Individual claims reserving using the Aalen-Johansen estimator

Gabriele Pittarello
gabriele.pittarello@unito.it
University of Turin

Martin Bladt
martinbladt@math.ku.dk
University of Copenhagen

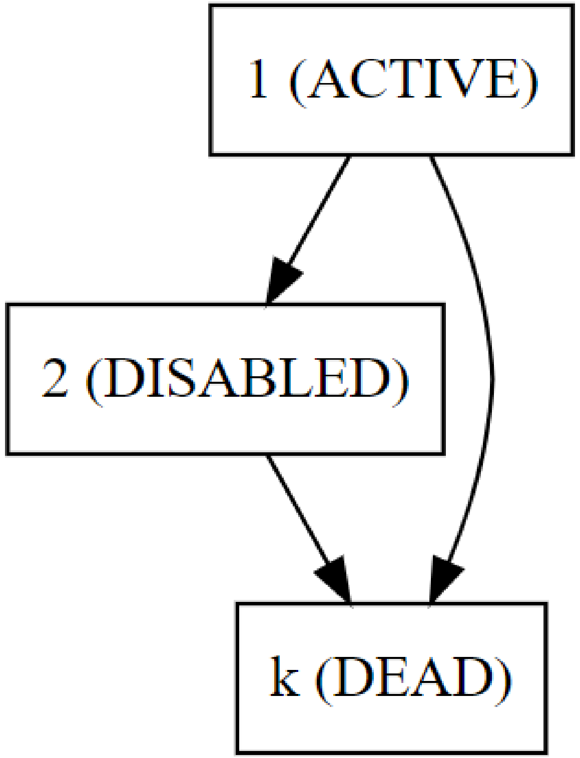
This presentation is about the work in Bladt and Pittarello (2023).

Martin Bladt, Associate Professor. University of Copenhagen.

Gabriele Pittarello, Postdoctoral Fellow. University of Turin.

Literature Review

- MSM have found widespread use in the domain of life insurance (Hoem 1969).

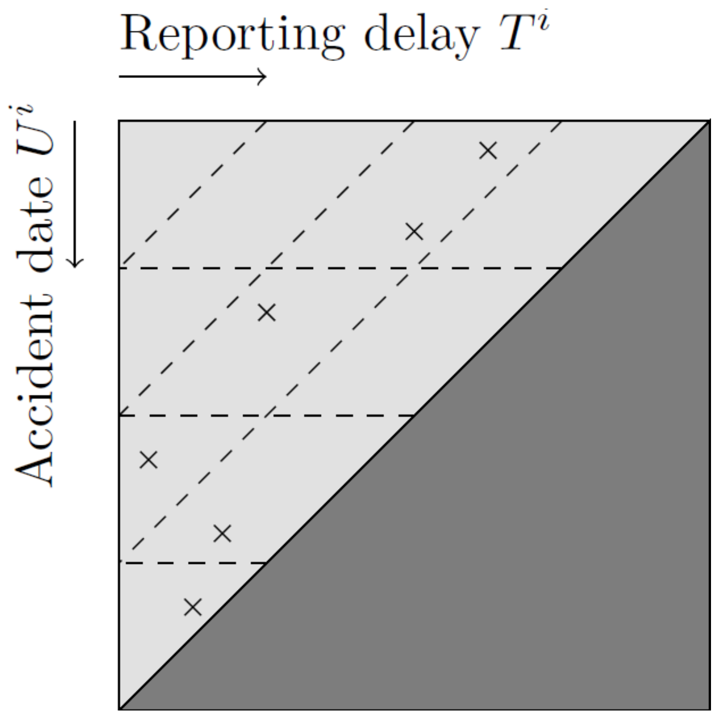


Example of MSM for modeling biometric states (active, deceased).

- Multi-state models have found widespread use in the domain of life insurance (Hoem 1969, 1972).
- Notable exceptions in non-life insurance (Hesselager 1994; Maciak, Mizera, and Pešta 2022).

- Multi-state models have found widespread use in the domain of life insurance (Hoem 1969, 1972).
- Notable exceptions in non-life insurance (Hesselager 1994; Maciak, Mizera, and Pešta 2022).
- The conditional Aalen-Johansen estimator (Bladt and Furrer (2023)) can be applied to estimate the distribution function of the **individual claim size** in an individual reserving model (Arjas 1989; Norberg 1993; Haastrup and Arjas 1996).
- Actuaries usually do reserving using aggregate data.

- Multi-state models have found widespread use in the domain of life insurance (Hoem 1969, 1972).
- Notable exceptions in non-life insurance (Hesselager 1994; Maciak, Mizera, and Pešta 2022).
- The conditional Aalen-Johansen estimator (Bladt and Furrer (2023)) can be applied to estimate the distribution function of the **individual claim size** in an individual reserving model (Arjas 1989; Norberg 1993; Haastrup and Arjas 1996).
- Actuaries usually do reserving using aggregate data.



The development triangle of reporting delays

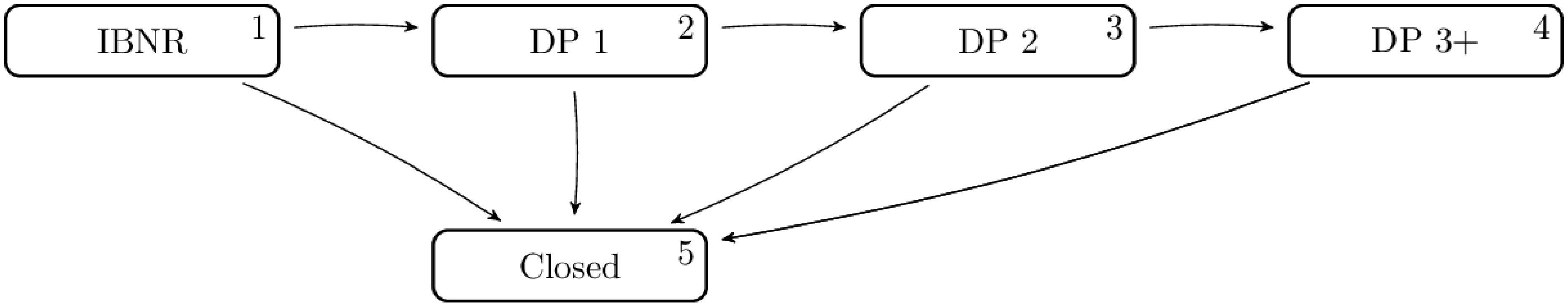
$$\mathcal{D}^{(k-1)} = \{d_{\ell,j} : \ell + j \leq k; \ell, j = 1, \dots, k-1\},$$

stems from the i.i.d. data $(U^i, T^i), i = 1, \dots, n$:

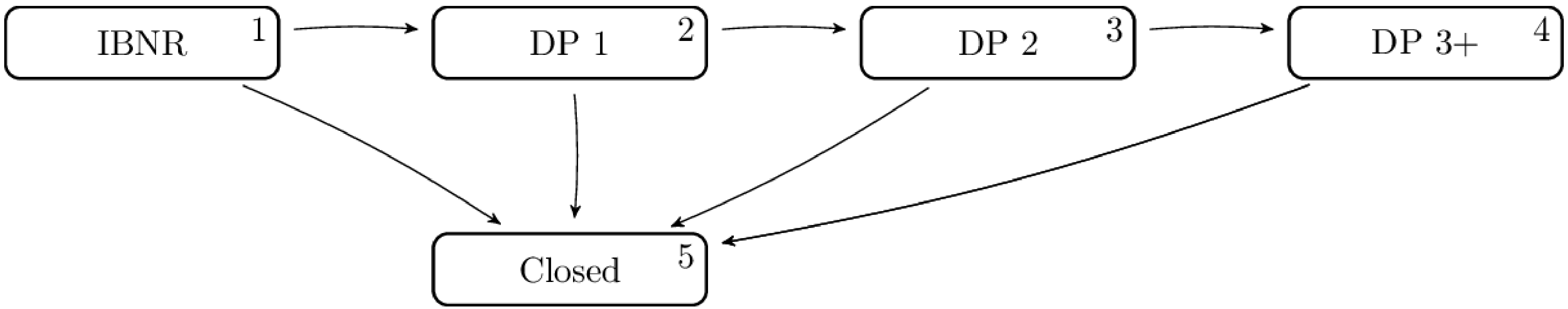
- U^i accident date.
- T^i the delay between accident and report.
- $d_{\ell,j} = \sum_{i=1}^{n} 1_{\{U^i=\ell \text{ and } T^i=j\}}$ denoting the total claims reported in accident period ℓ with delay j.

