non-negative components and* α *is positive and not an integer,*

(3) (3.10) *implies that* $\mathbf{X}$ *is regularly varying with a unique spectral measure if* $\mathbf{X}$ *has non-negative components and* α *is an odd integer.*

(4) (1) *and* (2) *cannot be extended to integer* α *without additional assumptions on the distribution of* $\mathbf{X}$. *There exist regularly varying* $\mathbf{X}_1$ *and* $\mathbf{X}_2$ *both satisfying* (3.10) *with the same function* C *but having different spectral measures.*

(5) *For integer* $\alpha > 0$, *there exist non-regularly varying* $\mathbf{X}$ *satisfying* (3.10).

3.6. Multivariate extensions. In this section we assume that $\mathbf{X}_1$ and $\mathbf{X}_2$ are random vectors with values in $\mathbb{R}^d$. The following result due to Hult and Lindskog [**24**], see also Resnick [**52**, Section 7.3], yields an extension of Lemma 3.1 to regularly varying vectors.

LEMMA 3.11. *Assume that* $\mathbf{X}_1$ *and* $\mathbf{X}_2$ *are independent regularly varying such that* $n\,P(c_n^{-1}\mathbf{X}_i \in) \stackrel{v}{\to} \mu_i$, $i=1,2$, *for some sequence* $c_n \to \infty$ *and Radon measures* μ_i, $i=1,2$. *Then* $\mathbf{X}_1+\mathbf{X}_2$ *is regularly varying and* $n\,P(c_n^{-1}(\mathbf{X}_1+\mathbf{X}_2) \in \cdot) \stackrel{v}{\to} \mu_1+\mu_2$.

The following lemma is often useful.

LEMMA 3.12. *Assume* $\mathbf{X}_1 \in \mathrm{RV}(\alpha,\mu)$ *and* $P(|\mathbf{X}_2|>x) = o(P(|\mathbf{X}_1|>x))$ *as* $x \to \infty$. *Then* $\mathbf{X}_1+\mathbf{X}_2 \in \mathrm{RV}(\alpha,\mu)$.

PROOF. It suffices to show that

$$P(x^{-1}(\mathbf{X}_1+\mathbf{X}_2) \in A) \sim P(x^{-1}\mathbf{X}_1 \in A)\,, \tag{3.11}$$

where A is any rectangle in $\mathbb{R}^d$ bounded away from zero. The latter class of sets generates vague convergence in $\mathcal{B}(\overline{\mathbb{R}}_0^d)$ and satisfies $\mu(\partial A)=0$. Assume that $A = [\mathbf{a},\mathbf{b}] = \{\mathbf{x} : \mathbf{a} \leqslant \mathbf{x} \leqslant \mathbf{b}\}$ for two vectors $\mathbf{a} < \mathbf{b}$, where $<$, $\leqslant$ are defined in the natural componentwise way. Write $\mathbf{a}^{\pm\varepsilon} = (a_1 \pm \varepsilon, \cdots, a_d \pm \varepsilon)$ and define $\mathbf{b}^{\pm\varepsilon}$ correspondingly. Define the rectangles $A^{-\varepsilon} = [\mathbf{a}^{-\varepsilon},\mathbf{b}^{\varepsilon}]$ and $A^{\varepsilon} = [\mathbf{a}^{\varepsilon},\mathbf{b}^{-\varepsilon}]$ in the same way as A. For small ε these sets are not empty, bounded away from zero and $A^{\varepsilon} \subset A \subset A^{-\varepsilon}$.

For small $\varepsilon > 0$,

$$\begin{aligned}
&P(x^{-1}(\mathbf{X}_1+\mathbf{X}_2) \in A)\\
&\quad = P(x^{-1}(\mathbf{X}_1+\mathbf{X}_2) \in A\,, x^{-1}|\mathbf{X}_2| > \varepsilon) + P(x^{-1}(\mathbf{X}_1+\mathbf{X}_2) \in A\,, x^{-1}|\mathbf{X}_2| \leqslant \varepsilon)\\
&\quad \leqslant P(|\mathbf{X}_2| > x\,\varepsilon) + P(x^{-1}\mathbf{X}_1 \in A^{-\varepsilon})\,.
\end{aligned}$$

Then

$$\begin{aligned}
\limsup_{x\to\infty} \frac{P(x^{-1}(\mathbf{X}_1+\mathbf{X}_2) \in A)}{P(|\mathbf{X}_1|>x)} &\leqslant \limsup_{x\to\infty} \frac{P(|\mathbf{X}_2| > x\,\varepsilon)}{P(|\mathbf{X}_1|>x)} + \limsup_{x\to\infty} \frac{P(x^{-1}\mathbf{X}_1 \in A^{-\varepsilon})}{P(|\mathbf{X}_1|>x)}\\
&= \mu(A^{-\varepsilon}) \downarrow \mu(A) \quad \text{as } \varepsilon \downarrow 0.
\end{aligned}$$

In the last step we used that A is a μ-continuity set. Similarly,

$$P(x^{-1}(\mathbf{X}_1+\mathbf{X}_2)\in A)\geqslant P(x^{-1}\mathbf{X}_1\in A^{\varepsilon}\,,x^{-1}|\mathbf{X}_2|\leqslant\varepsilon)$$
$$\geqslant P(x^{-1}\mathbf{X}_1\in A^{\varepsilon})-P(|\mathbf{X}_2|>\varepsilon\,x)\,.$$

Then

$$\liminf_{x\to\infty}\frac{P(x^{-1}(\mathbf{X}_1+\mathbf{X}_2)\in A)}{P(|\mathbf{X}_1|>x)}\geqslant\liminf_{x\to\infty}\frac{P(x^{-1}\mathbf{X}_1\in A^{\varepsilon})}{P(|\mathbf{X}_1|>x)}=\mu(A^{\varepsilon})\uparrow\mu(A)\quad\text{as }\varepsilon\downarrow 0.$$

In the last step we again used that A is a μ-continuity set.

