Klüppelberg, Severin:

Prediction of outstanding insurance claims

Sonderforschungsbereich 386, Paper 258 (2001)

Online unter: http://epub.ub.uni-muenchen.de/

Prediction of outstanding insurance claims*

Claudia Klüppelberg Martin Severin †

Abstract

Building reserves for outstanding liabilities is an important issue in the financial statement of any insurance company. In this paper we present a new model for delay in claim settlement and to predict IBNR (incurred but not reported) claims. The modelling is based on a data set of a portfolio of car liability data, which describes the claim settlement of a car insurance portfolio. The data consists of about 5 000 realisations of claims, all of which incurred in 1985 and were followed until the end of 1993.

In our model, the total claim amount process $(S(t))_{t\geq 0}$ is described by a Poisson shot noise model, i.e.

$$S(t) = \sum_{n=1}^{N(t)} X_n(t - T_n), \quad t \geq 0,$$

where $X_1(\cdot)$, $X_2(\cdot)$, ... are i.i.d. copies of the claim settlement process $X(\cdot)$ and the occurrence times $(T_i)_{i\in\mathbb{N}}$ of the consecutive claims are random variables such that $N(t) = \#\{n \in \mathbb{N};\ T_n \leq t\}$, $t \geq 0$, is a Poisson process, which is assumed to be independent of $X(\cdot)$.

The observed times of occurrences of claims are used to specify and to estimate the intensity measure of the Poisson process $N(\cdot)$. Motivated by results of an exploratory data analysis of the portolio under consideration, a hidden Markov model for $X(\cdot)$ is developed. This model is fitted to the data set, parameters are estimated by using an EM algorithm, and prediction of outstanding liabilities is done and compared with the real world outcome.

Keywords: EM algorithm, hidden Markov model, Poisson shot noise process, prediction, risk reserve, mixture model.

---

*Research supported by the German DFG Grant, Kl1041/1-2.

†Center of Mathematical Sciences, Munich University of Technology, D-80290 Munich, Germany, email: {cklu , severin}@ma.tum.de, http://www.ma.tum.de/stat/

1 Introduction

As emphasized in the Claims Reserving Manual [3], published by the Institute of Actuaries, London, the reserve for outstanding or for incurred but not fully reported (IBNFR) claims is an important issue in the financial statement of an insurance company. The problem in claims reserving is *to estimate the cost of claims - known or unknown to date -* to be paid ultimately (or yearly) by the company. A large variety of methods have been developed over time: based on deterministic and stochastic modelling, on macro- (modelling the whole portfolio) and micro-models (modelling each single claim); see Taylor [21] for more details.

In this paper, we

- suggest a new micro model, a shot noise model with a hidden Markov structure,
- fit a large portfolio by means of the EM algorithm,
- predict outstanding claims based on our model.

We model the *total claim amount process* by

$$S(t) = \sum_{n=1}^{N(t)} X_n(t - T_n)\,, \quad t \geq 0\,. \tag{1.1}$$

The different quantities of the model are as follows:

$N = (N(t))_{t\geq 0}$ is an inhomogeneous Poisson process with points T_n, $n \geq 1$, representing the time points, when the n-th claim occurred.

$X_n = (X_n(t))_{t\geq 0}$, $n \in \mathbb{N}$, are independent increasing stochastic processes on $\mathbb{R}_+$, representing the pay-off function of the insurance company for the n-th claim.

N is called the *shot process* and the X_n are called the *noise processes* of the model. We assume throughout that the processes $(X_n)_{n\in\mathbb{N}}$ are independent of N.

The model is a natural extension of the classical compound Poisson model taking into account that occurring claims may require payments over several years until they are finally settled. This model for i.i.d. noise processes has been suggested and investigated in Klüppelberg and Mikosch [10, 11]. Ruin probabilities for this model and for different premium principles have been derived by means of diffusion approximations in the finite variance case. The infinite variance case has been investigated in Klüppelberg, Mikosch and Schärf [12].

In this paper we want to see the model at work. To this end we analyse a large data set of a portfolio of motor vehicles liability data provided by a leading Swiss insurance company. The data are described in detail in Section 2.

Although $(X_n)_{n\in\mathbb{N}}$ are assumed to be i.i.d., as much structure is incorporated into the model as exhibited by our data. For instance, each single claim has a certain number of instalments and it turns out that claims with the same number of payoffs, we call them instalments are structurally similar, whereas those with different numbers of instalments have quite a different behaviour. Since the number of instalments of each claim is not known in advance we model this by a simple hidden Markov model; see Section 3.

The paper is organized as follows. After explaining our data in Section 2, in Section 3, we concentrate on the noise processes $X(\cdot)$ introducing the hidden Markov structure. We derive the prediction equations for this model in Section 3.2 on the micro level. In Section 4 the model is further specified. Each claim process is modelled as a compound process with a certain number of instalments and inter-instalment times. These are specified here and also the hidden Markov model is further specified. In Section 4.4 the claim number process is modelled. When appropriate, initial parameter estimates are derived. In Section 5 the parameters are estimated by MLE, where we use the independence of certain quantities, which factorizes the likelihood function. The instalments are then fitted by classical MLE, whereas the inter-instalment times require the use of the EM algorithm. Finally, in Section 6 the capital reserve for this portfolio for the following year is estimated as a one-step predictor and compared to the actual costs of the portfolio.

2 The data

We analyse a data set of a portfolio of motor vehicles liability data of a leading Swiss insurance company. Insurance coverage is given for both, material damage and casualty damage. The structure of the data set is shown in Table 1.

The data consists of claims which incurred during 1985 and were followed for a total

M	170785	250785	000000
Z	061092	610	
Z	260991	830	
Z	160786	30	
Z	070186	25	
M	110985	270985	101186
Z	311086	-5000	
Z	180886	-500	
Z	070786	26140	
M	250885	260885	090987
Z	200586	1250	
Z	200586	-955	
Z	130586	1000	

Table 1: Structure of the given data set: a capital M indicates the occurrence of a new claim, in the same line there is its date of occurrence, its date of reporting and its date of closing, if known already. If the claim has not yet been closed, this is indicated by 000000 in the third row; the first claim above has not yet been closed. Each capital Z in the lines below indicates an instalment of the claim, showing the date and the amount of payment made.

of 9 years until the end of 1993.

We have cleaned the data in the following way:

– several instalments on the same day have been summed up;

– negative payments are cleared with positive payments on the same day or on previous days;

– zero payments have been erased;

– periodical payments, as for instance annuities, have been erased; their deterministic behaviour allows a precise prediction, taking a possible mortality or recovery risk into account.

– most claims were closed by the end of 1993, the remaining 83 still open claims were erased.

After cleaning, our data consists of 5 681 realisations of single claims, all of which occurred during 1985 and were settled by the end of 1993. Payments are attributed to days: January 1, 1985 equals day 1 and December 31, 1993 equals day 3 287. All payments are made in Swiss Francs.

The delay between occurrence and reporting of the claim is not investigated in this paper; see Jewell [7, 8] for a careful analysis and consequences of this effect. We model and analyse a claim after it has occurred and it is assumed to be immediately reported to the insurance company.

3 The hidden Markov model

3.1 Modelling the claim size process

Each sample path of a single claim shows a finite number of instalments given at random times and after the last instalment the noise process remains on its level forever. Consequently, we model each single claim as

$$X(t) = \sum_{m=1}^{M(t)} Z_m , \quad t \geq 0, \tag{3.1}$$

where $M = (M(t))_{t\geq 0}$ is a counting process which counts the number of instalments up to time t, $(V_n)_{n\in\mathbb{N}}$ are the inter-instalment times and $(Z_n)_{n\in\mathbb{N}}$ are the sizes of instalments.

In predicting the costs of future payment for this portfolio, an important issue is the prediction of the ultimate number of instalments for a running, not yet closed claim. This number is unknown when a claim incurs and remains unobservable during claim settlement. We model this uncertainty by a simple hidden Markov model (HMM) governing each claim $X = (X(t))_{t\geq 0}$; for some background on HMM we refer to the monographs by Elliott, Aggoun and Moore [5], MacDonald and Zucchini [15] or Rabiner [19].

The HMM is given by a Markov chain $(\xi_n)_{n\in\mathbb{N}}$ with state space

$$\mathbb{S} = \{(n, m) | n \in \mathbb{N}, \, 1 \leq m \leq n\}$$

The index $n \in \mathbb{N}$ represents the total number of instalments of a single claim; once n has been fixed, the state (n, m) indicates that while the chain is in this state the m-th instalment is paid. For $m < n$, immediately after the payment the Markov chain changes to state $(n, m+1)$; for $m = n$, the final instalment is paid, and the Markov chain moves afterwards into its absorbing state Δ. Note that this means that the Markov chain behaves almost deterministically on its state space: $(\xi_n)_{n\in\mathbb{N}}$ starts in state $(n, 1)$ with probability $\alpha_{n,1}$ and then it passes the states $(n,1), (n,2), \dots, (n,n)$ deterministically and enters the absorbing state Δ after n transitions.

The hidden Markov chain ξ governs an observable sequence $(O_n)_{n\in\mathbb{N}}$ with

$$O_n = \begin{cases} (V_n, Z_n), & \text{if } \xi_n \neq \Delta, \\ (\infty, 0), & \text{if } \xi_n = \Delta, \end{cases} \tag{3.2}$$

where $(V_n)_{n\in\mathbb{N}}$ represent the inter-instalment times and $(Z_n)_{n\in\mathbb{N}}$ the sizes of the instalments.

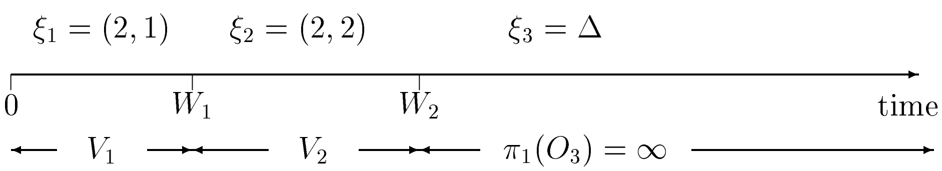

Figure 1: Relationship between the sequence of occurrence times $(W_k)_{k\in\mathbb{N}}$ of consecutive instalments and the Markov chain $(\xi_n)_{n\in\mathbb{N}}$, where π_1 means the projection of the first component of an observation vector.

The following distributional assumptions model our situation:

Assumption 3.1. *(1) Conditionally on the Markov chain $(\xi_n)_{n\in\mathbb{N}}$, the observations $(O_n)_{n\in\mathbb{N}}$ are mutually independent and O_j depends on ξ_j only, i.e. for $\xi_1, \dots, \xi_i$, such that $\xi_i \neq \Delta$,*

$$P\left(O_1 \in G_1, \dots, O_i \in G_i \mid \xi_1, \dots, \xi_i\right) = \prod_{j=1}^{i} P\left(O_j \in G_j \mid \xi_j\right), \quad i \in \mathbb{N},$$

for all Borel sets G_j in $\mathbb{R}_+ \cup \{\infty\} \times \mathbb{R}_+$.

(2) For $\xi_i \neq \Delta$, the distribution of O_i conditioned on ξ_i is given by

$$O_i \mid \{\xi_i = (n,i)\} \overset{d}{=} F_{n,i}, \quad n \in \mathbb{N}, 1 \leq i \leq n,$$

where for such ξ_i, *the distribution function* $F_{n,i}$ *has density* $f_{n,i}$ *and tail* $\overline{F}_{n,i}(v,z) = P_n(V_i > v, Z_i > z)$, $1 \le i \le n$, *where* P_n *denotes conditional probability with respect to* $\{\xi = n\}$.

3.2 Prediction of the shot noise model

Since the HMM only depends on the initial state of the Markov chain, we simplify the notation by setting $\alpha_{n,1} = \alpha_n$, $n \in \mathbb{N}$, and define a random variable ξ, which takes value n if and only if $\xi_1 = (n,1)$.

In the following theorem we show how to filter the value of ξ, i.e. the number of instalments of a single claim, using the information observed up to time t, summarized in the σ-field

$$\mathcal{F}_t = \sigma(\mathcal{H}_t), \quad t \ge 0\,, \tag{3.3}$$

where

$$\begin{aligned} \mathcal{H}_t &= \bigcup_{k=0}^{\infty} \mathcal{H}_{t,k} \\ &= \bigcup_{k=0}^{\infty} \Big\{ \{M(t) = k\,,\, O_j\,,\, 1 \le j \le k\} : O_j \in B\,,\, B \text{ Borel set in } \mathbb{R}_+ \times \mathbb{R}_+ \Big\} \qquad (3.4) \\ &= \bigcup_{k=0}^{\infty} \Big\{ \{W_k \le t < W_{k+1}\,,\, O_j\,,\, 1 \le j \le k\} : O_j \in B\,,\, B \text{ Borel set in } \mathbb{R}_+ \times \mathbb{R}_+ \Big\}\,. \end{aligned}$$

The observations $O_j = (V_j, Z_j)$ are defined in (3.2), $W_0 = 0$, $W_k = \sum_{j=1}^{k} V_k$, $k \in \mathbb{N}$, which is finite for $V_k < \infty$ and infinite else.
Furthermore, by definition $\mathcal{H}_{t,0} = \{\{M(t) = 0\}\} = \{\{V_1 > t\}\}$.
The following theorem is the main result of this section; it shows how to estimate the initial probabilities. For convenience we set $\prod_{j=1}^{0} a_j = 1$ and $\sum_{j=1}^{0} a_j = 0$.