A MSM for individual claims reserving

We model based on a continuous-time non-explosive pure jump process denoted by J on a finite state space, $\mathcal{S} = \{1, \ldots, k\}, k \in \mathbb{N}$. States correspond to the development periods (DP's) within a development triangle, and the "time spent" between state transitions corresponds to the claim size growth between DP's.



Example of multi-state model for claims reserving, with $k = 5$.



Example of multi-state model for claims reserving, with $k = 5$.

We introduce a strictly positive random variable W describing right-censoring. We are interested the triplet

$$\left(X, (J_z)_{0 \leqslant z \leqslant W}, Y \wedge W\right).$$

We denote by Y the possibly infinite absorption time of J, and J_z the state occupied by j in z.

For any x in the support of X we model the conditional occupation probabilities

$$p_k(z \mid x) = E\left[I(\{J_z = k\}) \mid X = x\right].$$

Our sample is constituted of the i.i.d. replicates

$$\left(X^i, (J_z^i)_{0 \leqslant z \leqslant W^i}, Y^i \wedge W^i\right)_{i=1}^n .$$

Define

Our sample is constituted of the i.i.d. replicates

$$\Big(X^i, (J^i_z)_{0\leqslant z\leqslant W^i}, Y^i \wedge W^i\Big)_{i=1}^{n}.$$

Define

- $\delta^i := I(Y^i \leq W^i)$, which equals 1 when the observation is absorbed (closed claim).
- $n = n^{\texttt{Closed}} + n^{\texttt{RBNS}}$

Our sample is constituted of the i.i.d. replicates

$$\left(X^i, (J_z^i)_{0\leqslant z\leqslant W^i}, Y^i \wedge W^i\right)_{i=1}^n .$$

Define

- $\delta^i := I(Y^i \leq W^i)$, which equals 1 when the observation is absorbed (closed claim).
- $n = n^{\texttt{Closed}} + n^{\texttt{RBNS}}$

Under certain regulatory conditions (Bladt and Pittarello 2023), we derive the conditional Aalen-Johansen estimator

$$p^{(n)}(z \mid x) = p^{(n)}(0 \mid x)\Pi^z \left(\mathrm{Id} + \Lambda^{(n)}(\mathrm{d}s \mid x)\right),$$

where $\mathrm{p}_j^{(n)}(0 \mid x) = I_j^{(n)}(0 \mid x)$, and $\Lambda^{(n)}(\mathrm{d}s \mid x)$ is the conditional Nelson-Aalen estimator for cumulative hazard Bladt and Furrer (2023). We denote the product integral with Π.

Predictors

We are presently interested in predicting the ultimate cost of our claims,

$$Y^{\texttt{Closed}} + Y^{\texttt{RBNS}} := \sum_{i=1}^{n} \delta^i Y^i + \sum_{i=1}^{n} (1 - \delta^i) Y^i,$$

Predictors

We are presently interested in predicting the ultimate cost of our claims,

$$Y^{\texttt{Closed}} + Y^{\texttt{RBNS}} := \sum_{i=1}^{n} \delta^i Y^i + \sum_{i=1}^{n} (1 - \delta^i) Y^i,$$

The predictor of the final total cost of RBNS claims is defined by

$$\hat{Y}^{\texttt{RBNS}} = \sum_{i}^{n} I(\delta^i = 0) \left(W^i + \hat{E}[Y|Y > W^i, X = x] \right)$$

Predictors

We are presently interested in predicting the ultimate cost of our claims,

$$Y^{\texttt{Closed}} + Y^{\texttt{RBNS}} := \sum_{i=1}^{n} \delta^i Y^i + \sum_{i=1}^{n} (1 - \delta^i) Y^i,$$

The predictor of the final total cost of RBNS claims is defined by

$$\hat{Y}^{\texttt{RBNS}} = \sum_{i}^{n} I(\delta^i = 0) \left(W^i + \hat{E}[Y|Y > W^i, X = x] \right)$$

The general formula for predicting the m-th moment is explicitly given by

$$\hat{E}\left[Y^m | Y > W^i, X = x\right] = \frac{1}{1 - \mathrm{p}_k^{(n)}(W^i \mid x)} \int_{W^i}^{+\infty} m y^{m-1} (1 - \mathrm{p}_k^{(n)}(y \mid x)) \mathrm{d}y.$$

A model for IBNR claims

We describe the total cost of IBNR claims with the collective risk model in (Klugman, Panjer, and Willmot 2012, 715:Ch.9):

$$Y^{\texttt{IBNR}} = \sum_{i'=1}^{n^{\texttt{IBNR}}} \tilde{Y}^{i'}.$$

We describe the total cost of IBNR claims with the collective risk model in (Klugman, Panjer, and Willmot 2012, 715:Ch.9):

$$Y^{\texttt{IBNR}} = \sum_{i'=1}^{n^{\texttt{IBNR}}} \tilde{Y}^{i'}.$$

The expected cost of IBNR claims is estimated by

$$\hat{Y}^{\texttt{IBNR}} = \hat{E}[Y^{\texttt{IBNR}}] = \hat{n}^{\texttt{IBNR}} \hat{E}[\tilde{Y}],$$

with the m-th moment of $\tilde{Y}$ estimated, using the **unconditional** Aalen-Johansen, by

$$\hat{E}[\tilde{Y}^m] = \int_0^\infty m y^{m-1}(1 - \mathrm{p}_k^{(n)}(y))\mathrm{d}y.$$

We describe the total cost of IBNR claims with the collective risk model in (Klugman, Panjer, and Willmot 2012, 715:Ch.9):

$$Y^{\texttt{IBNR}} = \sum_{i'=1}^{n^{\texttt{IBNR}}} \tilde{Y}^{i'}.$$

The expected cost of IBNR claims is estimated by

$$\hat{Y}^{\texttt{IBNR}} = \hat{E}[Y^{\texttt{IBNR}}] = \hat{n}^{\texttt{IBNR}} \hat{E}[\tilde{Y}],$$

with the m-th moment of $\tilde{Y}$ estimated, using the **unconditional** Aalen-Johansen, by

$$\hat{E}[\tilde{Y}^m] = \int_0^\infty m y^{m-1}(1 - \mathrm{p}_k^{(n)}(y))\mathrm{d}y.$$