Now collecting the upper and lower bounds, we arrive at the desired relation (3.11). □

4. Product-type functions

Products are more complicated objects than sums. Their asymptotic tail behavior crucially depends on which tail of the factors in the product dominates. If the factors have similar tail behavior the results become more complicated.

Assume for the moment $d=2$. The set $A=\{\mathbf{x}:x_1\,x_2>1\}$ is bounded away from zero and therefore regular variation of $\mathbf{X}$ implies that the limit

$$\frac{P(X_1\,X_2>x^2)}{P(|\mathbf{X}|>x)}=\frac{P(x^{-1}(X_1,X_2)\in A)}{P(|\mathbf{X}|>x)}\to\mu(A)$$

exists. However, the quantity $\mu(A)$ can be rather non-informative, for example, in the two extreme cases: $\mathbf{X}=(X,X)$ for a non-negative regularly varying X with index α and $\mathbf{X}=(X_1,X_2)$, where X_1 and X_2 are independent copies of X. In the former case, with the max-norm $|\cdot|$, $\mu(A)=1$, and in the latter case $\mu(A)=0$ since μ is concentrated on the axes.

Thus, the knowledge about regular variation of $\mathbf{X}$ is useful when $\mu(A)>0$, i.e., when the components of $\mathbf{X}$ are not (asymptotically) independent. However, if $\mu(A)=0$ the regular variation of $\mathbf{X}$ is too crude in order to determine the tails of the distribution of the products of the components.

4.1. One-dimensional results. In the following result we collect some of the well known results about the tail behavior of the product of two independent non-negative random variables.

LEMMA 4.1. *Assume that X_1 and X_2 are independent non-negative random variables and that X_1 is regularly varying with index $\alpha>0$.*

(1) *If either X_2 is regularly varying with index $\alpha>0$ or $P(X_2>x)=o(P(X_1>x))$ then X_1X_2 is regularly varying with index α.*

(2) *If X_1,X_2 are iid such that $E(X_1^{\alpha})=\infty$ then $P(X_1X_2>x)/P(X_1>x)\to\infty$.*

(3) *If X_1,X_2 are iid such that $E(X_1^{\alpha})<\infty$, then the only possible limit of $P(X_1\,X_2>x)/P(X_1>x)$ as $x\to\infty$ is given by $2E(X_1^{\alpha})$ which is attained under the condition*

$$\lim_{M\to\infty}\limsup_{x\to\infty}\frac{P(X_1X_2>x\,,M<X_1X_2\leqslant x/M)}{P(X_1>x)}=0\,.$$

(4) *Assume* $P(X_1 > x) \sim c^\alpha x^{-\alpha}$ *for some* $c > 0$. *Then for iid copies* $X_1, \ldots, X_n$ *of* X_1, $n \geqslant 1$,

$$P(X_1 \cdots X_n > x) \sim \frac{\alpha^{n-1}\, c^{n\alpha}}{(n-1)!}\; x^{-\alpha} \log^{n-1} x\,.$$

PROOF. (1) was proved in Embrechts and Goldie [**16**, p. 245].

(2) The following decomposition holds for any $M > 0$:

$$\frac{P(X_1X_2 > x)}{P(X_1 > x)} = \int_{(0,M]} \frac{P(X_2 > x/y)}{P(X_1 > x)}\, dP(X_1 \leqslant y) + \int_{(M,\infty)} \frac{P(X_2 > x/y)}{P(X_1 > x)}\, dP(X_1 \leqslant y) = I_1 + I_2\,. \tag{4.1}$$

By the uniform convergence theorem, $P(X_1 > x/y)/P(X_1 > x) \to y^{-\alpha}$ uniformly for $y \in (0, M]$. Hence

$$\begin{aligned} I_1 &\to \int_0^M y^\alpha\, dP(X_1 \leqslant y)\,, \quad x \to \infty\,,\\ &\to E(X_1^\alpha)\,, \quad M \to \infty\,. \end{aligned}$$

Hence, if $E(X_1^\alpha) = \infty$, (2) applies.

(3) It follows from Chover et al. [**6**] that the only possible limits of $P(X_1X_2 > x)/P(X_1 > x)$ are $2E(X_1^\alpha)$. The proof follows now from Davis and Resnick [**12**, Proposition 3.1].

(4) We start with the case when $P(Y_i/c > x) = x^{-\alpha}$, for $x \geqslant 1$ and an iid sequence (Y_i). Then $\sum_{i=1}^n \log(Y_i/c)$ is $\Gamma(\alpha, n)$ distributed:

$$P\left(\sum_{i=1}^n \log(Y_i/c) > x\right) = \frac{\alpha^n}{(n-1)!} \int_0^x y^{n-1}\, \mathrm{e}^{-\alpha\, y}\, dy\,, \quad x > 0\,.$$

Then, by Karamata's theorem,

$$\begin{aligned} P\left(\prod_{i=1}^n (Y_i/c) > x/c^n\right) &= \frac{\alpha^n}{(n-1)!} \int_0^{\log(x/c^n)} y^{n-1}\, \mathrm{e}^{-\alpha\, y}\, dy\\ &= \frac{\alpha^n}{(n-1)!} \int_1^{x/c^n} (\log z)^{n-1}\, z^{-\alpha-1}\, dz\\ &\sim \frac{\alpha^{n-1}}{(n-1)!} (\log(x/c^n))^{n-1}\, (x/c^n)^{-\alpha}\\ &\sim \frac{\alpha^{n-1}\, c^{n\alpha}}{(n-1)!} (\log x)^{n-1}\, x^{-\alpha}\,. \end{aligned} \tag{4.2}$$

