Article

Understanding Reporting Delay in General Insurance

Richard J. Verrall [1,†] and Mario V. Wüthrich [2,3,*,†]

[1] Cass Business School, City University London, 106 Bunhill Row, London EC1Y 8T2, UK; r.j.verrall@city.ac.uk
[2] ETH Zurich, RiskLab, Department of Mathematics, Zurich 8092, Switzerland
[3] Swiss Finance Institute, Walchestrasse 9, Zurich CH-8006, Switzerland
[*] Correspondence: mario.wuethrich@math.ethz.ch; Tel.: +41-44-632-33-90
[†] These authors contributed equally to this work.

Abstract: The aim of this paper is to understand and to model claims arrival and reporting delay in general insurance. We calibrate two real individual claims data sets to the statistical model of Jewell and Norberg. One data set considers property insurance and the other one casualty insurance. For our analysis we slightly relax the model assumptions of Jewell allowing for non-stationarity so that the model is able to cope with trends and with seasonal patterns. The performance of our individual claims data prediction is compared to the prediction based on aggregate data using the Poisson chain-ladder method.

1. Introduction

The aim of this paper is to understand the reporting delay function of general insurance claims. The reporting delay function is an important building block in claims reserving. We study this reporting delay function in continuous time based on real individual claims data. The data is used to calibrate a non-stationary version of the statistical model considered in Jewell [1,2]. This calibration then provides an estimation of the number of incurred but not yet reported (IBNYR) claims. The second building block in claims reserving is the modeling of the cost process of insurance claims. Jewell [1] stated in 1989: "Currently, the development of a good model for cost evolution over continuous time appears to require a long-term research effort, one that we believe will use the basic understanding of the event generation and reporting processes developed here, but will require much additional empirical effort to develop an understanding of cost-generating mechanisms and their evolution over time." Meanwhile, there have been some improvements into this direction, see Bühlmann et al. [3], Arjas [4], Norberg [5,6], Haastrup-Arjas [7], Taylor [8,9], Herbst [10], Larsen [11], Taylor et al. [12], Jessen et al. [13], Rosenlund [14], Pigeon et al. [15], Agbeko et al. [16], Antonio-Plat [17] and Badescu et al. [18,19]. But we believe that state-of-the-art modeling is still far from having a good statistical model that can be used in daily industry practice. This may partly be explained by the fact that it is rather difficult to get access to real individual claims data.

In this paper we study individual claims data of two different portfolios: a property insurance portfolio and a casualty insurance portfolio. We choose explicit distributional models for the individual claims arrival process modeling and we calibrate these models dynamically to the data. The calibration is back-tested against the observations and compared to the (Poisson) chain-ladder method (which is applied to aggregated data). The main conclusion is that the chain-ladder method has a good performance as long as the claims process is stationary, but in non-stationary environments our individual claims estimation approach clearly outperforms the chain-ladder method.

In the next two sections we introduce the underlying statistical model and we describe the available claims data. In Section 4 we model seasonality of the claims arrival process. In Section 5 we calibrate the reporting delay function and underpin this by statistical analysis. Finally, in Section 6 we compare the individual claims modeling estimate to the classical chain-ladder method on aggregated data and we back-test the results of these two approaches. The figures and the proofs are deferred to the appendix.

2. Individual Claims Arrival Modeling

We extend the model considered in Jewell [1,2] to an inhomogeneous marked Poisson point process. We define $\Lambda(t) \geq 0$ to be the instantaneous claims frequency and $w(t) \geq 0$ to be the instantaneous exposure at time $t \geq 0$. The total exposure W_Λ on time interval $(0, \tau_m]$ is given by

$$W_\Lambda = \int_0^{\tau_m} w(t)\Lambda(t)\,dt \tag{1}$$

We consider the run-off situation after time τ_m meaning that the exposure expires at time τ_m, i.e., $w(t) = w(t)1_{\{t \leq \tau_m\}}$ for $t \geq 0$. This implies that the total exposure is assumed to be finite $W_\Lambda < \infty$.

Assume that the claims of the total exposure W_Λ on $(0, \tau_m]$ occur at times T_ℓ (called accident dates or claims occurrence dates) and the claims counting process $(N(t))_{t \geq 0}$ is given by

$$N(t) = \sum_{\ell \geq 1} 1_{\{T_\ell \leq t\}}, \qquad \text{for } t \geq 0$$

and the total number of claims is given by

$$N = N(\tau_m) = \lim_{t \to \infty} N(t)$$

Model Assumptions 1. *We assume that the claims counting process $(N(t))_{t \geq 0}$ is an inhomogeneous Poisson point process with intensity $(w(t)\Lambda(t))_{t \geq 0}$.*

The second ingredient that we consider is the reporting date. Assume that a given claim ℓ occurs at time T_ℓ then we denote its reporting date at the insurance company by $S_\ell \geq T_\ell$. For accident date T_ℓ and corresponding reporting date S_ℓ of claim ℓ we define the reporting delay by

$$U_\ell = S_\ell - T_\ell \geq 0$$

This motivates the study of the following inhomogeneous marked Poisson point process.

Model Assumptions 2. *We assume that $((T_\ell, U_\ell)_{\ell=1,...,N(t)})_{t \geq 0}$ describes an inhomogeneous marked Poisson process with accident dates $(T_\ell)_{\ell \geq 1}$ generated by an inhomogeneous Poisson point process $(N(t))_{t \geq 0}$ having intensity $(w(t)\Lambda(t))_{t \geq 0}$ and with mutually independent reporting delays (marks) $U_\ell = U(T_\ell)$ having a time-dependent distribution $U = U(t) \sim F_{U|t,\Theta}$ for $t \geq 0$ and being independent of $(N(t))_{t \geq 0}$.*

From Jewell [1,2] and Norberg [5,6] we immediately obtain the following likelihood function for parameters Λ and Θ

$$\mathcal{L}_{N,(T_\ell,S_\ell)_{\ell=1,...,N}}(\Lambda, \Theta) = e^{-W_\Lambda} \frac{W_\Lambda^N}{N!} \prod_{\ell=1}^N \frac{w(T_\ell)\Lambda(T_\ell)}{W_\Lambda} f_{U|T_\ell,\Theta}(S_\ell - T_\ell) \tag{2}$$

where $f_{U|t,\Theta}$ denotes the density of $F_{U|t,\Theta}$ (for parameter Θ). Observe that we use a slight abuse of notation here. In Norberg [5,6] there is an additional factor $N!$ because strictly speaking Model Assumptions 2 consider ordered claims arrivals $T_{(\ell)} \leq T_{(\ell+1)}$ for all $\ell = 1,\ldots,N-1$, whereas in Jewell [1] and in Equation (2) claims arrivals are not necessarily ordered. The aim is to calibrate this

model to individual claims data, that is, we would like to calibrate claims frequency Λ and reporting delay distribution $F_{U|t,\Theta}$. One difficulty in this calibration lies in the fact that we have missing data, because information about occurred claims with reporting dates $S_\ell > \tau$ is not available at time $\tau \geq \tau_m$. Therefore, we only observe a thinned inhomogeneous Poisson process, we also refer to Norberg [5,6].

We choose $\tau \geq \tau_m$. This implies that all claims have occurred at time τ, providing $N = N(\tau) = N(\tau_m)$. By $M = M(\tau) \leq N$ we denoted the number of claims that are reported at time τ. The intractable likelihood (2) is then converted to, for details we refer to Jewell [1],

$$
\begin{aligned}
\mathcal{L}_{(T_\ell,S_\ell)_{\ell=1,\dots,M}}(\Lambda,\Theta) &= e^{-\pi_{\Lambda,\Theta}(\tau)W_\Lambda} \frac{W_\Lambda^M}{M!} \prod_{\ell=1}^{M} \frac{w(T_\ell)\Lambda(T_\ell)}{W_\Lambda} f_{U|T_\ell,\Theta}(S_\ell - T_\ell) \qquad (3) \\
&= e^{-\pi_{\Lambda,\Theta}(\tau)W_\Lambda} \frac{(\pi_{\Lambda,\Theta}(\tau)W_\Lambda)^M}{M!} \prod_{\ell=1}^{M} \frac{w(T_\ell)\Lambda(T_\ell)}{\pi_{\Lambda,\Theta}(\tau)W_\Lambda} f_{U|T_\ell,\Theta}(S_\ell - T_\ell)
\end{aligned}
$$

where we only consider reported claims ℓ with $S_\ell \leq \tau$ and $\pi_{\Lambda,\Theta}(\tau)$ denotes the probability that an incurred claim is reported by time $\tau \geq \tau_m$. This is given by

$$
\pi_{\Lambda,\Theta}(\tau) = \int_0^{\tau_m} \frac{w(t)\Lambda(t)}{W_\Lambda} F_{U|t,\Theta}(\tau - t)dt
$$

3. Description of the Data

For the statistical analysis we consider two different European insurance portfolios: (1) line of business (LoB) Property, and (2) LoB Casualty. For both portfolios data is available from 1/1/2001 until 31/10/2010. In Figures 1–3 we illustrate the data. Generally, LoB Property is colored blue and LoB Casualty is colored green (depending on the context special features may also be highlighted with other colors, this will be described in the corresponding captions). Figure 1 gives daily claims counts on the left-hand side (lhs) and monthly claims counts on the right-hand side (rhs). The following needs to be remarked for Figure 1:

- The monthly claims counts on the rhs show a clear annual seasonality.
- The daily claims counts on the lhs show a weekly seasonality with blue/green dots for weekdays, violet dots for Saturdays and orange dots for Sundays, in Table 1 we present the corresponding statistics.
- In general, these graphs are decreasing because of missing IBNYR claims (late reportings) that affect younger accident years more than older ones.

In Figure 2 (lhs) we give the daily claims reporting and Figure 2 (rhs) plots accident dates T_ℓ versus reporting delays $U_\ell = S_\ell - T_\ell$:

- Daily reporting differs between weekdays (blue/green) and weekends (violet for Saturdays and orange for Sundays). Basically there is no reporting on weekends because claims staff in insurance companies does not work on weekends, however there is a visible change in LoB Property after 2006.
- There is a change in reporting policy in LoB Property after 2006 (top, lhs), this is visible by the change of reportings on weekends (and will become more apparent in the statistical analysis below). We do not have additional information on this, but it may be caused by web-based reporting and needs special attention in modeling. We call 1/1/2006 "break point" in our analysis because it leads to non-stationarity, this will analyzed in detail below.
- Figure 2 (rhs) gives the accident dates T_ℓ versus the reporting delays $U_\ell = S_\ell - T_\ell$. We observe that the big bulk of the claims has a reporting delay of less than 1 year, and for both LoBs the resulting dots are located densely for $U_\ell \leq 1$. Bigger reporting delays are more sparse and LoB Casualty has more heavy-tailed reporting delays than LoB Property, the former having several claims with a reporting delay U_ℓ of more than 3 years.

Finally, Figure 3 gives the box plots on the yearly scale of the logged reporting delays $\log(U_\ell)$:

- LoB Property has a change in reporting policy that leads to a faster reporting after break point 1/1/2006.
- The graphs are generally decreasing because IBNYR claims (late reportings) are still missing, this corresponds to the upper-right (white) triangles in Figure 2 (rhs).

Table 1. Statistics per weekday: average daily claims counts and empirical standard deviation for LoB Property and LoB Casualty.

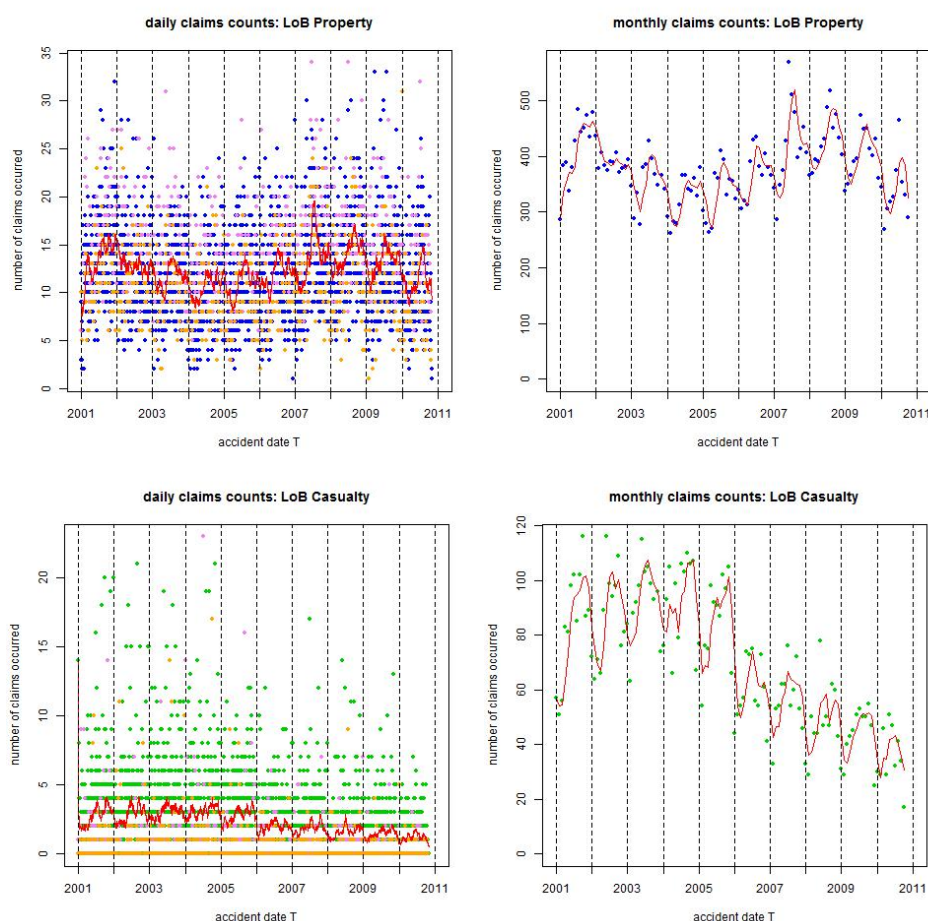|	**Mon**	**Tue**	**Wed**	**Thu**	**Fri**	**Sat**	**Sun**
LoB Property							
average	11.48	11.02	11.87	12.15	14.23	15.46	10.73
standard deviation	4.51	4.05	4.26	4.30	4.66	4.95	4.26
LoB Casualty							
average	2.89	2.92	2.89	2.65	2.61	1.09	0.82
standard deviation	2.72	2.82	2.63	2.40	2.45	2.08	1.80

Figure 1. Observed claims counts from 1/1/2001 until 31/10/2010: (top) LoB Property colored blue and (bottom) LoB Casualty colored green. The lhs gives daily claims counts and the rhs monthly claims counts; the red lines are the rolling averages over 30 days on the lhs and over 2 months on the rhs; violet dots show claims occurrence on Saturdays and orange dots claims occurrence on Sundays, the resulting statistics are provided in Table 1.
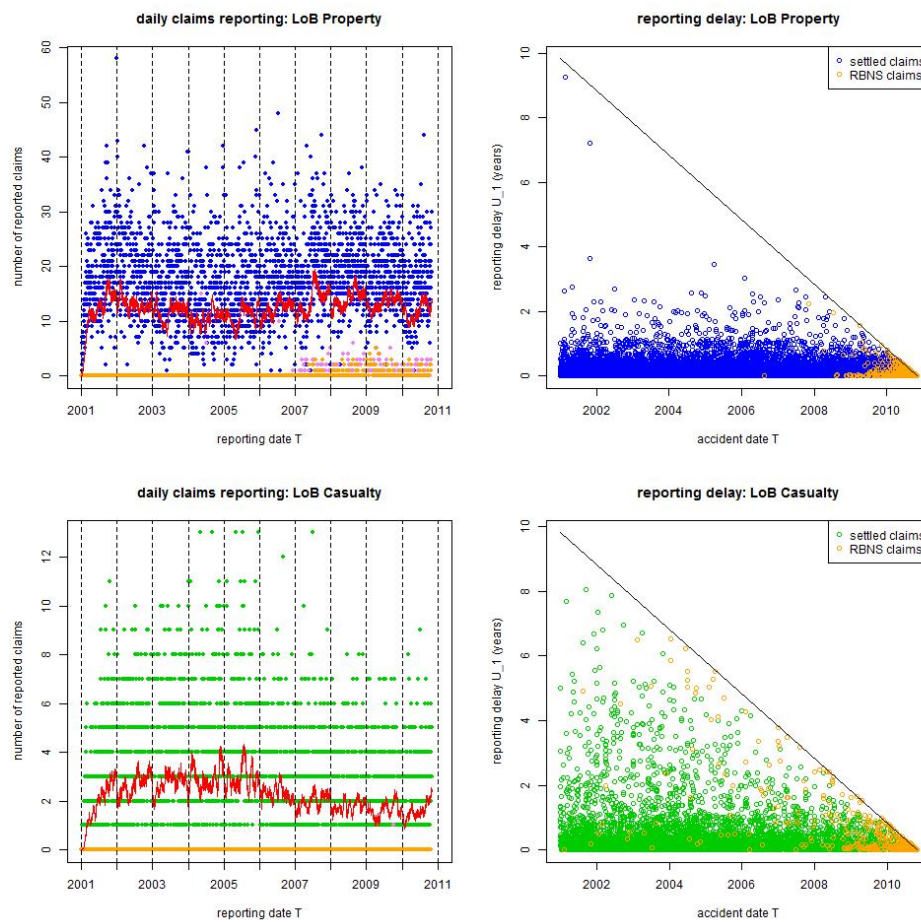
Figure 2. Observed and reported claims counts from 1/1/2001 until 31/10/2010: (top) LoB Property colored blue and (bottom) LoB Casualty colored green. The lhs gives daily claims reporting; the red lines are the 30 days rolling averages; violet dots show claims reporting on Saturdays and orange dots claims reporting on Sundays. The rhs plots accident dates T_ℓ versus reporting delays $U_\ell = S_\ell - T_\ell$; the upper-right (white) triangle corresponds to the missing data (IBNYR claims); the blue/green dots illustrate reported and settled claims; the orange dots reported but not settled (RBNS) claims.
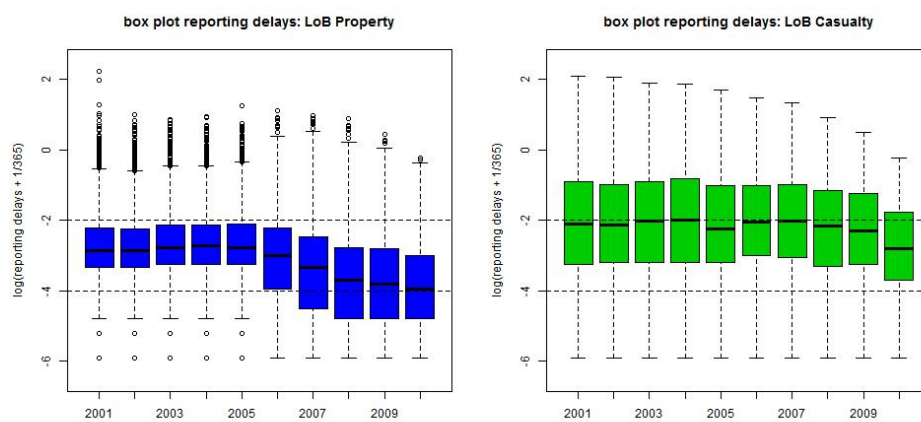

Figure 3. Box plots of the logged reporting delays $\log(U_\ell)$ on the yearly scale (lhs) LoB Property, (rhs) LoB Casualty.