Theorem 3.2. *Let* $n \in \mathbb{N}$ *and* $t \ge 0$. *For* $1 \le j \le n$ *set* $\overline{A}_{n,j}(v) = \overline{F}_{n,j}(v,0) = P_n\,(V_j > v)$, $v \ge 0$. *Then the conditional distribution* $P\,(\xi = n \,|\, \mathcal{F}_t)$ *is uniquely defined by its definition on all sets* $H_{t,k} = \{M(t) = k\,,\, O_j \in G_j\,, 1 \le j \le k\} \in \mathcal{H}_{t,k}$, *where* G_j, $1 \le j \le k$, *are*

Borel set in $\mathbb{R}_+ \times \mathbb{R}_+$*:*

$$\begin{aligned}
P\left(\xi = n \,|\, H_{t,k}\right) \;=\; & I_{\{M(t)=0\}} \, \frac{\alpha_n \, \overline{A}_{n,1}(t)}{\sum_{l=1}^{\infty} \, \alpha_l \, \overline{A}_{l,1}\,(t)} \\
& + \sum_{k=1}^{n-1} I_{\{M(t)=k\}} \, \frac{\alpha_n \, \prod_{j=1}^{k} \, f_{n,j}(O_j) \, \overline{A}_{n,k+1}\,(t - W_k)}{\alpha_k \, \prod_{s=1}^{k} \, f_{k,s}(O_s) + \sum_{l=k+1}^{\infty} \, \alpha_l \, \prod_{s=1}^{k} \, f_{l,s}(O_s) \, \overline{A}_{l,k+1}\,(t - W_k)} \\
& + I_{\{M(t)=n\}} \, \frac{\alpha_n \prod_{j=1}^{n} \, f_{n,j}(O_j)}{\alpha_n \prod_{j=1}^{n} \, f_{n,j}(O_j) + \sum_{l=n+1}^{\infty} \, \alpha_l \, \prod_{j=1}^{n} \, f_{l,j}(O_j) \, \overline{A}_{l,n+1}(t - W_n)}.
\end{aligned}$$

For the proof of this result we shall need several lemmata.

Lemma 3.3. *For* $t \geq 0$ *and* $n \in \mathbb{N}$*,*

$$P\left(H_{t,0} \,\cap\, \{\xi = n\}\right) \;=\; \alpha_n \int_{(t,\infty)\times\mathbb{R}_+} f_{n,1}(o)\, do, \tag{3.5}$$

$$P\left(H_{t,n} \,\cap\, \{\xi = n\}\right) \;=\; \alpha_n \prod_{j=1}^{n} \int_{G_j} f_{n,j}(o_j)\, do_j \tag{3.6}$$

and, denoting the first component of the observation O_j *by* $\pi_1(O_j) = V_j$ *and* $W_k = \sum_{i=1}^{k} V_i$*,* $k = 1, \ldots, n$*, we obtain for* $W_k \leq t < W_{k+1}$*,*

$$\begin{aligned}
& P\left(H_{t,k} \,\cap\, \{\xi = n\}\right) \\
&= \; \alpha_n \int_{G_1} \cdots \int_{G_k} \left(\int_{G_{t,k+1}} f_{n,k+1}(o)\, do \right) f_{n,k}(o_k)\, do_k \cdots f_{n,1}(o_1)\, do_1\,,
\end{aligned} \tag{3.7}$$

where $G_{t,k+1} = (t - \sum_{j=1}^{k} \pi_1(o_j), \infty) \times \mathbb{R}_+$*.*

Proof. Equation (3.5) follows by

$$P_n\left(H_{t,0}\right) \;=\; P_n\left(V_1 > t\right) = \int_{(t,\infty)\times\mathbb{R}_+} f_{n,1}(o)\, do.$$

Notice that for all $m \in \mathbb{N}$

$$H_{t,m} \;=\; \left(H_{t,m} \cap \{V_{m+1} < \infty\}\right) \cup \left(H_{t,m} \cap \{V_{m+1} = \infty\}\right). \tag{3.8}$$

To show (3.6), note that $\{V_{n+1} < \infty,\, \xi = n\} = \emptyset$ and thus

$$\begin{aligned}
P_n\left(H_{t,n}\right) \;&=\; P_n\left(H_{t,n} \cap \{V_{n+1} = \infty\}\right) \\
&=\; P_n\left(O_1 \in G_1, \ldots, O_n \in G_n\right) \\
&=\; \prod_{j=1}^{n} \int_{G_j} f_{n,j}(o_j)\, do_j,
\end{aligned}$$

where the last equality follows by Assumption 3.1.

Finally, for $k \in \{1, \ldots, n-1\}$ one obtains by properties of conditional expectation

$$\begin{aligned} P_n\left(H_{t,k}\right) &= E_n\Big[P_n\left(H_{t,k} \,\cap\, \{V_{k+1} < \infty\} \,|\, O_1, \ldots, O_k\right)\Big] \\ &= E_n\left[I_{\{O_1 \in G_1\}} \cdots I_{\{O_k \in G_k\}}\, P_n\left(V_{k+1} > t - \sum_{j=1}^{k} V_j \,\middle|\, O_1, \ldots, O_k\right)\right] \\ &= \int_{G_1 \times \ldots \times G_k} J \quad dF_{O_1,\ldots,O_k\,|\,\xi}\left(o_1, \ldots, o_k\,|\,n\right), \end{aligned}$$

where $F_{O_1,\ldots,O_k\,|\,\xi}$ is the joint distribution function of $O_1, \ldots, O_k$ given ξ and

$$\begin{aligned} J &= P_n\left(V_{k+1} > t - \sum_{j=1}^{k} v_j \,\middle|\, O_1 = (v_1, z_1), \ldots, O_k = (v_k, z_k)\right) \\ &= F_{O_{k+1}|O_1,\ldots,O_k,\xi}\left(t - \sum_{j=1}^{k} \pi_1(o_j) \,\middle|\, o_1, \ldots, o_k,\, n\right). \end{aligned}$$

By Assumption 3.1, $F_{O_1,\ldots,O_k\,|\,\xi} = \prod_{j=1}^{k} F_{O_j\,|\,\xi}$ and $F_{O_{k+1}\,|\,O_1,\ldots,O_k,\xi} = F_{O_{k+1}\,|\,\xi}$ and thus J simplifies to

$$J \;=\; \int_{(t-\sum_{j=1}^{k} v_j,\infty)\times\mathbb{R}_+} f_{n,k+1}(o)\, do$$

and the result follows. □

Lemma 3.4. *For $t \geq 0$, let Ψ be a $\mathcal{F}_t$-measurable random variable, and for $k \in \mathbb{N}$ let $\mathcal{H}_{t,k}$ be defined as in (3.4). Assume that Ψ is on $\mathcal{H}_{t,k}$ a function of $O_1, \ldots, O_k$ alone, i.e. $\Psi = \Psi(O_1, \ldots, O_k)$. Then for all $H_{t,k} \in \mathcal{H}_{t,k}$,*

$$\begin{aligned} \int_{H_{t,k}} \Psi\, dP \;=\; & \int_{G_1} \cdots \int_{G_k} \Psi(o_1, \ldots, o_k) \left[\alpha_k \prod_{s=1}^{k} f_{k,s}(o_s) \right. \\ & \left. + \sum_{l=k+1}^{\infty} \alpha_l\, \overline{A}_{l,k+1}\Big(t - \sum_{r=1}^{k} \pi_1(o_r)\Big) \prod_{s=1}^{k} f_{l,s}(o_s)\right] do_k \ldots do_1. \quad (3.9) \end{aligned}$$

If Ψ is a constant, then

$$\int_{H_{t,0}} \Psi\, dP \;=\; \Psi \sum_{l=1}^{\infty} \alpha_l\, \overline{A}_{l,1}(t). \quad (3.10)$$

Proof. Assume Ψ is a constant, then

$$\int_{H_{t,0}} \Psi \, dP \;=\; \Psi \, P\,(V_1 > t) = \Psi \sum_{l=1}^{\infty} \alpha_l \, P_l\,(V_1 > t)$$

and equation (3.10) follows by definition of $\overline{A}_{l,1}(t)$.

Next, equation (3.9) is established. Recall from (3.8) that

$$\int_{H_{t,k}} \Psi \, dP \;=\; \int_{H_{t,k} \cap \{V_{k+1}=\infty\}} \Psi \, dP + \int_{H_{t,k} \cap \{V_{k+1}<\infty\}} \Psi \, dP,$$

where

$$\begin{aligned}
\int_{H_{t,k} \cap \{V_{k+1}=\infty\}} \Psi \, dP &= \int_{\{O_j \in G_j,\, 1 \le j \le k,\, \xi = k\}} \Psi \, dP \\
&= \alpha_k \int_{\{O_j \in G_j,\, 1 \le j \le k\}} \Psi \, dP_k \\
&= \alpha_k \int_{G_1} \ldots \int_{G_k} \Psi(o_1, \ldots, o_k) \prod_{s=1}^{k} f_{k,s}(o_s) \, do_k \, \ldots \, do_1 \,, \qquad (3.11)
\end{aligned}$$

and, for $G_{t,k+1}(o_1, \ldots, o_k) = (t - \sum_{j=1}^{k} \pi_1(o_j), \infty) \times \mathbb{R}_+$,

$$\begin{aligned}
&\int_{H_{t,k} \cap \{V_{k+1}<\infty\}} \Psi \, dP \\
&= \sum_{l=k+1}^{\infty} \int_{\{O_j \in G_j,\, 1 \le j \le k,\, W_k \le t < W_{k+1},\, \xi = l\}} \Psi \, dP \\
&= \sum_{l=k+1}^{\infty} \alpha_l \int_{\{O_j \in G_j,\, 1 \le j \le k,\, W_k \le t < W_{k+1}\}} \Psi \, dP_l \\
&= \sum_{l=k+1}^{\infty} \alpha_l \int_{G_1} \ldots \int_{G_k} \int_{G_{t,k+1}(o_1,\ldots,o_k)} \Psi(o_1, \ldots, o_k) \, dF_{l,k+1}(o) \, dF_{l,k}(o_k) \, \ldots \, dF_{l,1}(o_1) \\
&= \sum_{l=k+1}^{\infty} \alpha_l \int_{G_1} \ldots \int_{G_k} \Psi(o_1, \ldots, o_k) \, \overline{A}_{l,k+1}\Big(t - \sum_{r=1}^{k} \pi_1(o_r)\Big) \, dF_{l,k}(o_k) \ldots dF_{l,1}(o_1) \\
&= \sum_{l=k+1}^{\infty} \alpha_l \int_{G_1} \ldots \int_{G_k} \Psi(o_1, \ldots, o_k) \, \overline{A}_{l,k+1}\Big(t - \sum_{r=1}^{k} \pi_1(o_r)\Big) \prod_{s=1}^{k} f_{l,s}(o_s) \, do_k \, \ldots \, do_1.
\end{aligned}$$

This together with (3.11) gives (3.9). □

Proof of Theorem 3.2. First notice that by Theorem 34.1 of Billingsley [2] it suffices to show the representation for all sets in the generator $\mathcal{H}_t$, which is closed under finite intersections.

First consider $t = 0$. Then $\mathcal{F}_0 = \{\emptyset, \Omega\}$ and on the left hand side we have $P(\xi = n \,|\, \mathcal{F}_0) = \alpha_n$. On the right hand side, because $I_{\{M(0)=k\}} = 0$ for $1 \le k \le n$, the second and third summand vanish. Furthermore, $I_{\{M(0)=0\}} = 1$, $\overline{A}_{l,1}(0) = P_l(V_1 > 0) = 1$ for all $l \in \mathbb{N}$ and $\sum_{l=1}^{\infty} \alpha_l = 1$, hence the result follows.

Now consider $t > 0$ and denote the right hand side by Ψ. To establish the theorem we have to show that

$$\int_H P(\xi = n \,|\, \mathcal{F}_t) \, dP \;\; = \;\; \int_H \Psi \, dP \,, \tag{3.12}$$

for all sets H in the generator $\mathcal{H}_t$ of $\mathcal{F}_t$.