Next consider an iid sequence (X_i) with $P(X_1 > x) \sim c^\alpha x^{-\alpha}$, independent of (Y_i), and assume without loss of generality that $c = 1$. Denote the distribution function of $\prod_{i=2}^n Y_i$ by G and let $h(x) \to \infty$ be any increasing function such that $x/h(x) \to \infty$. Then

$$\begin{aligned}
P\left(X_1 \prod_{i=2}^n Y_i > x\right) &= \int_0^\infty P(X_1 > x/y)\, dG(y) \\
&= \int_0^{h(x)} \frac{P(X_1 > x/y)}{P(Y_1 > x/y)}\, P(Y_1 > x/y)\, dG(y) \\
&\quad + \int_{h(x)}^\infty P(X_1 > x/y)\, dG(y) \\
&= I_1(x) + I_2(x)\,.
\end{aligned}$$

For any $\varepsilon > 0$, sufficiently large x and $y \in (0, h(x))$,

$$1 - \varepsilon \leqslant \frac{P(X_1 > x/y)}{P(Y_1 > x/y)} \leqslant 1 + \varepsilon\,.$$

Hence

$$I_1(x) \sim \int_0^{h(x)} P(Y_1 > x/y)\, dG(y)\,.$$

Now choose, for example, $h(x) = x/\log\log x$. Then

$$I_2(x) \leqslant \overline{G}(x/\log\log x) = O((x/(\log\log x))^{-\alpha} \log^{n-2} x) = o(x^{-\alpha} \log^{n-1} x)\,.$$

A similar argument yields

$$\int_{h(x)}^\infty P(Y_1 > x/y)\, dG(y) = o(x^{-\alpha}\, \log^{n-1} x)\,.$$

In view of (4.2) we conclude that

$$P\biggl(X_1 \prod_{i=2}^n Y_i > x\biggr) \sim I_1(x) \sim P\biggl(\prod_{i=1}^n Y_i > x\biggr)$$

A similar argument shows that we may replace in the left probability any Y_i, $i = 2, \ldots, n$, by X_i. This proves (4). □

Under the assumption $\limsup_{x\to\infty} x^\alpha P(X_i > x) < \infty$ upper bounds similar to (4) were obtained by Rosiński and Woyczyński [**54**]. The tail behavior of products of independent random variables is then also reflected in the tail behavior of polynomial forms of iid random variables with regularly varying tails and in multiple stochastic integrals driven by α-stable Lévy motion; see Kwapień and Woyczyński [**34**].

In the following results for the product $X_1 X_2$ of non-negative independent random variables X_1 and X_2 we assume that the tail of one of the factors dominates the tail of the other one.

LEMMA 4.2. *Assume X_1 and X_2 are non-negative independent random variables and that X_1 is regularly varying with index $\alpha > 0$.*

(1) *If there exists an $\varepsilon > 0$ such that $E(X_2^{\alpha+\varepsilon}) < \infty$, then*

$$P(X_1\, X_2 > x) \sim E(X_2^\alpha)\, P(X_1 > x)\,. \tag{4.3}$$

(2) *Under the assumptions of part* (1),

$$\sup_{x \geqslant y} \left| \frac{P(X_1\,X_2 > x)}{P(X_1 > x)} - E(X_2^\alpha) \right| \to 0\,, \quad \text{as } y \to \infty.$$

(3) *If* $P(X_1 > x) \sim c\,x^{-\alpha}$ *and* $E(X_2^\alpha) < \infty$ *then* (4.3) *holds.*

(4) *If* $P(X_2 > x) = o(P(X_1\,X_2 > x))$ *then* $X_1\,X_2$ *is regularly varying with index* α.

PROOF. Part (1) is usually attributed to Breiman [**4**] although he did not prove the result for general α. However, the proof remains the same for all $\alpha > 0$, and it also applies to the proof of (3): a glance at relation (4.1) shows that one has to prove $\lim_{M\to\infty} \limsup_{x\to\infty} I_2 = 0$ by applying a domination argument. An alternative proof of (1) is given in Cline and Samorodnitsky [**9**, Theorem 3.5(v)]. Part (3) is hardly available as an explicit result; but is has been implicitly used in various disguises e.g. in the books by Samorodnitsky and Taqqu [**56**] and in Ledoux and Talagrand [**36**]. Part (2) is Lemma 2.2 in Konstantinides and Mikosch [**33**]. Part (4) is due to Embrechts and Goldie [**16**], see also Theorem 3.5(iii) in Cline and Samorodnitsky [**9**]. □

Denisov and Zwart [**15**] give best possible conditions on the distributions of X_1 and X_2 such that Breiman's result (4.3) holds.

The lemma has applications in financial time series analysis. Indeed, financial time series are often assumed to be of the form $X_n = \sigma_n\,Z_n$, where the volatility σ_n is a measurable function of past Z_i's, (Z_i) is an iid sequence and (X_n) is strictly stationary. For example, this is the case for a strictly stationary GARCH(p, q) process, see e.g. Mikosch [**39**]. In many cases of interest, Z_n is light-tailed, e.g. standard normal, but σ_n is regularly varying with some positive index α. Breiman's result implies $P(X_1 > x) \sim E(Z_1^\alpha)\,P(\sigma_1 > x)$. Another case of interest is a stochastic volatility model, where the strictly stationary volatility sequence (σ_n) is independent of the iid noise sequence (Z_n). A convenient example is given when $\log \sigma_n$ constitutes a Gaussian stationary process. Then σ_n is log-normal. If Z_n is regularly varying with index α then Breiman's result yields $P(X_1 > x) \sim E(\sigma_1^\alpha)\,P(Z_1 > x)$.

The following results contain partial converses to Breiman's result, i.e., if we know that $X_1 X_2$ is regularly varying what can be said about regular variation of X_1 or X_2?

LEMMA 4.3. *Assume that* X_1 *and* X_2 *are independent non-negative random variables and that* $X_1 X_2$ *is regularly varying with positive index* α.