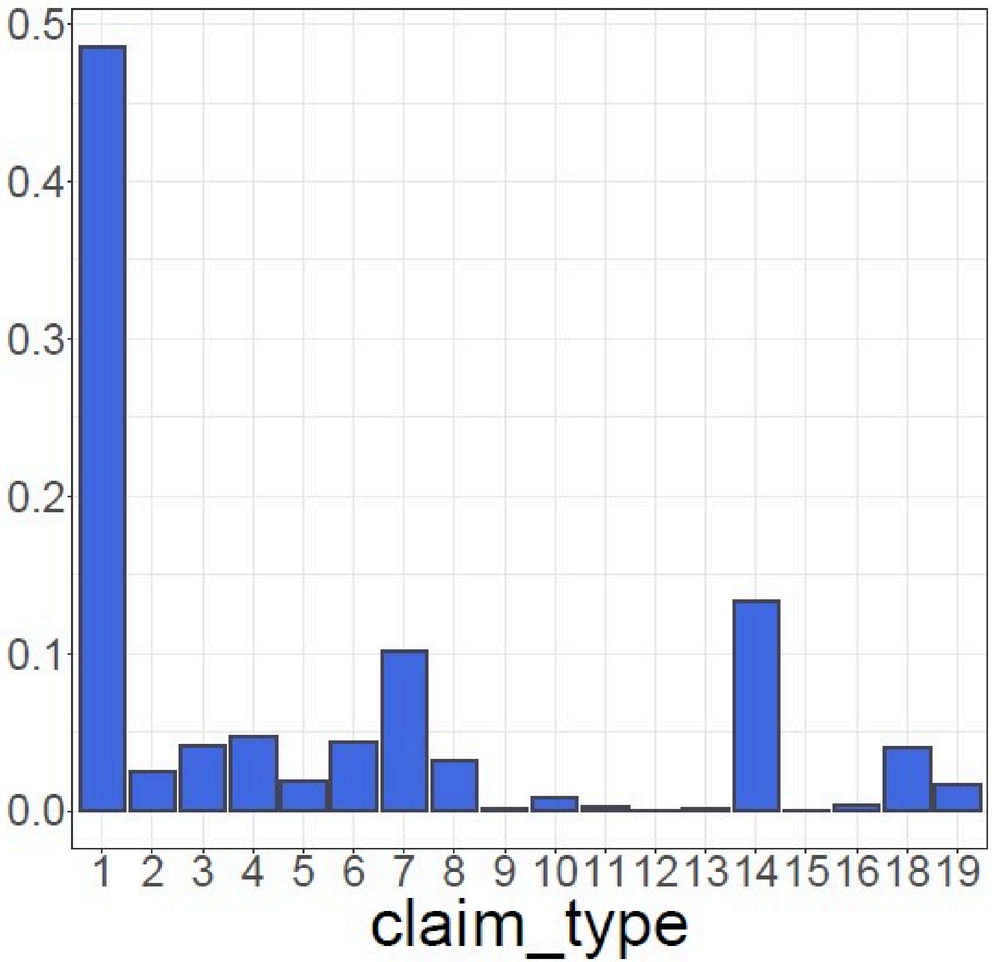
We calibrate the Mack Chain-Ladder estimator on $\mathcal{D}^{(k-1)}$ for $\hat{n}^{\texttt{IBNR}}$.

The total size of claims is $Y^{\texttt{TOT}} = Y^{\texttt{Closed}} + Y^{\texttt{RBNS}} + Y^{\texttt{IBNR}}$ and we estimate it with $\hat{Y}^{\texttt{TOT}} = Y^{\texttt{Closed}} + \hat{Y}^{\texttt{RBNS}} + \hat{Y}^{\texttt{IBNR}}$.

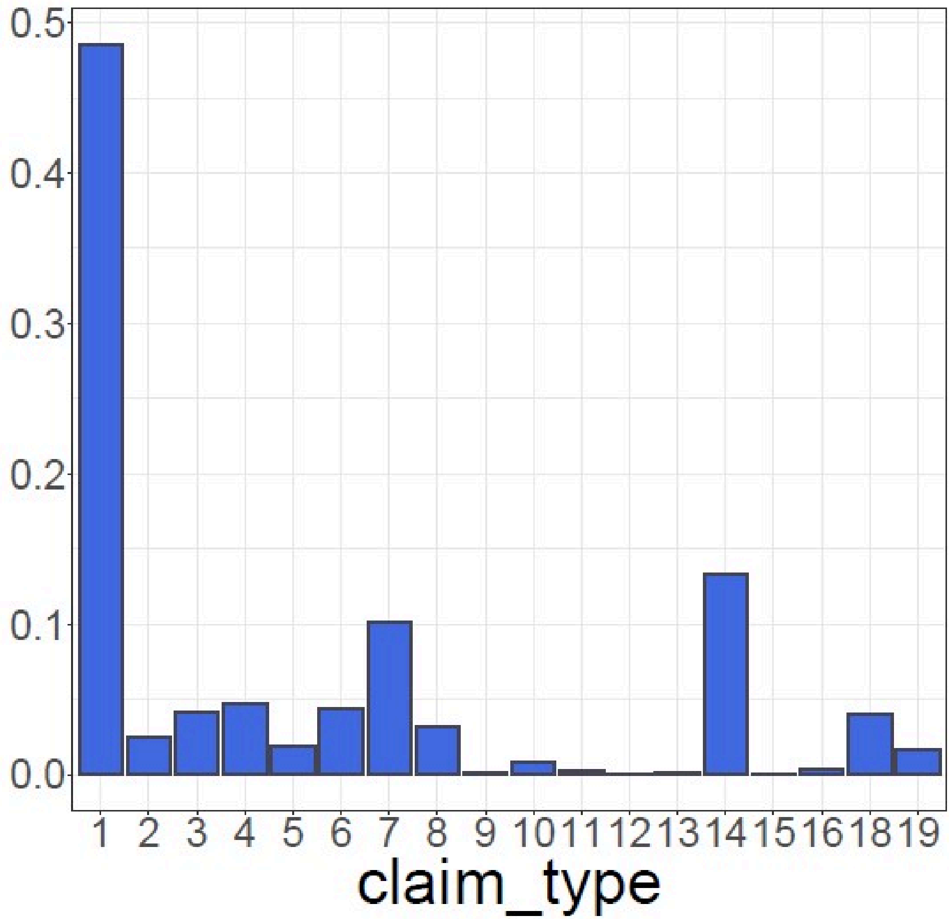
A data application on an insurance portfolio

A real data set from a Danish insurance company.

