Loss Reserving Analytics

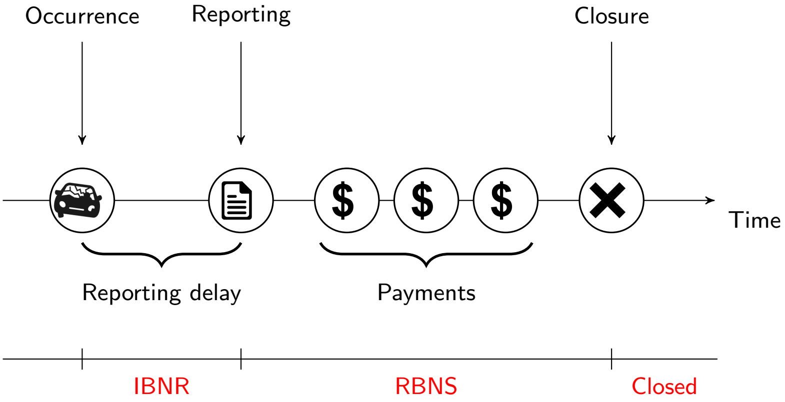
Micro-level IBNR reserving

Katrien Antonio
LRisk - KU Leuven and ASE - University of Amsterdam

April, 2019

Introduction

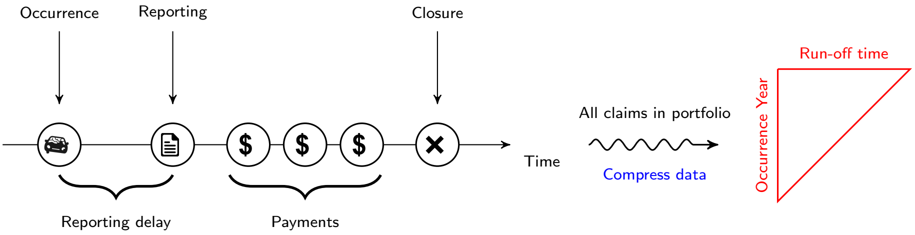
Development of a single claim



Introduction

Aggregated approach

We aggregate the data from the time line into a run-off triangle or claims development triangle:



Introduction

Pros and cons of aggregated approach

- Advantages of aggregating, pros of macro-level:
    - useful for accounting figures (audit)
    - established over years
    - low data requirements and computational power
    - simple and straightforward
    - ...

Inspired by Mario Wüthrich, 2017, New developments in claims reserving, 6th St. Petersburg Spring School.

Introduction

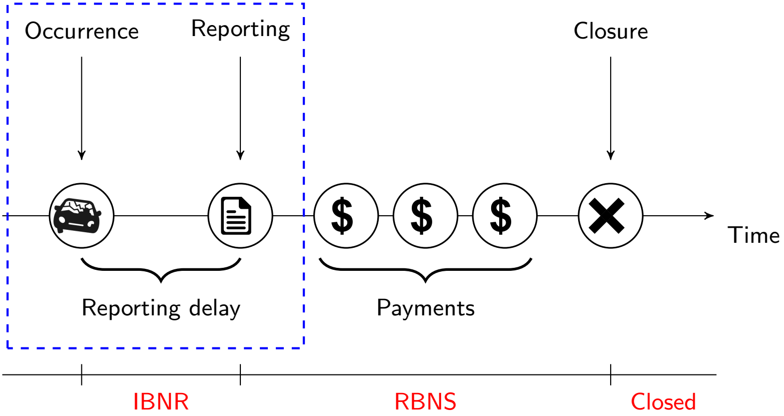
Pros and cons of aggregated approach

- Disadvantages of aggregating, pros of micro-level:
  - a lot of (detailed) data and insights gets lost
  - crude, brute force approach
  - individual claims (types) prediction is not available (viz. pricing of products)
  - case management (and early warning) is not possible
  - ...

Inspired by Mario Wüthrich, 2017, New developments in claims reserving, 6th St. Petersburg Spring School.