(1) *Assume that* X_2^p *for some* $p > 0$ *has a Lebesgue density of the form* $f(x) = c_0\,x^\beta\,\mathrm{e}^{-c\,x^\tau}$, $x > 0$, *for some constants* $\tau, c, c_0 > 0$, $\beta \in \mathbb{R}$, *such that* $x^\beta P(X_1 > x^{-1})$ *is ultimately monotone in* x. *Then* X_1 *is regularly varying with index* α *and Breiman's result* (4.3) *holds.*

(2) *Assume* $P(X_1 > x) = x^{-\alpha}$, $x \geqslant 1$. *Then* $X_2 \in \mathrm{RV}(\beta)$ *for some* $\beta < \alpha$ *if and only if* $X_1 X_2 \in \mathrm{RV}(\beta)$.

(3) *There exist* X_1, X_2 *such that* $E(X_1^\alpha) < \infty$, X_1 *and* X_2 *are not regularly varying and* $P(X_1 > x) = o(P(X_1 X_2 > x))$.

PROOF. (1) The idea is similar to the proof in Basrak et al. [**2**, Lemma 2.2], who assumed that X_2 is the absolute value of a normal random variable. Notice that if $X_1X_2 \in \mathrm{RV}(\alpha)$ then $(X_1X_2)^p \in \mathrm{RV}(\alpha/p)$ for $p > 0$. Therefore assume without loss of generality that $p = 1$ and we also assume for simplicity that $c = 1$. Since X_1X_2 is regularly varying there exists a slowly varying function L such that

$$\begin{aligned} L(x)\, x^{-\alpha} = P(X_1\, X_2 > x) &= \int_0^\infty P(X_1 > x/y)\, f(y)\, dy \\ &= c_0\, x^{1+\beta} \int_0^\infty P(X_1 > z^{-1})\, z^\beta \mathrm{e}^{-(z\,x)^\tau}\, dz \\ &= c_0\, \tau^{-1}\, x^{1+\beta} \int_0^\infty P(X_1 > v^{-1/\tau})\, v^{(\beta+1)/\tau-1}\, \mathrm{e}^{-v\, x^\tau}\, dv \\ &= x^{1+\beta} \int_0^\infty \mathrm{e}^{-r\, x^\tau}\, dU(r)\,, \end{aligned}$$

where

$$U(r) = \frac{c_0}{\tau} \int_0^r P(X_1 > v^{-1/\tau}) v^{(\beta+1)/\tau-1}\, dv = c_0 \int_0^{r^{1/\tau}} P(X_1 > z^{-1})\, z^\beta\, dz\,.$$

Thus

$$L(x^{1/\tau})\, x^{-(\alpha+\beta+1)/\tau} = \int_0^\infty \mathrm{e}^{-r\,x}\, dU(r)\,.$$

An application of Karamata's Tauberian theorem (see Feller [**21**, XIII, Section 5]) yields that

$$U(x) \sim \frac{L(x^{-1/\tau})\, x^{(\alpha+\beta+1)/\tau}}{\Gamma((\alpha+\beta+1)/\tau+1)}\,, \quad x \to \infty\,.$$

By assumption, $P(X_1 > z^{-1})\, z^\beta$ is ultimately monotone. Then the monotone density theorem (see Bingham et al. [**3**]) implies that

$$P(X_1 > x) \sim \frac{\tau}{c_0\, \Gamma((\alpha+\beta+1)/\tau)}\, \frac{L(x)}{x^\alpha}\,.$$

(2) This part is proved in Maulik and Resnick [**38**].

(3) An example of this kind, attributed to Daren Cline, is given in Maulik and Resnick [**38**]. □

Results for products of independent positive random variables can also be obtained by taking logarithms and then applying the corresponding results for regularly varying summands. The following example is in this line of thought.

LEMMA 4.4. *Let X_i be positive iid and such that $(\log X_1)_+ \in \mathrm{RV}(\alpha)$ for some $\alpha \geqslant 0$ and $P(X_1 \leqslant x^{-1}) = o(P(X_1 > x))$. Then for $n \geqslant 1$,*

$$P(X_1 \cdots X_n > x) \sim n\, P(X_1 > x)\,.$$

PROOF. We have for $x > 0$,

$$\begin{aligned} P(X_1 \cdots X_n > x) &= P(\log X_1 + \cdots + \log X_n > \log x) \\ &\sim n\, P(\log X_1 > \log x) = n\, P(X_1 > x)\,. \end{aligned}$$

This follows e.g. by an application of Lemma 3.3. Indeed, $(\log X_1)_+$ is regularly varying and by assumption, for $x > 0$,

$$P((\log X_1)_- > x) = P(X_1 < \mathrm{e}^{-x}) = o(P(X_1 > \mathrm{e}^{x})) = o(P((\log X_1)_+ > x)). \quad \square$$

Results for random products are rather rare. The following example is due to Samorodnitsky (personal communication). Extensions can be found in Cohen and Mikosch [**10**].

LEMMA 4.5. *Let (X_i) be an iid sequence with $P(X_1 > x) = c\,x^{-1}$ for some $x \geqslant c \geqslant 1$, K be Poisson(λ) distributed and independent of (X_i). Write $P_K = \prod_{i=1}^K X_i$. Then $P(P_K = 0) = \mathrm{e}^{-\lambda}$ and P_K has density f_P on (c,∞) satisfying as $x \to \infty$,*

$$f_P(x) \sim \frac{\mathrm{e}^{-\lambda} c^{-\lambda c} (\lambda c)^{1/4}}{2\sqrt{\pi}}\, x^{-2}\, (\log x)^{-3/4}\, \mathrm{e}^{2\,(\lambda c)^{1/2}\,(\log x)^{1/2}}\,,$$

and hence

$$P(P_K > x) \sim \frac{\mathrm{e}^{-\lambda} c^{-\lambda c} (\lambda c)^{1/4}}{2\sqrt{\pi}}\, x^{-1}\, (\log x)^{-3/4}\, \mathrm{e}^{2\,(\lambda c)^{1/2}\,(\log x)^{1/2}}\,.$$

Various results of this section can be extended to subexponential and even long-tailed distributions, see Cline and Samorodnitsky [**9**]. Resnick [**52**, Section 7.3.2] also treats the case of products with dependent regularly varying factors. Hult and Lindskog [**25**] extended Breiman's result in a functional sense to stochastic integrals $(\int_0^t \xi_{s-} d\eta_s)_{0\leqslant t\leqslant 1}$, where η is a Lévy process with regularly varying Lévy measure and ξ is a predictable integrand.