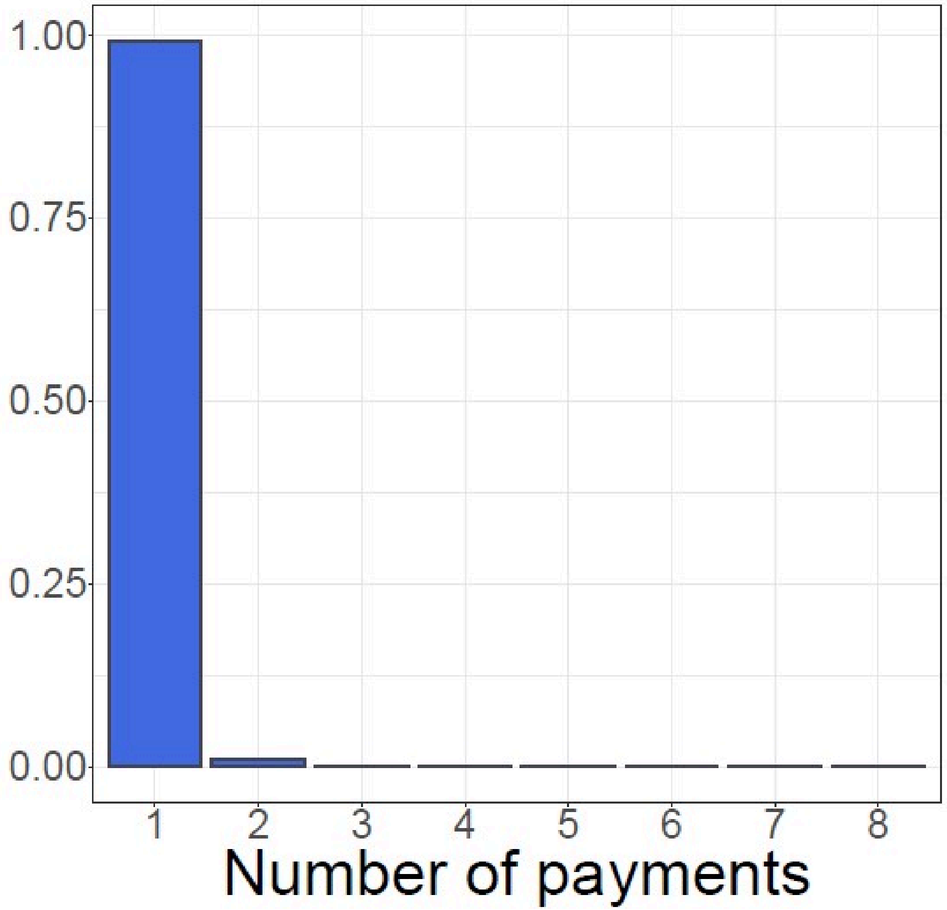
Covariates	**Description**
Claim_number	Policy identifier
claim_type $\in \{1, \ldots, 20\}$	Type of claim
AM	Accident month
CM	Calendar month of report
DM	Development month
incPaid	Incremental paid amount
Delta	Indicator, 0 when the claim is open



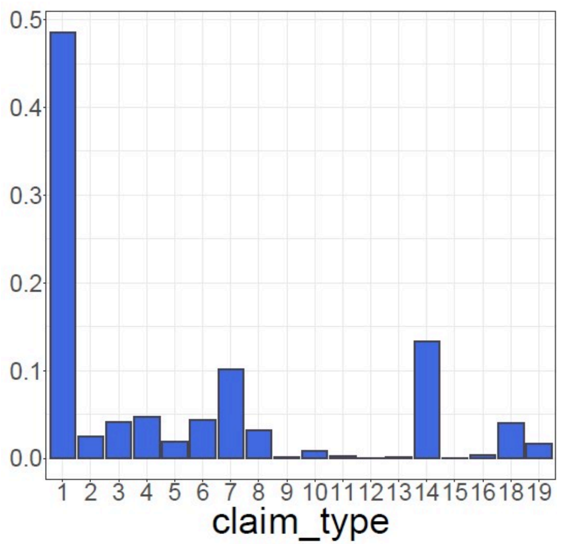
Frequency by `claim_type`.



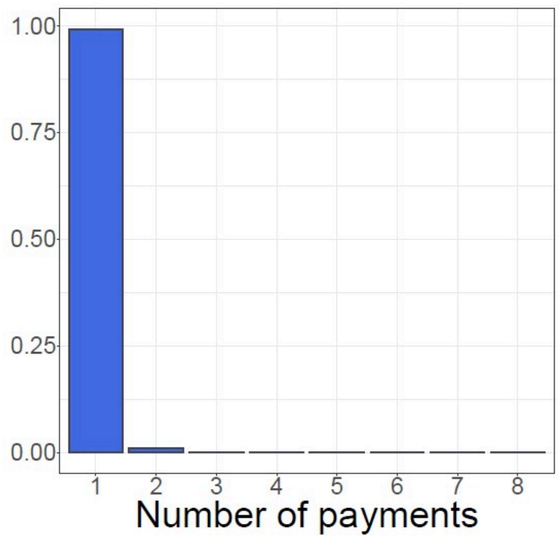
Frequency by `claim_type`.



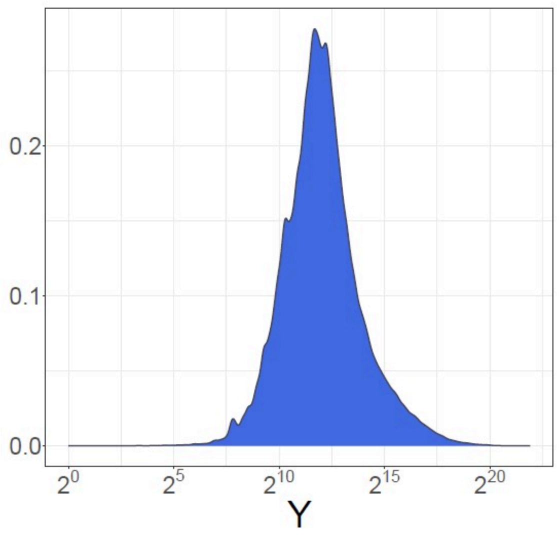
Distribution of the number of payments.

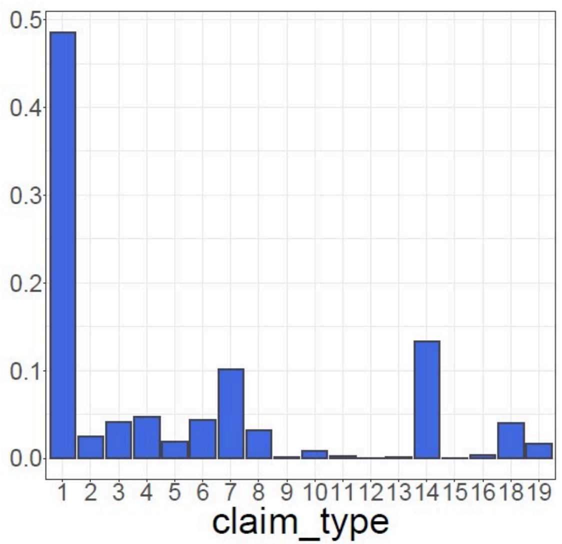


Frequency by `claim_type`.

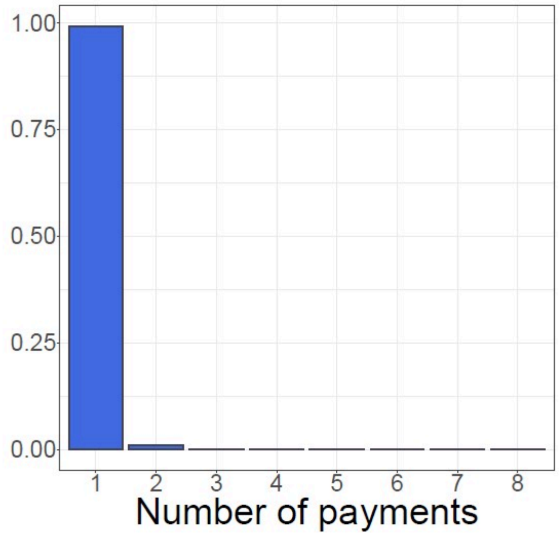


Distribution of the number of payments.