4. Seasonal Claims Frequency Modeling

4.1. Likelihood Function with Seasonality

We discuss the modeling of the claims frequency $\Lambda = (\Lambda(t))_{t \geq 0}$ which can be any measurable function having finite integral on interval $(0, \tau_m]$. Time $t \in \mathbb{R}$ is measured in daily units (unless stated otherwise). In order to get an appropriate model we study annual and weekly seasonality that both influence Λ, see Figure 1. For the weekly seasonality we choose a stationary periodic pattern. This is an appropriate choice unless the insurance product or the portfolio changes. For the annual seasonality we split the time interval $(0, \tau_m]$ into smaller sub-intervals on which statistical estimation is carried out. These smaller time intervals $(\tau_{i-1}, \tau_i]$ are given by finitely many integer-valued endpoints

$$0 = \tau_0 < \tau_1 < \ldots < \tau_{m-1} < \tau_m, \quad \text{and we set} \quad \Delta\tau_i = \tau_i - \tau_{i-1} \tag{4}$$

for $1 \leq i \leq m$. Typically, $\Delta\tau_i$ corresponds to a calendar month: this is naturally given if data becomes available on a monthly time scale. The following has to be considered: (i) the monthly time grid is not equidistant because months differ in the number of days; (ii) months and weekdays do not have the same periodicity; (iii) $\Delta\tau_i$ should be sufficiently large so that reliable estimation can be done, and sufficiently small so that we have homogeneity on these time intervals; and (iv) smoothing between neighboring time intervals can be applied later on. Such a (monthly) seasonal split is reasonable because often insurance claims are influenced by external factors (such as winter and summer) that (may) only affect bounded time intervals for claims occurrence. This approximation can be seen as a reasonable modeling assumption; if more randomness is involved then we should switch to a hidden Markov model, such as the Cox model presented in Badescu et al. [18,19].

On this time grid we then make the following assumptions: for all $1 \leq i \leq m$ and $t \in (\tau_{i-1}, \tau_i]$

$$w_i := w(\tau_i) = w(t), \qquad \Lambda_i \lambda_{\lceil t \rceil} = \Lambda_i \lambda_t = \Lambda(t) \tag{5}$$

with weekly periodic (piece-wise constant) pattern $\lambda_{\lceil t \rceil} = \lambda_t = \lambda_{t+7}$ for all $t > 0$ fulfilling normalization $\sum_{k=1}^{7} \lambda_k = 7$ and global parameter Λ_i for interval $(\tau_{i-1}, \tau_i]$. We remark the following:

- We have a weekly periodic piece-wise constant pattern that is assumed to be stationary and a (monthly) seasonal parameter Λ_i. The total exposure on $(\tau_{i-1}, \tau_i]$ is given by

$$W_\Lambda^{(i)} = w_i \Lambda_i \int_{\tau_{i-1}}^{\tau_i} \lambda_t dt = w_i \Lambda_i \sum_{k=\tau_{i-1}+1}^{\tau_i} \lambda_k =: w_i \Lambda_i \lambda_i^+ \tag{6}$$

  $\lambda_i^+ = \sum_{k=\tau_{i-1}+1}^{\tau_i} \lambda_k$ in general differs from $\Delta\tau_i$ because different months may have different weekday constellations.

- In the special case of $\lambda_t \equiv 1$ we obtain the piece-wise homogeneous case

$$w_i = w(\tau_i) = w(t), \qquad \Lambda_i = \Lambda(\tau_i) = \Lambda(t), \qquad \lambda_i^+ = \Delta\tau_i \tag{7}$$

  This is a step function for the claims frequency providing total exposure $W_\Lambda^{(i)} = w_i \Lambda_i \Delta\tau_i$. Note that model (7) was studied in Section 4.2 of Antonio-Plat [17]. For our real data examples, the influence of the additional weekly periodic parameter $(\lambda_k)_{k=1,\ldots,7}$ will be visualized in Figure 4, below.

- We could also choose a yearly seasonal pattern for $(\Lambda_i)_{1 \leq i \leq m}$ if, for instance, $\Delta\tau_i$ correspond to calendar months. This is supported by Figure 1 (rhs) and would reduce the number of parameters. This is particularly important for claims prediction, i.e., for predicting the number of claims of future exposure years. In our analysis we refrain from choosing additional structure for $(\Lambda_i)_{1 \leq i \leq m}$ because we will concentrate on inference of past exposures and because the volumes of the two LoBs are sufficiently large to get reliable inference results on a monthly scale.
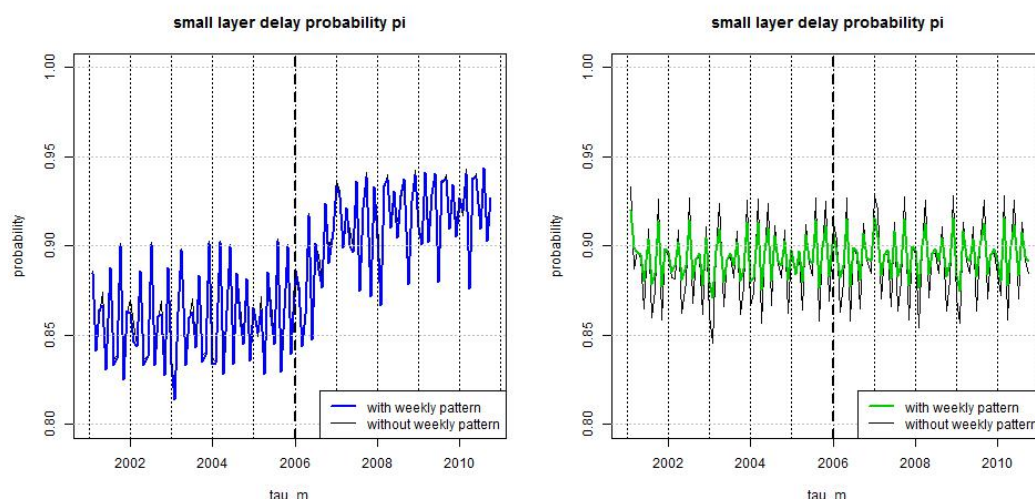
Figure 4. Estimated probability $\pi_{\hat{\Theta}_1^{(\tau_m)}}^{(m,1)}(\tau_m)$, see (23), for $\tau_m \in \{31/1/2001,\ldots,31/10/2010\}$ using weekly periodic pattern $(\lambda_k)_{k=1,\ldots,7}$ (blue/green) and setting $(\lambda_k)_{k=1,\ldots,7} \equiv 1$ (black) for (lhs) LoB Property under the full model (21), and (rhs) LoB Casualty under the null hypothesis reduced model.

Time grid (4) defines a natural partition on which the inhomogeneous marked Poisson point process decouples into independent components, see Theorem 2 in Norberg [6]. The (τ-observable) likelihood on time interval $(\tau_{i-1}, \tau_i]$ under the above assumptions is given by

$$\mathcal{L}_{(T_\ell^{(i)}, S_\ell^{(i)})_{\ell=1,\ldots,M_i}}^{(i)}(\Lambda, \Theta) = e^{-\pi_\Theta^{(i)}(\tau)W_\Lambda^{(i)}} \frac{(W_\Lambda^{(i)})^{M_i}}{M_i!} \prod_{\ell=1}^{M_i} \frac{\lambda_{T_\ell^{(i)}}}{\lambda_i^+} f_{U|T_\ell^{(i)}, \Theta}\left(S_\ell^{(i)} - T_\ell^{(i)}\right) \tag{8}$$

where $M_i = M_i(\tau)$ denotes the number of reported claims at time $\tau \geq \tau_m$ with accident dates $T_\ell^{(i)} \in (\tau_{i-1}, \tau_i]$ and corresponding reporting dates $S_\ell^{(i)} \leq \tau$ for $\ell = 1, \ldots, M_i$. The probability that an incurred claim with accident date in $(\tau_{i-1}, \tau_i]$ is reported at time $\tau \geq \tau_i$ simplifies to

$$\pi_\Theta^{(i)}(\tau) = \frac{1}{W_\Lambda^{(i)}} \int_{\tau_{i-1}}^{\tau_i} w_i \Lambda_i \lambda_t F_{U|t, \Theta}(\tau - t)\, dt = \frac{1}{\lambda_i^+} \int_{\tau_{i-1}}^{\tau_i} \lambda_t F_{U|t, \Theta}(\tau - t)\, dt$$

Observe that probability $\pi_\Theta^{(i)}(\tau)$ only depends on the (weekly-) seasonal pattern $(\lambda_t)_{t \in (\tau_{i-1}, \tau_i]}$ and on Θ, but not on global parameter Λ_i of time interval $(\tau_{i-1}, \tau_i]$.

Formula (8) specifies the likelihood on time interval $(\tau_{i-1}, \tau_i]$. Due to the independent splitting property of inhomogeneous marked Poisson processes under partitions the total likelihood function at time $\tau \geq \tau_m$ is given by

$$\mathcal{L}_{(T_\ell, S_\ell)_{\ell=1,\ldots,M}}(\Lambda, \Theta) = \prod_{i=1}^m \mathcal{L}_{(T_\ell^{(i)}, S_\ell^{(i)})_{\ell=1,\ldots,M_i}}^{(i)}(\Lambda, \Theta) \tag{9}$$

In the estimation procedure below we assume that the weekly periodic pattern $(\lambda_k)_{k=1,\ldots,7}$ is given and parameters $(\Lambda_i)_{1 \leq i \leq m}$ and Θ are estimated with maximum likelihood estimation (MLE) from (9), based on the knowledge of $(\lambda_k)_{k=1,\ldots,7}$. In fact, in the applications below we will use a plug-in estimate for $(\lambda_k)_{k=1,\ldots,7}$. We could also consider the full likelihood, including $(\lambda_k)_{k=1,\ldots,7}$, but for computational reasons we refrain from doing so.

4.2. Analysis of the MLE System

In this section we derive the maximum likelihood estimate (MLE) of $(\Lambda_i)_{1\leq i \leq m}$ and Θ based on the knowledge of the weekly periodic pattern $(\lambda_k)_{k=1,\ldots,7}$. We calculate the derivatives of the logarithms of (9) and (8), respectively, and set them equal to zero to find the MLE. This provides the following lemma.

Lemma 3. *Under Model Assumptions 2 with exposure* (5)*, the MLE* $((\widehat{\Lambda}_i^{(\tau)})_{1\leq i \leq m}, \widehat{\Theta}^{(\tau)})$ *of parameter* $((\Lambda_i)_{1\leq i\leq m}, \Theta)$ *at time $\tau \geq \tau_m$ for given weekly periodic pattern $(\lambda_k)_{k=1,\ldots,7}$ is obtained by the solution of*

$$\frac{\partial}{\partial\Theta}\sum_{i=1}^{m}\sum_{\ell=1}^{M_i}\log\left(\frac{\lambda_{T_\ell^{(i)}}f_{U|T_\ell^{(i)},\Theta}(S_\ell^{(i)}-T_\ell^{(i)})}{\lambda_i^+\pi_\Theta^{(i)}(\tau)}\right)=0$$

and

$$\Lambda_i=\frac{M_i}{\pi_\Theta^{(i)}(\tau)w_i\lambda_i^+}$$

The proof is given in Appendix A. Observe that for known weekly periodic pattern $(\lambda_k)_{k=1,\ldots,7}$ the MLE decouples in the sense that the MLE of Θ can be calculated independently of $(\Lambda_i)_{1\leq i\leq m}$. This substantially helps in the calibration below because it reduces complexity. Secondly, we remark that the function

$$(t,s)\ \mapsto\ f_{T,S|\{\tau_{i-1}<T\leq\tau_i\},\{S\leq\tau\},\Theta}(s-t)=\frac{\lambda_t f_{U|t,\Theta}(s-t)}{\lambda_i^+\pi_\Theta^{(i)}(\tau)}1_{\{s\geq t\}} \tag{10}$$

gives a density on $(\tau_{i-1},\tau_i]\times(0,\tau]$.

4.3. Calibration of the Weekly Periodic Pattern

The MLE in Lemma 3 assumes that the weekly periodic pattern $(\lambda_k)_{k=1,\ldots,7}$ is given (and known). In this section we compute a plug-in estimate $(\widehat{\lambda}_k)_{k=1,\ldots,7}$. This has the advantage that the MLE remains tractable. We estimate this weekly periodic pattern $(\lambda_k)_{k=1,\ldots,7}$ under one additional assumption which we *only* use for this purpose (and drop again thereafter): assume $\tau \geq \tau_m$ is fixed and that there exists $m^* \in \{1,\ldots,m\}$ such that

$$F_{U|t,\Theta}(\tau-t)=1 \tag{11}$$

for all $t \leq \tau_{m^*}$ and all Θ. Assumption Equation (11) is an approximation that we *only* use for choosing $(\lambda_k)_{k=1,\ldots,7}$. It has the advantage that all time points $t \leq \tau_{m^*}$ are fully experienced at time τ. Estimation of $(\lambda_k)_{k=1,\ldots,7}$ is then only done based on claims with $T_\ell \leq \tau_{m^*}$ because for these occurrence days there are no missing values. Of course, this neglects the latest information but often (in stationary cases, if τ_{m^*} is not too small and if late reportings do not distort the weekly periodic pattern) this estimation is sufficiently robust. The following lemma is proved in Appendix A.

Lemma 4. *Under Model Assumptions 2 with exposure* (5) *on a weekly time grid $\Delta\tau_i = 7$ (for all $1 \leq i \leq m^*$) and under assumption* (11)*, the MLE $(\widehat{\lambda}_k)_{k=1,\ldots,7}$ of the weekly periodic pattern $(\lambda_k)_{k=1,\ldots,7}$ with side constraint $\sum_{k=1}^{7}\lambda_k = 7$ based on claims with accident dates before τ_{m^*} is for $k = 1,\ldots,7$ given by*

$$\widehat{\lambda}_k=7\frac{\sum_{i=1}^{m^*}\sum_{\ell=1}^{M_i}1_{\{\lceil\widetilde{T}_\ell^{(i)}\rceil=k\}}}{\sum_{i=1}^{m^*}M_i}$$

where

$$\widetilde{T}_\ell^{(i)}=T_\ell^{(i)}-\tau_{i-1}$$

We may examine the robustness of these estimates by choosing different time horizons τ_{m^*}, this is done in Figure 5 (lhs); these graphs also show the confidence bounds we explain how to construct. We have

$$\widehat{\lambda}_k^{-1} = \frac{1}{7} \frac{\sum_{i=1}^{m^*} M_i}{\sum_{i=1}^{m^*} \sum_{\ell=1}^{M_i} 1_{\{\lceil \widetilde{T}_\ell^{(i)} \rceil = k\}}} = \frac{1}{7} \left(1 + \frac{\sum_{i=1}^{m^*} \sum_{\ell=1}^{M_i} 1_{\{\lceil \widetilde{T}_\ell^{(i)} \rceil \neq k\}}}{\sum_{i=1}^{m^*} \sum_{\ell=1}^{M_i} 1_{\{\lceil \widetilde{T}_\ell^{(i)} \rceil = k\}}} \right) =: \frac{1}{7} \left(1 + \frac{X_1}{X_2} \right)$$

The latter ratio considers the realization of two independent Poisson distributed random variables with means and variances, respectively,

$$\mu_1 = \mathbb{E}\left[X_1 \right] = \mathrm{Var}(X_1) = \frac{7 - \lambda_k}{7} \sum_{i=1}^{m^*} w_i \Lambda_i, \qquad \mu_2 = \mathbb{E}\left[X_1 \right] = \mathrm{Var}(X_1) = \frac{\lambda_k}{7} \sum_{i=1}^{m^*} w_i \Lambda_i$$

We set $v^* = \sum_{i=1}^{m^*} w_i \Lambda_i$. The central limit theorem provides for $i = 1, 2$

$$(X_i - \mu_i) / \sqrt{\mu_i} \Rightarrow \mathcal{N}(0,1) \qquad \text{for } v^* \to \infty$$

For this reason we approximate $X_i \overset{(d)}{\approx} Z_i \sim \mathcal{N}(\mu_i, \mu_i)$ and similarly

$$\widehat{\lambda}_k^{-1} = \frac{1}{7} \left(1 + \frac{X_1}{X_2} \right) \overset{(d)}{\approx} \frac{1}{7} \left(1 + \frac{Z_1}{Z_2} \right)$$

Following Hinkley [20] we can study the asymptotic behavior of the latter using $\mu_2 / \sqrt{\mu_2} \to \infty$ for $v^* \to \infty$. This provides approximation for large v^*

$$\mathbb{P}\left[\widehat{\lambda}_k \leq x \right] \approx 1 - \mathbb{P}\left[\frac{Z_1}{Z_2} \leq 7x^{-1} - 1 \right] \approx 1 - \Phi\left(\frac{\mu_2(7x^{-1} - 1) - \mu_1}{\sqrt{\mu_2(7x^{-1} - 1)^2 + \mu_1}} \right)$$

This allows us to derive approximate confidence bounds of the weekly periodic pattern estimates. Choose confidence level $\alpha \in (1/2, 1)$, then we get a two-sided confidence bound estimate

$$\widehat{\lambda}_k \in \left[x_- \left((1 - \alpha)/2 \right), x_+ \left((1 - \alpha)/2 \right) \right] \tag{12}$$

with for $p \in (0,1)$

$$x_\pm(p) = 7 \frac{(\Phi^{-1}(p))^2 \mu_2 - \mu_2^2}{(\Phi^{-1}(p))^2 \mu_2 - \mu_2^2 - \mu_1 \mu_2 \mp \Phi^{-1}(p) \sqrt{\mu_1 \mu_2} \sqrt{\mu_1 + \mu_2 - (\Phi^{-1}(p))^2}}$$

Replacing all parameters by their MLEs given by Lemma 4 and $w_i \widehat{\Lambda}_i = M_i$ (which is the MLE for $i \leq m^*$ under assumption (11)) we get an estimate for the confidence bounds (12). These are plotted in Figure 5.

In Figure 5 (rhs) we present the resulting MLEs $(\widehat{\lambda}_k)_{k=1,\ldots,7}$ and the corresponding (estimated) confidence bounds for confidence level $\alpha = 90\%$. We observe narrow confidence bounds and substantial daily differences. In particular, claims frequencies in LoB Casualty are much lower on weekends than on weekdays (this may suggest that we consider commercial casualty insurance business). For LoB Property we observe higher frequencies on Fridays and Saturdays, also this is directly related to the underlying business. Figure 5 (lhs) gives the corresponding time series as a function of τ_{m^*}. We observe convergence of the estimates after roughly 3 years of observations.
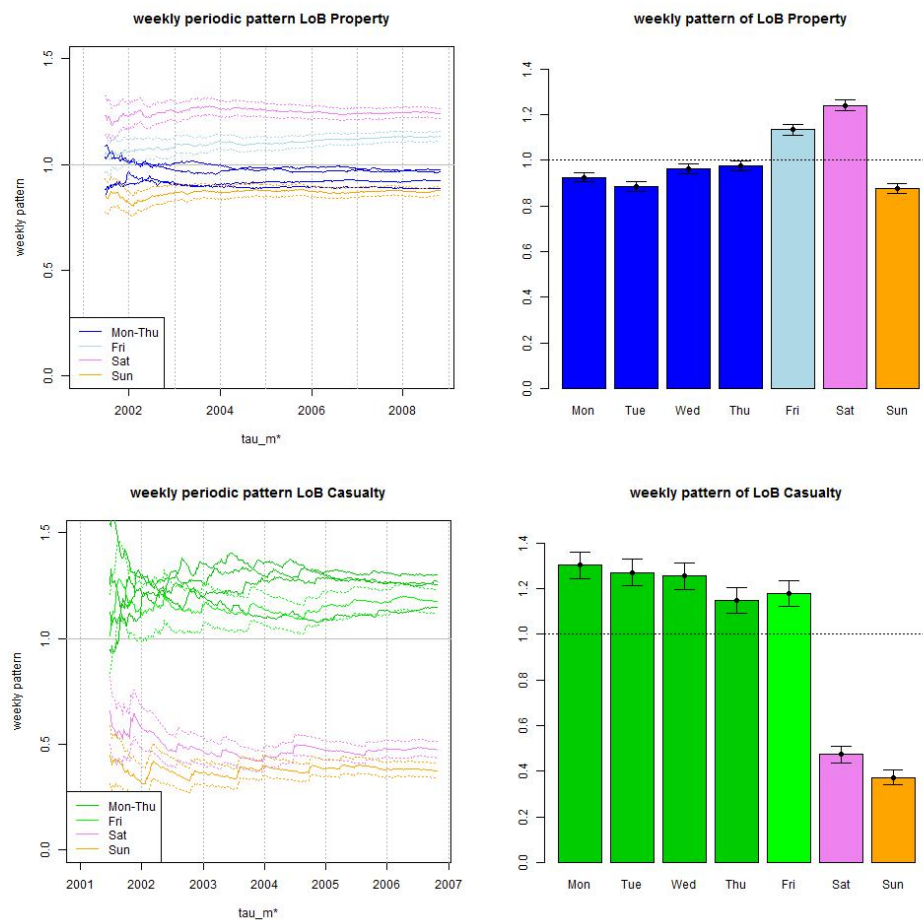
Figure 5. Weekly periodic pattern estimate $(\widehat{\lambda}_k)_{k=1,\dots,7}$ (top) LoB Property and (bottom) LoB Casualty: (lhs) time series as a function of τ_{m^*} and (rhs) for maximal τ_{m^*} such that (11) holds at time $\tau_m = 31/10/2010$ for a maximal reporting delay of 2 years (LoB Property) and of 4 years (LoB Casualty). The confidence bounds in all plots are given by (12) for confidence level $\alpha = 90\%$.

