Supplementary material for "Modelling the number of hidden events subject to an observation delay"

Jonas Crevecoeur[1,3], Katrien Antonio[1,2,3,4] and Roel Verbelen[1,3,4]

[1]Faculty of Economics and Business, KU Leuven, Belgium.
[2]Faculty of Economics and Business, University of Amsterdam, The Netherlands.
[3]LRisk, Leuven Research Center on Insurance and Financial Risk Analysis, KU Leuven, Belgium.
[4]LStat, Leuven Statistics Research Center, KU Leuven, Belgium.

February 18, 2019

A Maximum likelihood estimation of observation exposure parameters

We model a parameter vector $\boldsymbol{\gamma}$ which structures the observation exposures.

$$\begin{aligned}\ell(\boldsymbol{\gamma};\boldsymbol{\chi}) &= \sum_{t=1}^{\tau}\sum_{s=t}^{\tau} N_{t,s}\cdot\log(p_{t,s}) - \sum_{t=1}^{\tau} N_t^{\mathrm{R}}(\tau)\cdot\log(p_t^{\mathrm{R}}(\tau)) \\ &= \sum_{t=1}^{\tau}\sum_{s=t}^{\tau} N_{t,s}\cdot\log\left(F_{\tilde{U}}\left(\varphi_t(s-t+1)\right) - F_{\tilde{U}}\left(\varphi_t(s-t)\right)\right) \\ &\quad - \sum_{t=1}^{\tau} N_t^{\mathrm{R}}(\tau)\cdot\log\left(F_{\tilde{U}}\left(\varphi_t(\tau-t+1)\right)\right),\end{aligned} \tag{11}$$

where

$$\varphi_t(d) = \sum_{v=t}^{t+d-1}\exp(\boldsymbol{x}_{t,v}^{'}\boldsymbol{\gamma}).$$

No analytical solution exists for the optimal parameters $\boldsymbol{\gamma}$ and numerical optimization is required. We use the Newton-Raphson algorithm to maximize the likelihood (11). The Newton-Raphson algorithm updates the parameter estimates iteratively as follows

$$\hat{\boldsymbol{\gamma}}^{(k+1)} = \hat{\boldsymbol{\gamma}}^{(k)} - \boldsymbol{H}^{-1}(\hat{\boldsymbol{\gamma}}^{(k)})\cdot\boldsymbol{S}(\hat{\boldsymbol{\gamma}}^{(k)}). \tag{12}$$

In this formula $\boldsymbol{S}$ denotes the score vector and $\boldsymbol{H}$ is the Hessian of the loglikelihood in (11), i.e. the vector of first order and the matrix of second order partial derivatives respectively. Below we derive the expression for the first and second order derivatives of the loglikelihood when $F_{\tilde{U}}$ is a known twice continuously differentiable distribution function. The components of the score vector $\boldsymbol{S}$ are

$$\begin{aligned}\frac{\partial \ell(\boldsymbol{\gamma}, \boldsymbol{\xi}; \boldsymbol{\chi})}{\partial \gamma_i} =& \sum_{t=1}^{\tau}\sum_{s=t}^{\tau} \frac{N_{t,s}}{p_{t,s}} \cdot \left[f_{\tilde{U}}\left(\varphi_t(s-t+1)\right) \cdot \frac{\partial \varphi_t}{\partial \gamma_i}(s-t+1) - f_{\tilde{U}}\left(\varphi_t(s-t)\right) \cdot \frac{\partial \varphi_t}{\partial \gamma_i}(s-t)\right] \\ & - \sum_{t=1}^{\tau} \frac{N_t^{\mathrm{R}}(\tau)}{p_t^{\mathrm{R}}(\tau)} \cdot f_{\tilde{U}}\left(\varphi_t(\tau-t+1)\right) \cdot \frac{\partial \varphi_t}{\partial \gamma_i}(\tau-t+1),\end{aligned}$$

where $f_{\tilde{U}}(\,\cdot\,)$ denotes the density function of $F_{\tilde{U}}(\,\cdot\,)$ and

$$\begin{aligned} p_{t,s} &= F_{\tilde{U}}\left(\varphi_t(s-t+1)\right) - F_{\tilde{U}}\left(\varphi_t(s-t)\right) \\ p_{t,s}^{\mathrm{R}}(\tau) &= F_{\tilde{U}}\left(\varphi_t(\tau-t+1)\right).\end{aligned}$$