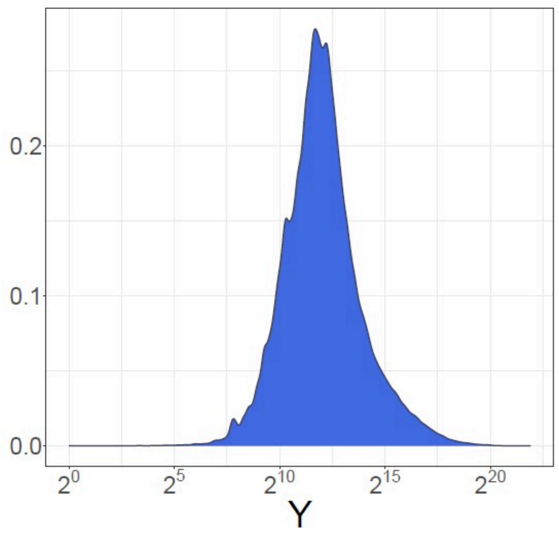


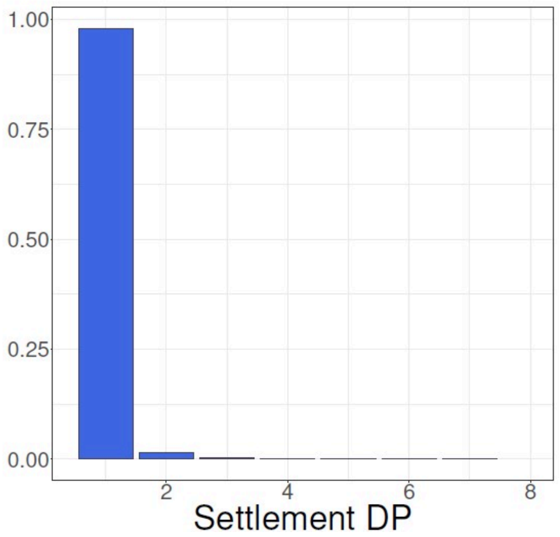


Frequency by `claim_type`.



Distribution of the number of payments.





Model comparison on different datasets

Model strategy:

1. We cut our time-serie at different depths ($k = 4, 5, 6, 7$).

Model strategy:

1. We cut our time-serie at different depths ($k = 4, 5, 6, 7$).
2. We define two performance metrics (next slide).

Model strategy:

1. We cut our time-serie at different depths ($k = 4, 5, 6, 7$).
2. We define two performance metrics (next slide).
3. For the different choices of k compare:
    - Our model (AJ) conditioning on the feature `claim_type`.
    - Our model (AJ) **without** conditioning on the feature `claim_type`.
    - The Chain-Ladder model (CL, Mack 1993).

- The (average) Continuously Ranked Probability Score (CRPS, Gneiting and Raftery (2007)).

$$\mathrm{CRPS}(\mathrm{p}_k^{(n)}(z \mid x), y) = \int_0^{+\infty} (\mathrm{p}_k^{(n)}(z \mid x) - 1_{\{y \leq z\}})^2 \mathrm{d}z$$

- The (average) Continuously Ranked Probability Score (CRPS, Gneiting and Raftery (2007)).
- The error incidence,

$$\mathtt{EI} = \frac{\hat{Y}^{\mathtt{TOT}}}{Y^{\mathtt{TOT}}} - 1.$$

Models comparison for $k = 4, 5, 6, 7$

k	`claim_type`	$Y^{\texttt{TOT}}$	$\mathrm{EI}^{\mathrm{TOT}}$ (AJ)	$\mathrm{EI}^{\mathrm{TOT}}$ (CL)	$\sqrt{\hat{\mathrm{Var}}(Y^{\texttt{TOT}})}/\hat{Y}^{\texttt{TOT}}$ (AJ)	$\sqrt{\hat{\mathrm{Var}}(Y^{\texttt{TOT}})}/\hat{Y}^{\texttt{TOT}}$ (CL)	CRPS
4	✓	616.1327	0.0035	0.0157	0.0029	0.0023	1.0000
	✗		-0.0029		0.0029		1.1403
5	✓	822.5956	0.0064	0.0209	0.0008	0.0024	1.0000
	✗		-0.0061		0.0007		0.4596
6	✓	999.6005	0.0059	0.0173	0.0017	0.0017	1.0000
	✗		-0.0052		0.0017		0.9987
7	✓	1190.9112	0.0146	0.0144	0.0011	0.0017	1.0000
	✗		-0.0142		0.0011		1.0022

Models comparison for $k = 4, 5, 6, 7$

k	`claim_type`	Y^{TOT}	EI^{TOT} (AJ)	EI^{TOT} (CL)	$\sqrt{\hat{\text{Var}}(Y^{\text{TOT}})}/\hat{Y}^{\text{TOT}}$ (AJ)	$\sqrt{\hat{\text{Var}}(Y^{\text{TOT}})}/\hat{Y}^{\text{TOT}}$ (CL)	CRPS
4	✓	616.1327	0.0035	0.0157	0.0029	0.0023	1.0000
	✗		-0.0029		0.0029		1.1403
5	✓	822.5956	0.0064	0.0209	0.0008	0.0024	1.0000
	✗		-0.0061		0.0007		0.4596
6	✓	999.6005	0.0059	0.0173	0.0017	0.0017	1.0000
	✗		-0.0052		0.0017		0.9987
7	✓	1190.9112	0.0146	0.0144	0.0011	0.0017	1.0000
	✗		-0.0142		0.0011		1.0022

Notes on the computation of $\text{Var}\widehat{(Y^{\text{TOT}})}$

We compute numerically the AJ estimator for the standard deviation,

$$\text{Var}\widehat{(Y^{\text{TOT}})} = \text{Var}\widehat{(Y^{\text{RBNS}})} + \text{Var}\widehat{(Y^{\text{IBNR}})},$$

as sum of the individual variability.

In the CL case, we use the Mack estimator for the process variance.