The left hand side of equation (3.12) is evaluated in Lemma 3.3:

$$\int_H P(\xi = n \,|\, \mathcal{F}_t) \, dP \;\; = \;\; P(H \cap \{\xi = n\}) \,, \quad H \in \mathcal{H}_t.$$

It remains to calculate the right hand side of equation (3.12). Let $H = H_{t,0} = \{V_1 > t\} \in \mathcal{H}_{t,0}$, then Ψ is constant on $H_{t,0}$, more precisely,

$$\Psi \;\; = \;\; \frac{\alpha_n \, \overline{A}_{n,1}(t)}{\sum_{l=1}^{\infty} \alpha_l \, \overline{A}_{l,1}(t)} \quad \text{on } H_{t,0} \,.$$

By equation (3.10) $\int_H \Psi dP = \alpha_n \overline{A}_{n,1}(t)$, which is the same as (3.5).

Next, let $n \in \mathbb{N}$ and $H = H_{t,k} \in \mathcal{H}_{t,k}$ for $1 \le k \le n$. Then on the set $H_{t,k}$ we have $M(t) = k$ and Ψ is a function in $O_1, \ldots, O_k$ alone, namely

$$\Psi \;\; = \;\; \begin{cases} \dfrac{\alpha_n \prod_{j=1}^{k} f_{n,j}(O_j) \, \overline{A}_{n,k+1}(t - W_k)}{\alpha_k \prod_{s=1}^{k} f_{k,s}(O_s) \, + \sum_{l=k+1}^{\infty} \alpha_l \prod_{s=1}^{k} f_{l,s}(O_s) \, \overline{A}_{l,k+1}(t - W_k)} \,, & 1 \le k < n \,, \\[2ex] \dfrac{\alpha_n \prod_{j=1}^{n} f_{n,j}(O_j)}{\alpha_n \prod_{s=1}^{n} f_{n,s}(O_s) \, + \sum_{l=n+1}^{\infty} \alpha_l \prod_{s=1}^{n} f_{l,s}(O_s) \, \overline{A}_{l,n+1}(t - W_n)} \,, & k = n \,. \end{cases}$$

The right hand side of (3.12) is then equal to (3.9). Taking into account that for $k \neq n$,

$$\overline{A}_{n,k+1}\Big(t - \sum_{j=1}^{k} \pi_1(o_k)\Big) \;\; = \;\; \int_{G_{t,k+1}(o_1,\ldots,o_k)} f_{n,k+1}(o) \, do \,,$$

where $G_{t,k+1(o_1,\ldots,o_k)} = (t - \sum_{j=1}^{k} \pi_1(o_j), \infty) \times \mathbb{R}_+$, then (3.9) coincides with equation (3.7) or (3.6), respectively. $\square$

The next theorem shows how to predict the noise process by means of conditional expectation.

Theorem 3.5. *Let $X(\cdot)$ be a claim process as defined in (3.1). For $t > 0$ and $0 \le s < t$ let $\mathcal{F}_s$ denote the σ-field defined in (3.3) representing the observations up to time s. Then the best predictor of $X(t)$ is given by*

$$E\left(X(t) - X(s) \,|\, \mathcal{F}_s\right) = \sum_{n=M(s)+1}^{\infty} \sum_{m=M(s)+1}^{n} P\left(\xi = n \,|\, \mathcal{F}_s\right) E_n\left(Z_m\, I_{\{W_m \le t\}} \,\middle|\, \mathcal{F}_s\right) \quad (3.13)$$

with mean square error

$$E\left([X(t) - X(s)]^2 \,\middle|\, \mathcal{F}_s\right) = \sum_{n=M(s)+1}^{\infty} P\left(\xi = n \,|\, \mathcal{F}_s\right) \times$$

$$\times \left(\sum_{m=M(s)+1}^{n} E_n\left(Z_m^2\, I_{\{W_m \le t\}} \,\middle|\, \mathcal{F}_s\right) + 2 \sum_{M(s)+1 \le i < j \le n} E_n\left(Z_i\, Z_j\, I_{\{W_j \le t\}} \,\middle|\, \mathcal{F}_s\right) \right). \quad (3.14)$$

Under the additional assumption that for $M(s)+1 \le m \le n$, V_m and Z_m are conditionally independent given $\xi = n$, then for $k \in \mathbb{N}$ such that the corresponding moment exists,

$$E_n\left(Z_m^k\, I_{\{W_m \le t\}} \,\middle|\, \mathcal{F}_s\right) = E_n\left(Z_m^k\right) P_n\left(W_m \le t \,|\, \mathcal{F}_s\right) \quad (3.15)$$

and, for $M(s) + 1 \le i < j$,

$$E_n\left(Z_i\, Z_j\, I_{\{W_j \le t\}} \,\middle|\, \mathcal{F}_s\right) = E_n\left(Z_i\right) E_n\left(Z_j\right) P_n\left(W_j \le t \,|\, \mathcal{F}_s\right). \quad (3.16)$$

Proof. First note that

$$\begin{aligned} E\left(X(t) - X(s) \,|\, \mathcal{F}_s\right) &= \sum_{n=M(s)+1}^{\infty} P\left(\xi = n \,|\, \mathcal{F}_s\right) E_n\left(X(t) - X(s) \,|\, \mathcal{F}_s\right) \\ &= \sum_{n=M(s)+1}^{\infty} P\left(\xi = n \,|\, \mathcal{F}_s\right) E_n\left(\sum_{m=M(s)+1}^{n} Z_m\, I_{\{W_m \le t\}} \,\middle|\, \mathcal{F}_s \right) \\ &= \sum_{n=M(s)+1}^{\infty} \sum_{m=M(s)+1}^{n} P\left(\xi = n \,|\, \mathcal{F}_s\right) E_n\left(Z_m\, I_{\{W_m \le t\}} \,\middle|\, \mathcal{F}_s\right). \end{aligned}$$

The proof of (3.14) is similar using the simple fact that $(\sum_{i=1}^{n} x_i)^2 = \sum_{i=1}^{n} x_i^2 + 2\sum_{1 \le i < j \le n} x_i x_j$. Now assume conditional independence of V_m and Z_m given ξ. Then for $M(s)+1 \le m \le n$ and $k \in \mathbb{N}$ sucht that $E_n(Z_m^k) < \infty$,

$$E_n\left(Z_m^k\, I_{\{W_m \le t\}} \,\middle|\, \mathcal{F}_s\right) = E_n\left(Z_m^k\right) P_n\left(W_m \le t \,|\, \mathcal{F}_s\right).$$

The term $E_n\left(Z_i\, Z_j\, I_{\{W_j\le t\}}\,\middle|\,\mathcal{F}_s\right)$ simplifies analogously for $M(s)+1\le i<j\le n$. $\square$

Theorem 3.5 for $s=0$ gives the following corollary.

Corollary 3.6. *Let $X(\cdot)$ be the process given by (3.1). Under the assumption of conditional independence of V_m and Z_m given ξ we have for $t\ge 0$,*

$$\begin{aligned} EX(t) &= \sum_{n=1}^{\infty}\alpha_n\sum_{m=1}^{n}E_n\left(Z_m\right)\,P_n\left(W_m\le t\right)\\ E(X(t))^2 &= \sum_{n=1}^{\infty}\alpha_n\left(\sum_{m=1}^{n}E_n\left(Z_m^2\right)\,P_n\left(W_m\le t\right)+2\sum_{1\le i<j\le n}E_n\left(Z_i\right)E_n\left(Z_j\right)\,P_n\left(W_j\le t\right)\right)\end{aligned}$$

Example 3.7. (a) Assume that all inter-instalment times are i.i.d. exponential random variables with same parameter $\lambda>0$, i.e. $P(V>v)=1-e^{-\lambda v}$, $v>0$, independent of the state of the HMM. Setting $H_{M(s)+1}=W_{M(s)+1}-s$ and $H_j=V_j$, $j=M(s)+2,\ldots,m$, by the no memory property of the exponential distribution, under the measure P_n for given $M(s)$ the random variables H_j are i.i.d. exponentially distributed with rate $\lambda>0$. Consequently equations (3.15) and (3.16) simplify due to

$$\begin{aligned} P_n(W_m\le t\,|\,\mathcal{F}_s) &= P_n\left(\sum_{j=1}^{m}V_j\le t\,\middle|\,\mathcal{F}_s\right)\\ &= P_n\left(W_{M(s)+1}+\sum_{j=M(s)+2}^{m}V_j\le t\,\middle|\,\mathcal{F}_s\right)\\ &= P_n\left(s+(W_{M(s)+1}-s)+\sum_{j=M(s)+2}^{m}V_j\le t\,\middle|\,\mathcal{F}_s\right)\\ &= P_n\left(\sum_{j=M(s)+1}^{m}H_j\le t-s\,\middle|\,M(s)\right).\end{aligned}$$

(b) Exponential distributions are special phase-type distributions. For this class of distributions $P_n(W_m\le t)$ can still be calculated explicitly. For details see Severin [20], Section 3.3.

4 Model specification

Prediction of outstanding liabilities requires specification of the initial distribution $(\alpha_n)_{n\in\mathbb{N}}$ of the HMM and the conditional distribution $f_{n,k}$ of the kth observations O_k (instalment and inter-instalment time), given $\xi = n$. For the rest of this paper, we assume conditional independence of the instalment and inter-instalment time, given $\xi = n$, i.e.

$$f_{n,k}(v,z) = h_{n,k}(v)\, g_{n,k}(z), \quad v, z > 0\,, \quad 1 \le k \le n\,, \tag{4.17}$$

where $g_{n,k}$ and $h_{n,k}$ are conditional densities of the instalments and inter-instalment times, respectively.

Next we investigate whether claims with different total numbers of instalments differ. We investigate the inter-instalment times and the instalments separately for the different classes.

4.1 The conditional distribution of the inter-instalment times

Figures 6 and 7 show the QQ-plots (versus the standard exponential distribution) of the inter-instalment times classified by the total number of instalments; i.e. for a total number n of instalments the subsamples of the k-th inter-instalment times for $1 \le k \le n$ are investigated separately. Thereby the intensities $\lambda_{n,k}$ of the exponential distributions are estimated by maximum likelihood estimators (MLEs).

As a compromise between simplicity and sophistication of the model we settled on the following. All inter-instalment times are modeled by an exponential distribution, the intensity, however changes. It seems at first sight natural to compare all first inter-instalment times, all second ones (provided there exists more than one) etc. However, the slopes of the fitted lines are quite different. There is a more systematic similiarity, when comparing all final inter-instalment times, all second last ones (provided there is more than one) etc. Table 2 is ordered to make this more transparent.

A further investigation of the estimated intensities in this table indicates, firstly, an increasing pattern in n for fixed k and, secondly, a decreasing pattern in k for fixed n. For fixed k, i.e. for the last inter-instalment times, for the second last etc., the mean

n	$\widehat{\lambda}_{n,n}$	$\widehat{\lambda}_{n,n-1}$	$\widehat{\lambda}_{n,n-2}$	$\cdots$		$\cdots$	$\widehat{\lambda}_{n,1}$
1	0.00501	–	–	–	–	–	–
2	0.00502	0.01013	–	–	–	–	–
3	0.00489	0.00966	0.01323	–	–	–	–
4	0.00503	0.00782	0.01403	0.01561	–	–	–
5	0.00405	0.00740	0.01082	0.01520	0.01809	–	–
6	0.00397	0.00581	0.01186	0.01302	0.02025	0.01943	–
7	0.00363	0.00643	0.00849	0.01337	0.01296	0.02182	0.02073
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$

Table 2: Estimated intensities $\widehat{\lambda}_{n,k}, 1 \leq k \leq n$, of inter-instalment times related to states of the Markov chain.

inter-instalment times decrease in n (the more instalments, the smaller are the mean inter-instalment times). For fixed n, i.e. the number of instalments fixed, the mean inter-instalment times increase in k (the first mean inter-instalment times are smaller than later ones). This leads to the following model:

$$\sqrt{\lambda_{n,k}} = a + bk + cn\,, \quad 1 \leq k \leq n\,. \tag{4.18}$$

Hence our model specification is

$$h_{n,k}(v) = (a + bk + cn)^2\, e^{-(a+bk+cn)^2\, v}, \quad v > 0.$$

Fitting the model to the intensities given in Table 2 by a simple regression estimate we obtain the estimates in Table 3, which we shall use as initial values for the more sophisticated estimation procedure in Section 5.

Note that none of the 95% confidence intervals contains the value 0, hence a reduction of the model seems not feasible.

4.2 The conditional distribution of the instalments

Figures 8 and 9 show the histograms of the log-instalments classified by the total number of instalments; i.e. for a total number n of instalments the subsamples of the k-th logarithmic instalments for $1 \leq k \leq n$, are investigated separately.

	LSE	std. dev.	$CI_{0.95}$
$\widehat{a}$	0.0867	0.0044	(0.0776, 0.0958)
$\widehat{b}$	-0.0164	0.0010	(-0.0185, -0.0143)
$\widehat{c}$	0.0128	0.0010	(0.0107, 0.0149)

Table 3: Least squares estimates (LSE) of the intensity parameters a, b and c with standard errors (std. dev.) and asymptotic 95% confidence intervals ($CI_{0.95}$).