4.2. Multivariate extensions. Breiman's result (4.3) has a multivariate analog. It was proved in the context of regular variation for the finite-dimensional distributions of GARCH processes where multivariate products appear in a natural way; see Basrak et al. [**2**].

LEMMA 4.6. *Let $\mathbf{A}$ be an $m \times d$ random matrix such that $E(\|\mathbf{A}\|^{\alpha+\varepsilon}) < \infty$ for some matrix norm $\|\cdot\|$ and $\varepsilon > 0$. If $\mathbf{X} \in \mathrm{RV}(\alpha,\mu)$ assumes values in $\mathbb{R}^d$ and is independent of $\mathbf{A}$, then $\mathbf{AX}$ is regularly varying with index α and*

$$\frac{P(\mathbf{AX} \in \cdot)}{P(|\mathbf{X}| > x)} \xrightarrow{v} E\left(\mu\{\mathbf{x} : \mathbf{A}\,\mathbf{x} \in \cdot\}\right)\,.$$

5. Other functions

5.1. Powers. Let $\mathbf{X} \geqslant \mathbf{0}$ be a regularly varying random vector with index $\alpha > 0$. It is straightforward from the definition of multivariate regular variation that for $p > 0$, $\mathbf{X}^p = (X_1^p, \ldots, X_d^p)$ is regularly varying with index α/p. This can be seen from the polar coordinate representation of regular variation with $|\cdot|$ the max-norm, see (2.2):

$$\frac{P(|\mathbf{X}^p| > t\,x\,, \widetilde{\mathbf{X}^p} \in \cdot)}{P(|\mathbf{X}^p| > x)} = \frac{P(|\mathbf{X}| > (t\,x)^{1/p}\,, (\widetilde{\mathbf{X}})^p \in \cdot)}{P(|\mathbf{X}| > x^{1/p})} \to t^{-\alpha/p}\, P(\mathbf{\Theta}^p \in \cdot)\,.$$

5.2. Polynomials. We consider a sum $S_n = X_1 + \cdots + X_n$ of iid non-negative random variables X_i. Assume that X_1 is regularly varying with index $\alpha > 0$. By virtue of Lemma 3.2 this is equivalent to the fact that S_n is regularly varying and $P(S_n > x) \sim n\, P(X_1 > x)$. Then S_n^p for $p > 0$ is regularly varying with index α/p and

$$P(S_n^p > x) \sim n\, P(X_1^p > x) \sim P(X_1^p + \cdots + X_n^p > x)\,.$$

The latter relation has an interesting consequence for integers $k > 1$. Then one can write

$$S_n^k = \sum_{i=1}^n X_i^k + \sum X_{i_1} \cdots X_{i_k}\,,$$

where the second sum contains the off-diagonal products. It follows from the results in Section 4 that this sum consists of regularly varying summands whose index does not exceed $\alpha/(k-1)$. Hence, by Lemma 3.12, the influence of the off-diagonal sum on the tail of S_n^k is negligible. The regular variation of polynomial functions of the type

$$\sum_{1 \leqslant i_1, \ldots, i_k \leqslant n} c_{i_1 \ldots i_k}\, X_{i_1}^{p_{i_1}} \cdots X_{i_k}^{p_{i_k}}$$

for non-negative coefficients $c_{i_1 \ldots i_k}$ and integers $p_i \geqslant 0$ can be handled by similar ideas.

5.3. Maxima. Assume that $\mathbf{X} \in \mathrm{RV}(\alpha, \mu)$ and write $M_d = \max_{i=1,\ldots,d} X_i$ for the maximum of the components of $\mathbf{X}$. The set $A = \{\mathbf{x} : x_i > 1 \text{ for some } i\}$ is bounded away from zero and $\mu(\partial A) = 0$. Then

$$\frac{P(M_d > x)}{P(|\mathbf{X}| > x)} = \frac{P(x^{-1}\mathbf{X} \in A)}{P(|\mathbf{X}| > x)} \to \mu(A)\,.$$

If $\mu(A) > 0$, M_d is regularly varying with index α. In particular, if $\mathbf{X}$ has non-negative components and $|\cdot|$ is the max-norm, then $M_d = |\mathbf{X}|$ which is clearly regularly varying.

If $X_1, \ldots, X_n$ are independent, direct calculation with

$$\frac{P(X_i > x)}{P(|X_i| > x)} \to p_i \quad \text{and} \quad \frac{P(|X_i| > x)}{P(|\mathbf{X}| > x)} \to c_i\,,$$

yields the following limits

$$\frac{P(M_d > x)}{P(|\mathbf{X}| > x)} \sim \sum_{i=1}^d p_i\, \frac{P(|X_i| > x)}{P(|\mathbf{X}| > x)} \to \sum_{i=1}^d c_i\, p_i\,.$$

For iid X_i we obtain $\sum_{i=1}^d c_i\, p_i = d\, p$.

Next we consider maxima with a random index.

Lemma 5.1. *Assume that K is independent of the sequence (X_i) of iid random variables with distribution function F and right endpoint x_F.*

(1) *If $EK < \infty$ then*

$$P(M_K > x) \sim EK\, P(X_1 > x)\,, \quad x \uparrow x_F\,.$$

Hence X_1 is regularly varying with index α if and only if M_K is regularly varying with index α.

(2) *If $EK = \infty$ assume that $P(K > x) = L(x)x^{-\alpha}$ for some $\alpha \in [0,1)$ and a slowly varying function L. Then*

$$P(M_K > x) \sim (\overline{F}(x))^{\alpha} L(1/\overline{F}(x))\,, \quad x \uparrow x_F\,. \tag{5.1}$$

Hence X_1 is regularly varying with index $p > 0$ if and only if M_K is regularly varying with index $p\alpha$.