The derivatives of the time change operator φ_t with respect to $\boldsymbol{\gamma}$ are

$$\frac{\partial}{\partial \gamma_i}\varphi_t(s-t+1) = \sum_{v=t}^{s} x_{t,v,i} \cdot \alpha_{t,v}$$

where $x_{t,s,i}$ is the covariate value of the i-th parameter for reporting on date s for a claim that occurred on date t. The Hessian $\boldsymbol{H}$ is given by

$$\begin{aligned}&\frac{\partial \ell(\boldsymbol{\gamma}; \boldsymbol{\chi})}{\partial \gamma_i \partial \gamma_j} = \\ &\quad \sum_{t=1}^{\tau}\sum_{s=t}^{\tau} \frac{N_{t,s}}{p_{t,s}} \cdot \Bigg[f_{\tilde{U}}^{'}\left(\varphi_t(s-t+1)\right) \cdot \frac{\partial \varphi_t}{\partial \gamma_i}(s-t+1) \cdot \frac{\partial \varphi_t}{\partial \gamma_j}(s-t+1) \\ &\qquad - f_{\tilde{U}}^{'}\left(\varphi_t(s-t)\right) \cdot \frac{\partial \varphi_t}{\partial \gamma_i}(s-t) \cdot \frac{\partial \varphi_t}{\partial \gamma_j}(s-t) \\ &\qquad + f_{\tilde{U}}\left(\varphi_t(s-t+1)\right) \cdot \frac{\partial \varphi_t}{\partial \gamma_i \partial \gamma_j}(s-t+1) - f_{\tilde{U}}\left(\varphi_t(s-t)\right) \cdot \frac{\partial \varphi_t}{\partial \gamma_i \partial \gamma_j}(s-t)\Bigg] \\ &\quad - \sum_{t=1}^{\tau}\sum_{s=t}^{\tau} \frac{N_{t,s}}{p_{t,s}^2} \cdot \Bigg[f_{\tilde{U}}\left(\varphi_t(s-t+1)\right)^2 \cdot \frac{\partial \varphi_t}{\partial \gamma_i}(s-t+1) \cdot \frac{\partial \varphi_t}{\partial \gamma_j}(s-t+1) \\ &\qquad + f_{\tilde{U}}\left(\varphi_t(s-t)\right)^2 \cdot \frac{\partial \varphi_t}{\partial \gamma_i}(s-t) \cdot \frac{\partial \varphi_t}{\partial \gamma_j}(s-t) \\ &\qquad - f_{\tilde{U}}\left(\varphi_t(s-t+1)\right) \cdot f_{\tilde{U}}\left(\varphi_t(s-t)\right) \cdot \frac{\partial \varphi_t}{\partial \gamma_i}(s-t+1) \cdot \frac{\partial \varphi_t}{\partial \gamma_j}(s-t) \\ &\qquad - f_{\tilde{U}}\left(\varphi_t(s-t+1)\right) \cdot f_{\tilde{U}}\left(\varphi_t(s-t)\right) \cdot \frac{\partial \varphi_t}{\partial \gamma_i}(s-t) \cdot \frac{\partial \varphi_t}{\partial \gamma_j}(s-t+1)\Bigg] \\ &\quad - \sum_{t=1}^{\tau} \frac{N_t^{\mathrm{R}}(\tau)}{p_t^{R}(\tau)} \cdot \Bigg[f_{\tilde{U}}^{'}\left(\varphi_t(\tau-t+1)\right) \cdot \frac{\partial \varphi_t}{\partial \gamma_i}(\tau-t+1) \cdot \frac{\partial \varphi_t}{\partial \gamma_j}(\tau-t+1)\end{aligned}$$