One common pattern is the bimodality of most of the histograms. At first we supposed that a rather regular small sum might occur right at the occurrence of most claims, as e.g. a fine requested by the police. However, this suggestion is obviously not true as one glance at the histograms $(3,2),(4,2),(5,2),(5,3)$ etc. shows. ¿From the histograms it is certainly not obvious, what models to choose. A further problem is that for claims with larger numbers of instalments there are not enough data available to provide a sensible model.

Assuming all instalments to be i.i.d., a histogram of all log-instalments in Figure 2 shows clearly the bimodality, but also suggests a normal model for the log-instalments, i.e. we model the instalments by a lognormal mixture distribution. Lognormal models are very common in the statistical analysis of insurance data. Consequently, we choose a normal mixture model for the log-instalments, where we investigated further the choice of parameters for claims with different numbers of instalments. Figure 10 visualises the outcome: log-means and log-variances of the different instalments seem not to vary so much as to justify a much more complicated model.

Hence, also for reasons of parsimony, we choose the same lognormal mixture model for all instalments; i.e. for all $1 \le i \le n$,

$$g_{n,i}(z) = g(z) = \frac{1}{z}\left(p\,\varphi_{\mu_1,\sigma_1^2}(\log z) + (1-p)\,\varphi_{\mu_2,\sigma_2^2}(\log z)\right), \quad z > 0, \tag{4.19}$$

where φ_{μ,σ^2} is the density function of a normal distribution with mean μ and variance σ^2.

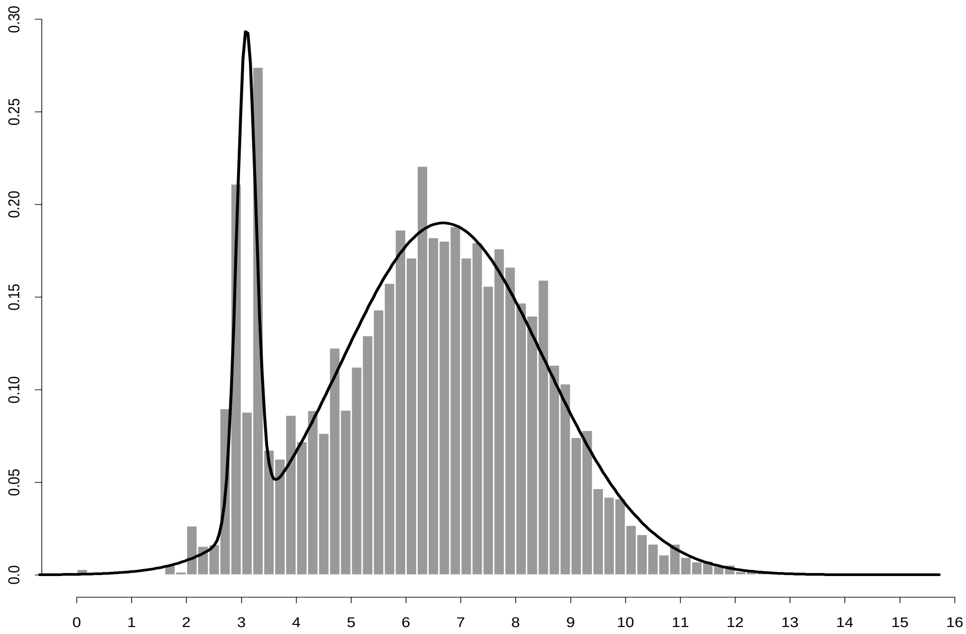

Figure 2: Histogram of all log-instalments and fitted density of bimodal normal mixture model.

4.3 The initial distribution of the Markov chain

For our data, each claim has a certain total number of instalments reaching from 1 to 48, the frequencies of the number of instalments are given in Table 4. Figure 3 shows the histogram of the total number of instalments for 1 to 10 instalments.

# of instalments	1	2	3	4	5	6
abs. frequency	1966	1343	840	525	313	184
rel. frequency	0.3461	0.2364	0.1479	0.0924	0.0551	0.0324
# of instalment	7	8	9	10	...	48
abs. frequency	115	75	61	52	...	1
rel. frequency	0.0202	0.0132	0.0107	0.0092	...	0.0002

Table 4: Partition of the portfolio by the number of instalments.

In predicting the costs of future payment for this portfolio, an important issue is the prediction of the ultimate number of instalments for a running, not yet closed claim. Due

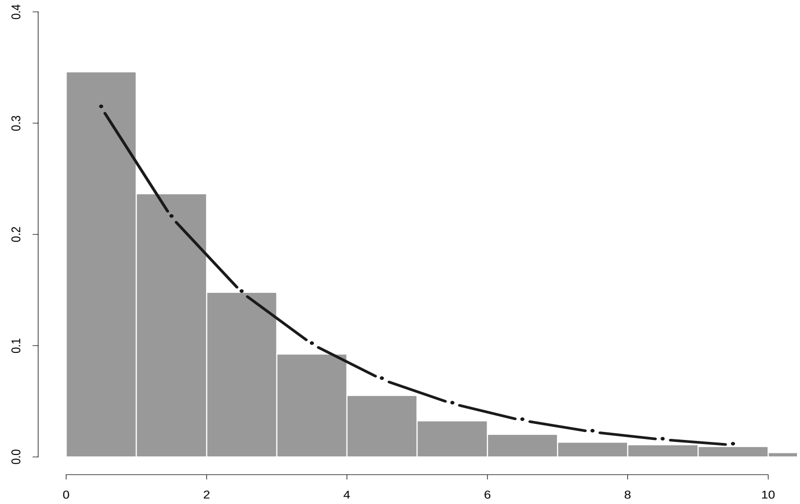

Figure 3: Histogram, corresponding to Table 4 with fitted density of truncated geometric distribution (4.20).

to the fact that we have only claims which occurred during one single year, we do not have enough data to fit a sophisticated model. Consequently, we model the total number of instalments by a truncated geometric random variable, which fits quite well at least to that part of the histogram, where enough data are available; see Figure 3:

$$\alpha_n = P(\xi = n) = \frac{q^{n-1}(1-q)}{1-q^{48}}\,, \quad n = 1, \ldots, 48\,, \quad 0 < q < 1\,. \tag{4.20}$$

We are well aware of the fact that fixing the support of the distribution to the realized $N = 48$ instalments is not optimal. Joint maximum likelihood estimation of q and N, however, leads to highly unstable parameters, moreover, the estimator of N will not be an integer. There are remedies for this problem, see e.g. the discussion in Johnson, Kotz and Kemp [9], Section 8.2, in connection to the binomial distribution. We decided at this point to restrict ourselves to the observed $N = 48$ and estimated q by maximum likelihood estimation, resulting in an initial estimate $\widehat{q} = 0.5348$.

4.4 The intensity function of the claim number process

Figure 4 shows the daily numbers of claims per day during the year 1985 of occurrency of the claims.

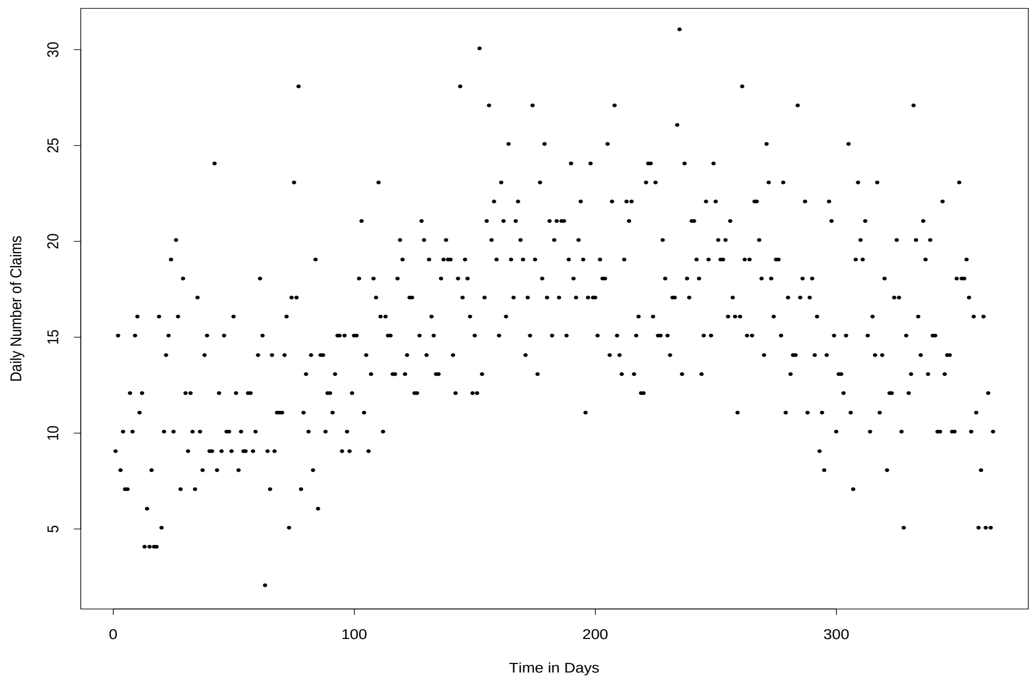

Figure 4: Time series plot of the daily number of claims Δ_i, $i = 1, \ldots, 365$, during 1985.

We think of these data as (over the day) aggregated realisations of the inhomogeneous Poisson process N; i.e. if

$$\Delta_i := \#\{j \mid T_j \in (i-1, i]\}, \quad i = 1, \ldots, 365,$$

then we interpret $\Delta_i = N(i) - N(i-1)$ as the increment on day i. The obvious seasonality over the year will be captured by a deterministic *intensity function* $\lambda : \mathbb{R}_+ \to \mathbb{R}_+$; i.e. we assume for $0 < s < t$,

$$P(N(t) - N(s) = n) = \exp(-(\Lambda(t) - \Lambda(s))) \frac{(\Lambda(t) - \Lambda(s))^n}{n!}, \quad n \in \mathbb{N}_0,$$

where

$$\Lambda(t) = \int_0^t \lambda(s) ds, \quad t \geq 0.$$

Since we have only one year of aggregated data to fit the model, the choice of model is rather limited and we have fitted a quadratic model for the intensity function. The problem of estimating the unknown model paramters by observations Δ_i with mean $\mu_i = \int_{i-1}^{i} \lambda(t)dt$ is circumvented by the quadratic model and by the following result, whose proof is a simple application of the binomial formula.

Proposition 4.1. *For some $n \in \mathbb{N}$ let $\lambda(t) = \sum_{j=0}^{n} \beta_j \, t^j$, $t \geq 0$ and define $\mu_i = \int_{i-1}^{i} \lambda(t)\,dt$, $i \in \mathbb{N}$. Then $\mu_i = \sum_{k=0}^{n} \alpha_k \, i^k$ and the relationship of the coefficient vectors $\alpha = (\alpha_0,\, \alpha_1, \ldots, \alpha_n)'$ and $\beta = (\beta_0,\, \beta_1, \ldots, \beta_n)'$ is given by*

$$\alpha \;=\; A_n\, \beta$$

with the regular matrix $A_n = (a_{k,j})_{0 \leq k,j \leq n}$ given by

$$a_{k,j} \;=\; \begin{cases} \binom{j}{k} \dfrac{(-1)^{j-k}}{j-k+1}, & k \leq j, \\ 0, & k > j. \end{cases}$$

We analysed the data, however, additionally for weekday effects. To this end we fitted a model capturing a basic quadratic time effect, but we allowed also for additional weekday effects; i.e. we investigated the model

$$\lambda(t) = \beta_0 + \beta_1 t + \beta_2 t^2 + \sum_{wd=1}^{6} (\beta_{0wd} + \beta_{1wd} t + \beta_{2wd} t^2) \delta_{wd}(t)\,, \quad t \geq 0\,, \tag{4.21}$$

where the dummy variables $\delta_{wd}(t)$ are given by

$$\delta_{wd}(t) \;=\; \begin{cases} 1, & \text{if time } t \text{ is at weekday } wd, \\ 0, & \text{else.} \end{cases}$$

We fitted the model using a linear regression model for Δ_i, $i = 1, \ldots, 365$:

$$\begin{aligned} \Delta_i &= \mu_i + \varepsilon_i \\ &= \alpha_0 + \alpha_1 i + \alpha_2 i^2 + \sum_{wd=1}^{6} (\alpha_{0wd} + \alpha_{1wd} i + \alpha_{2wd} i^2)\, \delta_{wd}(i) + \varepsilon_i\,, \end{aligned}$$

where the ε_i are i.i.d. $N(0, \sigma^2)$.

By a series of F-tests and a careful analysis of residuals using standard model selection criteria the model (4.21) simplifies to

$$\lambda(t) = \beta_0 + \beta_1 t + \beta_2 t^2 + \sum_{wd=1}^{3} \beta_{0wd}\delta_{wd}(t)\,, \quad t \geq 0\,,$$

where we have used Proposition 4.1 for the transition from Δ_i to $\lambda(t)$. The value of the variable wd identifies the following weekdays

$$wd = \begin{cases} 1, & \text{Mondays or Tuesdays,} \\ 2, & \text{Wednesday or Thursdays} \\ 3, & \text{Fridays or Saturdays.} \end{cases}$$

The regression coefficients and standard errors are shown in Table 5.