PROOF. (1) Write $F(x) = P(X_i \leqslant x)$. Then by monotone convergence, as $x \uparrow x_F$,

$$P(M_K > x) = \overline{F}(x)\, E\left[1 + F(x) + \cdots + F^{K-1}(x)\right] \sim EK\, \overline{F}(x)\,.$$

(2) By Karamata's Tauberian theorem (see Feller [**21**, XIII, Section 5]) and a Taylor expansion argument as $x \uparrow x_F$

$$\begin{aligned} P(M_K > x) &= 1 - E(F^K(x)) = 1 - E\big(\mathrm{e}^{\log F(x)\,K}\big) \\ &\sim (-\log F(x))^{\alpha} L(1/(-\log F(x))) \\ &\sim (\overline{F}(x))^{\alpha} L(1/\overline{F}(x))\,. \end{aligned}$$

Finally, if X_1 is regularly varying, $L(1/\overline{F}(x))$ is slowly varying and therefore $(\overline{F}(x))^{\alpha} L(1/\overline{F}(x))$ is regularly varying with index $-p\alpha$. □

5.4. Minima. For the minimum $m_d = \min(X_1, \ldots, X_d)$ of $\mathbf{X} \in \mathrm{RV}(\alpha, \mu)$ similar calculations apply by observing that $m_d = -\max(-X_1, \ldots, -X_d)$. This observation is not useful if some of the X_i's do not assume negative values. Nevertheless, in this situation

$$P(m_d > x) = P(X_1 > x\,, \ldots, X_d > x) = P(x^{-1}\mathbf{X} \in B)\,,$$

where $B = \{\mathbf{x} : \min_{i=1,\ldots,d} x_i > 1\}$ which is bounded away from zero and $\mu(\partial B) = 0$, and therefore m_d is regularly varying with index α if $\mu(B) > 0$. However, for independent X_i, m_d is not regularly varying with index α since $\mu(B) = 0$ and

$$P(m_d > x) = \prod_{i=1}^{d} P(X_i > x)\,.$$

In particular, if all $X_i \in \mathrm{RV}(\alpha)$, then $m_d \in \mathrm{RV}(d\alpha)$.

For an integer-valued non-negative random variable K independent of the sequence (X_i) of iid non-negative regularly varying random variables we have

$$P(m_K > x) = \sum_{n=1}^{\infty} P(K = n)\, [P(X_1 > x)]^n\,.$$

Let n_0 be the smallest positive integer such that $P(K = n_0) > 0$. Then

$$P(m_K > x) \sim P(K = n_0)\, [P(X_1 > x)]^{n_0}\,,$$

implying that m_K is regularly varying with index $n_0\alpha$.

5.5. Order statistics. Let $X_{(1)} \leqslant \cdots \leqslant X_{(d)}$ be the *order statistics* of the components of the vector $\mathbf{X} \in \mathrm{RV}(\alpha, \mu)$. The tail behavior of the order statistics has been studied in some special cases, including infinite variance α-stable random vectors which are regularly varying with index $\alpha < 2$, see Theorem 4.4.8 in Samorodnitsky and Taqqu [**56**]. It is shown in Samorodnitsky [**55**] (cf. Theorem 4.4.5 in Samorodnitsky and Taqqu [**56**]) that each $X_{(i)}$ as well as the order statistics of the $|X_i|$'s are regularly varying with index α.

For a general regularly varying vector $\mathbf{X}$ with index α similar results can be obtained. We assume that $\mathbf{X}$ has non-negative components. Write $x_{(1)} \leqslant \cdots \leqslant x_{(d)}$ for the ordered values of $x_1, \ldots, x_d$. Notice that the sets $A_i = \{\mathbf{x} : x_{(i)} > x\}$ are bounded away from zero. Hence the limits

$$\lim_{x \to \infty} \frac{P(X_{(i)} > x)}{P(|\mathbf{X}| > x)} = \mu(A_i)$$

exist and if $\mu(A_i) > 0$ then $X_{(i)}$ is regularly varying. This statement can be made more precise by the approach advocated in Samorodnitsky and Taqqu [**56**], Theorem 4.4.5, which also works for general regularly varying vectors:

$$\begin{aligned} &\frac{P(X_{(d-i+1)} > x)}{P(|\mathbf{X}| > x)} \\ &= \sum_{j=i}^{d} (-1)^{j-i} \binom{j-1}{i-1} \sum_{1 \leqslant i_1 < \cdots < i_j \leqslant d} \frac{P(X_{i_1} > x\,, \ldots, X_{i_j} > x)}{P(|\mathbf{X}| > x)} \end{aligned} \tag{5.2}$$

$$\to \sum_{j=i}^{d} (-1)^{j-i} \binom{j-1}{i-1} \sum_{1 \leqslant i_1 < \cdots < i_j \leqslant d} \mu(\{\mathbf{x} : x_{i_1} > 1\,, \ldots, x_{i_j} > 1\}). \tag{5.3}$$

In the same way one can also show the joint regular variation of a vector of order statistics.