5. Calibration of the Reporting Delay Distribution

In this section we study the choice of the reporting delay distribution $F_{U|t,\Theta}$. In an empirical analysis we identify three regimes of reporting delays which we will model separately. In short, (1) small reporting delay layer where we consider a weekday structure, see Figure 6; (2) middle reporting delay layer with the main bulk of reportings, see Figure 14 (top); and (3) large reporting delay layer that should have an appropriate tail for late reportings, see Figure 14 (bottom). We call these the small, middle and large layers, and label them by $n = 1, 2, 3$.
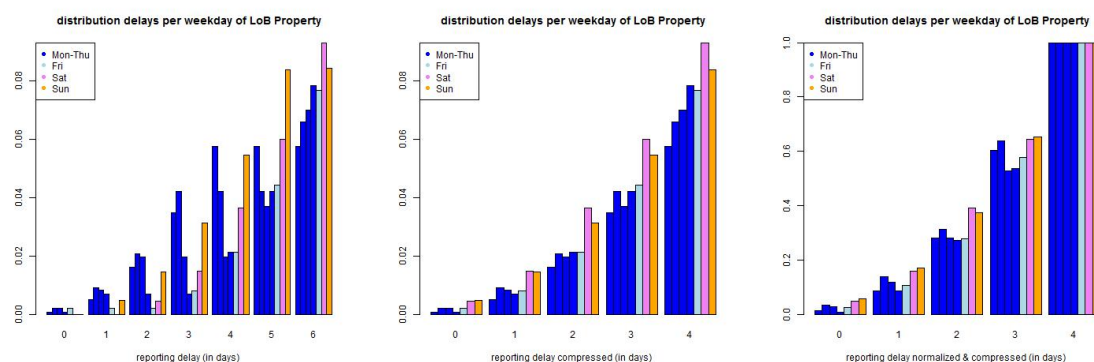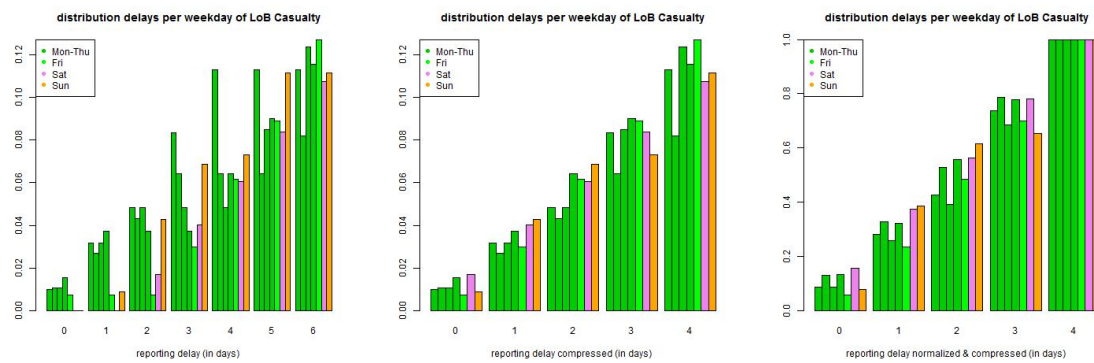

Figure 6. *Cont.*

Figure 6. (top) LoB Property and (bottom) LoB Casualty: empirical distribution of reporting delays separated by weekdays of claims occurrence of all data with accident date prior to 01/2006 and maximal reporting delay of $U_\ell \leq 365$ days (lhs) per day, (middle) compressed by weekends, and (rhs) compressed and normalized to 1 after a delay of one week.

5.1. Decoupling of the Reporting Delay Distribution

As introduced in (4), we choose a monthly time grid $0 = \tau_0 < \tau_1 < \ldots < \tau_m$ with τ_m being the end point of the last observed calendar month. The monthly time grid is naturally given because available data is provided on that time scale. From a statistical point of view also finer or wider time grids are possible. In LoB Property we have about 400 claims per month which gives a coefficient of variation of 5% (for the Poisson distribution) and in LoB Casualty we have between 80 and 100 claims which gives a coefficient of variation of roughly 10%, see Figure 1 (rhs). If we would have stationarity we could (or even should) take bigger time intervals, but because of the yearly seasonal pattern, monthly time intervals are preferred to capture these seasonal differences.

The end point τ_m of the last observed calendar month will also be considered as a variable that evolves when more and more information becomes available. First (rather limited) information is available at 31/1/2001 and latest available information is as of 31/10/2010. Thus, τ_m will run from 31/1/2001 to 31/10/2010 and we perform dynamic calibration based on actual information.

We make assumption (5) on that monthly time grid and we remark that the monthly time grid is not equally spaced in number of days. Therefore it is convenient to measure time t in daily units. We then assume that the weekly seasonal pattern $(\lambda_k)_{k=1,\ldots,7}$ is estimated (and fixed, see Figure 5) through Lemma 4 (and we drop the upper hat in the notation of λ_k). Note that fixing the weekly periodic pattern reduces the computational complexity in the sequel.

Next we need to choose the reporting delay distribution $F_{U|t,\Theta}$. We choose three layers with thresholds $0 = u^{(0)} < u^{(1)} < u^{(2)} < u^{(3)} = \infty$ and density

$$f_{U|t,\Theta}(u) = \sum_{n=1}^{3} p_{U|t,\Theta}^{(n)} f_{U|t,\Theta}^{(n)}(u) 1_{\{u^{(n-1)} \leq u < u^{(n)}\}} \tag{13}$$

where the probability weights $p_{U|t,\Theta}^{(n)} \geq 0$ are normalized $\sum_{n=1}^{3} p_{U|t,\Theta}^{(n)} = 1$, and $f_{U|t,\Theta}^{(n)}(\cdot)$ are densities supported on $[u^{(n-1)}, u^{(n)})$ for $n = 1, 2, 3$. We make the following assumptions:

Assumption 5. *We choose time units in days and make the following (additional) assumptions for the density in (13): For $t \in (\tau_{i-1}, \tau_i]$ and $n = 1, 2, 3$ we assume*

$$p_{U|t,\Theta}^{(n)} = p_{U|\tau_i,\Theta'}^{(n)} \qquad f_{U|t,\Theta}^{(1)} = f_{U|\lceil t \rceil,\Theta'}^{(1)} \qquad f_{U|t,\Theta}^{(2)} = f_{U|\tau_i,\Theta'}^{(2)} \qquad f_{U|t,\Theta}^{(3)} = f_{U|\tau_i,\Theta}^{(3)}$$

This assumption says that the density of the small layer depends on weekdays $\lceil t \rceil$ and the remaining terms only depend on the accident month $(\tau_{i-1}, \tau_i]$. The cumulative distribution function is under Assumption 5 for $t \in (\tau_{i-1}, \tau_i]$ given by

$$
F_{U|t,\Theta}(u) = \begin{cases} p^{(1)}_{U|\tau_i,\Theta} F^{(1)}_{U|\lceil t \rceil,\Theta}(u) & \text{for } u^{(0)} \le u < u^{(1)} \\ p^{(1)}_{U|\tau_i,\Theta} + p^{(2)}_{U|\tau_i,\Theta} F^{(2)}_{U|\tau_i,\Theta}(u) & \text{for } u^{(1)} \le u < u^{(2)} \\ p^{(1)}_{U|\tau_i,\Theta} + p^{(2)}_{U|\tau_i,\Theta} + p^{(3)}_{U|\tau_i,\Theta} F^{(3)}_{U|\tau_i,\Theta}(u) & \text{for } u^{(2)} \le u < u^{(3)} \end{cases} \tag{14}
$$

This split in layers again defines a partition and the likelihood decouples into independent parts; see also Theorem 2 of Norberg [6]. The log-likelihood of (9) is then at time τ_m given by

$$
\sum_{i=1}^{m} \sum_{n=1}^{3} \ell^{(i,n)}_{(T^{(i,n)}_\ell, S^{(i,n)}_\ell)_{\ell=1,\dots,M_{i,n}}}(\Lambda, \Theta) \quad \propto \quad \sum_{i=1}^{m} \sum_{n=1}^{3} -\pi^{(i,n)}_\Theta(\tau_m) W^{(i,n)}_{\Lambda,\Theta} + M_{i,n} \log(W^{(i,n)}_{\Lambda,\Theta})
$$
$$
+ \sum_{\ell=1}^{M_{i,n}} \log \left(\frac{\lambda_{T^{(i,n)}_\ell}}{\lambda^+_i} f^{(n)}_{U|T^{(i,n)}_\ell,\Theta}(U^{(i,n)}_\ell) \right) \tag{15}
$$

with $M_{i,n} = M_{i,n}(\tau_m)$ being the number of reported claims at time τ_m with accident dates $T^{(i,n)}_\ell \in (\tau_{i-1}, \tau_i]$ and reporting delays $U^{(i,n)}_\ell = S^{(i,n)}_\ell - T^{(i,n)}_\ell \in [u^{(n-1)}, u^{(n)})$. The total exposures for this partition are given by

$$
W^{(i,n)}_{\Lambda,\Theta} = w_i \Lambda_i \int_{\tau_{i-1}}^{\tau_i} \lambda_t \int_{u^{(n-1)}}^{u^{(n)}} f_{U|t,\Theta}(u) du \, dt = w_i \Lambda_i \, p^{(n)}_{U|\tau_i,\Theta} \int_{\tau_{i-1}}^{\tau_i} \lambda_t dt = w_i \Lambda_i \lambda^+_i p^{(n)}_{U|\tau_i,\Theta} \tag{16}
$$

The probability that these claims are reported at time τ_m is given by

$$
\pi^{(i,n)}_\Theta(\tau_m) = \frac{1}{\lambda^+_i} \int_{\tau_{i-1}}^{\tau_i} \lambda_t F^{(n)}_{U|t,\Theta}(\tau_m - t) \, dt \tag{17}
$$

Note that $F^{(n)}_{U|t,\Theta}(\tau_m - t) = 1$ for $\tau_m - t \ge u^{(n)}$ because in that case all claims have been reported at time τ_m with reporting delay less than $u^{(n)}$. This may substantially simplify the analysis in the lower and middle layers $n = 1, 2$, and is similar to (11).

5.2. Calibration of the Small Reporting Delay Layer

5.2.1. Model in the Small Reporting Delay Layer

We start by considering the small reporting delay layer $[u^{(0)}, u^{(1)})$. Data shows that reporting delays have a weekly pattern because claims divisions do not (necessarily) work at weekends and a claim occurring, for instance, on a Saturday can only be reported on Monday. This is illustrated in Figure 6. This indicates that we need a (week-) daily modeling approach. In order to not over-parametrize our model we try to keep this (week-) daily modeling layer as small as possible. The canonical choice then is to set $u^{(1)} = 7$ because after one week all claims have experienced a full weekly cycle and reporting should be on a similar level for all weekdays, this is supported by Figure 6 (middle), though not fully.

We could now try to maximize the log-likelihood (15) by brute force. Observe that this includes a coupling between all layers through exposures $W^{(i,n)}_{\Lambda,\Theta}$ because we have $p^{(3)}_{U|t,\Theta} = 1 - p^{(1)}_{U|t,\Theta} - p^{(2)}_{U|t,\Theta}$. This is unpleasant from a computational point of view, and we therefore propose an approximation.

For $i < m$ we have $\tau_m \geq u^{(1)} + \tau_i$ which implies $\pi_{\Theta}^{(i,1)}(\tau_m) = 1$ for all $i < m$. For $i = m$ we have from (17) and under Assumption 5, assumption $\lambda_t = \lambda_{\lceil t \rceil}$ and a change of variable $u = \tau_m - t$

$$\pi_{\Theta}^{(m,1)}(\tau_m) = \frac{1}{\lambda_m^+} \int_0^{\Delta\tau_m} \lambda_{\tau_m - u} F_{U|\tau_m-u,\Theta}^{(1)}(u)\,du \tag{18}$$

$$= 1 - \frac{1}{\lambda_m^+} \sum_{v=0}^{u^{(1)}-1} \lambda_{\tau_m-v} \int_v^{v+1} 1 - F_{U|\tau_m-v,\Theta}^{(1)}(u)\,du$$

This shows that if parameter $\Theta = (\Theta_1, \Theta_2, \Theta_3, p_1, p_2)$ splits into parameters Θ_n for $f_{U|t,\Theta}^{(n)} = f_{U|t,\Theta_n}^{(n)}$, $n = 1, 2, 3$, and $p_{U|t,\Theta}^{(1)} = p_1$ and $p_{U|t,\Theta}^{(2)} = p_2$ (we omit possible time dependence in the notation of Θ_n and p_n), then the coupling of the lower layer with the other two layers happens through $\pi_{\Theta_1}^{(m,1)}(\tau_m)$ and $p_{U|t,\Theta}^{(1)} = p_1$ and $p_{U|t,\Theta}^{(2)} = p_2$ in exposures $W_{\Lambda,\Theta}^{(i,n)}$. If we neglect in (18) the latest 7 days of observations, i.e., claims with occurrence dates $T_\ell^{(m,1)} \in (\tau_m - u^{(1)}, \tau_m]$, then the calibration of the lower layer density $f_{U|t,\Theta}^{(1)} = f_{U|t,\Theta_1}^{(1)}$ completely decouples from the other two layers because we have full information (no missing data) for accidents with occurrence dates in $(\tau_0, \tau_m - u^{(1)}]$ at time τ_m. This is indicated by the dashed red line in Figure 7 (lhs). In most cases this provides a reasonable approximation because the last $u^{(1)} = 7$ days will not completely change the calibration of the lower reporting delay distribution $F_{U|t,\Theta_1}^{(1)}$ if we have observations over, say, 10 years ≈ 3652 days. For this reason we shorten the last time interval to $(\tau_{m-1}, \tau_m - u^{(1)}]$ which provides in view of (15) MLE for Θ_1

$$0 \overset{!}{=} \frac{\partial}{\partial \Theta_1} \left[\sum_{i=1}^{m-1} \sum_{\ell=1}^{M_{i,1}} \log f_{U|T_\ell^{(i,1)},\Theta_1}^{(1)}(U_\ell^{(i,1)}) + \sum_{\ell=1}^{M'_{m,1}} \log f_{U|T_\ell^{(m,1)},\Theta_1}^{(1)}(U_\ell^{(m,1)}) \right] \tag{19}$$

where $M'_{m,1}$ denotes the number of reported claims at time τ_m with accident dates $T_\ell^{(m,1)} \in (\tau_{m-1}, \tau_m - u^{(1)}]$ and reporting delays $U_\ell^{(m,1)} \in [0, u^{(1)})$. The first component Θ_1 of Θ is assumed to fully characterize density $f_{U|t,\Theta_1}^{(1)}$ but no other part of the reporting delay distribution.
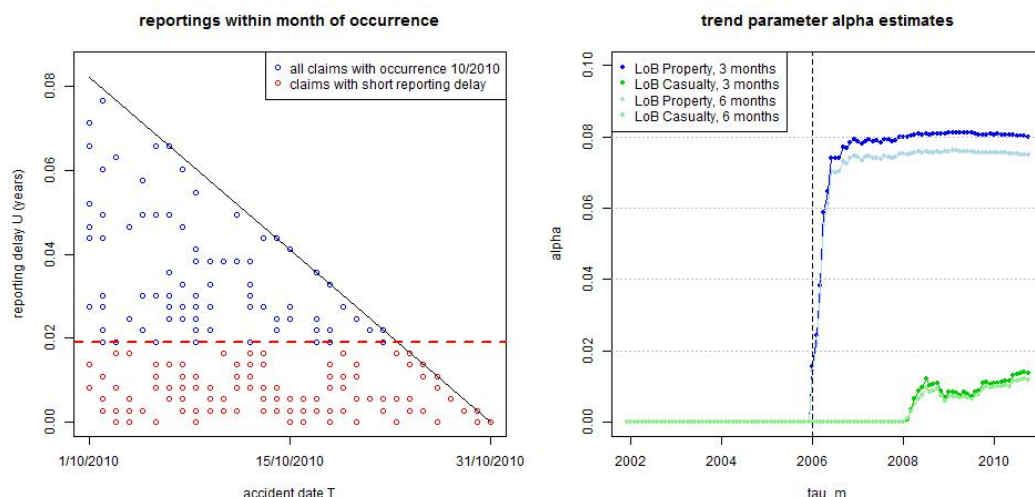

Figure 7. (lhs) small reporting delay layer as of $31/10/2010$ for occurrence dates in $10/2010$ of LoB Property, and (rhs) estimated trend parameter α at times $\tau_m \in \{31/12/2001, \ldots, 31/10/2010\}$ for LoB Property (blue/light blue) and LoB Casualty (green/light green) with dark colors are for $u^{(2)} = 3$ months and light colors for $u^{(2)} = 6$ months.

Next we discuss the explicit choice of $f_{U|t,\Theta_1}^{(1)}$. In Figure 6 (lhs) we plot the empirical distribution of all claims with accident dates before 01/2006 and a maximal reporting delay of 365 days. Figure 6 only shows the claims with reporting delays $U_\ell < u^{(1)} = 7$, i.e., belonging to the small layer. The lhs shows the individual delay distributions per weekday of occurrence, the middle pictures the same distributions but compressed by the weekends (because there are (almost) no reportings on Saturdays and Sundays) and the rhs shows the compressed graph that is normalized to 1 at time $u^{(1)}$. The graphs indicate that we should start by modeling weekdays individually. We make the following Ansatz: choose discrete distributions with, for $u = 0, \ldots, u^{(1)} - 1 = 6$ and $t \geq 0$,

$$f_{U|t,\Theta_1}^{(1)}(u) = \theta_{\lceil t \rceil}(u) = \theta_{\lceil t \rceil + u^{(1)}}(u) \tag{20}$$

with $\theta_{\lceil t \rceil}(u) \geq 0$ such that $\sum_{v=0}^{u^{(1)}-1} \theta_{\lceil t \rceil}(v) = 1$. The second identity in (20) implies that we obtain a weekly periodic reporting delay distribution with 42 parameters $\Theta_1 = (\theta_s(u))_{0 \leq u \leq u^{(1)}-1, 1 \leq s \leq u^{(1)}-1}$, if we assume stationarity (20) on the weekly time grid. We may even ask for more parameters because of potential non-stationarity of the reporting behavior, see Figure 2. We refrain from doing so but we use a rolling window to detect and capture non-stationarity. We should also remark that the 42 parameters raise questions about over-parametrization. We will investigate this question below and we will find that we can reduce the number of parameters in LoB Casualty, in LoB Property we will work with parametrization (20) which will provide rather stable results due to sufficient volume in this LoB.