$$+ f_{\tilde{U}}(\varphi_t(\tau - t + 1)) \cdot \frac{\partial \varphi_t}{\partial \gamma_i \partial \gamma_j}(\tau - t + 1)\Bigg]$$

$$+ \sum_{t=1}^{\tau} \frac{N_t^{\mathrm{R}}(\tau)}{p_t^{\mathrm{R}}(\tau)^2} \cdot f_{\tilde{U}}\left(\varphi_t(\tau - t + 1)\right)^2 \cdot \frac{\partial \varphi_t}{\partial \gamma_i}(\tau - t + 1) \cdot \frac{\partial \varphi_t}{\partial \gamma_j}(\tau - t + 1),$$

where the second order derivatives of φ_t with respect to $\boldsymbol{\gamma}$ are

$$\frac{\partial}{\partial \gamma_i \partial \gamma_j} \varphi_t(s - t + 1) = \sum_{v=t}^{s} x_{t,v,i} \cdot x_{t,v,j} \cdot \alpha_{t,v}$$

The Newton-Raphson algorithm in (12) models the observation exposure parameters $\boldsymbol{\gamma}$. Together with the observation parameters, the simulation study of Section 3.5 estimates the variance parameter σ in the lognormal time-changed distribution. The Newton-Raphson algorithm in (12) can easily be extended to this case, where the distribution function of $F_{\tilde{U}}$ depends on parameters.

B Simulation procedure

We outline the algorithm that was used to generate data sets from the four scenarios specified in Section 3.5.1. This algorithm combines a model for the occurrence of events with a model for the observation delay as described in Section 2. We divide the algorithm in three steps.

Step 1. Occurrence We first generate the number of occurred events. The number of daily events follows a Poisson distribution

$$N_t \sim \text{Poisson}(\lambda_t),$$

where the intensity λ_t is obtained from the occurrence process specification for the scenarios in Section 3.5.

Step 2. Observation We now simulate the observation date for each occurred event. Combining equation (6) and (7), we can write the probability that an event from date t is observed on date s as

$$p_{t,s} = P\left(\tilde{U} \in \left[\sum_{v=t}^{s-1} \alpha_{t,v}, \sum_{v=t}^{s} \alpha_{t,v}\right)\right).$$

We define the observation date random variable

$$S_t = \min_s \left\{ s \in \mathbb{N} \,\middle|\, \sum_{v=t}^{s} \alpha_{t,v} > \tilde{U} \right\}. \tag{13}$$

This expression transforms the time-changed observation delay random variable into the associated observation date. Consequently S_t satisfies $P(S_t = s) = p_{t,s}$. For each event that occurred on date t we generate a realization from the distribution of $\tilde{U}$. We obtain the corresponding observation date by replacing the random variable $\tilde{U}$ in (13) by this sampled value.

Step 3. Truncation With steps 1 and 2 we have simulated an observation date for each occurred event. We split this data set into observed and hidden events. We use the data set with observed events to calibrate the model and to predict the number of hidden events. The hidden events are kept only for evaluating the prediction accuracy.

C A standard distribution for the time changed observation delay

Modeling the time-changed observation delay with an exponential distribution has significant computational benefits. Therefore, this section puts focus on the use of the exponential distribution as a standard distribution for modeling the time-changed observation delay $\tilde{U}$. Since the exponential distribution is light-tailed it is less suited for long or heavy-tailed delays. We outline a strategy for addressing this weakness of the exponential distribution.

Our strategy bins the possible observation delays ($s-t = 0, 1, \ldots$) and categorizes these bins with a delay covariate $x^{\texttt{delay}}_{s-t}$. This covariate is then included in the observation exposure specification. For each bin we estimate a parameter to capture its effect on observation exposure. These parameters can strongly reshape the distribution, hereby overcoming many of the disadvantages of the exponential distribution. We present a maximum likelihood driven binning strategy in Appendix C.1 and then Appendix C.2 derives the same bins by linking our approach to the non-parametric Kaplan-Meier estimator (Kaplan and Meier, 1958).