For iid positive X_i's the limits of the ratios $P(X_{(i)} > x)/P(|\mathbf{X}| > x)$ are zero with the exception of $i = 1$. However, one can can easily derive that $X_{(d-i+1)}$ is regularly varying with index $i\,\alpha$. Indeed, by virtue of (5.2),

$$\frac{P(X_{(d-i+1)} > x)}{[P(X_1 > x)]^i} \sim \frac{d \cdots (d-i+1)}{i!}\,.$$

5.6. General transformations. Since the notion of regular variation bears some resemblance with weak convergence it is natural to apply the continuous mapping theorem to a regularly varying vector $\mathbf{X}$ with index α. Assume that $f : \overline{\mathbb{R}}_0^d \to \overline{\mathbb{R}}_0^m$ for some $d, m \geqslant 1$ is an a.e. continuous function with respect to the limit measure μ such that the inverse image with respect to f of any set $A \in \mathcal{B}(\overline{\mathbb{R}}_0^m)$ which is bounded away from zero is also bounded away from zero in $\overline{\mathbb{R}}_0^d$. Then we may conclude that

$$\frac{P(f(x^{-1}\mathbf{X}) \in A)}{P(|\mathbf{X}| > x)} = \frac{P(x^{-1}\mathbf{X} \in f^{-1}(A))}{P(|\mathbf{X}| > x)} \to \mu(f^{-1}(A))\,,$$

provided $\mu(\partial f^{-1}(A)) = 0$.

This means that $f(x^{-1}\mathbf{X})$ can be regularly varying in $\overline{\mathbb{R}}_0^m$, usually with an index different from α. Think for example of the functions $f(\mathbf{x}) = x_1 \cdots x_d$, $\min_{i=1,\ldots,d} x_i$, $\max_{i=1,\ldots,d} x_i$, $(x_1^p, \ldots, x^p)$, $c_1 x_1 + \cdots + c_d x_d$. These are some of the examples of the previous sections. These functions have in common that they are homogeneous, i.e., $f(t\,\mathbf{x}) = t^q f(\mathbf{x})$ for some $q > 0$, all $t > 0$. Then $f(\mathbf{X})$ is regularly varying with index α/q.

References

[1] B. Basrak, R. A. Davis, and T. Mikosch, (2002) *A characterization of multivariate regular variation*, Ann. Appl. Probab. **12**, 908–920.

[2] B. Basrak, R. A. Davis, and T. Mikosch, (2002) *Regular variation of GARCH processes*, Stoch. Proc. Appl. **99**, 95–116.

[3] N. H. Bingham, C. M. Goldie, and J. L. Teugels, (1987) *Regular Variation*, Cambridge University Press, Cambridge.

[4] L. Breiman, (1965) *On some limit theorems similar to the arc-sin law*, Theory Probab. Appl. **10**, 323–331.

[5] P. J. Brockwell, and R. A. Davis, (1991) *Time Series: Theory and Methods*, 2nd edition. Springer-Verlag, New York.

[6] J. Chover, P. Ney, and S. Wainger, (1973) *Functions of probability measures*, J. Analyse. Math. **26**, 255–302.

[7] D. B. H. Cline, (1983) *Estimation and linear prediction for regression, autoregression and ARMA with infinite variance data*, Ph.D. dissertation, Colorado State University.

[8] D. B. H. Cline, (1983) *Infinite series or random variables with regularly varying tails*, Technical Report 83-24, Inst. Appl. Math. Statist., University of British Columbia, Vancouver, B. C.

[9] D. B. H. Cline and G. Samorodnitsky, (1994) *Subexponentiality of the product of independent random variables*, Stoch. Proc. Appl. **49**, 75–98.

[10] S. Cohen and T. Mikosch, (2006) *Tail behavior or random products and stochastic exponentials*, Work in progress.

[11] R. A. Davis, and S. I. Resnick, (1985) *Limit theory for moving averages of random variables with regularly varying tail probabilities*, Ann. Probab. **13**, 179–195.

[12] R. A. Davis, and S. I. Resnick, (1985) *More limit theory for the sample correlation function of moving averages*, Stoch. Proc. Appl. **20**, 257–279.

[13] R. A. Davis, and S. I. Resnick, (1986) *Limit theory for the sample covariance and correlation functions of moving averages*, Ann. Statist. **14**, 533–558.

[14] R. A. Davis and S. I. Resnick, (1996) *Limit theory for bilinear processes with heavy tailed noise*, Ann. Appl. Probab. **6**, 1191–1210.

[15] D. Denisov and B. Zwart, (2005) *On a theorem of Breiman and a class of random difference equations*, EURANDOM Technical Report 2005-039.

[16] P. Embrechts, and C. M. Goldie, (1980) *On closure and factorization theorems for subexponential and related distributions*, J. Austral. Math. Soc. Ser. A **29**, 243–256.

[17] P. Embrechts, C. M. Goldie, and N. Veraverbeke, (1979) *Subexponentiality and infinite divisibility*, Z. Wahrscheinlichkeitstheorie verw. Gebiete **49**, 335–347.

[18] P. Embrechts, C. Klüppelberg, and T. Mikosch, (1997) *Modelling Extremal Events for Insurance and Finance*, Springer-Verlag, Berlin.

[19] P. Embrechts and E. Omey (1983) *On subordinated distributions and random record processes*, Proc. Camb. Phil. Soc. **93**, 339-353.

[20] G. Faÿ, B. González-Arévalo, T. Mikosch, and G. Samorodnitsky, (2006) *Modeling teletraffic arrivals by a Poisson cluster process*, Queueing Syst. Theory Appl., To appear. Available under www.math.ku.dk/∼mikosch.

[21] W. Feller, (1971) *An Introduction to Probability Theory and Its Applications.* Vol. II. Second edition. Wiley, New York.

[22] C. M. Goldie, (1991) *Implicit renewal theory and tails of solutions of random equations*, Ann. Appl. Probab. **1**, 126–166.

[23] D. Heath, S. Resnick, and G. Samorodnitsky, (1998) *Heavy tails and long range dependence in ON/OFF processes and associated fluid models*, Math. Oper. Res. **23**, 145–165.

[24] H. Hult and F. Lindskog, (2005) *Extremal behavior of regularly varying stochastic processes*, Stoch. Proc. Appl. **115**, 249–274; See also: H. Hult, (2003) *Topics on Fractional Brownian Motion and Regular Variation for Stochastic Processes*, PhD thesis, Department of Mathematics, KTH Stockholm.

[25] H. Hult and F. Lindskog, (2006) *On Kesten's counterexample to the Cramér–Wold device for regular variation*, Bernoulli **12**, 133–142.