	LSE		std. dev.			LSE	std. dev.	
α_0	5.778		8.760	10^{-1}	α_{01}	1.353	7.280	10^{-1}
α_1	1.018	10^{-1}	8.553	10^{-3}	α_{03}	2.173	7.292	10^{-1}
α_2	-2.410	10^{-4}	2.263	10^{-3}	α_{05}	3.116	7.292	10^{-1}

Table 5: Least squares estimates (LSE) for the αs.

The fitted model $\widehat{\Delta}_i$, $i = 1, \ldots, 365$, is shown in Figure 5. For more details see Severin [20], Chapter 6.

5 Parameter estimation

5.1 Likelihood specification

By the model specification of the previous sections the parameter vector θ of the hidden Markov model is given by

$$\theta = (q, a, b, c, p, \mu_1, \sigma_1, \mu_2, \sigma_2) = (\vartheta, \vartheta_g), \tag{5.22}$$

where q characterizes the initial probabilities $(\alpha_n)_{n\in\mathbb{N}}$ of the Markov chain $(\xi_n)_{n\in\mathbb{N}}$; see (4.20), a, b and c specify the intensities $\lambda_{n,i}$, $1 \leq i \leq n$, of the conditional distributions of inter-instalment times; see (4.18), and, finally, $\vartheta_g = (p, \mu_1, \sigma_1, \mu_2, \sigma_2)$ contains the parameters of the density of instalments; see (4.19).

Since according to model (1.1) we suppose that all claims occur independently, the joint likelihood with respect to all claims factorizes into the product of the single claims likelihood contributions; so it is sufficient to specify the likelihood contribution of a single claim. Obviously, by Assumption 3.1, the likelihood contribution based on the complete data is given by

$$\begin{aligned} L_c(\theta) &= L_c(\theta; M(t), O_1, \ldots, O_{M(t)}, \xi_1, \ldots, \xi_{M(t)+1}) \\ &= \alpha_{\xi_1} \left\{ I_{\{\xi > M(t)\}} \prod_{i=1}^{M(t)} f_{\xi_i}(O_i)\, \overline{A}_{\xi_{M(t)+1}} \left(t - W_{M(t)}\right) + I_{\{\xi = M(t)\}} \prod_{i=1}^{M(t)} f_{\xi_i}(O_i) \right\}, \end{aligned}$$

where, for instance, the function $I_{\{\xi=M(t)\}}$ indicates that the claim is settled. By (4.17), this factorizes in two independent parts:

$$L_c(\theta) = L_c^{(1)}(\vartheta_g; M(t), Z_1, \ldots, Z_{M(t)})\, L_c^{(2)}(\vartheta; M(t), V_1, \ldots, V_{M(t)}, \xi_1, \ldots, \xi_{M(t)+1}),$$

where (we suppress the random variables)

$$L_c^{(1)}(\vartheta_g) = \prod_{i=1}^{M(t)} \frac{1}{Z_i} \left(p\, \varphi_{\mu_1, \sigma_1^2}(\log Z_i) + (1-p)\, \varphi_{\mu_2, \sigma_2^2}(\log Z_i) \right) \tag{5.23}$$

is the likelihood contribution of the observed instalments and the complete data likelihood of the observed inter-instalment times is given by

$$\begin{aligned} L_c^{(2)}(\vartheta) &= \alpha_{\xi_1}(q) \left\{ I_{\{\xi > M(t)\}} \prod_{i=1}^{M(t)} h_{\xi_i}(V_i; a, b, c)\, \overline{A}_{\xi_{M(t)+1}} \left(t - W_{M(t)}; a, b, c\right) \right. \\ &\qquad \left. + I_{\{\xi = M(t)\}} \prod_{i=1}^{M(t)} h_{\xi_i}(V_i; a, b, c) \right\}. \end{aligned}$$

5.2 Fitting the instalments

We estimate the distribution of the instalment sizes by maximizing the likelihood function (5.23). Based on a total of $N = 8\,387$ instalments the MLE of the mixture parameter p, the two means μ_i and standard deviations σ_i, $i = 1, 2$, are given in Table 6.

The normal mixture model satisfies the regularity conditions giving asymptotic normality and efficiency of the estimated parameters; see Lehmann [13], Theorem 4.1, p.429

	MLE	std. dev	$CI_{0.95}$
$\widehat{p}$	0.1770	0.0049	(0.1674, 0.1866)
$\widehat{\mu}_1$	3.0980	0.0058	(3.0866, 3.1094)
$\widehat{\sigma}_1$	0.1877	0.0057	(0.1765, 0.1989)
$\widehat{\mu}_2$	6.3958	0.0227	(6.3513, 6.4403)
$\widehat{\sigma}_2$	1.6889	0.0162	(1.6571, 1.7207)

Table 6: MLEs and observed standard deviations (std. dev.) of the parameters of the instalment distribution with asymptotic 95% confidence intervals.

and McLachlan and Basford [16], Section 2.1. More precisely,

$$\sqrt{N}(\widehat{\vartheta}_g - \vartheta_g) \to N(0, I(\vartheta_g)^{-1}), \quad N \to \infty,$$

where $I(\vartheta_g)^{-1}$ is the inverse of the (expected) information matrix $I(\vartheta_g)$. By the law of large numbers we may replace $I(\vartheta_g)^{-1}$ by the (observed) information matrix

$$H(\widehat{\vartheta}_g)^{-1} = \frac{1}{10^7}\begin{pmatrix} 237.36 & -3.31 & 80.86 & 255.86 & -187.90 \\ -3.31 & 337.83 & -10.28 & -34.88 & 52.04 \\ 80.86 & -10.28 & 321.64 & 326.22 & -243.55 \\ 255.86 & -34.88 & 326.22 & 5165.33 & -764.12 \\ -187.90 & 52.04 & -243.55 & -764.12 & 2638.61 \end{pmatrix}. \tag{5.24}$$

Correlations deduced from (5.24) are given in Table 7.

	$\widehat{p}$	$\widehat{\mu}_1$	$\widehat{\sigma}_1$	$\widehat{\mu}_2$	$\widehat{\sigma}_2$
$\widehat{p}$	1.0000	-0.0117	0.2926	0.2311	-0.2374
$\widehat{\mu}_1$	-0.0117	1.0000	-0.0312	-0.0264	0.0551
$\widehat{\sigma}_1$	0.2926	-0.0312	1.0000	0.2531	-0.2644
$\widehat{\mu}_2$	0.2311	-0.0264	0.2531	1.0000	-0.2070
$\widehat{\sigma}_2$	-0.2374	0.0551	-0.2644	-0.2070	1.0000

Table 7: Correlations of estimated parameters given in Table 6.

5.3 Fitting the inter-instalment times by the EM algorithm

Since the distributions of the inter-instalment times depend on the state of the non-observable Markov chain, an EM algorithm is used to estimate the remaining model parameters $\vartheta = (q, a, b, c)$ simultaneously. For details on the EM algorithm we refer to Dempster, Laird and Rubin [4], McLachlan and Krishnan [17] and Wu [22]; the EM algorithm in a HMM setting is also known as Baum-Welch algorithm; cf. Baum et al. [1].

The idea of the EM alogrithm is to provide a sequence $(\vartheta^{(j)})_{j\geq 0}$ of estimates which converges to a MLE, say $\vartheta^\star$. Each of the $\vartheta^{(j)}$ is constructed by performing a so-called expectation step (E-step) and a maximization step (M-step) iteratively, where we take the initial values $\vartheta^{(0)} = (\widehat{q}, \widehat{a}, \widehat{b}, \widehat{c})$ from the end of Section 4.3 and Table 3.

To be more specific, let $V = (M(t), V_1, \ldots, V_{M(t)})$ be the random vector of observable data, having likelihood function $L(\vartheta)$, where ϑ is the corresponding parameter vector in the parameter space $[0,1] \times \mathbb{R}^3$. The observation vector V is viewed as being incomplete in the sense that maximum likelihood estimation is made difficult by the absence of some part of the data and is thus called the incomplete data part. Rather than performing a maximization of this incomplete data likelihood $L(\vartheta)$, one augments the observed data V with unobservable data ξ (also called latent data or hidden data). The augmented vector $X = (V, \xi)$ is called the complete data vector and its likelihood function is denoted by $L_c(\vartheta)$.

Formally there exist two sample spaces $\mathcal{X}$ and $\mathcal{V}$ and a many-to-one mapping V from $\mathcal{X}$ to $\mathcal{V}$. Instead of observing the complete data X in $\mathcal{X}$ one observes the incomplete data $V = V(X)$ in $\mathcal{V}$ and thus the likelihoods of X and V are related by

$$\begin{aligned}
L(\vartheta; v) &= \int_{\mathcal{X}(v)} L_c(\vartheta; x)\, dx, \qquad (5.25)\\
&= \alpha_{M(t)} \prod_{i=1}^{M(t)} h_{M(t),i}(V_i) + \sum_{n=M(t)+1}^{\infty} \alpha_n \prod_{i=1}^{M(t)} h_{n,i}(V_i)\, \overline{A}_{n+1,i}(t - W_{M(t)})\\
&= \alpha_{M(t)} \prod_{i=1}^{M(t)} \lambda_{M(t),i} \exp\left(-\sum_{i=1}^{M(t)} \lambda_{M(t),i} V_i\right)\\
&\quad + \sum_{n=M(t)+1}^{\infty} \alpha_n \prod_{i=1}^{M(t)} \lambda_{n,i} \exp\left(-\sum_{i=1}^{M(t)} \lambda_{n,i} V_i - \lambda_{n+1,i}(t - W_{M(t)})\right),
\end{aligned}$$

where $\mathcal{X}(v) = \{x \mid v(x) = v\}$.

If the hidden variable ξ was observable, the complete data likelihood as a function in the unknown parameter ϑ would be given by $L_c(\vartheta)$. The EM algorithm solves the problem of maximizing the incomplete data likelihood $L(\vartheta)$ indirectly and iteratively in terms of the complete data likelihood $L_c(\vartheta)$. This means the EM algorithm finds the value of ϑ, say $\vartheta^\star$, that maximizes $L(\vartheta)$, that is the MLE for ϑ based on the observed data V. Because $L_c(\vartheta)$ is not observable it is replaced by its conditional expectation given the observations and the latest fit for ϑ.

On the basis of $\vartheta^{(j)}$, the so-called updated parameter estimator $\vartheta^{(j+1)}$ is calculated as follows: in the E-step, the conditional expectation of the complete data log-likelihood, $\log L_c(\vartheta)$, with respect to the distribution $\left(P_{\vartheta^{(j)}}\left(\xi = n \mid \mathcal{F}_t\right)\right)_{n\geq 1}$ is calculated (as usual we work with the log-likelihood function):

$$\begin{aligned} E_{\vartheta^{(j)}}\left[\log L_c(\vartheta) \middle| \mathcal{F}_t\right] &= \sum_{k=0}^{\infty} I_{\{M(t)=k\}} \sum_{n=k}^{\infty} P_{\vartheta^{(j)}}\left(\xi = n \middle| \mathcal{F}_t\right) \Bigg\{ \log \alpha_n(q) \\ &\quad + \sum_{i=1}^{k} \log \lambda_{n,i} - \lambda_{n,i} V_i - (1-\delta_{n,k})\, \lambda_{n,k+1}\,(t - W_k) \Bigg\}. \end{aligned} \tag{5.26}$$

According to Theorem 3.2, the conditional distribution of the initial state ξ of the Markov-chain, given the information up to time t, and the initial estimator $\vartheta^{(j)}$ modifies to

$$\begin{aligned} P_{\vartheta^{(j)}}\left(\xi = n \mid \mathcal{F}_t\right) &= I_{\{M(t)=0\}} \frac{\alpha_n\, e^{-\lambda_{n,1} t}}{\sum_{l=1}^{\infty} \alpha_l\, e^{-\lambda_{l,1} t}} \\ &+ \sum_{k=1}^{n-1} I_{\{M(t)=k\}} \frac{\alpha_n \prod_{l=1}^{k} h_{n,l}(V_l)\, \overline{A}_{n,k+1}\,(t - W_k)}{\alpha_k \prod_{s=1}^{k} h_{k,s}(V_s) + \sum_{l=k+1}^{\infty} \alpha_l \prod_{s=1}^{k} h_{l,s}(V_s)\, \overline{A}_{l,k+1}\,(t - W_k)} \\ &+ I_{\{M(t)=n\}} \frac{\alpha_n \prod_{l=1}^{n} h_{n,l}(V_l)}{\alpha_n \prod_{l=1}^{n} h_{n,l}(V_l) + \sum_{l=n+1}^{\infty} \alpha_l \prod_{s=1}^{n} h_{l,s}(V_s)\, \overline{A}_{l,n+1}(t - W_k)}, \end{aligned}$$

where all the α, λ, h and A depend on the current fit of q, a, b, c, i.e. they depend on $\vartheta^{(j)}$. Moreover,

$$\prod_{l=1}^{k} h_{n,l}(V_l) = \prod_{l=1}^{k} \lambda_{n,l}\, \exp\left(-\sum_{l=1}^{k} \lambda_{n,l}\, V_l \right), \quad k = 1, \ldots, n,$$

and

$$\prod_{l=1}^{k} h_{n,l}(V_l)\, \overline{A}_{n,k+1}\,(t - W_k) = \prod_{l=1}^{k} \lambda_{n,l}\, \exp\left(-\sum_{l=1}^{k} \lambda_{n,l}\, V_l - \lambda_{n,k+1}\,(t - W_k) \right).$$

Expectation (5.26) gives the contribution of one single claim and hence, because all claims are assumed to occur independently, the conditional expectation of the complete data log-likelihood for all observations is given by

$$Q(\vartheta, \vartheta^{(j)}) = \sum_{i=1}^{5681} E^{(i)}_{\vartheta^{(j)}} \left[\log L_c(\vartheta) \,|\, \mathcal{F}_t \right], \tag{5.27}$$

where the upper index i indicates the contribution of the ith claim.