Optimization problem (19) provides Lagrangian with Lagrange multipliers $\chi = (\chi_s)_{1 \leq s \leq u^{(1)}}$

$$\mathfrak{L}_{\tau_m}(\Theta_1, \chi) = \sum_{i=1}^{m-1} \sum_{\ell=1}^{M_{i,1}} \log f_{U|T_\ell^{(i,1)},\Theta_1}^{(1)}(U_\ell^{(i,1)}) + \sum_{\ell=1}^{M'_{m,1}} \log f_{U|T_\ell^{(m,1)},\Theta_1}^{(1)}(U_\ell^{(m,1)})$$
$$- \sum_{s=1}^{u^{(1)}} \chi_s \left(\sum_{v=0}^{u^{(1)}-1} \theta_s(v) - 1 \right)$$

The MLE of (19) is then found by setting the derivatives of $\mathfrak{L}_{\tau_m}(\Theta_1, \chi)$ w.r.t. Θ_1 and χ equal to zero and solving this system of equations. For $s = 1, \ldots, u^{(1)}$ and $u = 0, \ldots, u^{(1)} - 1$ we define $M''_{s,u} = M''_{s,u}(\tau_m)$ to be the number of reported claims at time τ_m with accident dates in $(\tau_0, \tau_m - u^{(1)}]$, reported on weekday s and having reporting delay u (we think of $s = 1$ being Mondays and $s = 7$ being Sundays). The MLE of $\theta_s(u)$ at time τ_m is then given by

$$\widehat{\theta}_s^{(\tau_m)}(u) = \frac{M''_{s,u}(\tau_m)}{\sum_{u=0}^{u^{(1)}-1} M''_{s,u}(\tau_m)}$$

The lower reporting delay layer distribution is at time τ_m estimated by

$$f_{U|t,\widehat{\Theta}_1^{(\tau_m)}}^{(1)}(u) = \widehat{\theta}_{\lceil t \rceil}^{(\tau_m)}(u) = \frac{M''_{\lceil t \rceil,u}(\tau_m)}{\sum_{u=0}^{u^{(1)}-1} M''_{\lceil t \rceil,u}(\tau_m)} \tag{21}$$

To capture potential non-stationarity we will choose a fixed window length K and consider this estimate based on observations in $(\tau_{(m-K)\vee 0}, \tau_m - u^{(1)}]$.

5.2.2. Empirics and Fitting the Small Reporting Delay Layer Distributions

We estimate distributions (21) in the lower reporting delay layer $[u^{(0)}, u^{(1)})$ for the 2 LoBs.

In Tables 2 and 3 we give the observed number of reported claims $M''_{s,u}(\tau_m)$ for weekdays $1 \leq s \leq 7$ and reporting delays $0 \leq u \leq 6 = u^{(1)} - 1$ at time $\tau_m = 31/10/2010$ for claims with accident dates before 26/10/2010. We see that there is a weekly pattern with no reportings on weekends in LoB Casualty and fewer reportings on weekends in LoB Property. For the latter we estimate all parameters

$\theta_s(u)$ individually, for the former we may also discuss other approaches, for instance, compressing weekends and shift the weekdays in Table 3 to the left (which is done on the rhs of Table 3). Moreover, the numbers in Table 3 are rather small and of similar size which may also suggest that we should not distinguish different reporting delays. We investigate this more formally below.

Table 2. Observed number of reported claims $M''_{s,u}(\tau_m)$ for weekdays $1 \le s \le 7$ and reporting delays $0 \le u \le 6 = u^{(1)} - 1$ at time $\tau_m = 31/10/2010$ for claims with accident dates before $26/10/2010$ of LoB Property.

$u =$	0	1	2	3	4	5	6
Mon $s = 1$	164	482	254	220	172	5	3
Tue $s = 2$	172	504	255	221	3	5	207
Wed $s = 3$	178	545	258	17	4	246	248
Thu $s = 4$	163	542	20	9	355	239	257
Fri $s = 5$	193	88	15	651	369	296	287
Sat $s = 6$	49	33	754	445	314	266	266
Sun $s = 7$	20	470	307	211	195	197	6

Table 3. Observed number of reported claims $M''_{s,u}(\tau_m)$ for weekdays $1 \le s \le 7$ and reporting delays $0 \le u \le 6 = u^{(1)} - 1$ at time $\tau_m = 31/10/2010$ for claims with accident dates before $26/10/2010$ of LoB Casualty, the rhs is compressed by weekends.

$u =$	0	1	2	3	4	5	6	0*	1*	2*	3*	4*
Mon $s = 1$	17	31	33	41	34	0	0	17	31	33	41	34
Tue $s = 2$	15	36	25	42	0	0	30	15	36	25	42	30
Wed $s = 3$	16	40	25	0	0	43	53	16	40	25	43	53
Thu $s = 4$	19	43	0	0	33	27	38	19	43	33	27	38
Fri $s = 5$	12	0	0	29	40	33	40	12	29	40	33	40
Sat $s = 6$	0	0	9	10	12	10	15	9	10	12	10	15
Sun $s = 7$	0	5	12	7	5	10	0	5	12	7	5	10

We start with LoB Property. We show in Figure 8 the resulting estimates (see (20)) with a rolling window of length $2 \cdot 365$ days (solid lines) which is compared to the estimate considering all observations (dotted lines). We clearly see the non-stationarity after the break point at $1/1/2006$, and the rolling window seems to capture it rather well. Therefore, we do not use any other measures here, but work with the rolling window of length $2 \cdot 365$ days (solid lines).
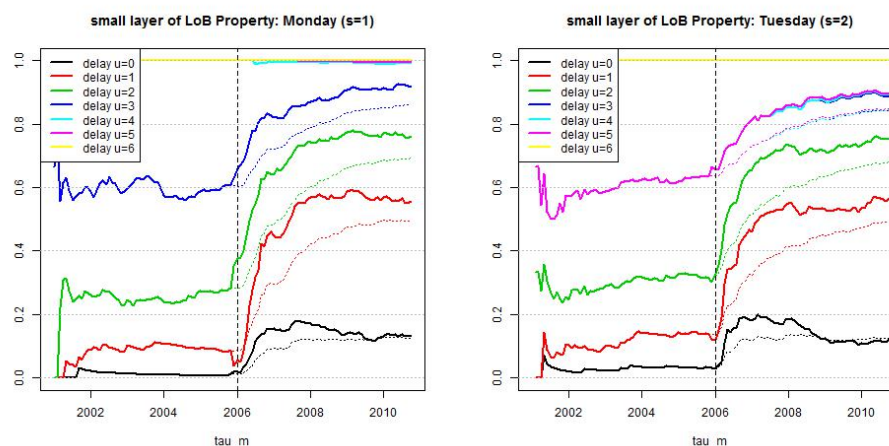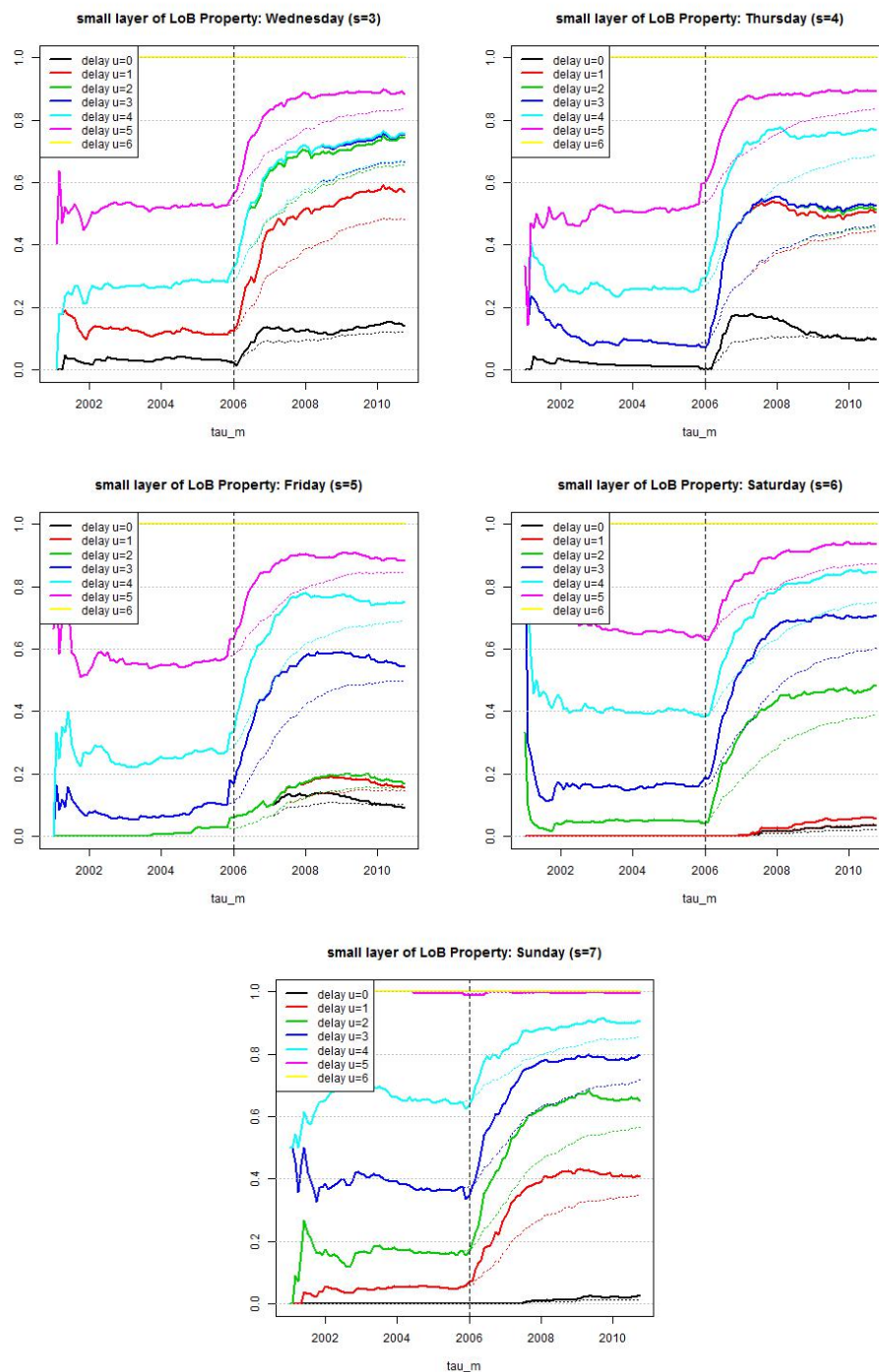

Figure 8. *Cont.*

Figure 8. LoB Property, small reporting delay layer: estimated cumulative distribution $F_{U|t,\Theta}^{(1)}$ per weekday $s = \lceil t \rceil$. Dotted lines show calibration based on all observations since τ_0 and lines show calibration based on a rolling window of length $2 \cdot 365$ days.

For LoB Casualty the non-stationarity is less obvious, see Figure 9. In fact, the resulting estimates with a rolling window of length $2 \cdot 365$ days (solid lines) are rather volatile which is a clear sign of over-parametrization. The dotted lines show the estimates based on all available observations, these are much smoother with a slight positive trend for some of the weekdays. At this point we could investigate more thoroughly this non-stationarity, we refrain from doing so because in this case study it may only marginally influence the estimation of the number of IBNYR claims: the potential trend

has a very moderate slope which affects less then 10% of the claims in LoB Casualty (small reporting delay layer). For this reason, we simply choose a stationary model and we mainly aim at studying whether we can further reduce the number of parameters in Θ_1. First, we compress the weekends (in Table 3 we go from the lhs denoted $0, \ldots, 6$ to the rhs denoted $0^*, \ldots, 4^*$). Then we test the null hypothesis whether for all weekdays $s = 1, \ldots, 7$ and compressed reporting delays $u^* = 1^*, \ldots, 4^*$ we can choose the same (empirical) probability, and for all weekdays $s = 1, \ldots, 7$ we can choose the same probability for reporting delay $u^* = 0^*$, that is, we test the null hypothesis $\theta_1(0^*) = \ldots = \theta_7(0^*)$ (delay 0^* does not differ between weekdays $s = 1, \ldots, 7$) and $\theta_1(1^*) = \ldots = \theta_7(4^*)$ (delays $1^*, \ldots, 4^*$ and weekdays $s = 1, \ldots, 7$ do not differ). We perform for every weekday $s = 1, \ldots, 7$ a Pearson's χ^2-test (for the information at time $\tau_m = 31/10/2010$). The corresponding test statistics is

$$
\chi_s^2 = \sum_{u^*=0^*}^{4^*} \frac{\left(M''_{s,u^*}(\tau_m) - M''_{s,\bullet}(\tau_m)\, \widehat{\theta}_s^{(\tau_m)}(u^*) \right)^2}{M''_{s,\bullet}(\tau_m)\, \widehat{\theta}_s^{(\tau_m)}(u^*)}, \tag{22}
$$

with $M''_{s,u^*}(\tau_m)$ being the number of reported claims with occurrence day s and compressed reporting delay u^*, $M''_{s,\bullet}(\tau_m) = \sum_{u^*=0^*}^{4^*} M''_{s,u^*}(\tau_m)$ and $\widehat{\theta}_s^{(\tau_m)}(u^*)$ being the corresponding MLE under the null hypothesis.
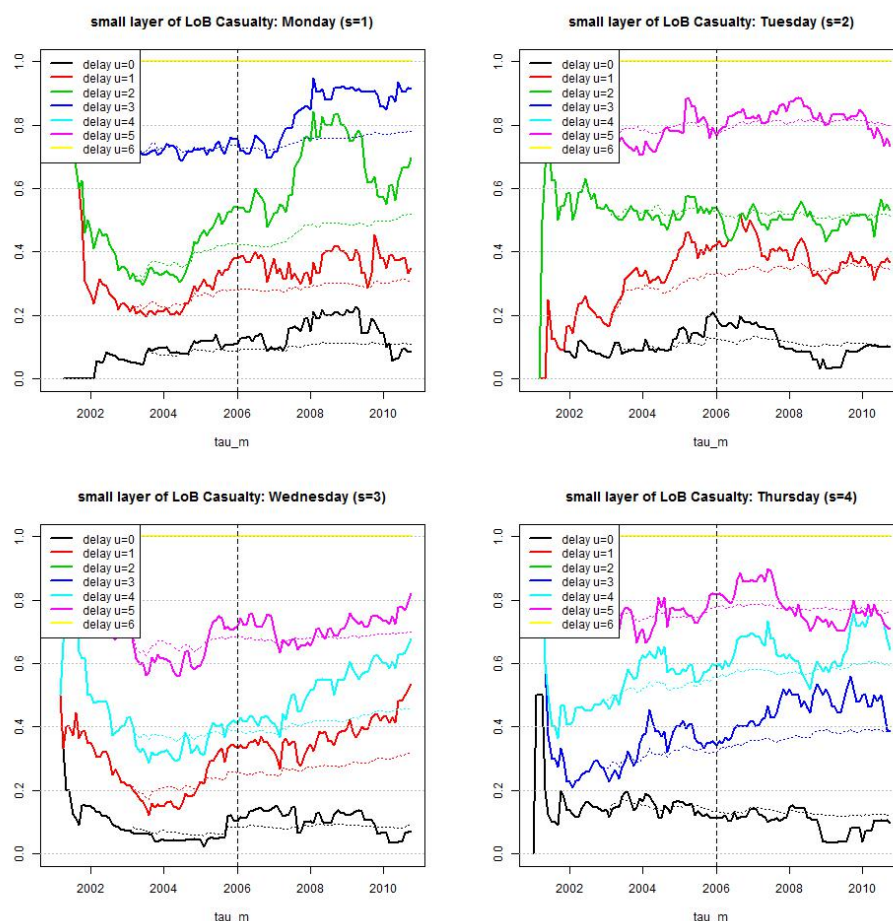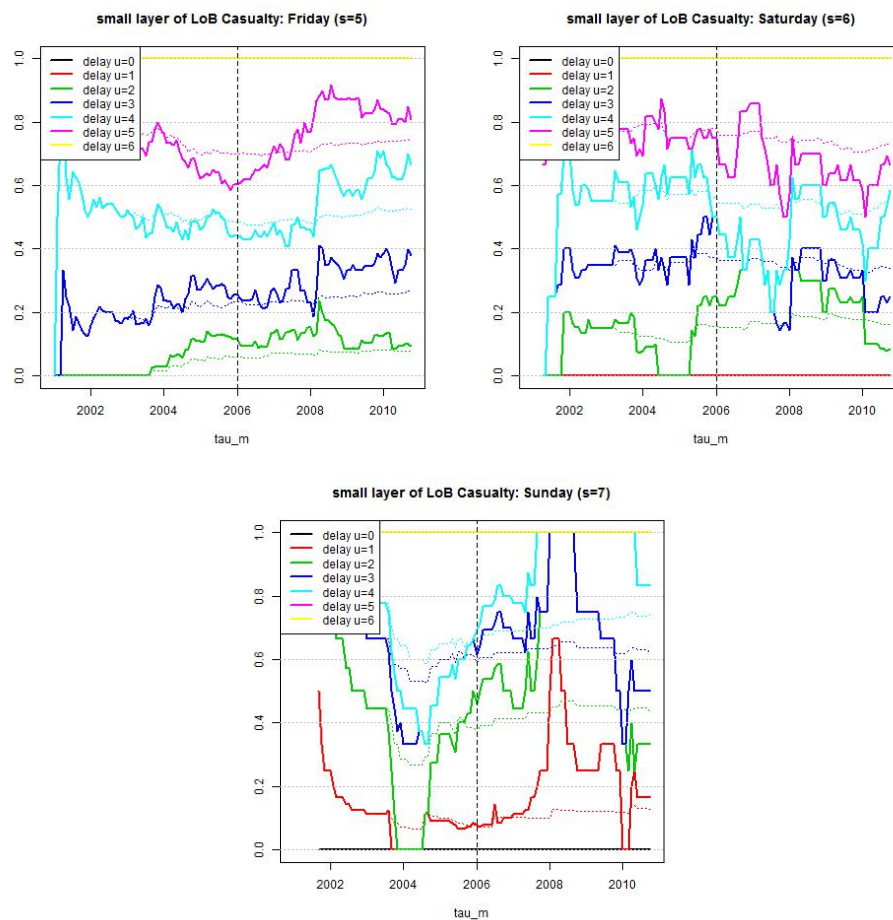

Figure 9. *Cont.*

Figure 9. LoB Casualty, small reporting delay layer: estimated cumulative distribution $F^{(1)}_{U|t,\Theta}$ per weekday $s = \lceil t \rceil$. Dotted lines show calibration based on all observations since τ_0 and lines show calibration based on a rolling window of length $2 \cdot 365$ days.

We provide the resulting *p*-values in Table 4. The resulting *p*-value for claims with accident dates on Wednesdays $s = 3$ is 3.2% and for all other weekdays s we obtain *p*-values bigger than 20%. We consider these *p*-values to be sufficiently large so that we do not reject the null hypothesis. This leads to a substantial reduction in the number of parameters, and we only choose three different values for $\Theta_1 = (\theta_s(u))_{0 \le u \le u^{(1)}-1, 1 \le s \le u^{(1)}}$ at any time point τ_m in this reduced case (the third one being 0% for weekends). In Figure 10 we show the results, the solid line gives the estimates under the null hypothesis and the dotted lines the estimates of the model with 42 parameters. For LoB Casualty we choose this reduced model (under the null hypothesis).

Using (18) and the fact that we choose a step function for $F^{(1)}_{U|t,\Theta_1}$ we obtain estimated probability in the small reporting delay layer given by

$$\pi^{(m,1)}_{\widehat{\Theta}^{(\tau_m)}_1}(\tau_m) = 1 - \frac{1}{\lambda^+_m} \sum_{v=0}^{u^{(1)}-1} \lambda_{\tau_m-v} \left(1 - \sum_{u=0}^{v} \widehat{\theta}^{(\tau_m)}_{\tau_m-v}(u)\right). \tag{23}$$

The results are presented in Figure 4 and they are compared to the case where we do not choose a weekly periodic pattern, that is, where we set $(\lambda_k)_{k=1,\dots,7} \equiv 1$. This latter model is the one used in Section 4.2 of Antonio-Plat [17]. We see that the weekly periodic pattern essentially smooths the estimates, in particular, for weekends in LoB Casualty. This confirms the findings of Section 4.3 and, in particular, of Figure 5. For this reason we continue with the model allowing us for the modeling of a

weekly periodic pattern $(\lambda_k)_{k=1,\ldots,7}$. We fix the resulting estimates $\pi_{\widehat{\Theta}_1^{(\tau_m)}}^{(m,1)}(\tau_m)$ and then calibrate the middle and large layers, this is explained next.

Table 4. p-values of the χ^2-tests under the corresponding null hypotheses for test statistics χ_s^2, see Equation (22), for weekdays $s = 1,\ldots,7$ (with 4 degrees of freedom).