C.1 Binning observation delay

Our binning strategy maximizes the loglikelihood in (8) when the observation exposures depend only on the time elapsed since the event occurred, i.e.

$$\alpha_{t,s} = \exp(\boldsymbol{\gamma}^{\texttt{delay}} \cdot x^{\texttt{delay}}_{s-t}) = \exp(\gamma^{\texttt{s-t}}),$$

where we estimate for each delay $s - t$ a separate parameter $\gamma^{\texttt{s-t}}$. Furthermore we neglect the last term in (8), capturing the effect of the right truncation. Under these restrictions, the loglikelihood to optimize is

$$\ell(\boldsymbol{\gamma};\boldsymbol{\chi}) = -\sum_{t=1}^{\tau}\sum_{v=t}^{\tau-1}\left(\sum_{s=v+1}^{\tau} N_{t,s}\right)\cdot \exp(\gamma^{\texttt{v-t}}) + \sum_{t=1}^{\tau}\sum_{s=t}^{\tau} N_{t,s}\cdot \log\left(1-\exp(-\exp(\gamma^{\texttt{s-t}})\right)$$

We compute the derivatives of $\ell(\boldsymbol{\gamma};\boldsymbol{\chi})$ with respect to the observation exposure parameter $\gamma^{\texttt{d}}$ for positive delays $d \in \mathbb{N}$

$$\frac{\partial \ell(\boldsymbol{\gamma};\boldsymbol{\chi})}{\partial \gamma^{\texttt{d}}} = -\exp(\gamma^{\texttt{d}})\cdot \sum_{t=1}^{\tau-d-1}\sum_{s=t+d+1}^{\tau} N_{t,s} + \frac{\exp(\gamma^{\texttt{d}})}{\exp(\exp(\gamma^{\texttt{d}}))-1}\cdot\sum_{t=1}^{\tau-d} N_{t,t+d}.$$

Both sums in this expression have a logical interpretation. The first sum $(\sum_{t=1}^{\tau-1-d}\sum_{s=d+t+1}^{\tau} N_{t,s})$ counts the number of observed events with a delay longer than d days, whereas the second sum $(\sum_{t=1}^{\tau-d} N_{t,t+d})$ counts all events with a delay of exactly d days. These derivatives are zero when

$$\exp(\gamma^{\texttt{d}}) = -\log\left(1-\frac{|\texttt{delay = d}|}{|\texttt{delay} \geqslant \texttt{d}|}\right), \tag{14}$$

where $|\texttt{delay = d}|$ denotes the number of events observed with a delay of d days and $|\texttt{delay > d}|$ the number of events with a delay of more than d days.

We propose to bin the observation delay by grouping delays for which (14) is approximately constant. Figure 16 visualizes this approach for the liability insurance data set discussed in Section 3. This figure shows in red the estimated delay parameters using approximation (14). The top panel shows the estimates for delays up to 31 days, whereas the parameters for larger delays (up to 400 days) are shown in the bottom panel. Based on this knowledge observation delay is grouped in 23 bins, separated by vertical gray bars in Figure 16. We use more bins for short delays, since for these delays (14) differs strongly. Moreover, many accidents have a short observation delay, which makes these first delays more important. As expected, this binning strategy identifies an increase in observation probability after exactly one year. In Section 3 we structure these bins in a categorical delay covariate $x_{s-t}^{\texttt{delay}}$ and estimate observation delay in a maximum likelihood framework. In Figure 16 the fitted parameters are plotted in blue. These parameters deviate from those found using approximation (14), since other covariate effects were estimated simultaneously. However, the maximum likelihood estimates are close to the approximate values which makes this approximation suitable for choosing initial values in the calibration.

C.2 A link with the Kaplan-Meier estimator

We show that under the binning strategy of Appendix C.1 the time changed model has the same flexibility as the Kaplan-Meier estimator and is as such suitable for modelling a wide range of portfolios.