[26] H. Hult and F. Lindskog, (2006) *Extremal behavior of stochastic integrals driven by regularly varying Lévy processes*, Ann. Probab., to appear.

[27] M. Jacobsen, T. Mikosch, and G. Samorodnitsky, (2006) *Regular variation of linear processes*, Work in progress.

[28] H. Kesten, (1973) *Random difference equations and renewal theory for products of random matrices*, Acta Math. **131**, 207–248.

[29] C. Klüppelberg and T. Mikosch, (1993) *Spectral estimates and stable processes*, Stoch. Proc. Appl. **47**, 323–344.

[30] C. Klüppelberg and T. Mikosch, (1996) *Self–normalised and randomly centred spectral estimates*, In: C. C. Heyde, Yu. V. Prokhorov, R. Pyke, and S. T. Rachev, (Eds.) *Proceedings of the Athens International Conference on Applied Probability and Time Series,* vol. 2: Time Series, pp. 259–271. Springer-Verlag, Berlin.

[31] C. Klüppelberg and T. Mikosch, (1996) *The integrated periodogram for stable processes*, Ann. Statist. **24**, 1855-1879.

[32] P. Kokoszka and M. Taqqu, (1996) *Parameter estimation for infinite variance fractional ARIMA*, Ann. Statist. **24**, 1880–1913.

[33] D. Konstantinides and T. Mikosch, (2005) *Stochastic recurrence equations with heavy-tailed innovations*, Ann. Probab. **33**, 1992–2035.

[34] S. Kwapień and W. A. Woyczyński (1992) *Random Series and Stochastic Multiple Integrals: Single and Multiple*, Birkhauser, Boston.

[35] W. E. Leland, M. S. Taqqu, W. Willinger, and D. V. Wilson, (1993) *On the self-similar nature of Ethernet traffic*, ACM/SIGCOMM Computer Communications Review, 183–193.

[36] M. Ledoux and M. Talagrand, (1991) *Probability in Banach Spaces. Isoperimetry and Processes.* Springer-Verlag, Berlin.

[37] E. Lukacs, (1983) *Developments in Characteristic Function Theory*. Griffin & Company. London.

[38] K. Maulik and S. I. Resnick, (2006) *Characterization and examples of hidden regular variation*, Extremes, to appear.

[39] T. Mikosch, (2003) *Modeling dependence and tails of financialtime series*, In: B. Finkenstädt and H. Rootzén (Eds.) (2003) *Extreme Values in Finance, Telecommunications, and the Environment,* Chapman & Hall, London, pp. 187–286.

[40] T. Mikosch, S. Resnick, H. Rootzén, and A. Stegeman, (2002) *Is network traffic approximated by stable Lévy motion or fractional Brownian motion?*, Ann. Appl. Probab. **12**, 23–68.

[41] T. Mikosch and G. Samorodnitsky, (2000) *The supremum of a negative drift random walk with dependent heavy-tailed steps*, Ann. Appl. Probab. **10**, 1025–1064.

[42] T. Ostrogorski, (1984) *Abelian type theorems for some integral operators in* $\mathbb{R}^n$, Publ. Inst. Math. (Beograd) (N. S.) **35**, 93–103.

[43] T. Ostrogorski, (1984) *Asymptotic behaviour of Fourier transforms in* $\mathbb{R}^n$, Publ. Inst. Math. (Beograd) (N. S.) **35**, 105–117.

[44] T. Ostrogorski, (1995) *Regular variation on homogeneous cones*, Slobodan Aljančić memorial volume of Publ. Inst. Math. (Beograd) (N. S.) **58**, 51–70.

[45] T. Ostrogorski, (1997) *Regular variation in* $\mathbb{R}^n_+$, Mathematica **39**, 265–276.

[46] T. Ostrogorski, (1998) *Homogeneous cones and Abelian theorems*, Internat. J. Math. Math. Sci. **21**, 643–652.

[47] T. Ostrogorski, (1998) *The Laplace transform on homogeneous cones*, In: *Generalized Functions—Linear and Nonlinear Problems (Novi Sad, 1996).* Integral Transform. Spec. Funct. **6**, 265–276.

[48] V. V. Petrov, (1975) *Sums of Independent Random Variables.* Springer-Verlag, Berlin.

[49] V. V. Petrov, (1995) *Limit Theorems of Probability Theory*, Oxford University Press, Oxford.

[50] S. I. Resnick, (1986) *Point processes, regular variation and weak convergence*, Adv. Appl. Prob. **18**, 66–138.

[51] S. I. Resnick, (1987) *Extreme Values, Regular Variation, and Point Processes*, Springer-Verlag, New York.

[52] S. I. Resnick, (2006) *Probabilistic and Statistical Modeling of Heavy Tailed Phenomena*, Birkhauser, Boston. To appear.

[53] S. I. Resnick and E. Willekens, (1991) *Moving averages with random coefficients and random coefficient autoregressive models*, Stoch. Models **7**, 511–525.

[54] J. Rosiński and W. A. Woyczyński, (1987) *Multilinear forms in Pareto-like random variables and product random measures*, Coll. Math. (S. Hartmann Festschrift) **51**, 303–313.

[55] G. Samorodnitsky, (1988) *Extrema of skewed stable processes*, Stoch. Proc. Appl. **30**, 17–39.

[56] G. Samorodnitsky and M. S. Taqqu, (1994) *Stable Non-Gaussian Random Processes. Stochastic Models with Infinite Variance*, Chapman and Hall, London.

[57] A. J. Stam, (1973) *Regular variation of the tail of a subordinated probability distribution*, Adv. Appl. Probab. **5**, 287–307.

(Received 07 07 2006)
(Revised 04 10 2006)

Laboratory of Actuarial Mathematics
University of Copenhagen
Universitetsparken 5
DK-2100 Copenhagen
Denmark
ahj@math.ku.dk
mikosch@math.ku.dk
www.math.ku.dk/∼mikosch