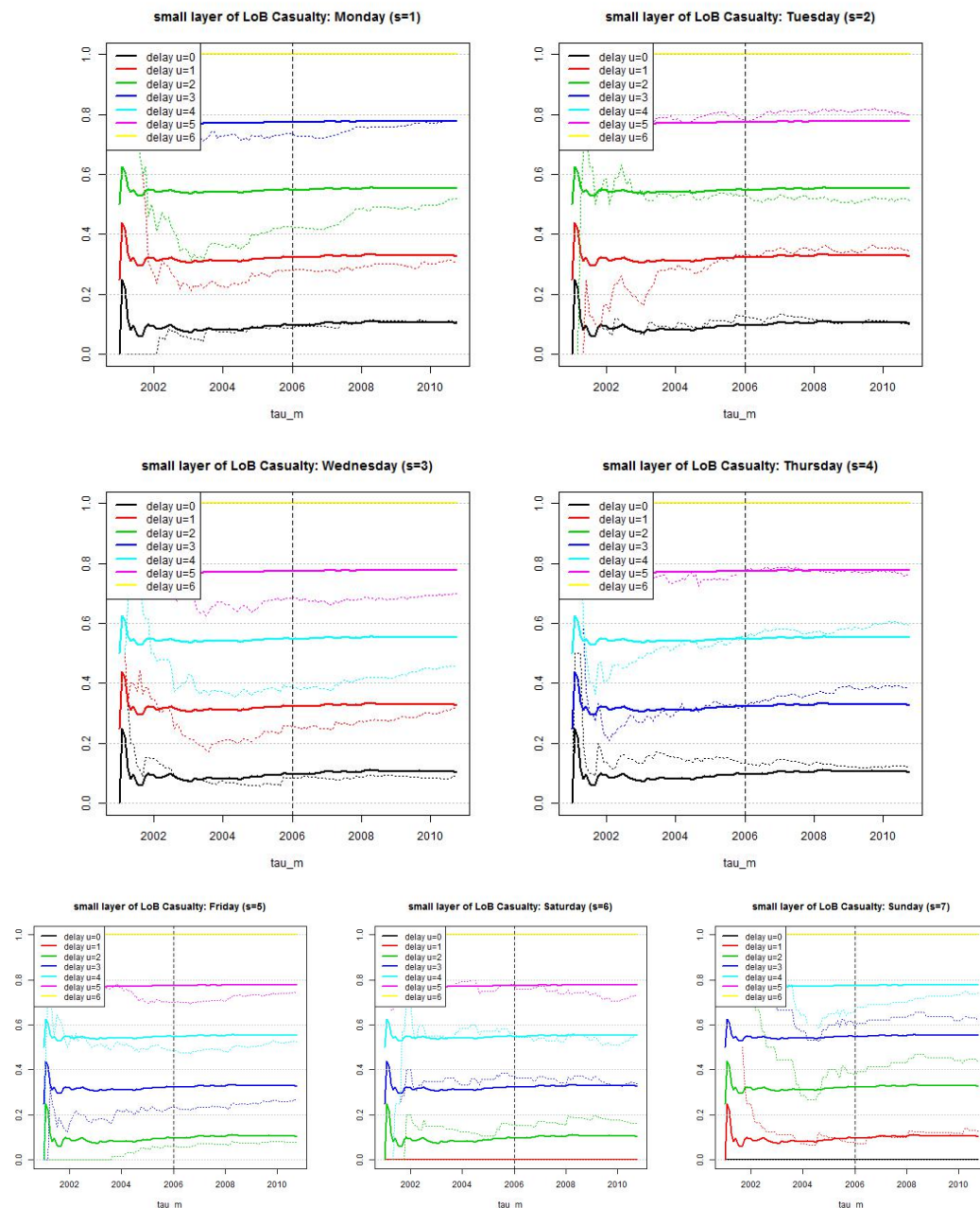|	Mon $s = 1$	Tue $s = 2$	Wed $s = 3$	Thu $s = 4$	Fri $s = 5$	Sat $s = 6$	Sun $s = 7$
p-values	79%	29%	3.2%	37%	44%	52%	47%

Figure 10. LoB Casualty, small reporting delay layer: estimated cumulative distribution $F_{U|t,\Theta}^{(1)}$ per weekday $s = \lceil t \rceil$. Dotted lines show calibration based on individual weekdays and reporting delays and lines show calibration under compressed weekends and the null hypothesis that we only need three parameters.

5.3. Calibration of Middle and Large Reporting Delay Layers

We come back to log-likelihood (15). We replace in the small reporting delay layer parameter Θ_1 by its estimate $\widehat{\Theta}_1^{(\tau_m)}$ derived in the previous subsection. This provides the log-likelihood at time τ_m, we only show the terms including the unknown parameters $(\Lambda, \Theta_{-1}) := (\Lambda, \Theta_2, \Theta_3, p_1, p_2)$,

$$
\begin{aligned}
\ell_{\tau_m}(\Lambda, \Theta_{-1}) \quad \propto \quad & \sum_{i=1}^{m} \left[-\pi_{\widehat{\Theta}_1^{(\tau_m)}}^{(i,1)}(\tau_m) W_{\Lambda,\Theta}^{(i,1)} + \sum_{n=2}^{3} -\pi_{\Theta_n}^{(i,n)}(\tau_m) W_{\Lambda,\Theta}^{(i,n)} + \sum_{n=1}^{3} M_{i,n} \log(W_{\Lambda,\Theta}^{(i,n)}) \right] \\
& + \sum_{i=1}^{m} \sum_{n=2}^{3} \sum_{\ell=1}^{M_{i,n}} \log \left(\frac{\lambda_{T_\ell^{(i,n)}}}{\lambda_i^+} f_{U|T_\ell^{(i,n)},\Theta_n}^{(n)}(U_\ell^{(i,n)}) \right)
\end{aligned}
\tag{24}
$$

where $\pi_{\widehat{\Theta}_1^{(\tau_m)}}^{(i,1)}(\tau_m) = 1$ for $i < m$. From this we compute the MLE of $\Lambda = (\Lambda_i)_{i=1,\dots,m}$ (see also Lemma 3):

$$
\Lambda_i = \frac{M_i(\tau_m)}{w_i \lambda_i^+ \left(\pi_{\widehat{\Theta}_1^{(\tau_m)}}^{(i,1)}(\tau_m) p_{U|\tau_i,\Theta}^{(1)} + \sum_{n'=2}^{3} \pi_{\Theta_{n'}}^{(i,n')}(\tau_m) p_{U|\tau_i,\Theta}^{(n')} \right)} \qquad \text{for } i = 1, \dots, m
\tag{25}
$$

If we insert this back into (24) we get (only stating relevant terms for parameter estimation)

$$
\begin{aligned}
\ell_{\tau_m}(\Theta_{-1}) \quad \propto \quad & \sum_{i=1}^{m} \sum_{n=1}^{3} M_{i,n} \log \left(\frac{p_{U|\tau_i,\Theta}^{(n)}}{\pi_{\widehat{\Theta}_1^{(\tau_m)}}^{(i,1)}(\tau_m) p_{U|\tau_i,\Theta}^{(1)} + \sum_{n'=2}^{3} \pi_{\Theta_{n'}}^{(i,n')}(\tau_m) p_{U|\tau_i,\Theta}^{(n')}} \right) \\
& + \sum_{i=1}^{m} \sum_{n=2}^{3} \sum_{\ell=1}^{M_{i,n}} \log \left(f_{U|\tau_i,\Theta_n}^{(n)}(U_\ell^{(i,n)}) \right)
\end{aligned}
$$

where we have used Assumption 5 for $f_{U|t,\Theta_n}^{(n)}$ with $n = 2, 3$. Recall that we have normalization $\sum_{n'=1}^{3} p_{U|\tau_i,\Theta}^{(n')} = 1$ of the layer probabilities because the reporting delays need to be in one of the three layers. Assuming $p_{U|\tau_i,\Theta}^{(3)} > 0$ we can normalize these probabilities by dividing by this third probability and setting $q_{\tau_i}^{(n)} = p_{U|\tau_i,\Theta}^{(n)} / p_{U|\tau_i,\Theta}^{(3)} \geq 0$. From this we see that we can rewrite the last log-likelihood in terms of $q = (q_{\tau_i}^{(1)}, q_{\tau_i}^{(2)}, q_{\tau_i}^{(3)} = 1)_{1 \leq i \leq m}$. This provides the log-likelihood

$$
\begin{aligned}
\ell_{\tau_m}(\Theta_2, \Theta_3, q) \quad \propto \quad & -\sum_{i=1}^{m} M_i \log \left(\pi_{\widehat{\Theta}_1^{(\tau_m)}}^{(i,1)}(\tau_m) q_{\tau_i}^{(1)} + \sum_{n'=2}^{3} \pi_{\Theta_{n'}}^{(i,n')}(\tau_m) q_{\tau_i}^{(n')} \right) \\
& + \sum_{i=1}^{m} \sum_{n=1}^{2} M_{i,n} \log \left(q_{\tau_i}^{(n)} \right) + \sum_{i=1}^{m} \sum_{n=2}^{3} \sum_{\ell=1}^{M_{i,n}} \log \left(f_{U|\tau_i,\Theta_n}^{(n)}(U_\ell^{(i,n)}) \right)
\end{aligned}
\tag{26}
$$

To implement MLE of (26) there remains the calculation of $\pi_{\Theta_{n'}}^{(i,n')}(\tau_m)$ for $n' = 2, 3$. This is what we are going to discuss next.

5.3.1. Choice of Layers and Approximate Log-likelihood

We still need to specify threshold $u^{(2)}$. The lower limit was chosen to be $u^{(1)} = 7$ days. For $u^{(2)}$ we test different reporting delays $\kappa = 3, 6, 9$ or 12 months. Note that κ months is not well-defined in terms of number of days. We set $u^{(2)}$ equal to $89, 181, 273$ or 365 which is the minimal number of days

that κ consecutive calendar months can have. The accident periods $(\tau_{i-1}, \tau_i]$ with $i \leq m - \kappa$ are then fully observed in the middle layer at time τ_m, and we have

$$\pi_{\Theta_2}^{(i,2)}(\tau_m) = 1 \qquad \text{for } 1 \leq i \leq m - \kappa$$

Thus, we only need to study $m - \kappa + 1 \leq i \leq m$ in more detail for the middle reporting delay layer. As for the large reporting delay layer we see that there are no observations possible at time τ_m for accident periods $(\tau_{i-1}, \tau_i]$ with $m - \kappa + 2 \leq i \leq m$ and, therefore,

$$\pi_{\Theta_3}^{(i,3)}(\tau_m) = 0 \qquad \text{for } m - \kappa + 2 \leq i \leq m$$

The remaining layers and probabilities are more involved, and we consider these next.

Large reporting delay layer $[u^{(2)}, \infty)$ *with* $n = 3$. For $i \leq m - \kappa + 1$ we have under Assumption 5

$$\pi_{\Theta_3}^{(i,3)}(\tau_m) = \frac{1}{\lambda_i^+} \int_{\tau_{i-1}}^{\tau_i} \lambda_t F_{U|\tau_i,\Theta_3}^{(3)}(\tau_m - t)\, dt$$

To simplify optimization (26) we assume that for reportings in the large layer the weekly periodic pattern $(\lambda_k)_{k=1,\ldots,7}$ only has a marginal influence. This is justified by the argument that between claims occurrence and reporting there are at least κ months of delay, and therefore the specific weekday of the accident should only marginally influence the late reporting (as long as a particular claim type cannot only occur on one specific weekday). For $i \leq m - \kappa$ this provides the approximation

$$\pi_{\Theta_3}^{(i,3)}(\tau_m) \approx \widetilde{\pi}_{\Theta_3}^{(i,3)}(\tau_m) = \frac{1}{\Delta\tau_i} \int_{\tau_m - \tau_i}^{\tau_m - \tau_{i-1}} F_{U|\tau_i,\Theta_3}^{(3)}(u)\, du$$

The situation $i = m - \kappa + 1$ is more delicate. We have $u^{(2)} \in \{89, 181, 273, 365\}$ which is the minimal number of days that κ consecutive calendar months can have, the maximal number of days being 92, 184, 276 or 366 days, respectively. Therefore, the following integral may also be non-zero on time interval $(\tau_{m-\kappa}, \tau_{m-\kappa+1}]$ at time τ_m

$$
\begin{aligned}
\pi_{\Theta_3}^{(m-\kappa+1,3)}(\tau_m) \approx \widetilde{\pi}_{\Theta_3}^{(m-\kappa+1,3)}(\tau_m) &= \frac{1}{\Delta\tau_{m-\kappa+1}} \int_{\tau_m - \tau_{m-\kappa+1}}^{\tau_m - \tau_{m-\kappa}} F_{U|\tau_{m-\kappa+1},\Theta_3}^{(3)}(u)\, du \\
&= \frac{1}{\Delta\tau_{m-\kappa+1}} \int_{u^{(2)}}^{\tau_m - \tau_{m-\kappa}} F_{U|\tau_{m-\kappa+1},\Theta_3}^{(3)}(u)\, du
\end{aligned}
$$

Note that $\Delta\tau_{m-\kappa+1} \geq 28$ and $\tau_m - \tau_{m-\kappa} - u^{(2)} \leq 3$ which implies that $\widetilde{\pi}_{\Theta_3}^{(m-\kappa+1,3)}(\tau_m) \leq 3/28 < 1/9$ for all m. Therefore, this term only marginally influences the results (and it could also be skipped for parameter estimation but we will keep it).

Middle reporting delay layer $[u^{(1)}, u^{(2)})$ *with* $n = 2$. We still need to treat the cases $m - \kappa + 1 \leq i \leq m - 1$ and $i = m$. We have for $m - \kappa + 1 \leq i \leq m$

$$
\begin{aligned}
\pi_{\Theta_2}^{(i,2)}(\tau_m) &= \frac{1}{\lambda_i^+} \int_{\tau_{i-1}}^{\tau_i} \lambda_t F_{U|\tau_i,\Theta_2}^{(2)}(\tau_m - t)\, dt \\
&= \frac{1}{\lambda_i^+} \sum_{v=0}^{\Delta\tau_i - 1} \lambda_{\tau_i - v} \int_v^{v+1} F_{U|\tau_i,\Theta_2}^{(2)}(\tau_m - \tau_i + u)\, du
\end{aligned}
$$

As we indicate below, this can be implemented and MLE can be performed. In order to speed up the MLE optimization we also approximate $\pi_{\Theta_2}^{(i,2)}(\tau_m)$. However, this approximation is *only* used for parameter estimation, for the number of IBNYR claims estimation we will use the exact form

$\pi_{\Theta_2}^{(i,2)}(\tau_m)$. For the approximation in MLE we also neglect weekday differences which provides for $m - \kappa + 1 \le i \le m - 1$

$$\pi_{\Theta_2}^{(i,2)}(\tau_m) \approx \widetilde{\pi}_{\Theta_2}^{(i,2)}(\tau_m) = \frac{1}{\Delta \tau_i} \int_{\tau_m - \tau_i}^{\tau_m - \tau_{i-1}} F_{U|\tau_i,\Theta_2}^{(2)}(u)\,du$$

For $i = m$ we have $F_{U|\tau_m,\Theta_2}^{(2)}(u) = 0$ for $u \le u^{(1)}$. This provides the approximation

$$\pi_{\Theta_2}^{(m,2)}(\tau_m) \approx \widetilde{\pi}_{\Theta_2}^{(m,2)}(\tau_m) = \frac{1}{\Delta \tau_m} \int_{u^{(1)}}^{\Delta \tau_m} F_{U|\tau_m,\Theta_2}^{(2)}(u)\,du$$

because otherwise reporting delays belong to the small layer.

This allows us to approximate the log-likelihood (26) by (we also refer to Corollary 7, below)

$$
\widetilde{\ell}_{\tau_m}(\Theta_2, \Theta_3, q) \quad \propto \quad -\sum_{i=1}^{m} M_i \log \left(\pi_{\widehat{\Theta}_1^{(\tau_m)}}^{(i,1)}(\tau_m) q_{\tau_i}^{(1)} + \sum_{n'=2}^{3} \widetilde{\pi}_{\Theta_{n'}}^{(i,n')}(\tau_m) q_{\tau_i}^{(n')} \right) \tag{27}
$$
$$
+ \sum_{i=1}^{m} \sum_{n=1}^{2} M_{i,n} \log \left(q_{\tau_i}^{(n)} \right) + \sum_{i=1}^{m} \sum_{n=2}^{3} \sum_{\ell=1}^{M_{i,n}} \log \left(f_{U|\tau_i,\Theta_n}^{(n)}(U_\ell^{(i,n)}) \right)
$$

where several of the π's and $\widetilde{\pi}$'s are either 0 or 1 (we give them in detail below). To calculate them we need the following lemma.

Lemma 6. *Assume $X \sim F$ and choose $x_1 < x_2$. We have*

$$\int_{x_1}^{x_2} F(x)\,dx = x_2 F(x_2) - x_1 F(x_1) - \mathbb{E}\left[X 1_{\{x_1 < X \le x_2\}} \right]$$

Proof of Lemma 6. The proof follows by applying integration by parts. $\square$

Corollary 7. *We choose $U^{(i,n)} \sim F_{U|\tau_i,\Theta_n}^{(n)}$ and denote the corresponding expectation by $E_{U|\tau_i,\Theta_n}^{(n)}$.*

1. *Small layer $[0, u^{(1)}) = [0, 7)$. For $1 \le i \le m - 1$ we have probability $\pi_{\widehat{\Theta}_1^{(\tau_m)}}^{(i,1)}(\tau_m) = 1$ and the case $i = m$ is given by (23).*

2. *Middle layer $[u^{(1)}, u^{(2)}) = [7, u^{(2)})$. For $1 \le i \le m - \kappa$ we have probability $\pi_{\Theta_2}^{(i,2)}(\tau_m) = \widetilde{\pi}_{\Theta_2}^{(i,2)}(\tau_m) = 1$. For $m - \kappa + 1 \le i \le m - 1$ we have*

$$
\widetilde{\pi}_{\Theta_2}^{(i,2)}(\tau_m) = \frac{\tau_m - \tau_{i-1}}{\Delta \tau_i} F_{U|\tau_i,\Theta_2}^{(2)}(\tau_m - \tau_{i-1}) - \frac{\tau_m - \tau_i}{\Delta \tau_i} F_{U|\tau_i,\Theta_2}^{(2)}(\tau_m - \tau_i)
$$
$$
- \frac{1}{\Delta \tau_i} E_{U|\tau_i,\Theta_2}^{(2)} \left[U^{(i,2)} 1_{\{\tau_m - \tau_i < U^{(i,2)} \le \tau_m - \tau_{i-1}\}} \right]
$$

and for $i = m$

$$
\widetilde{\pi}_{\Theta_2}^{(m,2)}(\tau_m) = F_{U|\tau_m,\Theta_2}^{(2)}(\Delta \tau_m) - \frac{1}{\Delta \tau_m} E_{U|\tau_m,\Theta_2}^{(2)} \left[U^{(m,2)} 1_{\{u^{(1)} < U^{(m,2)} \le \Delta \tau_m\}} \right]
$$

3. *Large layer $[u^{(2)}, \infty)$. For $m - \kappa + 2 \le i \le m$ we have probability $\pi_{\Theta_3}^{(i,3)}(\tau_m) = \widetilde{\pi}_{\Theta_3}^{(i,3)}(\tau_m) = 0$. For $1 \le i \le m - \kappa$ we have*

$$
\widetilde{\pi}_{\Theta_3}^{(i,3)}(\tau_m) = \frac{\tau_m - \tau_{i-1}}{\Delta \tau_i} F_{U|\tau_i,\Theta_3}^{(3)}(\tau_m - \tau_{i-1}) - \frac{\tau_m - \tau_i}{\Delta \tau_i} F_{U|\tau_i,\Theta_3}^{(3)}(\tau_m - \tau_i)
$$
$$
- \frac{1}{\Delta \tau_i} E_{U|\tau_i,\Theta_3}^{(3)} \left[U^{(i,3)} 1_{\{\tau_m - \tau_i < U^{(i,3)} \le \tau_m - \tau_{i-1}\}} \right]
$$

and for $i = m - \kappa + 1$

$$
\begin{aligned}
\widetilde{\pi}_{\Theta_3}^{(m-\kappa+1,3)}(\tau_m) \;=\; & \frac{\tau_m - \tau_{m-\kappa}}{\Delta \tau_{m-\kappa+1}} F_{U|\tau_{m-\kappa+1},\Theta_3}^{(3)}(\tau_m - \tau_{m-\kappa}) \\
& - \frac{1}{\Delta \tau_{m-\kappa+1}} E_{U|\tau_{m-\kappa+1},\Theta_3}^{(3)}\left[U^{(m-\kappa+1,3)} 1_{\{u^{(2)} < U^{(m-\kappa+1,3)} \le \tau_m - \tau_{m-\kappa}\}} \right]
\end{aligned}
$$

Finally, we need to choose explicit distribution functions $F_{U|\tau_i,\Theta_n}^{(n)}$ for the reporting delay layers $n = 2, 3$. This is what we are going to do in the next subsection.