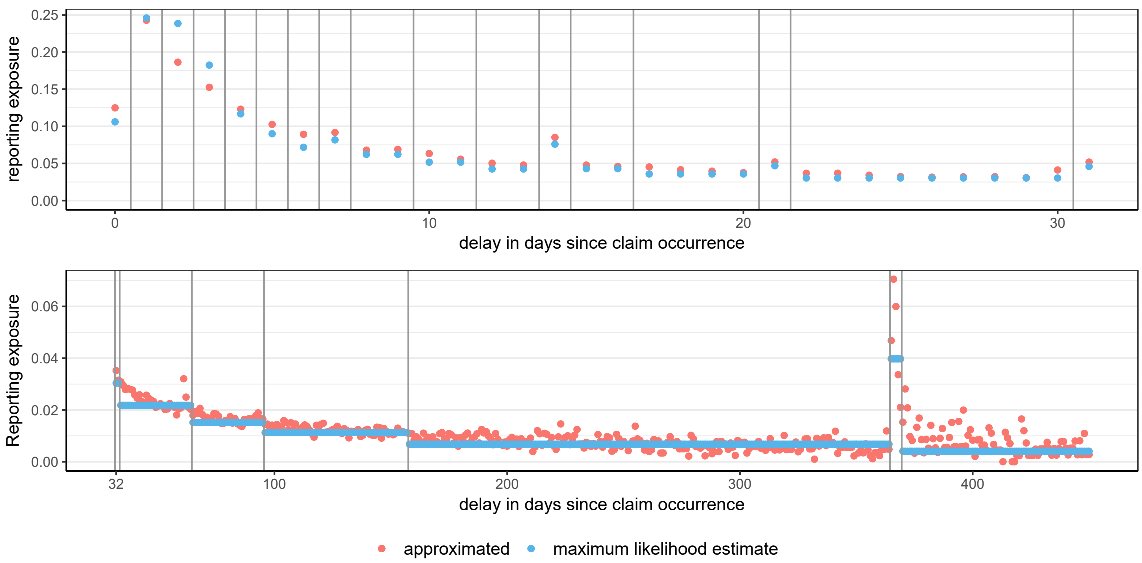

Figure 16: Observation exposure estimates for the delay effect during the first month after the accident occurrence (top) and longer delays (bottom). In red, we show estimates obtained for each delay using (14). The vertical lines indicate the chosen bins. Maximum likelihood estimates for the observation delay parameter corresponding to each bin in the regression structure proposed in Section 3.2 are plotted in blue.

The Kaplan-Meier estimator for the survival function of the observation delay random variable is

$$\widehat{P(\texttt{delay} > d)} = \prod_{i=0}^{d} \left(1 - \frac{|\texttt{delay} = i|}{|\texttt{delay} \geqslant i|}\right), \tag{15}$$

When we model the time-changed observation delay distribution $\tilde{U}$ using an exponential distribution then the survival probability for an event from occurrence day t is

$$\begin{aligned} P\left(\texttt{delay} > d \mid \texttt{occ. day} = t\right) &= P\left(\tilde{U} \geqslant \varphi_t(d+1)\right) \\ &= 1 - F_{\tilde{U}}\left(\sum_{i=1}^{d+1} \alpha_{t,t+i-1}\right) \\ &= \prod_{i=0}^{d} \exp\left(-\alpha_{t,t+i}\right). \end{aligned} \tag{16}$$

Notice the similarity between this expression and the Kaplan-Meier estimator in (15). When the observation exposure only depends on the time passed since the occurrence of the event, i.e. $\alpha_{t,t+i} := \alpha_i$, then

$$P\left(\texttt{delay} > d\right) = \prod_{i=0}^{d} \exp\left(-\alpha_i\right),$$

where α_i is the observation exposure at delay i. This expression no longer depends on the occurrence date t of the event. The Kaplan-Meier estimator is retrieved when

$$\alpha_i = -\log\left(1 - \frac{|\texttt{delay} = i|}{|\texttt{delay} \geqslant i|}\right). \tag{17}$$

Since $\alpha_i = \exp(\gamma^i)$, this is the same estimator we found in (14) through maximum likelihood estimation. This show that by estimating a separate delay parameter for each delay ($d = 0, 1, \ldots$) we obtain a model with the same flexibility as the non-parametric Kaplan-Meier estimator.