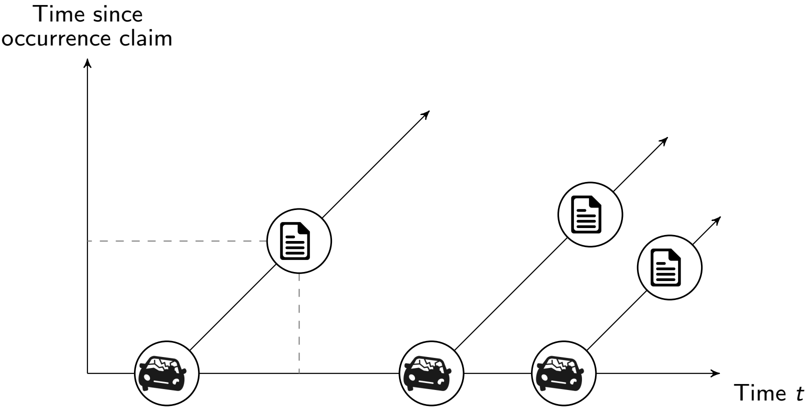
Research focus

IBNR claim counts



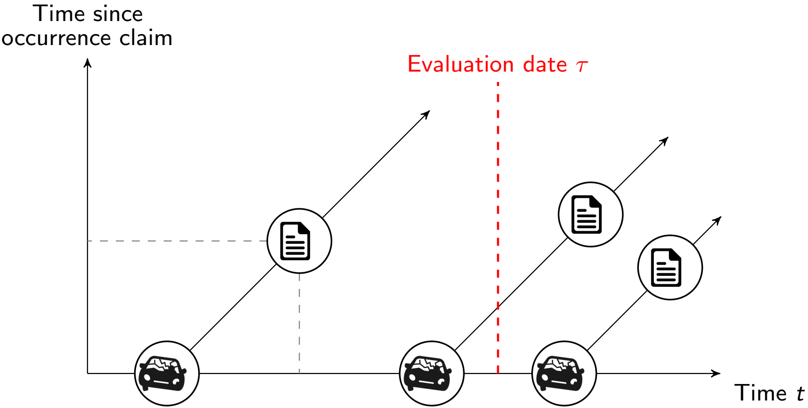
Research focus

IBNR claim counts



Research focus

IBNR claim counts



The insurance company is not aware (yet) of claims related to past exposures that are not (yet) reported!

Research questions

- Research questions with focus on IBNR?
  - How many claims occurred but are not yet reported, because reporting delay is subject to right truncation?
  - When will these IBNR claims be reported?
  - How to specify claim occurrences and reporting delay at daily level (=natural time unit)?
  - How to incorporate covariate information?
- Pioneering work by Ragnar Norberg (1993, 1999), (basic, first) implementation in Antonio & Plat (2014).

A closer look at the micro-level data!

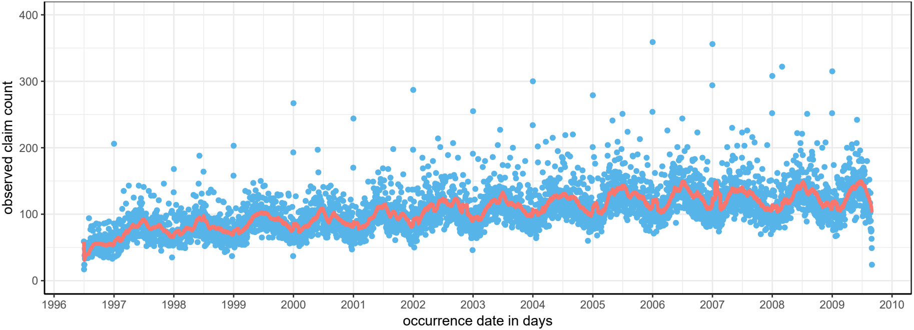
Case study
Structure of the data

- Large European dataset of liability claims (from private individuals).
- Two essential variables:
  - occurrence date
  - reporting date

  $+$ exposure, if available.
- Observation window: July 1, 1996 to August 31, 2009.