An individual model for the claim size

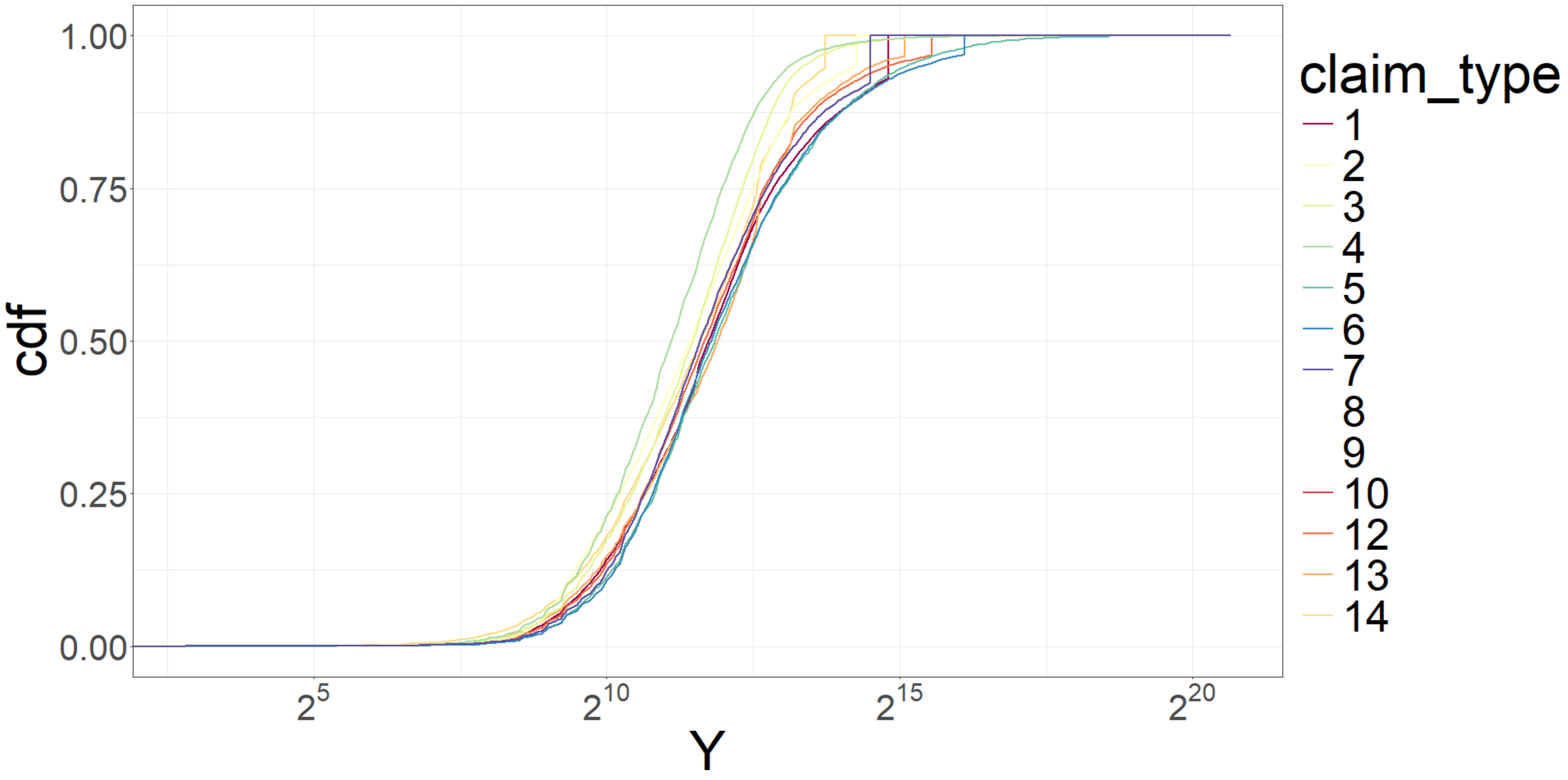


Figure 1: Severity curve for a k=4.

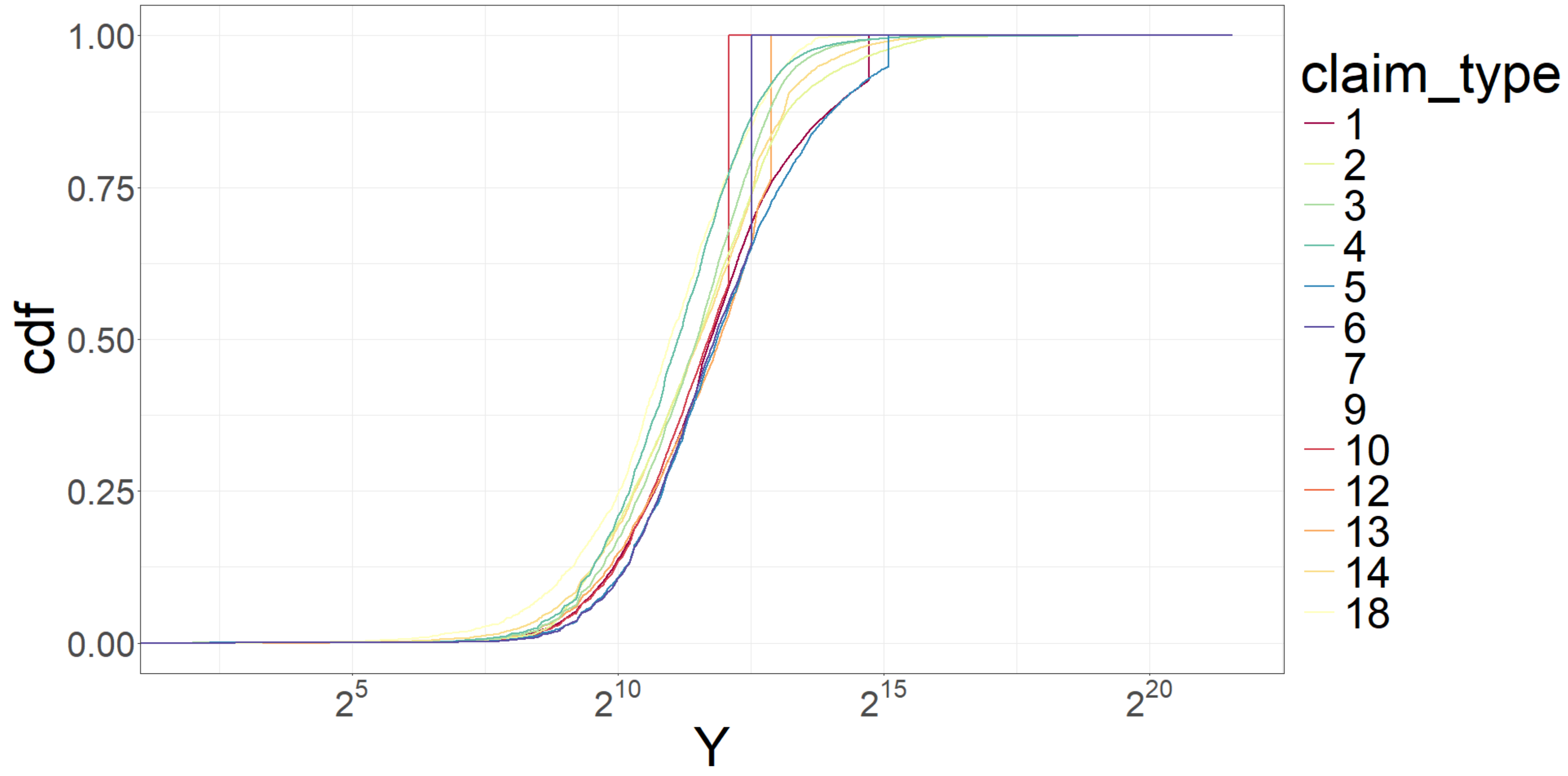


Figure 2: Severity curve for a k=5.