5.3.2. Choice of Explicit Distributions and Layer Probabilities

There remains the modeling of the reporting delay distributions $F_{U|\tau_i,\Theta_n}^{(n)}$ in the middle and large layers as well as the relative layer probabilities $q_{\tau_i}^{(n)}$ for $n = 1, 2$. We have considered different models, compared them to each other, checked them for robustness and applied statistical model selection criteria such as Akaike's information criterion. Our favorite model that is at the same time not too difficult and gives appropriate results is the following: (i) for the middle layer $n = 2$ we choose a stationary truncated log-normal distribution, (ii) for the upper layer $n = 3$ we choose a stationary shifted log-normal distribution, and (iii) we choose non-stationary relative layer probabilities $q_{\tau_i}^{(n)}$. We justify these choices by some statistical analysis below.

Lemma 8. *Assume $X^{(2)}$ has a truncated log-normal distribution with parameters $\mu_2 \in \mathbb{R}$ and $\sigma_2 > 0$ supported in a non-empty interval $[\nu_{i-1}, \nu_i] \subset \mathbb{R}_+$. The density of $X^{(2)}$ is given by*

$$
f^{(2)}(x) = \frac{1}{\Phi\left(\frac{\log \nu_i - \mu_2}{\sigma_2}\right) - \Phi\left(\frac{\log \nu_{i-1} - \mu_2}{\sigma_2}\right)} \frac{1}{\sqrt{2\pi}\sigma_2} \frac{1}{x} \exp\left\{ -\frac{(\log x - \mu_2)^2}{\sigma_2^2} \right\} 1_{\{\nu_{i-1} \le x \le \nu_i\}}
$$

The distribution of $X^{(2)}$ is given by

$$
F^{(2)}(x) = \frac{\Phi\left(\frac{\log(x \wedge \nu_i) - \mu_2}{\sigma_2}\right) - \Phi\left(\frac{\log \nu_{i-1} - \mu_2}{\sigma_2}\right)}{\Phi\left(\frac{\log \nu_i - \mu_2}{\sigma_2}\right) - \Phi\left(\frac{\log \nu_{i-1} - \mu_2}{\sigma_2}\right)} 1_{\{\nu_{i-1} \le x\}}
$$

The expectation of $X^{(2)}$ on layer $(x_1, x_2] \subset [\nu_{i-1}, \nu_i]$ is given by

$$
\mathbb{E}\left[X^{(2)} 1_{\{x_1 < X^{(2)} \le x_2\}} \right] = \exp\{\mu_2 + \sigma_2^2/2\} \frac{\Phi\left(\frac{\log x_2 - (\mu_2 + \sigma_2^2)}{\sigma_2}\right) - \Phi\left(\frac{\log x_1 - (\mu_2 + \sigma_2^2)}{\sigma_2}\right)}{\Phi\left(\frac{\log \nu_i - \mu_2}{\sigma_2}\right) - \Phi\left(\frac{\log \nu_{i-1} - \mu_2}{\sigma_2}\right)}
$$

Assume $X^{(3)} = \nu + Z \sim F^{(3)}$ has a shifted log-normal distribution with Z being log-normally distributed with parameters $\mu_3 \in \mathbb{R}$ and $\sigma_3 > 0$. We have on the layer $(x_1, x_2] \subset [\nu, \infty)$, set $z_1 = x_1 - \nu$ and $z_2 = x_2 - \nu$,

$$
\begin{aligned}
\int_{x_1}^{x_2} F^{(3)}(x)dx \;=\; & z_2 \Phi\left(\frac{\log z_2 - \mu_3}{\sigma_3}\right) - z_1 \Phi\left(\frac{\log z_1 - \mu_3}{\sigma_3}\right) \\
& - \exp\left\{ \mu_3 + \frac{\sigma_3^2}{2} \right\} \left[\Phi\left(\frac{\log z_2 - (\mu_3 + \sigma_3^2)}{\sigma_3}\right) - \Phi\left(\frac{\log z_1 - (\mu_3 + \sigma_3^2)}{\sigma_3}\right) \right]
\end{aligned}
$$

Proof of Lemma 8. The proof is a straightforward consequence of calculations with log-normal distributions, see also Section 3.2.3 in [21]. □

The model of Lemma 8 will be chosen below. We will compare it to the situation where we also have a truncated log-normal distribution for $F^{(3)}$ in the upper layer, and we also compare it to the case of replacing the log-normal by gamma distributions. More details are provided below.

There remains the choice of $q_{\tau_i}^{(n)} = p_{U|\tau_i,\Theta}^{(n)}/p_{U|\tau_i,\Theta}^{(3)} \geq 0$ for $n = 1, 2$. We consider for break point $\tau_{m_0} = 1/1/2006$, $\gamma > 0$ and $t \in (\tau_{i-1}, \tau_i]$ the functional forms

$$p_{U|\tau_i,\Theta}^{(1)}/p_{U|\tau_i,\Theta}^{(3)} \;=\; q_{\tau_i}^{(1)} \;=\; \overline{q}^{(1)} \exp\left\{\alpha(i-m_0)_+^{\gamma}\right\} \tag{28}$$

$$p_{U|\tau_i,\Theta}^{(2)}/p_{U|\tau_i,\Theta}^{(3)} \;=\; q_{\tau_i}^{(2)} \;=\; \overline{q}^{(2)} \exp\left\{-\alpha\frac{\overline{q}^{(1)}}{\overline{q}^{(2)}}(i-m_0)_+^{\gamma}\right\}$$

with given trend parameter $\alpha \geq 0$ after break point τ_{m_0}. For $\alpha \to 0$ we have

$$p_{U|t,\Theta}^{(1)} \;=\; \overline{q}^{(1)} p_{U|\tau_i,\Theta}^{(3)} + \alpha\,\overline{q}^{(1)} p_{U|\tau_i,\Theta}^{(3)}(i-m_0)_+^{\gamma} + o(\alpha)$$

$$p_{U|t,\Theta}^{(2)} \;=\; \overline{q}^{(2)} p_{U|\tau_i,\Theta}^{(3)} - \alpha\,\overline{q}^{(1)} p_{U|\tau_i,\Theta}^{(3)}(i-m_0)_+^{\gamma} + o(\alpha)$$

This shows that we roughly model a γ-power increase/decrease after the break point τ_{m_0}. Normalization $p_{U|\tau_i,\Theta}^{(1)} + p_{U|\tau_i,\Theta}^{(2)} + p_{U|\tau_i,\Theta}^{(3)} = 1$ implies

$$p_{U|\tau_i,\Theta}^{(3)} \;=\; \left(\overline{q}^{(1)} \exp\{\alpha(i-m_0)_+^{\gamma}\} + \overline{q}^{(2)} \exp\{-\alpha(i-m_0)_+^{\gamma}\} + 1\right)^{-1}$$

$$\;=\; \left(\overline{q}^{(1)} + \overline{q}^{(2)} + 1 + o(\alpha)\right)^{-1} \qquad \text{as } \alpha \to 0$$

This shows that $p_{U|\tau_i,\Theta}^{(3)}$ is almost constant for α small, i.e., the break point only marginally influences late reportings (because it mainly speeds up immediate reporting after claims occurrence, this will be seen below and corresponds to the red graphs in Figure 11).
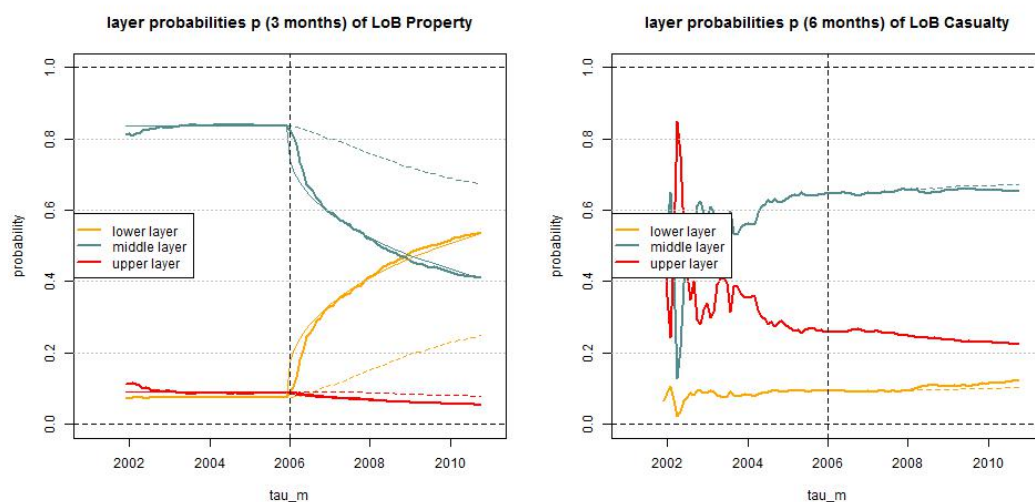

Figure 11. Estimated layer probabilities $p_{U|\tau_m,\Theta}^{(n)}$, $n = 1, 2, 3$, in the truncated/shifted log-normal model (lhs) LoB Property with $u^{(2)} = 3$ and (rhs) LoB Casualty with $u^{(2)} = 6$ months for $\tau_m \in \{31/12/2010, \dots, 31/10/2010\}$; the solid lines give the dynamic versions $\alpha > 0$ with $\gamma = 1/4$ and the dotted lines the static versions $\alpha = 0$.

With these choices we revisit log-likelihood (27) which is now given by, set $\bar{q} = (\bar{q}^{(1)}, \bar{q}^{(2)}, \alpha, \gamma)$,

$$\bar{\ell}_{\tau_m}(\Theta_2, \Theta_3, \bar{q}) \quad \propto \quad -\sum_{i=1}^{m} M_i \log \left(\pi_{\widehat{\Theta}_1^{(\tau_m)}}^{(i,1)}(\tau_m) q_{\tau_i}^{(1)} + \widetilde{\pi}_{\Theta_2}^{(i,2)}(\tau_m) q_{\tau_i}^{(2)} + \widetilde{\pi}_{\Theta_3}^{(i,3)}(\tau_m) \right) \tag{29}$$

$$+ \sum_{i=1}^{m} \sum_{n=1}^{2} M_{i,n} \log \left(q_{\tau_i}^{(n)} \right) + \sum_{i=1}^{m} \sum_{n=2}^{3} \sum_{\ell=1}^{M_{i,n}} \log \left(f_{U|\tau_i,\Theta_n}^{(n)}(U_\ell^{(i,n)}) \right)$$

If we use the explicit truncated/translated log-normal distributions introduced in Lemma 8 there are 9 parameters $(\bar{q}^{(1)}, \bar{q}^{(2)}, \alpha, \gamma, \mu_2, \sigma_2, \mu_3, \sigma_3, u^{(2)})$ involved in this optimization. Note that we use canonical translation $\nu = u^{(2)}$ for the upper layer distribution $F_{U|t,\Theta_3}^{(3)}$.

5.3.3. Preliminary Model Selection

We need to choose optimal parameters $(\bar{q}^{(1)}, \bar{q}^{(2)}, \alpha, \gamma, \mu_2, \sigma_2, \mu_3, \sigma_3, u^{(2)})$ for the model presented in Lemma 8. This can be achieved by applying MLE to log-likelihood (29). Since, eventually, we would like to do this for any time point $\tau_m \in \{31/12/2001, \ldots, 31/10/2010\}$, this would be computationally too expensive, and also not sufficiently robust (over time). For this reason we do a preliminary model selection based on the data as of 31/10/2010. In this preliminary model selection analysis we determine (a) the explicit distributions, (b) the upper layer threshold $u^{(2)}$ as well as (c) the parameter $\gamma > 0$ of power function (28). Based on these three choices we then calibrate the model for the time series $\tau_m \in \{31/12/2001, \ldots, 31/10/2010\}$. These three choices are done based on Tables 5–7.

Table 5. AIC and BIC as of 31/10/2010 (top) $\gamma = 1/4$ and $u^{(2)} = 6$ months and (bottom) LoB Property truncated/truncated dynamic ($\alpha > 0$) log-normal model with $\gamma = 1/4$ and LoB Casualty truncated/truncated static ($\alpha = 0$) log-normal model.

Method: Log-normal Distribution	AIC	BIC
LoB Property		
truncated/shifted (static $\alpha = 0$)	347'532	347'592
truncated/truncated (static $\alpha = 0$)	347'461	347'522
truncated/shifted (dynamic $\alpha > 0$)	339'270	339'348
truncated/truncated (dynamic $\alpha > 0$)	339'204	339'274
LoB Casualty		
truncated/shifted (static $\alpha = 0$)	87'979	88'028
truncated/truncated (static $\alpha = 0$)	87'941	87'990
truncated/shifted (dynamic $\alpha > 0$)	87'973	88'028
truncated/truncated (dynamic $\alpha > 0$)	87'934	87'990

Threshold	AIC	BIC
LoB Property		
$u^{(2)} = 3$ months	339'033	339'103
$u^{(2)} = 6$ months	339'204	339'274
$u^{(2)} = 9$ months	339'291	339'361
$u^{(2)} = 12$ months	339'376	339'445
LoB Casualty		
$u^{(2)} = 3$ months	87'908	87'957
$u^{(2)} = 6$ months	87'941	87'990
$u^{(2)} = 9$ months	87'963	88'013
$u^{(2)} = 12$ months	87'988	88'037

Table 6. AIC and BIC as of 31/10/2010 for LoB Property truncated/truncated dynamic log-normal model with $u^{(2)} = 6$ months.

Threshold	AIC	BIC
LoB Property		
$\gamma = 1/2$ months	339'594	339'663
$\gamma = 1/3$ months	339'267	339'337
$\gamma = 1/4$ months	339'204	339'274
$\gamma = 1/5$ months	339'222	339'292

Table 7. AIC and BIC as Table 5 (top) but with log-normal distributions replaced by gamma distributions.

Method: Gamma Distribution	AIC	BIC
LoB Property		
truncated/shifted (static $\alpha = 0$)	348'121	348'174
truncated/truncated (static $\alpha = 0$)	348'109	348'161
truncated/shifted (dynamic $\alpha > 0$)	339'935	340'005
truncated/truncated (dynamic $\alpha > 0$)	339'856	339'926
LoB Casualty		
truncated/shifted (static $\alpha = 0$)	88'000	88'042
truncated/truncated (static $\alpha = 0$)	88'887	88'929
truncated/shifted (dynamic $\alpha > 0$)	87'995	88'051
truncated/truncated (dynamic $\alpha > 0$)	88'888	88'943

We start by comparing the following models for the pair $(F^{(2)}, F^{(3)})$: (i) truncated/shifted log-normal (as in Lemma 8); and (ii) truncated/truncated log-normal. For these two models we consider the static version $\alpha = 0$ in (28) and the dynamic version $\alpha > 0$. We set $\gamma = 1/4$ and $u^{(2)} = 6$ months (this is further considered below) and then we calculate the MLE of $(\overline{q}^{(1)}, \overline{q}^{(2)}, \alpha = 0, \mu_2, \sigma_2, \mu_3, \sigma_3)$ (for the static version) and the MLE of $(\overline{q}^{(1)}, \overline{q}^{(2)}, \alpha, \mu_2, \sigma_2, \mu_3, \sigma_3)$ (for the dynamic version). Finally, we calculate Akaike's information criterion (AIC) and the Bayesian information criterion (BIC) for these model choices. The model with the smallest AIC and BIC, respectively, should be preferred. The results are presented in Table 5 (top). LoB Property: the dynamic ($\alpha > 0$) truncated/truncated log-normal model should be preferred, closely followed by the dynamic ($\alpha > 0$) truncated/shifted log-normal one. LoB Casualty: the truncated/truncated log-normal model should be preferred, the static and the dynamic versions are judged by BIC rather similarly. We have decided in favor of the static version because this reduces the number of parameters to be selected.

Next we analyze AIC and BIC of the optimal layer limit $u^{(2)}$ for the dynamic truncated/truncated log-normal model in LoB Property (for given $\gamma = 1/4$) and the static truncated/truncated log-normal model in LoB Casualty. The results are presented in Table 5 (bottom). We observe that we prefer for both LoBs a small threshold of $u^{(2)} = 3$ months.

Next we do the same analysis for the parameter γ in the dynamic ($\alpha > 0$) version of LoB Property, see Table 6 and we compare the log-normal model to the gamma model, see Table 7. From this preliminary analysis our conclusions are:

▷ LoB Property: We choose the dynamic ($\alpha > 0$) truncated/truncated log-normal model with $\gamma = 1/4$ and $u^{(2)} = 3$ months.

▷ LoB Casualty: We choose the static ($\alpha = 0$) truncated/truncated log-normal model with $u^{(2)} = 3$ months.

These are our preferred models as of $\tau_m = 31/10/2010$ and the remaining model parameters $(\overline{q}^{(1)}, \overline{q}^{(2)}, \alpha, \mu_2, \sigma_2, \mu_3, \sigma_3)$ are obtained by MLE from (29).