In the M-step $Q(\vartheta, \vartheta^{(j)})$ is maximized with respect to ϑ over the parameter space $\Theta = [0,1] \times \mathbb{R}^3$. This means finding a parameter $\vartheta^{(j+1)} \in \Theta$ such that

$$Q(\vartheta^{(j+1)}, \vartheta^{(j)}) \;\; \geq \;\; Q(\vartheta, \vartheta^{(j)}), \quad \text{for all } \vartheta \in \Theta.$$

In the next step $\vartheta^{(j)}$ is replaced by the updated EM iterate $\vartheta^{(j+1)}$ and $Q(\vartheta, \vartheta^{(j+1)})$ is set up. The E- and M-steps are performed until, for instance the change in the sequence of EM iterates is sufficiently small.

Table 8 gives the EM iterates of q, a, b and c by applying the EM algorithm to the portfolio after a one year observation period. The initial values for the EM algorithm correspond to step 0 in Table 8. After the 54th iteration the values are stable up to the

jth step	$q^{(j)}$	$a^{(j)}$	$b^{(j)}$	$c^{(j)}$
0	0.5348	0.0867	-0.0164	0.0128
1	0.7442	0.0929	-0.0148	0.0143
2	0.7680	0.0843	-0.0126	0.0144
3	0.7741	0.0799	-0.0116	0.0146
4	0.7761	0.0777	-0.0112	0.0147
⋮	⋮	⋮	⋮	⋮
54	0.7642	0.0749	-0.0104	0.0158

Table 8: Iterates of model parameters after a one year observation period.

fourth decimal. Since $Q(\vartheta, \vartheta^{(j)})$ of (5.27) is continuous in both variables, we can apply Wu [22], Theorem 2, to show that the limit point of $\left(\vartheta^{(j)}\right)_{j \geq 0}$ is a MLE of (5.25).

For this result, Wu [22] assumes mild regularity (continuity and compactness) conditions. By (5.25) and the fact that all conditional densities of inter-instalment times are continuous functions in a, b, c, the likelihood function $L(\vartheta)$ is a continuous function in ϑ. Morover, the EM iterates given in Table 8 are evidently converging to some finite values, not approaching the boundary of Θ. For more details see Severin [20], Chapter 5.

A drawback of the EM algorithm in practice is that asymptotic variance-covariance matrices for parameters (e.g. standard errors) are not automatic by-products as they are in the standard situation. Louis [14] showed, by applying the "missing information principle" of Orchard and Woodbury [18], how to compute the observed infomation matrix for missing information problems when using the EM algorithm.

To be more precise, let $S(\vartheta) = \partial \log L(\vartheta)/\partial\vartheta$ and $S_c(\vartheta) = \partial \log L_c(\vartheta)/\partial\vartheta$ be the score statistic based on the observed (incomplete) data and complete data, respectively. Louis [14] proved that $S(\vartheta)$ can be expressed as the conditional expectation of $S_c(\vartheta)$ given the observed information $\mathcal{F}_t$, i.e.

$$S(\vartheta) = E_\vartheta \left(S_c(\vartheta) \mid \mathcal{F}_t \right).$$

Let $H(\vartheta)$, with $H(\vartheta)_{i,j} = -\partial^2 \log L(\vartheta)/\partial\vartheta_i\partial\vartheta_j$ and $H_c(\vartheta)$, with $H_c(\vartheta)_{i,j} = -\partial^2 \log L_c(\vartheta)/\partial\vartheta_i\partial\vartheta_j$ be the matrices of the negatives of the second-order derivatives of the incomplete data log-likelihood and complete data log-likelihood, respectively. Again by Louis [14],

$$H(\vartheta) = E_\vartheta \left(H_c(\vartheta) \mid \mathcal{F}_t \right) - E_\vartheta \left(S_c(\vartheta)\, S_c(\vartheta)' \mid \mathcal{F}_t \right) + S(\vartheta)\, S(\vartheta)'.$$

Hence, because the MLE satisfies $S(\vartheta^\star) = 0$, the observed (incomplete) information matrix of the MLE $\vartheta^\star$ is given by $H(\vartheta^\star)^{-1}$, where

$$H(\vartheta^\star) = E_{\vartheta^\star} \left(H_c(\vartheta^\star) \mid \mathcal{F}_t \right) - E_{\vartheta^\star} \left(S_c(\vartheta^\star)\, S_c(\vartheta^\star)' \mid \mathcal{F}_t \right),$$

In our case of independent claim observations, $S_c(\vartheta^\star) = \sum_{i=1}^{5681} S_c^{(i)}(\vartheta^\star)$, where $S_c^{(i)}(\vartheta^\star)$ denotes the contribution of the ith claim to the score at $\vartheta^\star$, computation of the quantity $E_{\vartheta^\star} \left(S_c(\vartheta^\star)\, S_c(\vartheta^\star)' \mid \mathcal{F}_t \right)$ becomes rather costly, cf. Louis [14].

Jamshidian and Jennrich [6] suggest an alternative algorithms to solve this problem by applying a Richardson extrapolation method to differentiate $S(\vartheta)$ at $\vartheta^\star$ numerically in order to get an estimate of $H(\vartheta^\star)$ and call this the *NDS (Numerical Differentiation of Score) algorithm.*

If we set

$$\widehat{\vartheta} = \vartheta^{(54)} = (q^{(54)}, a^{(54)}, b^{(54)}, c^{(54)})', \tag{5.28}$$

the *NDS estimate of the variance-covariance* matrix of $\widehat{\vartheta}$ according to Jamshidian and Jennrich [6] is given by

$$\widetilde{H}(\widehat{\vartheta})^{-1} = \frac{1}{10^5} \begin{pmatrix} 11.13 & 0.80 & 0.06 & -1.01 \\ 0.80 & 0.56 & 0.01 & -0.18 \\ 0.06 & 0.01 & 0.16 & -0.08 \\ -1.01 & -0.18 & -0.08 & 0.17 \end{pmatrix}. \tag{5.29}$$

Table 9 shows the EM estimates of q, a, b and c with their standard errors.

	MLE	std. dev.	$CI_{0.95}$
$\widehat{q}$	0.7642	0.0105	(0.7436, 0.7848)
$\widehat{a}$	0.0749	0.0024	(0.0702, 0.0796)
$\widehat{b}$	-0.0104	0.0013	(-0.0129, -0.0079)
$\widehat{c}$	0.0158	0.0013	(0.0133, 0.0183)

Table 9: EM estimates of q, a, b and c with standard errors (std. dev.) and asymptotic 95% confidence intervals $CI_{0.95}$.

Correlations deduced from (5.29) are given in Table 10. By our model selection the parameters a, b and c of the inter-instalment times are highly correlated to the parameter of the Markov chain q.

	$\widehat{q}$	$\widehat{a}$	$\widehat{b}$	$\widehat{c}$
$\widehat{q}$	1.0000	0.3213	0.0492	-0.7348
$\widehat{a}$	0.3213	1.0000	0.0443	-0.5953
$\widehat{b}$	0.0492	0.0443	1.0000	-0.4820
$\widehat{c}$	-0.7348	-0.5953	-0.4820	1.0000

Table 10: Correlations of estimated parameters given in Table 9.

6 Prediction of outstanding insurance claims

Assume $0 \le s \le t$ and set $S = \sum_{n=1}^{N(s)} X_n(t - T_n) - X_n(s - T_n)$, that is the sum of all payoffs in the time interval $(s, t]$ for all claims occurred during $[0, s]$ and let

$$\mathcal{F} = \sigma\Bigg((N(\tau))_{0\le\tau\le s}\,,\ \Big((X_n(\tau - T_n))_{0\le\tau\le s} \Big)_{1\le n\le N(s)} \Bigg)$$

be the σ-field containing the observed claim number process and all observed payoffs during $[0, s]$. Obviously, because we assume all claims to occur independently of each other S is predicted by

$$E(S\,|\,\mathcal{F}) = \sum_{i=1}^{N(s)} E(X_n(t - T_n) - X_n(s - T_n)\,|\,\mathcal{F}).$$

For each single claim, the conditional expectations of $X_n(t-T_n)-X_n(s-T_n)$ are computed, by Theorem 3.5, equations (3.13) and (3.15). For the model parameter θ, cf. (5.22), we use the estimated parameters $\widehat{\theta} = (\widehat{\vartheta}, \widehat{\vartheta}_g)$ given in Table 9 and Table 6.

In our situation we set $s = 365$ and $t = 730$ and hence for all claims occurred in 1985, S gives the sum of all payoffs of these claims until the end of 1986. This yields

$$E\left(S \mid \mathcal{F}\right) = 18\,127\,950.$$

The realized costs for the portfolio in the same period is

$$S = 17\,493\,881.$$

For a square integrable random variable S and a σ-field $\mathcal{F}$, Jensen's inequality provides an almost sure upper bound for the expected prediction error, given the observed infomation, i.e.

$$E(\,|S - E(S\,|\,\mathcal{F})|\,|\,\mathcal{F}) \le \sqrt{\mathrm{var}(S\,|\,\mathcal{F})}.$$

For the specified portfolio, by (3.14) and the estimated parameter $\widehat{\theta}$, $\sqrt{\mathrm{var}\left(S \mid \mathcal{F}\right)} = 900\,039$. The observed prediction error is given by $|S - E(S\,|\,\mathcal{F})| = 634\,069$.

7 Summary

Given one year of data is certainly a restriction which we had to live with, but insurance companies do not have to. Thus certain drawbacks of our analysis can certainly be improved in real life, as for instance the rather crude inhomogeneous Poisson model for the counting process. Furthermore, we have only been able to predict the reserves for exactly those claims which appeared during one year. However, of course, the following year claims occur within the very same portfolio, giving more data and better estimation possibilities, also for years further ahead. We have been considering sampling future years from our model, thus having the opportunity to larger steps of prediction, however, decided to leave this to future work. We think of this investigation as a first step to a more realistic stochastic modelling of an insurance portfolio and to a more precise estimation of future capital reserves.

Acknowledgment

We want to express our sincere thanks to the leading Swiss insurance company who provided the data set. We very much appreciate their additional work making the data accessible to us, which allowed us to test our model. The second author thanks Søren Asmussen for his invitation to the Centre for Mathematical Science of Lund University in Sweden and his hospitality. He also thanks Søren Asmussen and Tobias Rydén for fruitful discussions on hidden Markov models.

References

[1] Baum, L. E., Petrie, T., Soules, G., Weiss, N. (1970) A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. *Ann. Math. Statist.,* **41***, 164-171.*

[2] Billingsley, P. (1995) *Probability and Measure.* Wiley & Sons, New York.

[3] The Institute of Actuaries (1987) *Claims Reserving Manual, Vol. I & II.* Institute of Actuaries, London.

[4] Dempster, A. P., Laird, N. M., Rubin, D. B. (1977) Maximum likelihood from incomplete data via the EM algorithm (with discussion). *J. Roy. Statist. Soc. Ser. B* **39***, 1-38.*

[5] Elliott, R. J., Aggoun, L., Moore, J. B. (1995) *Hidden Markov Models: Estimation and Control.* Springer, New York.

[6] Jamshidian, M., Jennrich, R. I. (2000) Standard errors for EM estimation. *J. Roy. Statist. Soc. Ser. B* **62***, 257-270.*

[7] Jewell, W. S. (1989) Predicting IBNYR events and delays — I. continuous time. *Astin Bull.* **19***, 25-55.*

[8] Jewell, W. S. (1990) Predicting IBNYR events and delays — II. discrete time. *Astin Bull.* **20***, 93-111.*

[9] Johnson, N. L., Kotz, S., Klemp, A. W. (1993) *Univariate Discrete Distributions.* Wiley & Sons, New York.