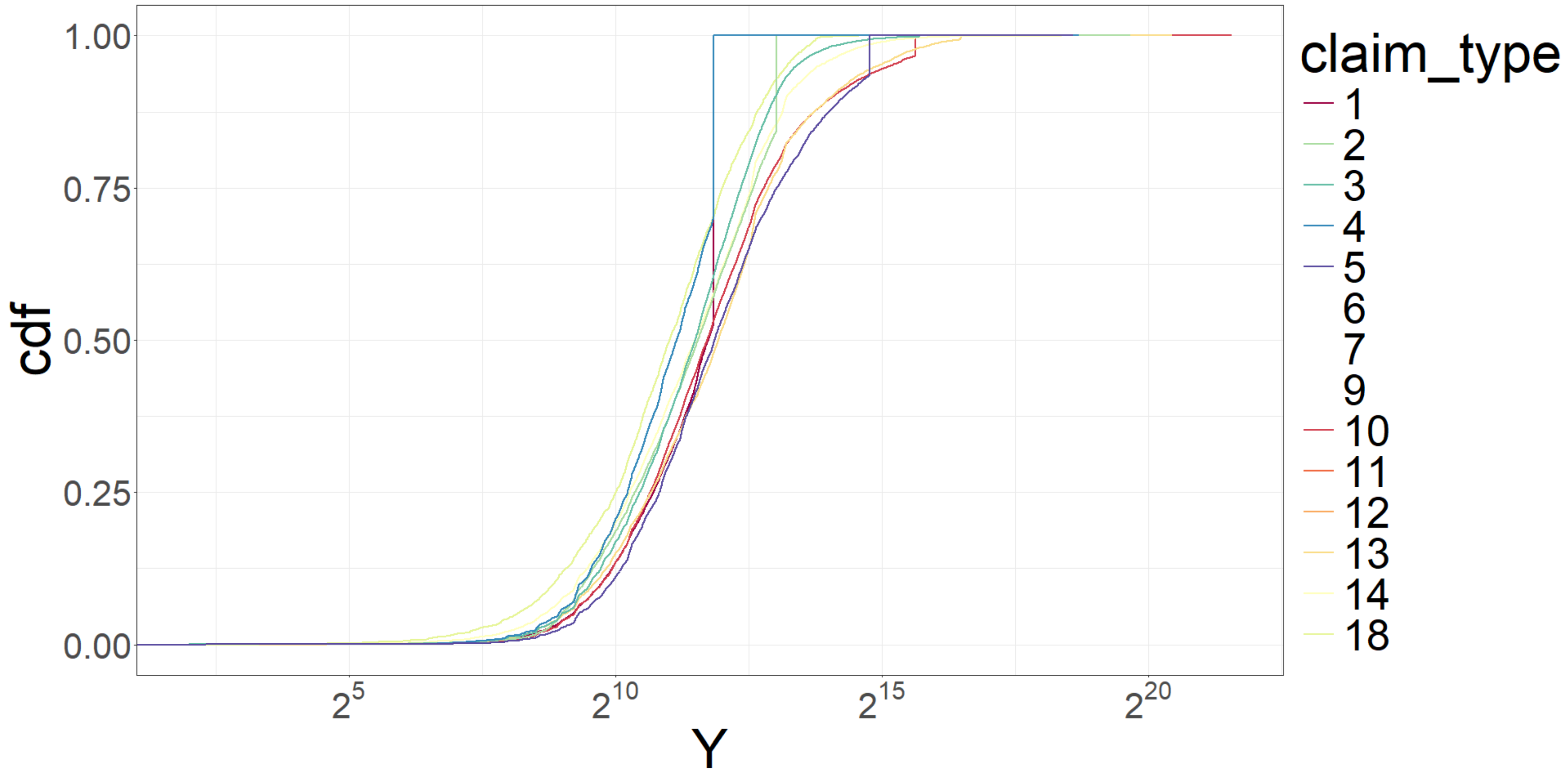


Figure 3: Severity curve for a k=6.

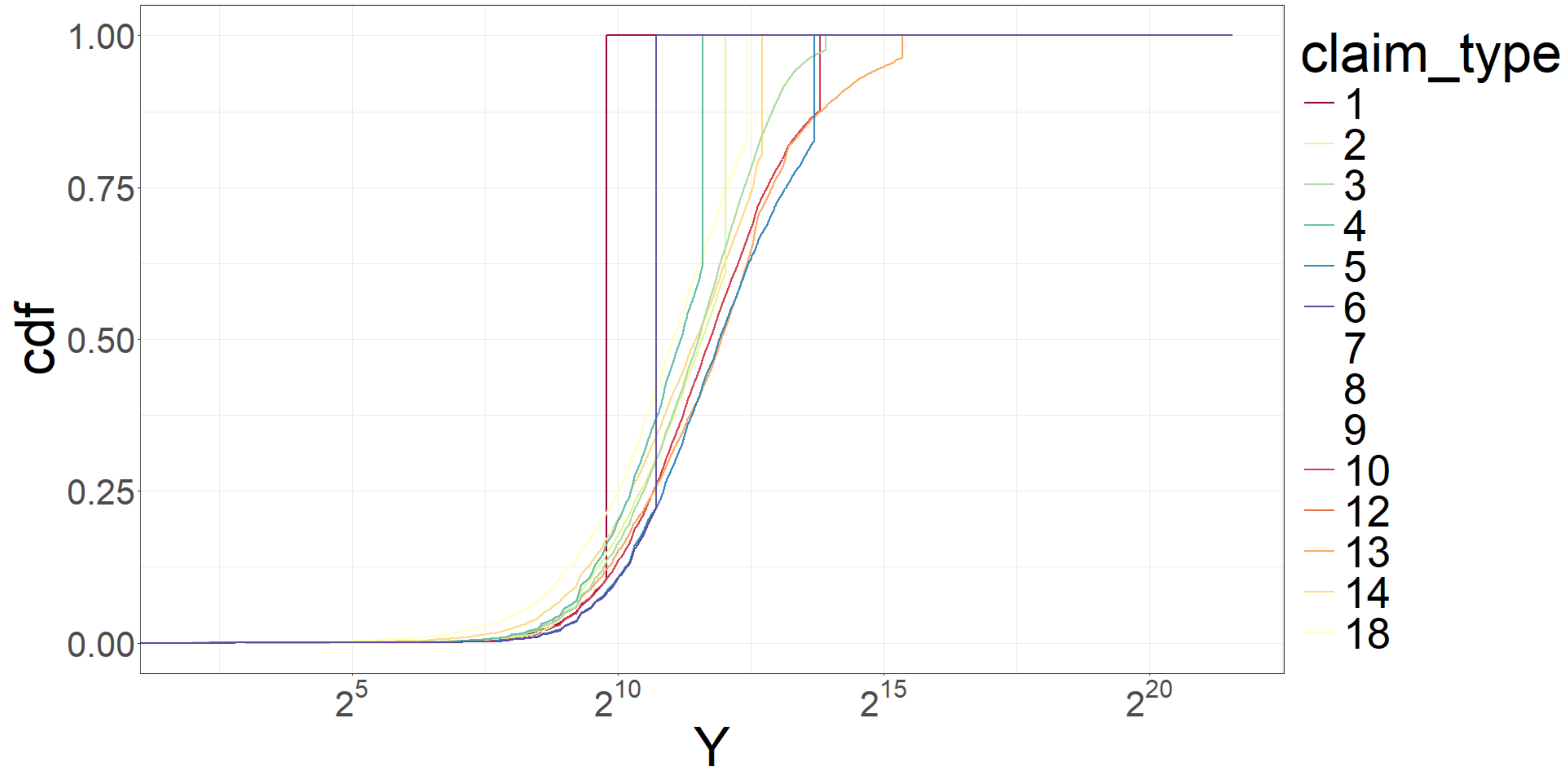


Figure 4: Severity curve for a k=7.

Model comparison on a single dataset

Model strategy:

1. We cut our time-serie at maximum of 5 calendar periods.

Model strategy:

1. We cut our time-serie at maximum of 5 calendar periods.
2. We compare, using CRPS and `EI`:
    - The AJ model conditioning on the feature `claim_type` for $k = 4, 5, 6$.
    - The AJ model **without** conditioning on the feature `claim_type` for $k = 4, 5, 6$.
    - The Chain-Ladder model (CL, Mack 1993).