5.3.4. Model Calibration Over Time

In the previous subsection we have identified the optimal models as of $\tau_m = 31/10/2010$ using AIC and BIC. Observe that this is the optimal model selection with having maximal available data. This selection will be revised in this subsection because we study the models when incoming information increases over $\tau_m \in \{31/12/2001, \ldots, 31/10/2010\}$. In Figures 12 and 13 we present the MLEs of parameters $(\mu_2, \sigma_2, \mu_3, \sigma_3)$ using (29) for (lhs) the truncated/truncated log-normal models and (rhs) the truncated/shifted log-normal models. LoB Property considers the dynamic version $\alpha > 0$ with $\gamma = 1/4$ and LoB Casualty considers the static version $\alpha = 0$. From this analysis we see that the parameters behave much more robust over time in the truncated/shifted log-normal model, see Figures 12 (rhs) and 13 (rhs). For this reason we abandon our previous choice, and we select the truncated/shifted log-normal model for both LoBs. This is in slight contrast to the AIC and BIC analysis in the previous subsection, but the differences between the two models in Table 5 (lhs) are rather small which does not severely contradict the truncated/shifted model selection. In addition, for LoB Property we choose threshold $u^{(2)} = 3$ months (Figure 12 (top, rhs)) and for LoB Casualty we decide for the bigger threshold $u^{(2)} = 6$ months (Figure 13 (bottom, rhs)), also because estimation over time is more robust for this latter choice.
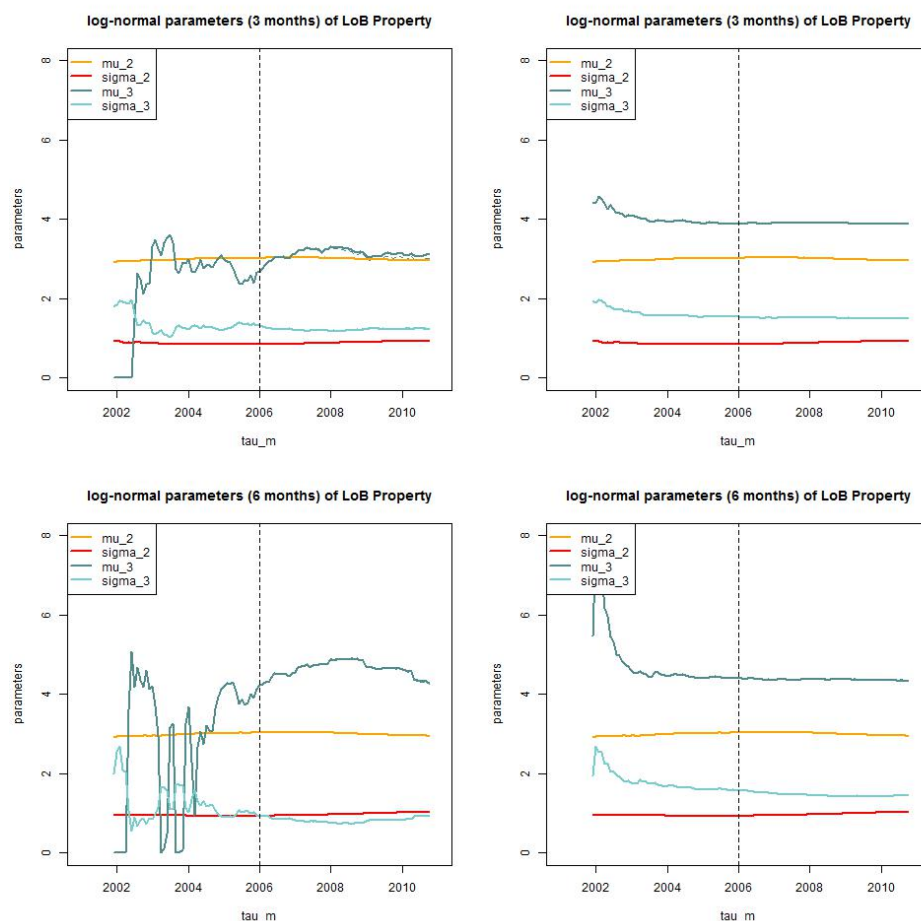

Figure 12. LoB Property: parameter estimates of $(\mu_2, \sigma_2, \mu_3, \sigma_3)$ for the dynamic model $\alpha > 0$ with $\gamma = 1/4$ for $\tau_m \in \{31/12/2010, \ldots, 31/10/2010\}$ (top, lhs) $u^{(2)} = 3$ months truncated/truncated and (top, rhs) and truncated/shifted; and (bottom, lhs) $u^{(2)} = 6$ months truncated/truncated and (bottom, rhs) and truncated/shifted.
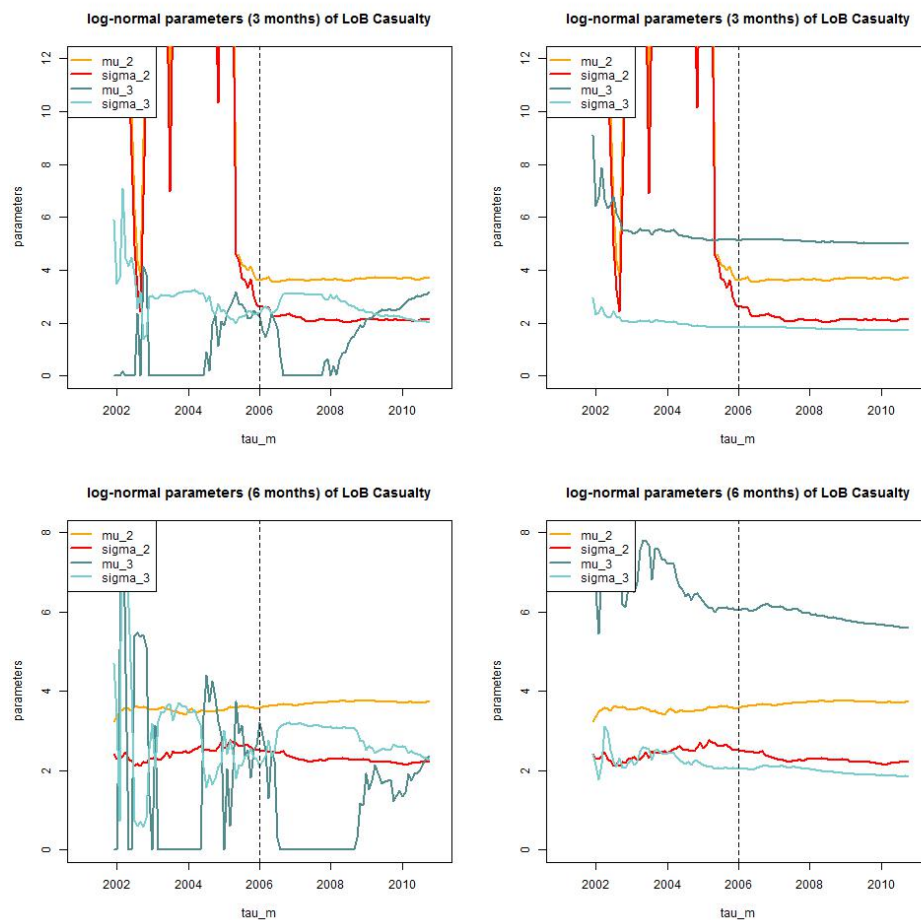
Figure 13. LoB Casualty: parameter estimates of $(\mu_2, \sigma_2, \mu_3, \sigma_3)$ for the static model $\alpha = 0$ for $\tau_m \in \{31/12/2010, \ldots, 31/10/2010\}$ (top, lhs) $u^{(2)} = 3$ months truncated/truncated and (top, rhs) and truncated/shifted; and (bottom, lhs) $u^{(2)} = 6$ months truncated/truncated (bottom, rhs) and truncated/shifted.

In Figure 11 we provide the estimated layer probabilities $p_{U|\tau_m,\Theta}^{(n)}$, $n = 1, 2, 3$, in the truncated/shifted log-normal model for LoB Property (with $u^{(2)} = 3$ months) and for LoB Casualty (with $u^{(2)} = 6$ months) for $\tau_m \in \{31/12/2010, \ldots, 31/10/2010\}$. The solid lines give the dynamic versions $\alpha > 0$ with $\gamma = 1/4$ and the dotted lines the static versions $\alpha = 0$. In LoB Property we observe stationarity of these layer probabilities up to break point $\tau_{m_0} = 1/1/2006$, and our modeling approach Equation (28) with $\gamma = 1/4$ seems to capture the non-stationarity after the break point rather well (note that the thin solid line shows an exact function $x^{1/4}$ after the break point by considering the map $i \mapsto p_{U|\tau_i,\Theta}^{(n)}$ at time $\tau_m = 31/10/2010$ and the dotted lines show the stationary case $\alpha = 0$). In this analysis we also see that LoB Casualty may be slightly non-stationary after 1/1/2008, see Figure 11 (rhs). This was not detected previously, but is also supported by the time series of the trend parameter α estimates given in Figure 7 (rhs). However, since the resulting trend parameter α is comparably small we remain with the static version for LoB Casualty.

Our conclusions are as follows. For LoB Property we choose the dynamic ($\alpha > 0$) truncated/shifted log-normal model with $\gamma = 1/4$ and $u^{(2)} = 3$ months; for LoB Casualty we choose the static ($\alpha = 0$) truncated/shifted log-normal model with $u^{(2)} = 6$ months. In Figure 14 we present the resulting calibration as of $\tau_m = 31/10/2010$. In the middle layer the fitted distribution looks convincing, see Figure 14 (top). In the upper layer the fitted distribution is more conservative than the observations, see Figure 14 (bottom), this is supported by the fact that the empirical distribution is

not sufficiently heavy-tailed because of missing information about late reportings (IBNYR claims in upper left triangles in Figure 2). This missing information has a bigger influence in casualty insurance because reporting delays are more heavy-tailed. Moreover, in both LoBs the shifted modeling approach is more conservative than the truncated one.
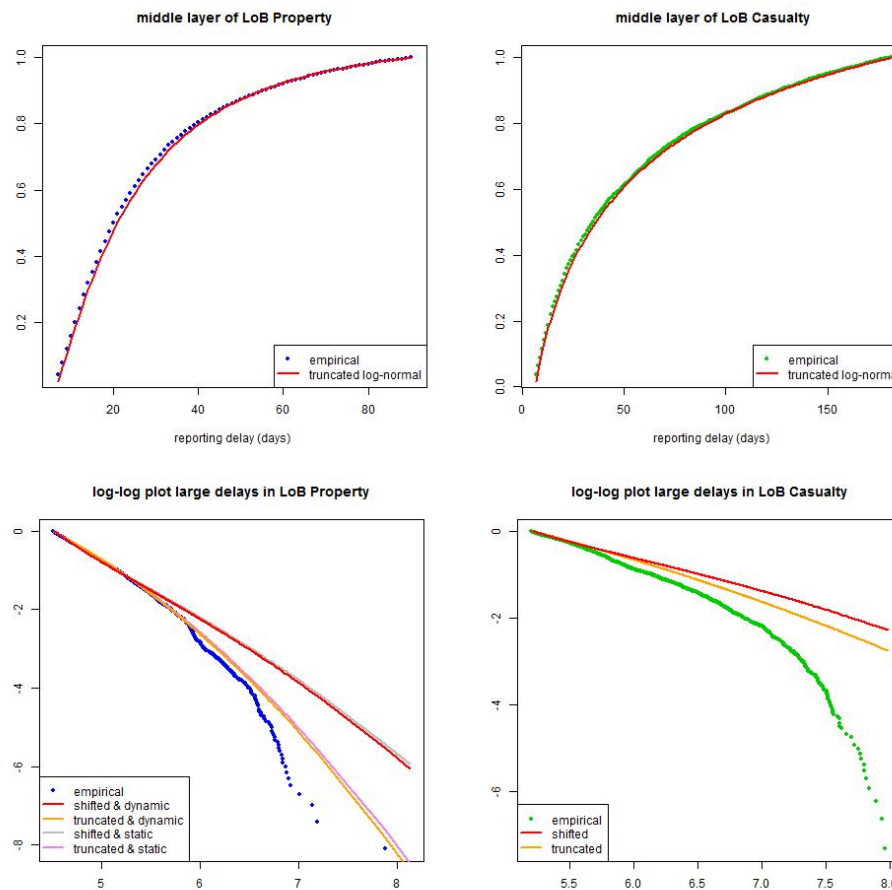

Figure 14. (top) calibration of truncated log-normal distribution in the middle layer $[u^{(1)}, u^{(2)})$ for (lhs) LoB Property with $u^{(2)} = 3$ months and (rhs) LoB Casualty with $u^{(2)} = 6$ months; (bottom) calibration of shifted and truncated log-normal distributions in the large layer $[u^{(2)}, \infty)$ for (lhs) LoB Property with $u^{(2)} = 3$ months (static and dynamic versions) and (rhs) LoB Casualty with $u^{(2)} = 6$ months (only static versions).

The model is now fully specified and we can estimate the number of IBNYR claims (late reportings). Using Equations (6), (25) and replacing all parameters Θ by their MLEs $\widehat{\Theta}^{(\tau_m)}$ at time τ_m we receive estimate for the total number of incurred claims in time interval $(\tau_{i-1}, \tau_i]$

$$\widehat{N}_i^{(\tau_m)} = w_i \widehat{\Lambda}_i^{(\tau_m)} \lambda_i^+ = \frac{M_i(\tau_m)}{\sum_{n=1}^3 \pi_{\widehat{\Theta}_n^{(\tau_m)}}^{(i,n)}(\tau_m) p_{U|\tau_i, \widehat{\Theta}^{(\tau_m)}}^{(n)}} \tag{30}$$

The number of estimated IBNYR claims is given by the difference $\widehat{\text{IBNYR}}_i^{(\tau_m)} = \widehat{N}_i^{(\tau_m)} - M_i(\tau_m)$. For illustration we choose three time points $\tau_m \in \{31/03/2004, 31/07/2007, 31/10/2010\}$, and we always use the relevant available information at those time points τ_m. The results are presented in Figure 15, the blue/green line gives the number of reported claims $M_i(\tau_m)$ and the red line the estimated number of incurred claims $\widehat{N}_i^{(\tau_m)}$ (the spread being the number of estimated IBNYR claims). We note that the spread is bigger for LoB Casualty than LoB Property (which is not surprising in view

of our previous analysis). For LoB Property in Figure 15 (top) we also compare the static (gray) to the dynamic (red) estimation. The static version seems inappropriate since it cannot sufficiently capture the non-stationarity and estimation is too conservative after the break point $\tau_{m_0} = 1/1/2006$.
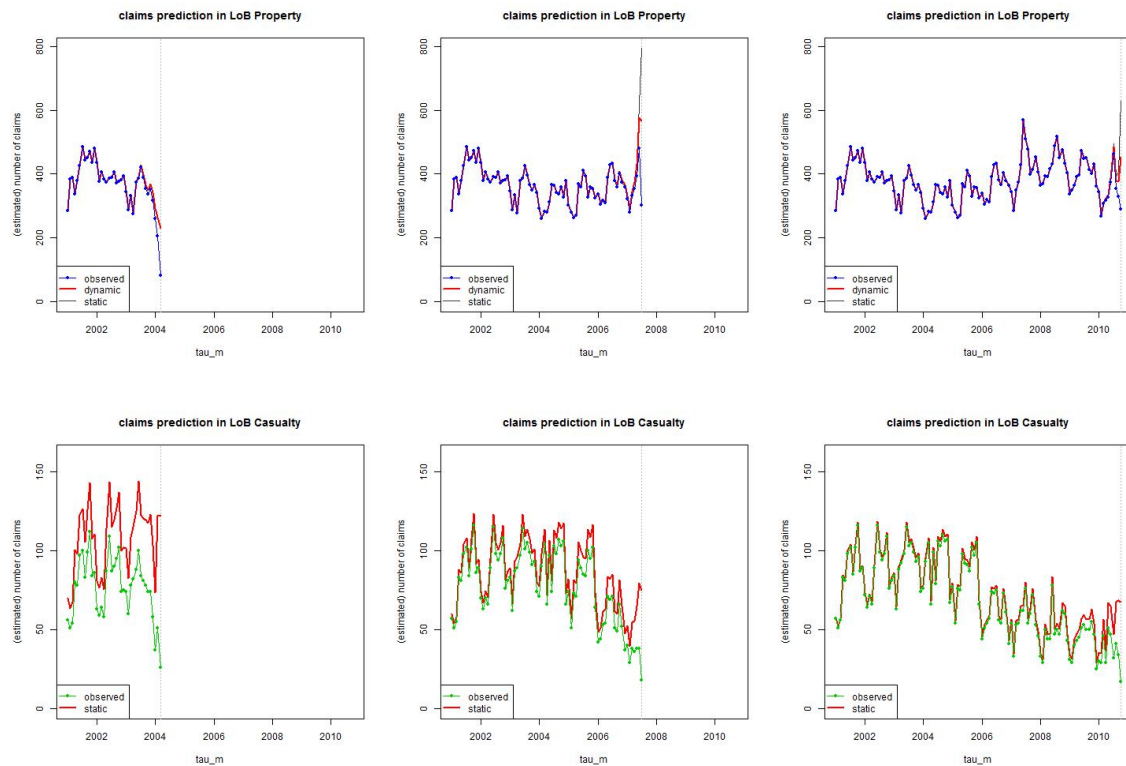

Figure 15. Estimation of the number of incurred claims $\widehat{N}_i^{(\tau_m)}$ for (top) LoB Property and (bottom) LoB Casualty at times $\tau_m \in \{31/03/2004, 31/07/2007, 31/10/2010\}$ using the truncated/shifted log-normal model with $u^{(2)} = 3$ months and $u^{(2)} = 6$ months, respectively; "observed" (blue/green) gives the number of reported claims $M_i(\tau_m)$ in each period $(\tau_{i-1}, \tau_i]$ at time τ_m, "dynamic/static" (red) gives the total number of estimated claims $\widehat{N}_i^{(\tau_m)}$ (the spread $\widehat{N}_i^{(\tau_m)} - M_i(\tau_m)$ giving the estimated number of IBNYR claims).

In the final section of this paper, we back-test our calibration and estimation, and compare it to the chain-ladder estimation.

6. Homogeneous (Poisson) Chain-Ladder Case and Back-Testing

In this section we compare our individual claims modeling calibration to the classical chain-ladder method on aggregate data. The chain-ladder method is probably the most popular method in claims reserving. Interestingly, the cross-classified chain-ladder estimation can be derived under Model Assumptions 2 and additional suitable homogeneity assumptions.

6.1. Cross-Classified Chain-Ladder Model

We choose an equidistant grid

$$0 = \tau_0 < \tau_1 < \tau_2 < \ldots \qquad \text{with } \Delta\tau_i = \tau_i - \tau_{i-1} = \tau_1 \text{ for all } i \in \mathbb{N}.$$

Denote by $M_{i,j}$ the number of claims with accident dates in $(\tau_{i-1}, \tau_i]$ and reporting dates in $(\tau_{i+j-1}, \tau_{i+j}]$, for $i \in \mathbb{N}$ and $j \in \mathbb{N}_0$. Under Model Assumptions 2 the random variables $M_{i,j}$ are independent and Poisson distributed with exposures

$$W_{i,j} = \int_{\tau_{i-1}}^{\tau_i} w(t)\Lambda(t) \int_{\tau_{i+j-1}-t}^{\tau_{i+j}-t} f_{U|t,\Theta}(u)du\,dt$$

We now make the following homogeneity assumptions for $t \in (\tau_{i-1}, \tau_i]$ and $i \in \mathbb{N}$

$$w(t)\Lambda(t) = w_{\tau_i}\Lambda_{\tau_i}, \qquad f_{U|t,\Theta} = f_{U|0,\Theta} = f_{U|\Theta} \tag{31}$$

i.e., we may drop time index t in $f_{U|t,\Theta}$ and $F_{U|t,\Theta}$, respectively. We define $W_i = \Delta\tau_i w_{\tau_i}\Lambda_{\tau_i}$. Assumptions (31) imply

$$W_{i,j} = W_i \frac{1}{\Delta\tau_i} \int_{\tau_{i-1}}^{\tau_i} \int_{\tau_{i+j-1}-t}^{\tau_{i+j}-t} f_{U|\Theta}(u)du\,dt = W_i \frac{1}{\Delta\tau_1} \int_0^{\tau_1} \int_{\tau_j-t}^{\tau_{j+1}-t} f_{U|\Theta}(u)du\,dt$$

If we now define reporting pattern $(\gamma_j)_{j\geq 0}$ by

$$\gamma_j = \gamma_j(\Theta) = \frac{1}{\Delta\tau_1} \int_0^{\tau_1} \int_{\tau_j-t}^{\tau_{j+1}-t} f_{U|\Theta}(u)du\,dt = \frac{1}{\Delta\tau_1} \int_0^{\tau_1} F_{U|\Theta}(\tau_{j+1} - t)) - F_{U|\Theta}(\tau_j - t))dt$$

we see that under (31) the random variables $M_{i,j}$ are independent and Poisson distributed with

$$M_{i,j} \sim \text{Poisson}\left(W_i\gamma_j\right) \tag{32}$$

Moreover, we have normalization $\sum_{j\geq 0} \gamma_j = 1$. This provides the cross-classified Poisson version of the chain-ladder model. Under the additional assumption that $\sum_{j=0}^{J} \gamma_j = 1$ for a finite J the MLE exactly provides the chain-ladder estimator. This result goes back to Hachemeister-Stanard [22], Kremer [23] and Mack [24], and for more details and the calculation of the chain-ladder estimator $\widehat{M}_{i,j}^{CL(\tau_m)}$ of $M_{i,j}$ with $i + j > m$ at time $\tau_m \geq \tau_J$ we refer to Theorem 3.4 in Wüthrich-Merz [25]. Using these chain-ladder estimators we get estimate

$$\widehat{N}_i^{CL(\tau_m)} = \sum_{j=0}^{m-i} M_{i,j} + \sum_{j\geq m-i+1} \widehat{M}_{i,j}^{CL(\tau_m)} = M_i(\tau_m) + \sum_{j\geq m-i+1} \widehat{M}_{i,j}^{CL(\tau_m)} \tag{33}$$

for the estimated number of claims in period $(\tau_{i-1}, \tau_i]$ with $\tau_i \leq \tau_m$. This chain-ladder estimate $\widehat{N}_i^{CL(\tau_m)}$ is compared to the estimate $\widehat{N}_i^{(\tau_m)}$ provided in (30).