[10] Klüppelberg, C., Mikosch, T. (1995) Explosive Poisson shot-noise processes with applications to risk reserves. *Bernoulli* **1***, 125-147.*

[11] Klüppelberg, C., Mikosch, T. (1995) Delay in claim settlement and ruin probability approximations. *Scand. Actuar. J., 154-168.*

[12] Klüppelberg, C., Mikosch, T., Schärf, A. (2001) Regular variation in the mean and stable limits for Poisson shot noise. *Submitted for publication.*

[13] Lehmann, E. L. (1991) *Theory of Point Estimation.* Wadsworth & Brooks/Cole, Belmont, California, U.S.A.

[14] Louis, T. A. (1982) Finding the observed information matrix when using the EM algorithm. *J. Roy. Statist. Soc. Ser. B* **44**, *226-233.*

[15] MacDonald, I. L., Zucchini, W. (1997) *Hidden Markov and Other Models for Discrete-Valued Time Series.* Monographs on Statistics and Applied Probability **70**, Chapman & Hall, London.

[16] McLachlan, G. J., Basford, K. E. (1988) *Mixture Models: Inference and Applications to Clustering.* Marcel Dekker, New York.

[17] McLachlan, G. J., Krishnan, T. (1997) *The EM Algorithm and Extensions.* Wiley Series in Probability and Statistics, New York.

[18] Orchard, T., Woodbury M. A. (1972) A missing information principle: theroy and application. *Proc. 6th Berkley Symp. on Mathematical Statistics and Probability* **1**, *697-715.*

[19] Rabiner, L. R. (1989) A tutorial on hidden Markov models and selected applications in speech recognition. *Proc. IEEE* **77**, 257-285.

[20] Severin, M. (2001) *Modelling Delay in Claim Settlement: Estimation and Prediction of IBNR claims.* PhD thesis, Center of Mathematical Sciences, Munich University of Technology.

[21] Taylor, G. C. (1986) *Claims Reserving in Non-Life Insurance.* North-Holland, Amsterdam.

[22] Wu, C. F. J. (1983) On the convergence properties of the EM algorithm. *Ann. Stat.* **11**, *95-103.*

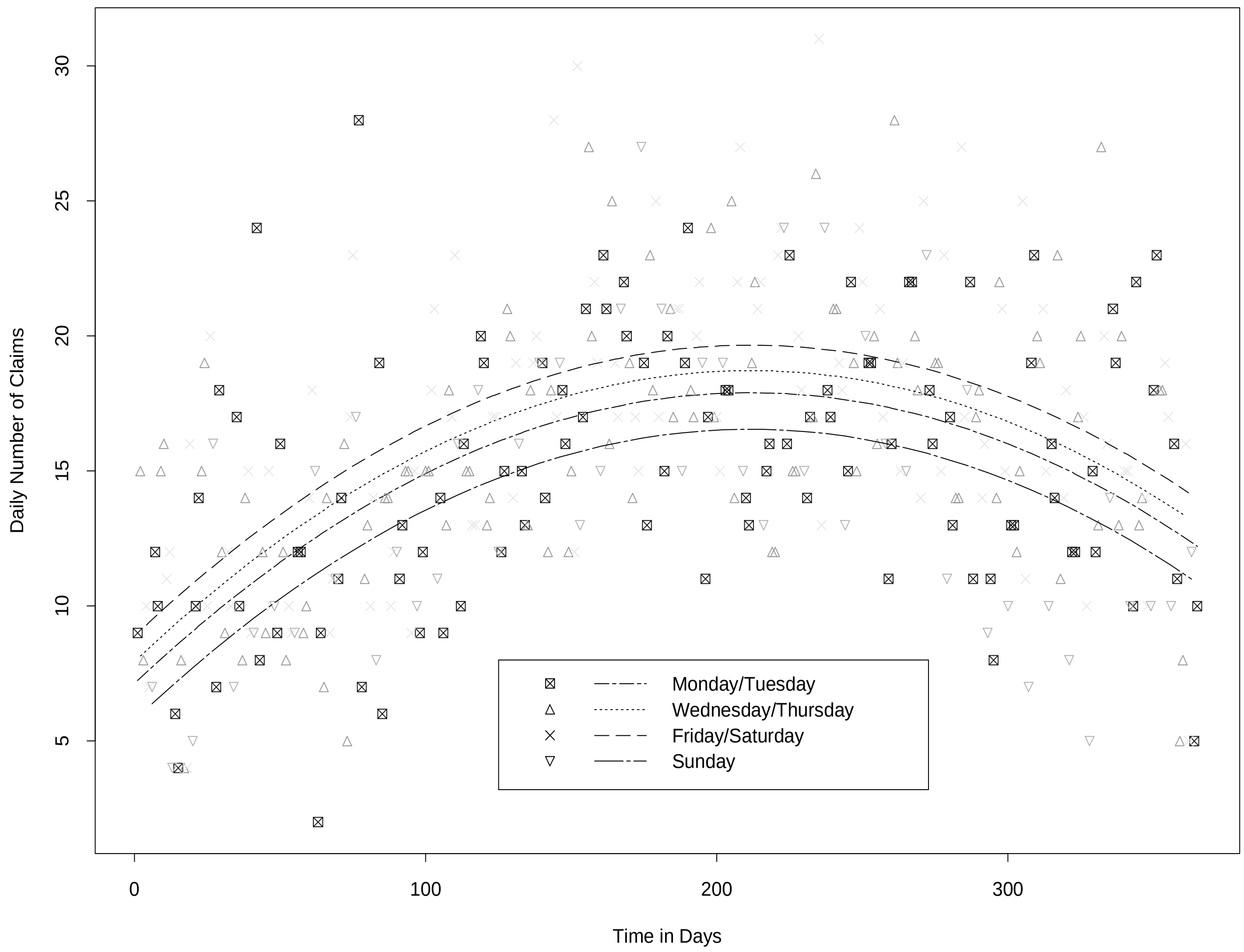

Figure 5: Fitted values $\widehat{\Delta}_i$.

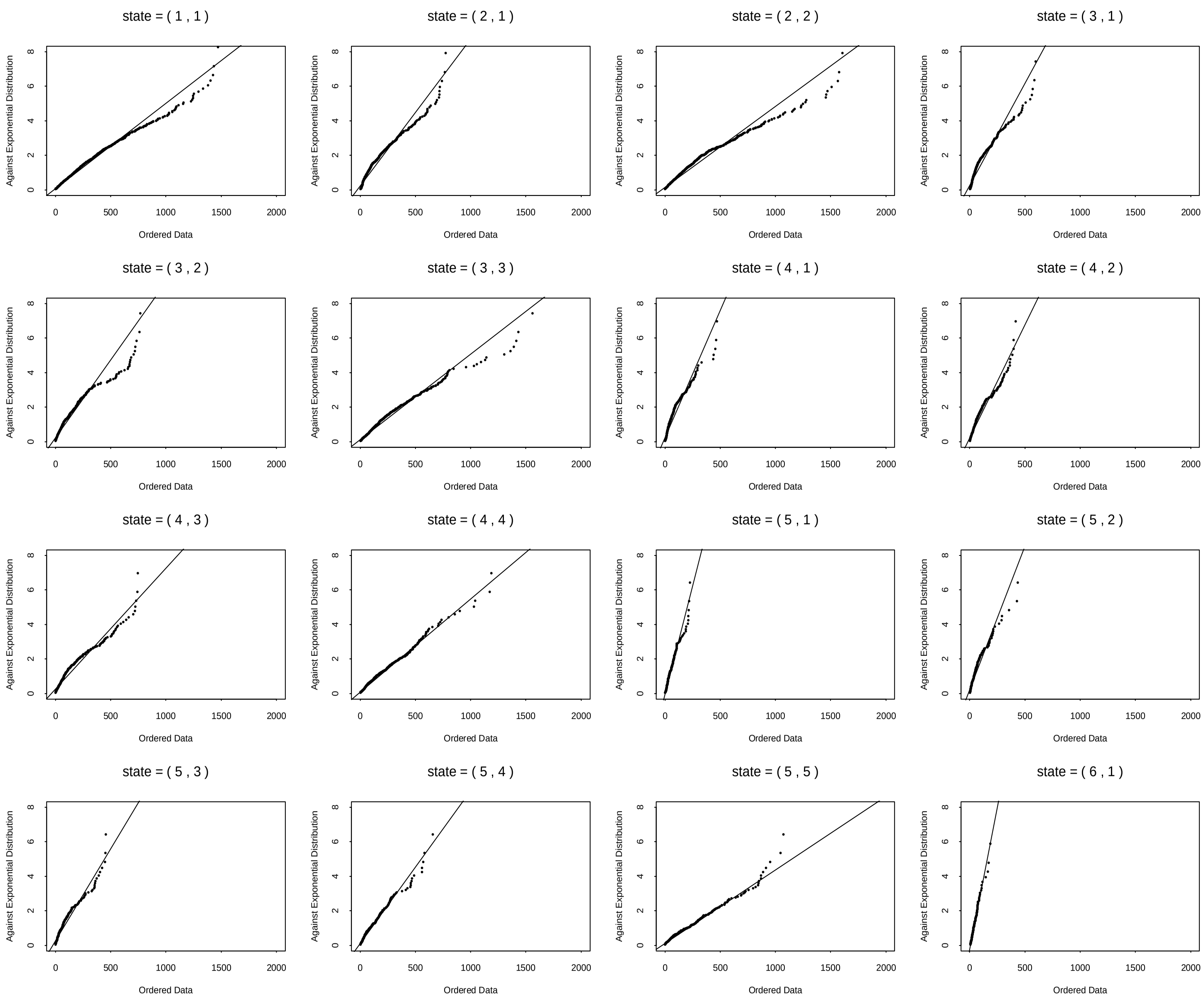

Figure 6: QQ-plot of fitted exponential distributions against selected samples of inter-instalment times.

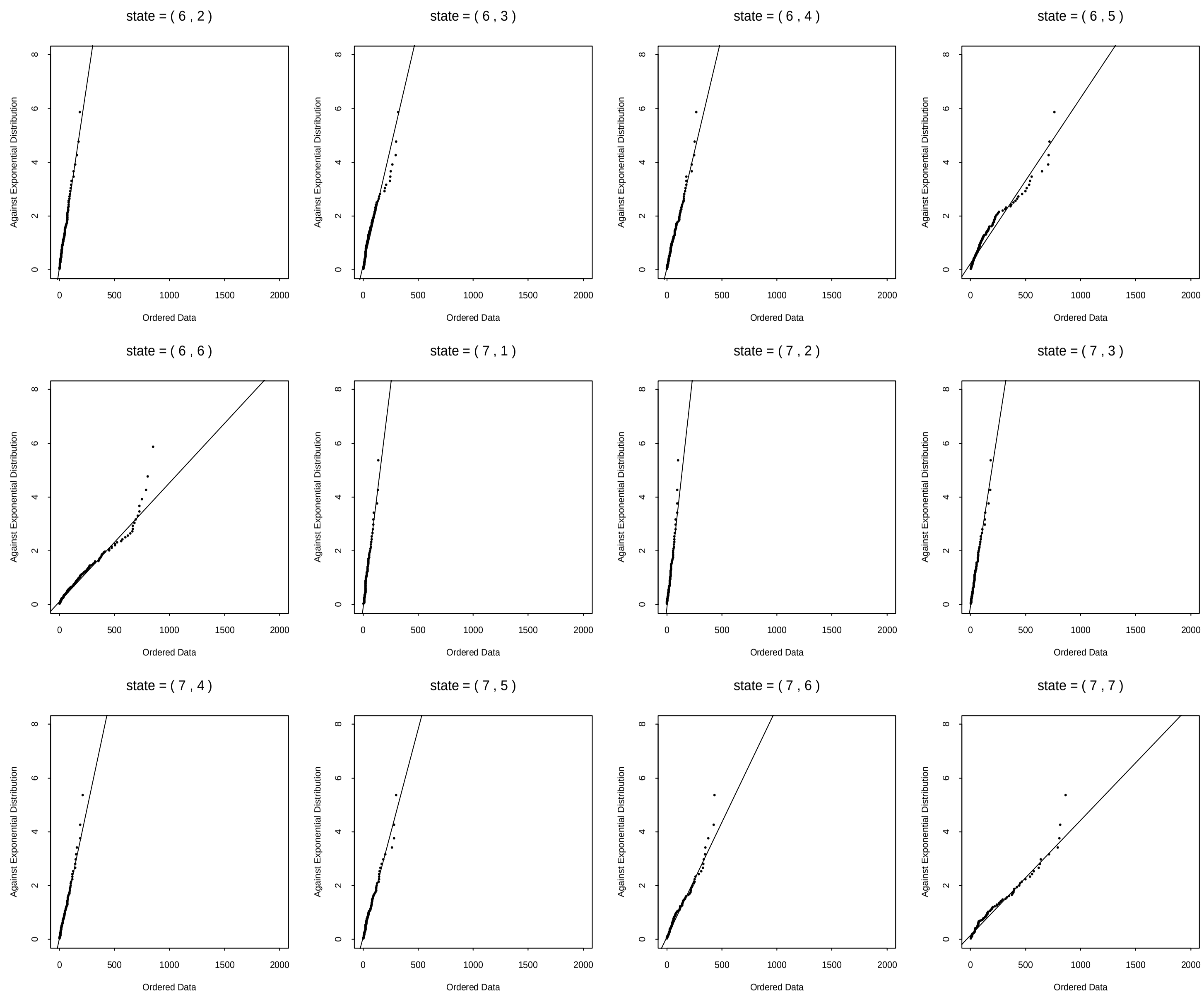

Figure 7: QQ-plot of fitted exponential distributions against selected samples of inter-instalment times.