Results for the data set with $k = 6$.

k	`claim_type`	`EI` (AJ)	$\sqrt{\hat{\text{Var}}(Y^{\texttt{TOT}})}/\hat{Y}^{\texttt{TOT}}$ (CL) (AJ)	CRPS (average, relative)
4	✗	-0.0059	0.0016	1.0143
	✓	-0.0040	0.0020	1.0000
5	✗	-0.0069	0.0015	0.9916
	✓	-0.0052	0.0018	1.0000
6	✗	-0.0052	0.0017	0.9987
	✓	-0.0059	0.0017	1.0000

- $Y^{\texttt{TOT}} = 999.6005$ (Actual, millions).
- $\sqrt{\hat{\text{Var}}(Y^{\texttt{TOT}})}/\hat{Y}^{\texttt{TOT}} = 0.0017$ (CL).
- $\texttt{EI} = 0.0173$ (CL).
- Using the CRPS we can select the model with features for $k = 4$:
  - Claims reserve 1.485 millions.
  - Standard Deviation 2.0385 millions.

Replicable results

- Replication material can be found in our GitHub folder.

README

DOI 10.5281/zenodo.10118896

Individual claims reserving using the Aalen-Johansen estimator

This repository contains the code to replicate the manuscript Individual claims reserving using the Aalen-Johansen estimator.

The replication material is organized as described in this file and it can be used for replicating exactly the results of our manuscript.

The script `elaborate_latex_tables.R` contains the code to print the tables in the manuscript from the the results in the `results_csv` folder.

The script `helper_functions_ajr.R` contains functions that we used to perform the calculations.

The real data that we used in our application will not be shared, but we provide the results that we obtained.

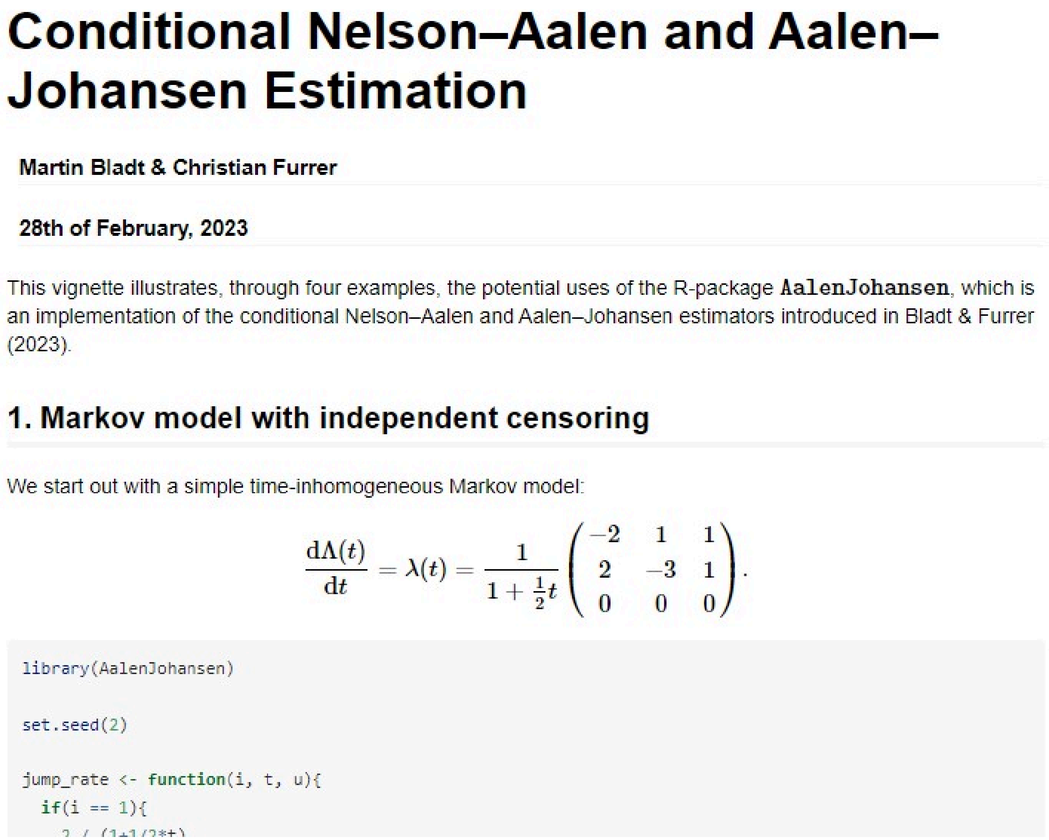
For each section, we describe below the relevant replication material.

Section 4. Simulation of RBNS claims

Section 4.1. Implementation

gpitt71/conditional-aj-reserving folder ReadME.md.

- Replication material can be found in our GitHub folder.
- The R package `AalenJohansen` for the conditional Aalen-Johansen estimation is available on CRAN.



Conditional Nelson–Aalen and Aalen–Johansen Estimation

Martin Bladt & Christian Furrer

28th of February, 2023

This vignette illustrates, through four examples, the potential uses of the R-package **AalenJohansen**, which is an implementation of the conditional Nelson–Aalen and Aalen–Johansen estimators introduced in Bladt & Furrer (2023).

1. Markov model with independent censoring

We start out with a simple time-inhomogeneous Markov model:

$$\frac{\mathrm{d}\Lambda(t)}{\mathrm{d}t} = \lambda(t) = \frac{1}{1+\frac{1}{2}t}\begin{pmatrix} -2 & 1 & 1 \\ 2 & -3 & 1 \\ 0 & 0 & 0 \end{pmatrix}.$$

```
library(AalenJohansen)

set.seed(2)

jump_rate <- function(i, t, u){
  if(i == 1){
    2 / (1+1/2*t)
```

`AalenJohansen` package vignette.

Thank you for your attention!

References

Arjas, Elja. 1989. "The Claims Reserving Problem in Non-Life Insurance: Some Structural Ideas." *ASTIN Bulletin: The Journal of the IAA* 19 (2): 139–52.

Bladt, Martin, and Christian Furrer. 2023. "Conditional Aalen–Johansen Estimation." *arXiv Preprint arXiv:2303.02119*.

Bladt, Martin, and Gabriele Pittarello. 2023. "A Novel Approach for Individual Claims Reserving Using the Aalen-Johanson Estimation." *Preprint*.

Gneiting, Tilmann, and Adrian E Raftery. 2007. "Strictly Proper Scoring Rules, Prediction, and Estimation." *Journal of the American Statistical Association* 102 (477): 359–78.

Haastrup, Svend, and Elja Arjas. 1996. "Claims Reserving in Continuous Time; a Nonparametric Bayesian Approach." *ASTIN Bulletin: The Journal of the IAA* 26 (2): 139–64.

Hesselager, Ole. 1994. "A Markov Model for Loss Reserving." *ASTIN Bulletin: The Journal of the IAA* 24 (2): 183–93.

Hoem, Jan M. 1969. "Markov Chain Models in Life Insurance." *Blätter Der DGVFM* 9 (2): 91–107.

———. 1972. "Inhomogeneous Semi-Markov Processes, Select Actuarial Tables, and Duration-Dependence in Demography." In *Population Dynamics*, 251–96. Elsevier.

Klugman, Stuart A, Harry H Panjer, and Gordon E Willmot. 2012. *Loss Models: From Data to Decisions*. Vol. 715. John Wiley & Sons.

Maciak, Matúš, Ivan Mizera, and Michal Pešta. 2022. "Functional Profile Techniques for Claims Reserving." *ASTIN Bulletin: The Journal of the IAA* 52 (2): 449–82.

Mack, Thomas. 1993. "Distribution-Free Calculation of the Standard Error of Chain Ladder Reserve Estimates." *ASTIN Bulletin: The Journal of the IAA* 23 (2): 213–25.

Norberg, Ragnar. 1993. "Prediction of Outstanding Liabilities in Non-Life Insurance." *ASTIN Bulletin: The Journal of the IAA* 23 (1): 95–115.