6.2. Back-Testing

In this section we back-test the estimations obtained by calibration (21) and (29) and compare it to the homogeneous chain-ladder case (33). We therefore calculate for each exposure period $(\tau_{i-1}, \tau_i]$ with $\tau_i \leq \tau_m \in \{31/12/2001, \ldots, 31/10/2010\}$ the estimates $\widehat{N}_i^{(\tau_m)}$ and $\widehat{N}_i^{CL(\tau_m)}$. These estimates are compared to the latest estimates $\widehat{N}_i^{(\tau_e)}$ and $\widehat{N}_i^{CL(\tau_e)}$ at time $\tau_e = 31/10/2010$. In particular, we back-test the estimates with time lags $j = 0, 1$ against the latest available estimation. We therefore define the ratios

$$\chi_{i,0} = \frac{\widehat{N}_i^{(\tau_i)}}{\widehat{N}_i^{(\tau_e)}} \quad \text{and} \quad \chi_{i,1} = \frac{\widehat{N}_i^{(\tau_{i+1})}}{\widehat{N}_i^{(\tau_e)}}$$

$$\chi_{i,0}^{CL} = \frac{\widehat{N}_i^{CL(\tau_i)}}{\widehat{N}_i^{CL(\tau_e)}} \quad \text{and} \quad \chi_{i,1}^{CL} = \frac{\widehat{N}_i^{CL(\tau_{i+1})}}{\widehat{N}_i^{CL(\tau_e)}}$$

The first ratio $\chi_{i,0}$ compares the estimation of the number of claims in period $(\tau_{i-1}, \tau_i]$ based on the information available at time τ_i to the latest available estimation at time $\tau_e = 31/10/2010$, and the second variable $\chi_{i,1}$ considers the same ratio but with the estimation based on the information at time τ_{i+1} (i.e., after one period of development). Moreover, we calculate the relative process uncertainties for $j = 0, 1$ defined by

$$\varsigma_{i,j} = \frac{\left(\widehat{N}_i^{(\tau_{i+j})} - M_i(\tau_{i+j}) \right)^{1/2}}{\widehat{N}_i^{(\tau_e)}}$$

this is the relative standard deviation of the number of IBNYR claims at time τ_{i+j} for a Poisson random variable, normalized with $\widehat{N}_i^{(\tau_e)}$ to make it comparable to $\chi_{i,j}$. We present the results in Figure 16:

- Figure 16 (top), LoB Property: We see that the chain-ladder estimate $\widehat{N}_i^{CL(\tau_{i+j})}$, $j = 0, 1$, clearly over-estimates the number of claims after the break point $\tau_{m_0} = 1/1/2006$, whereas our non-stationary approach (28) can capture this change rather well and estimations $\chi_{i,j}$ are centered around 1. Remarkable is that after 5 years of observations, the volatility of the estimation can almost completely be explained by process uncertainty (we plot confidence bounds of $2\varsigma_{i,j}$), which means that model uncertainty is comparably low. We also see that the faster reporting behavior after break point τ_{m_0} has substantially decreased the uncertainty and the volatility in the number of IBNYR claims. Before the break point, uncertainty for $j = 0$ is comparably high, this can partly be explained by the fact that claims history is too short for model calibration.

- Figure 16 (bottom), LoB Casualty: After roughly 5 years of claims experience the estimation in the individual claims model and the chain-ladder model are very similar. This indicates that in a stationary situation, the performance of the (simpler) chain-ladder model is sufficient. However, this latter statement needs to be revised because non-stationarity can be detected more easily in the individual claims history modeling approach, in particular, the individual claims model reacts more sensitively to structural changes. The confidence bounds have a reasonable size because the back-testing exercise does not violate them too often (i.e., the observations are mostly within the confidence bounds), however, in practice they should be chosen slightly bigger because they do not consider model uncertainty.

We close this section with a brief discussion of two related modeling approaches. Our approach can be viewed as a refined model of the one used in Antonio-Plat [17]. The first refinement we use is that we consider the weekly periodic pattern $(\lambda_k)_{k=1,\dots,7}$ which was supported by the statistical analysis given in Figure 5. Neglecting this weekly periodic pattern would lead to less smooth small layer probabilities, in particular, in LoB Casualty, see Figure 4. The second refinement is that we choose a reporting delay distribution that depends on the weekday of the claims occurrence. This is especially important for the small reporting delay layer because weekday configuration essentially influences the short reporting delays. In our analysis, this then leads to a mixture model with three different layers. Antonio-Plat [17] use a mixture of a Weibull distribution and 9 degenerate components which fits their purposes well. Our approach raises the issue of over-parametrization which we analyze graphically, i.e., we observe rather stable parameter estimates after 4 years of observations, see for instance Figures 12 (rhs) and 13 (rhs).

The issue of over-parametrization is of essential relevance if we would like to do prediction for future exposure periods. That is, in our analysis we have mainly concentrated on making statistical inference of the instantaneous claims frequency $\Lambda(t)$ of past exposures $t \leq \tau_m$ at a given time point $\tau \geq \tau_m$. Going forward, we may also want to model and predict the claims occurrence and the claims reporting processes, respectively, of future exposures. An over-parametrized model will have a low predictive power because it involves too much model uncertainty, therefore we should choose as few parameters as necessary. Moreover, for predictive modeling it will also be necessary to model stochastically the instantaneous claims frequency process Λ, i.e., our statistical inference method is

not sufficient for predicting claims of future exposures, but it will help to calibrate a stochastic model for Λ. A particularly interesting model is the marked Cox model proposed in Badescu et al. [18,19]. Similarly to Antonio-Plat [17], Badescu et al. [18,19] consider the piece-wise homogeneous case (7), but with $(\Lambda_i)_i$ being a state-dependent process which is driven by a time-homogeneous hidden Markov process (in the sense of a state-space model). We believe that this is a promising modeling approach for claims occurrence and reporting prediction, if merged with our weekday dependent features (both for claims occurrence and claims reporting delays).
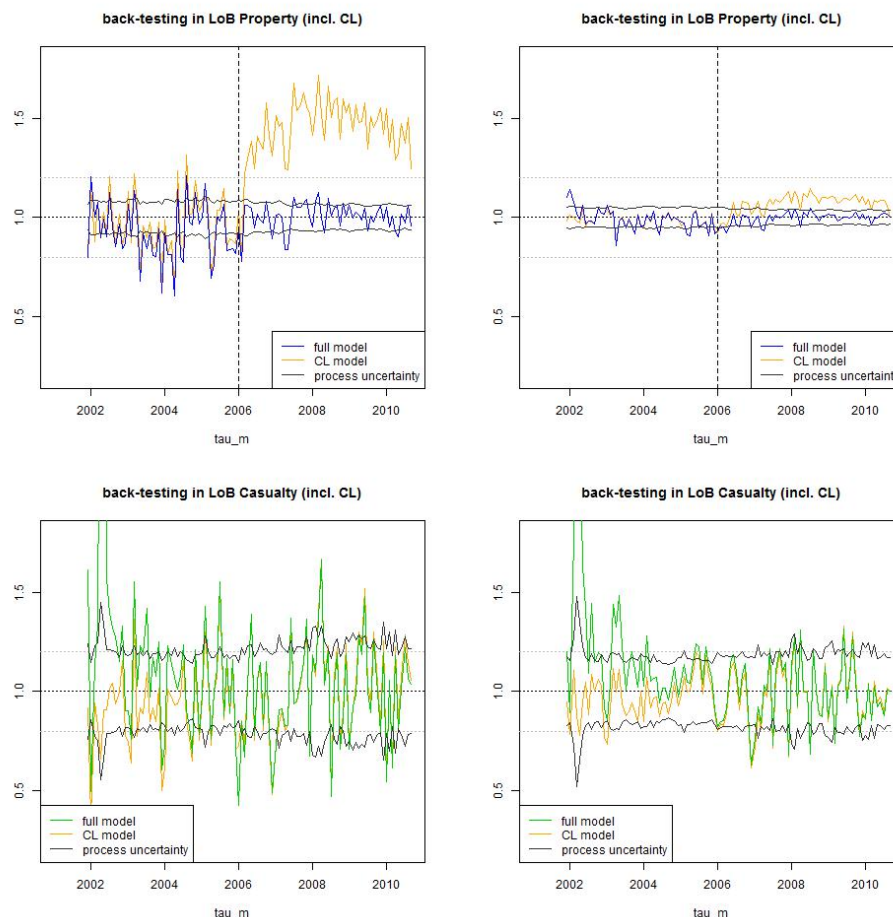

Figure 16. Back-test LoB Property (top) and LoB Casualty (bottom): we compare the (non-stationary) estimate $\chi_{i,j}$ (blue/green) to the (stationary) chain-ladder estimate $\chi_{i,j}^{CL}$ (orange) for (lhs) time lag $j = 0$ and (rhs) time lag $j = 1$. The black line shows the process uncertainty confidence bounds of 2 (relative) standard deviations $\varsigma_{i,j}$.

6.3. Conclusions

We have provided an explicit calibration of a reporting delay model to individual claims data. For the two LoBs considered it takes about 5 years of observations to have sufficient information to calibrate the model (this is true for the stationary case and needs to be analyzed in more detail in a non-stationary situation). As long as the claims reporting process is stationary the individual claims reporting model and the aggregate chain-ladder model provide very similar estimates for the number of IBNYR claims, but as soon as we have non-stationarity the chain-ladder model fails to provide reliable estimates and one should use individual claims modeling. Moreover, our individual claims modeling approach is able to detect non-stationarity more quickly than the aggregate chain-ladder method. Going forward, there are two different directions that need to be considered. First, for the

evaluation of parameter uncertainty one can embed our individual claims reporting model into a Bayesian framework (this is similar to the Cox model proposed in Badescu et al. [18,19]) or one can use bootstrap methods. Secondly, the far more difficult problem is the modeling of the cost evolution and the claims cash flow process. We believe that this is still an open problem and it is the next building block of individual claims reserving that should be studied on real data examples.

Author Contributions: Both authors have contributed equally to this paper.

Conflicts of Interest: The author declares no conflict of interest.

Appendix A. Proofs

Proof of Lemma 3. We maximize likelihood function (9). The derivatives of its logarithm provide for $i = 1, \ldots, m$ the requirements (denoted by $\overset{!}{=}$)

$$\frac{\partial \log \mathcal{L}_{(T_\ell, S_\ell)_{\ell=1,\ldots,M}}(\Lambda, \Theta)}{\partial \Lambda_i} = -\pi_\Theta^{(i)}(\tau) w_i \lambda_i^+ + \frac{M_i}{\Lambda_i} \overset{!}{=} 0$$

and for the derivative w.r.t. the reporting delay parameter Θ we obtain

$$\frac{\partial \log \mathcal{L}_{(T_\ell, S_\ell)_{\ell=1,\ldots,M}}(\Lambda, \Theta)}{\partial \Theta} = \sum_{i=1}^{m} \left[-w_i \Lambda_i \lambda_i^+ \frac{\partial}{\partial \Theta} \pi_\Theta^{(i)}(\tau) + \sum_{\ell=1}^{M_i} \frac{\frac{\partial}{\partial \Theta} f_{U|T_\ell^{(i)},\Theta}(S_\ell^{(i)} - T_\ell^{(i)})}{f_{U|T_\ell^{(i)},\Theta}(U_\ell^{(i)} - S_\ell^{(i)})} \right] \overset{!}{=} 0$$

The first requirements provide identity

$$\Lambda_i = \frac{M_i}{\pi_\Theta^{(i)}(\tau) w_i \lambda_i^+}$$

Plugging this into the second requirement provides identity

$$\begin{aligned}
0 &= \sum_{i=1}^{m} \left[-M_i \frac{\partial}{\partial \Theta} \log(\pi_\Theta^{(i)}(\tau)) + \sum_{\ell=1}^{M_i} \frac{\partial}{\partial \Theta} \log(f_{U|T_\ell^{(i)},\Theta}(S_\ell^{(i)} - T_\ell^{(i)})) \right] \\
&= \sum_{i=1}^{m} \sum_{\ell=1}^{M_i} \frac{\partial}{\partial \Theta} \left[\log \left(f_{U|T_\ell^{(i)},\Theta}(S_\ell^{(i)} - T_\ell^{(i)}) \right) - \log(\pi_\Theta^{(i)}(\tau)) \right] \\
&= \sum_{i=1}^{m} \sum_{\ell=1}^{M_i} \frac{\partial}{\partial \Theta} \log \left(\frac{\lambda_{T_\ell^{(i)}} f_{U|T_\ell^{(i)},\Theta}(S_\ell^{(i)} - T_\ell^{(i)})}{\lambda_i^+ \pi_\Theta^{(i)}(\tau)} \right)
\end{aligned}$$

where in the last identity we have added constants in Θ which vanish under the derivative (note that these constants were added to indicate that we obtain the densities (10)). □

Proof of Lemma 4. We maximize likelihood function (9) for $i \leq m^*$ on weekly time grid $\Delta \tau_i = 7$ under side constraint $\sum_{k=1}^{7} \lambda_k = 7$ and under assumption (11) which implies $\pi_\Theta^{(i)}(\tau) = 1$ and $\frac{\partial}{\partial \lambda_k} \pi_\Theta^{(i)}(\tau) = 0$ for $i \leq m^*$. The corresponding Lagrangian is given by

$$\sum_{i=1}^{m^*} \log \mathcal{L}_{(T_\ell^{(i)}, S_\ell^{(i)})_{\ell=1,\ldots,M_i}}^{(i)}(\Lambda, \Theta) - \chi \left(\sum_{k=1}^{7} \lambda_k - 7 \right)$$

Under the above assumptions we calculate the derivative w.r.t. λ_k, $1 \leq k \leq 7$, of the Lagrangian

$$\begin{aligned}
\frac{\partial}{\partial \lambda_k} \left(\sum_{i=1}^{m^*} \log \mathcal{L}_{(T_\ell^{(i)}, S_\ell^{(i)})_{\ell=1,\ldots,M_i}}^{(i)}(\Lambda, \Theta) - \chi \left(\sum_{k=1}^{7} \lambda_k - 7 \right) \right) &= \sum_{i=1}^{m^*} -w_i \Lambda_i + \sum_{\ell=1}^{M_i} \frac{1}{\lambda_k} \mathbb{1}_{\{\lceil \tilde{T}_\ell^{(i)} \rceil = k\}} - \chi \\
&\overset{!}{=} 0
\end{aligned}$$

and the derivative w.r.t. Λ_i are given by

$$\frac{\partial}{\partial \Lambda_i}\left(\sum_{i=1}^{m^*}\log \mathcal{L}^{(i)}_{(T_\ell^{(i)},S_\ell^{(i)})_{\ell=1,\dots,M_i}}(\Lambda,\Theta) - \chi\left(\sum_{k=1}^{7}\lambda_k - 7\right)\right) = -w_i\lambda_i^+ + \frac{M_i}{\Lambda_i} = -7w_i + \frac{M_i}{\Lambda_i} \stackrel{!}{=} 0$$

where we have used $\lambda_i^+ = 7$ on the weekly time grid $\Delta\tau_i = 7$. The latter implies $\Lambda_i = M_i/(7w_i)$ and plugging this into the former requirement provides

$$\sum_{i=1}^{m^*} -M_i/7 + \sum_{\ell=1}^{M_i}\frac{1}{\lambda_k}1_{\{\lceil \widetilde{T}_\ell^{(i)}\rceil = k\}} - \chi \stackrel{!}{=} 0$$

This implies that

$$\lambda_k = \frac{\sum_{i=1}^{m^*}\sum_{\ell=1}^{M_i}1_{\{\lceil \widetilde{T}_\ell^{(i)}\rceil = k\}}}{\chi + \sum_{i=1}^{m^*}M_i/7}$$

Lagrange multiplier χ is found from normalization $\sum_{k=1}^{7}\lambda_k = 7$ which provides the claim. $\quad\square$

References

1. Jewell, W.S. Predicting IBNYR events and delays I. Continuous time. *ASTIN Bull.* **1989**, *19*, 25–55.
2. Jewell, W.S. Predicting IBNYR events and delays II. Discrete time. *ASTIN Bull.* **1990**, *20*, 93–111.
3. Bühlmann, H.; Schnieper, R.; Straub, E. Claims reserves in casualty insurance based on a probabilistic model. *Bull. Swiss Assoc. Actuaries* **1980**, *1980*, 21–45.
4. Arjas, E. The claims reserving problem in non-life insurance: Some structural ideas. *ASTIN Bull.* **1989**, *19*, 139–152.
5. Norberg, R. Prediction of outstanding liabilities in non-life insurance. *ASTIN Bull.* **1993**, *23*, 95–115.
6. Norberg, R. Prediction of outstanding liabilities II. Model variations and extensions. *ASTIN Bull.* **1999**, *29*, 5–25.
7. Haastrup, S.; Arjas, E. Claims reserving in continuous time; a nonparametric Bayesian approach. *ASTIN Bull.* **1996**, *26*, 139–164.
8. Taylor, G. *The Statistical Distribution of Incurred Losses and Its Evolution Over Time I: Non-Parametric Models*; Casualty Actuarial Society: Arlington, VA, USA, 1999; working paper.
9. Taylor, G. *The Statistical Distribution of Incurred Losses and Its Evolution Over Time II: Parametric Models*; Casualty Actuarial Society: Arlington, VA, USA, 1999; working paper.
10. Herbst, T. An application of randomly truncated data models in reserving IBNR claims. *Insur. Math. Econ.* **1999**, *25*, 123–131.
11. Larsen, C.R. An individual claims reserving model. *ASTIN Bull.* **2007**, *37*, 113–132.
12. Taylor, G.; McGuire, G.; Sullivan, J. Individual claim loss reserving conditioned by case estimates. *Ann. Actuar. Sci.* **2008**, *3*, 215–256.
13. Jessen, A.H.; Mikosch, T.; Samorodnitsky, G. Prediction of outstanding payments in a Poisson cluster model. *Scand. Actuar. J.* **2011**, *2011*, 214–237.
14. Rosenlund, S. Bootstrapping individual claims histories. *ASTIN Bull.* **2012**, *42*, 291–324.
15. Pigeon, M.; Antonio, K.; Denuit, M. Individual loss reserving with the multivariate skew normal framework. *ASTIN Bull.* **2013**, *43*, 399–428.
16. Agbeko, T.; Hiabu, M.; Martínez-Miranda, M.D.; Nielsen, J.P.; Verrall, R. Validating the double chain ladder stochastic claims reserving model. *Variance* **2014**, *8*, 138–160.
17. Antonio, K., Plat, R. Micro-level stochastic loss reserving for general insurance. *Scand. Actuar. J.* **2014**, *2014*, 649–669.
18. Badescu, A.L.; Lin, X.S.; Tang, D. A marked Cox model for the number of IBNR claims: Theory. *Insur. Math. Econ.* **2016**, *69*, 29–37.
19. Badescu, A.L.; Lin, X.S.; Tang, D. *A Marked Cox Model for the Number of IBNR Claims: Estimation and Application*, Version 14 March 2016; SSRN: Rochester, NY, USA, 2016.

20. Hinkley, D.V. On the ratio of two correlated normal random variables. *Biometrika* **1969**, *56*, 635–639.

21. Wüthrich, M.V. *Non-Life Insurance: Mathematics & Statistics*, Version 15 April 2016; SSRN: Rochester, NY, USA, 2016; doi:10.2139/ssrn.2319328.

22. Hachemeister, C.A.; Stanard, J.N. IBNR claims count estimation with static lag functions. *ASTIN Colloq.* 1975.

23. Kremer, E. *Einführung in die Versicherungsmathematik*; Vandenhoek & Ruprecht: Göttingen, Germany, 1985.

24. Mack, T. A simple parametric model for rating automobile insurance or estimating IBNR claims reserves. *ASTIN Bulletin* **1991**, *21*, 93–109.

25. Wüthrich, M.V.; Merz, M. *Stochastic Claims Reserving Manual: Advances in Dynamic Modeling*, Version 21 August 2015; Swiss Finance Institute Research Paper No. 15-34; SSRN: Rochester, NY, USA, 2015; doi:10.2139/ssrn.2649057.