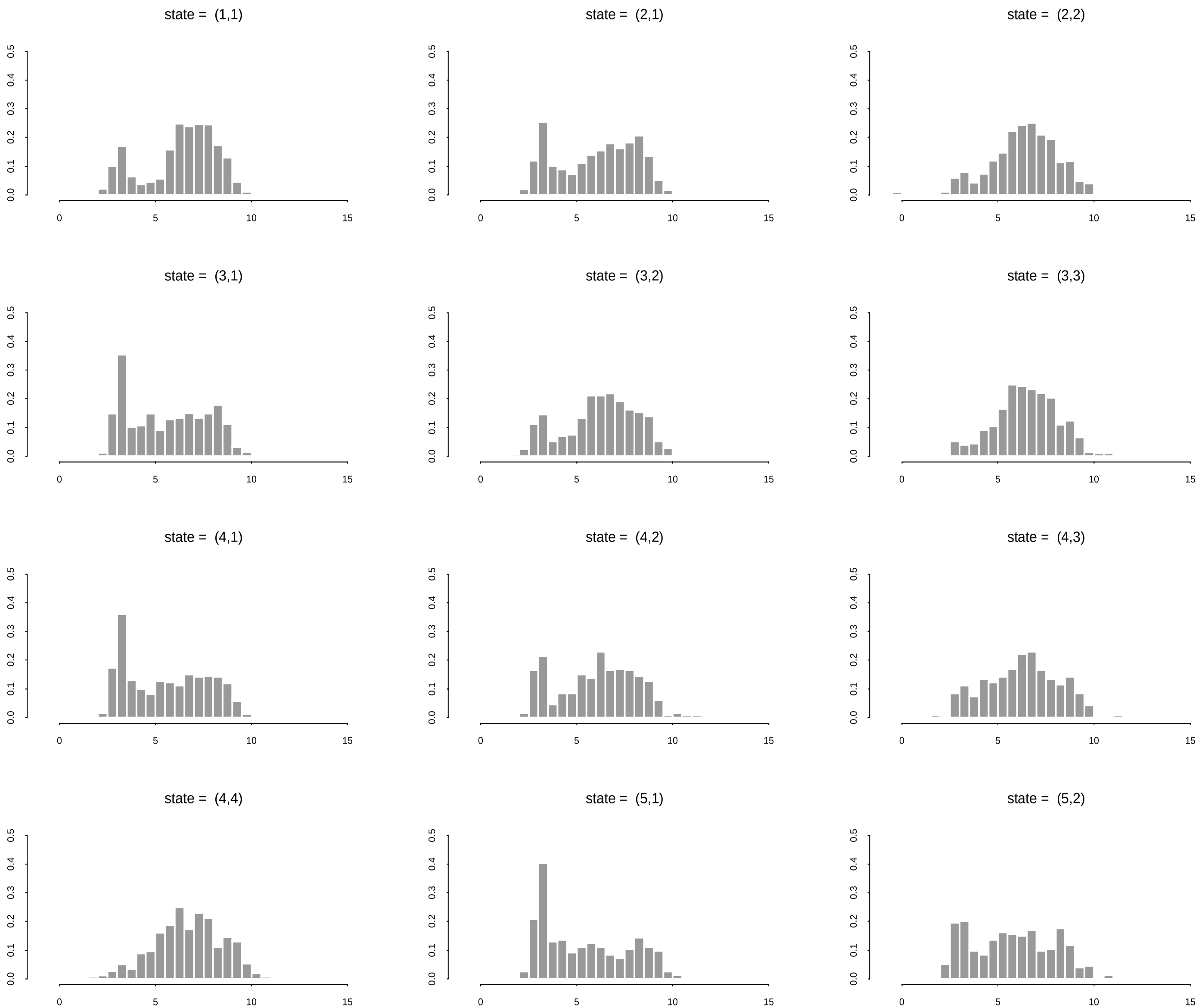

Figure 8: Histogram of selected samples of logarithmic instalments according to the subdivision of the portfolio by the number of instalments.

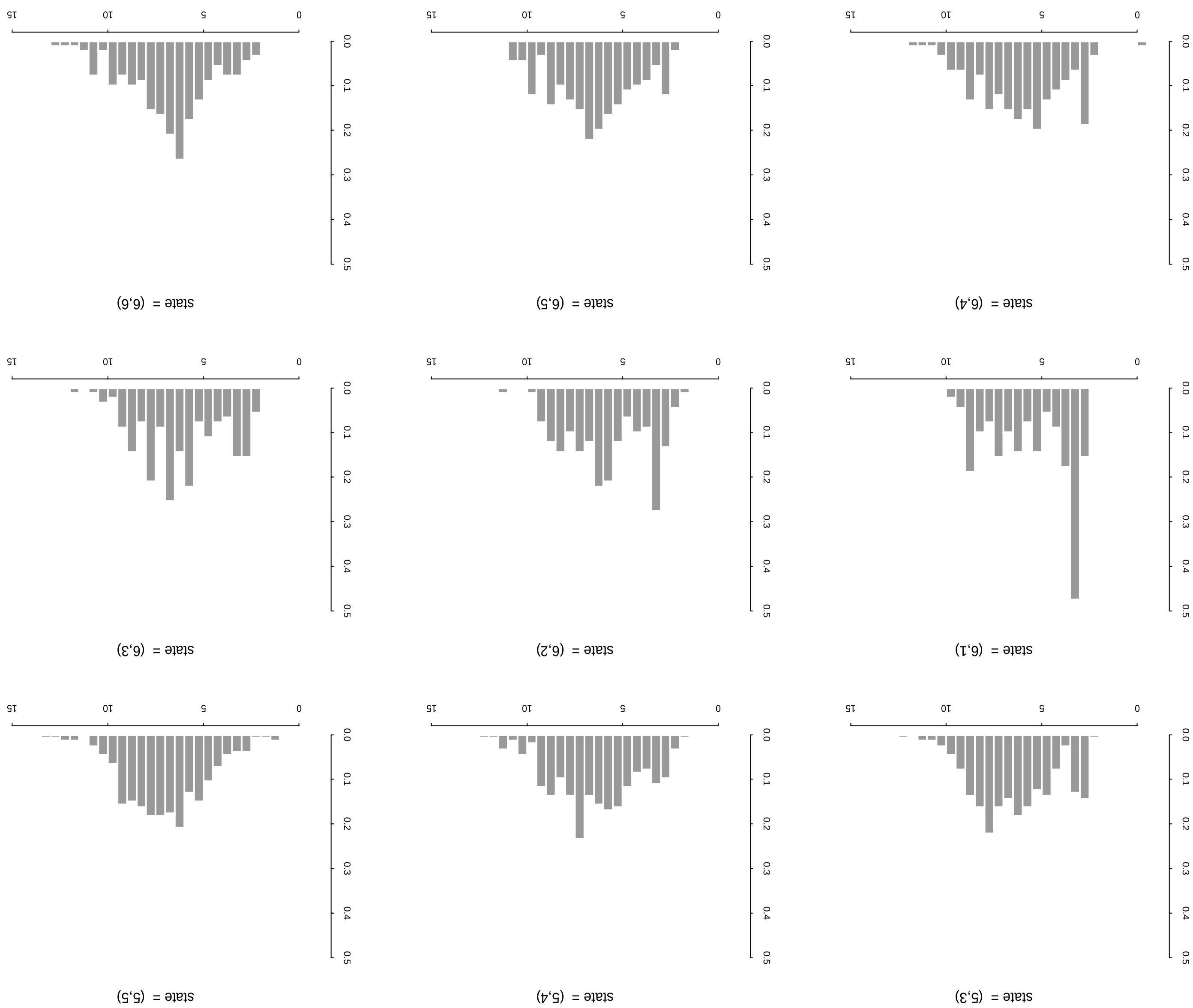

Figure 9: Histogram of selected samples of logarithmic instalments arranged according to the subdivision of the portfolio by the number of instalments.

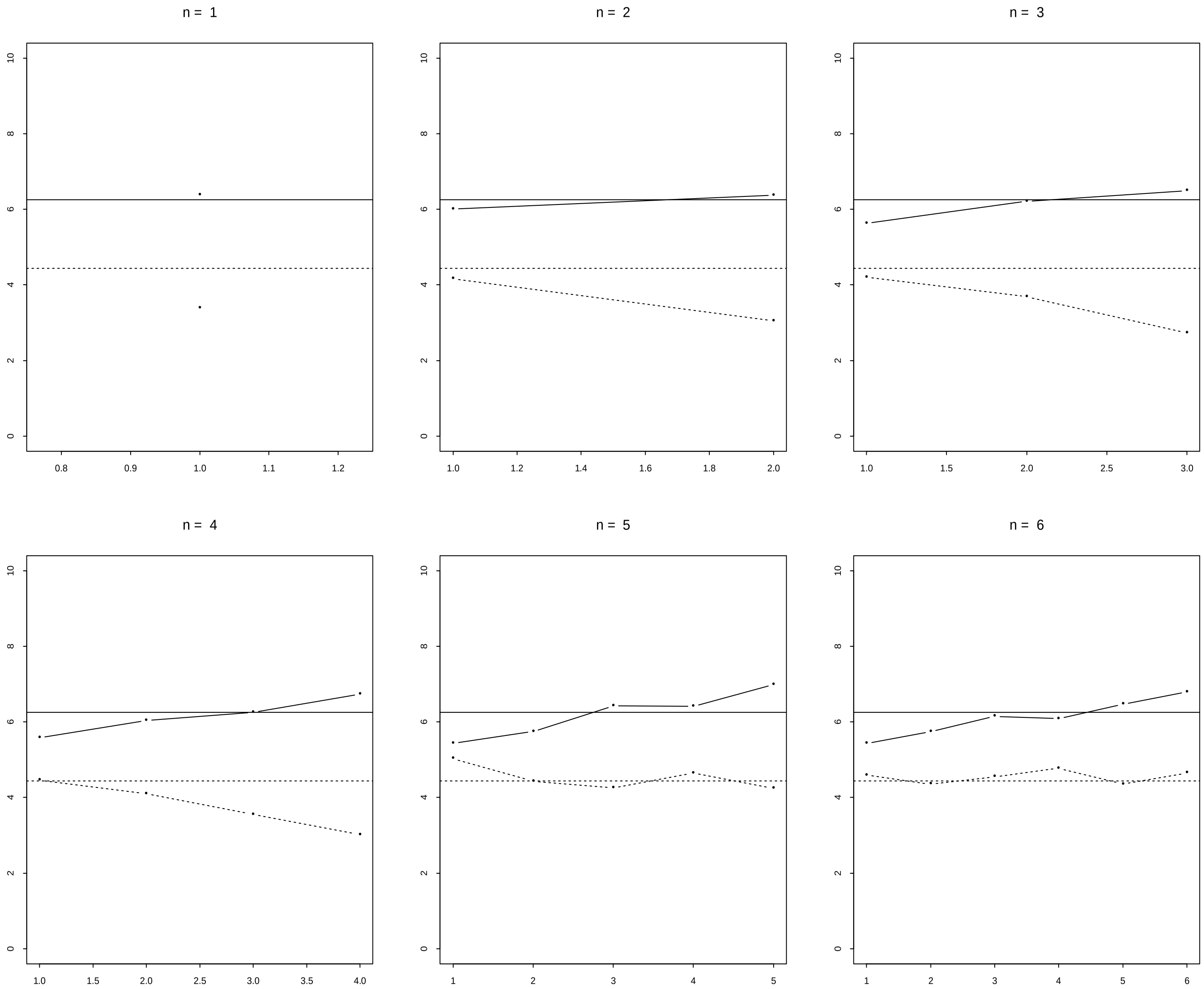

Figure 10: Log-means and log-variances of instalments in their subclasses. The horizontal lines indicate the log-mean (solid lines) and log-variance (dashed lines) estimated from all instalments. The dots give the corresponding estimates when claims are classified by their number of instalments. For instance, for $n = 3$ we get 3 point estimators corresponding to log-mean and log-variance of their first, second and third instalment. The points are connected giving a piecewise linear function for log-means (solid lines) and log-variances (dashed lines).