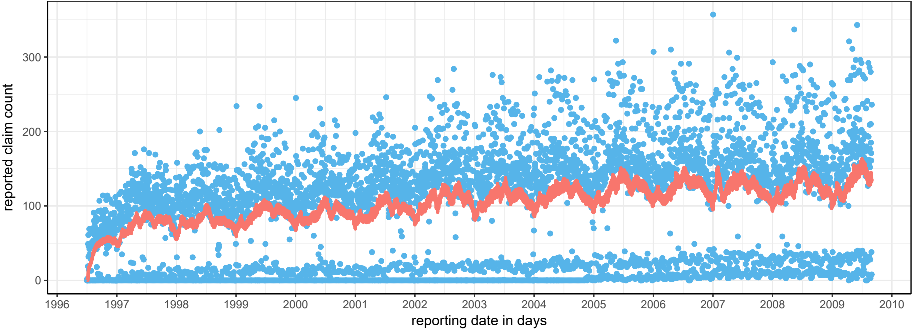
Case study

Claim occurrence process



Case study

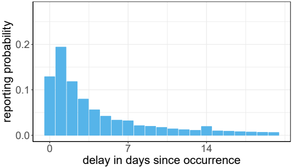
Reporting process



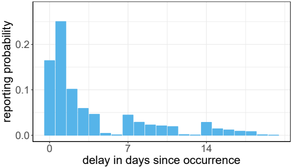
Case study

Reporting delay

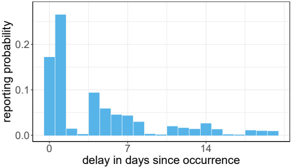
Declining pattern in reporting delay + intra-week pattern, depending on the occurrence day of the week.



All claims



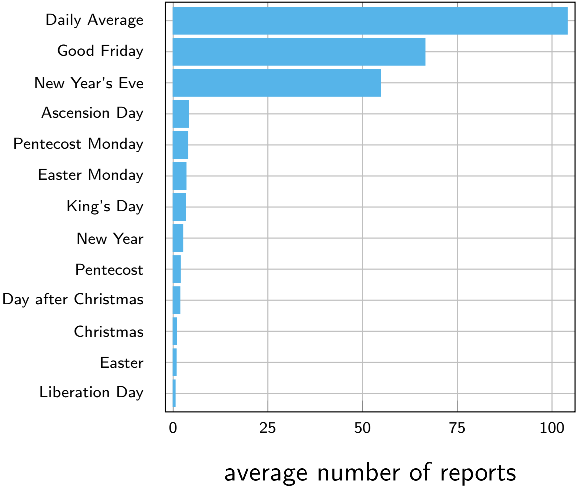
Monday



Thursday

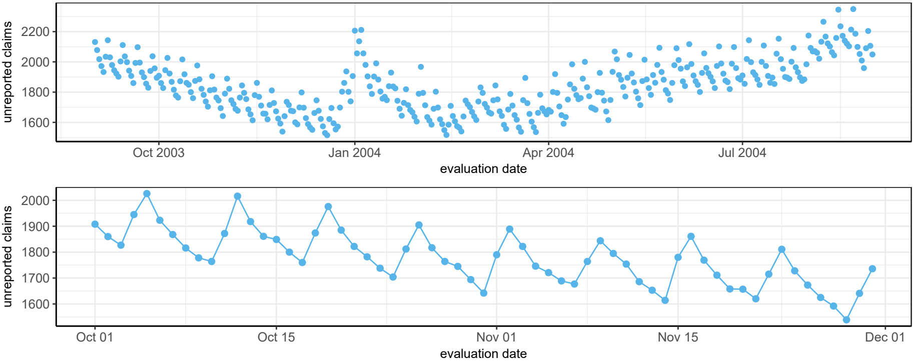
Case study

Holidays



Case study

Total IBNR counts



The statistical building blocks!

The statistical model

- Two (in progress) contributions:
  1. Verbelen, R., Antonio, K., Claeskens, G & Crèvecoeur, J. 2018, submitted.
     - joint estimation of occurrence process and reporting delay distribution
     - regression approach.
  2. Crèvecoeur, J., Antonio., K. & Verbelen, R. 2019, EJOR.
     - incorporate calendar day effects in reporting delay distribution

       cfr. national holidays and during weekend, reporting at specific delays (e.g. 14 days, 1 year)
     - regression approach.

Notations

- N_t: the (total) number of claims that occurred on day t.
- $N_{t,s}$: the number of claims from day t that are reported on day s.
- Each claim gets reported eventually, thus

$$N_t = \sum_{s=t}^{\infty} N_{t,s}.$$

Notations

A daily run-off triangle with reported claims

occurrence day	reporting delay (in days)				
	0	$\cdots$	$\tau - t$	$\cdots$	$\tau - 1$
1	N_{11}	$\cdots$	$N_{1,1+\tau-t}$	$\cdots$	$N_{1,\tau}$
$\vdots$					
t	N_{tt}	$\cdots$	$N_{t,\tau}$		
$\vdots$				IBNR	
τ	$N_{\tau\tau}$				

The statistical model
Assumptions

(A1) The daily total claim counts N_t for $t = 1, \ldots, \tau$ are independently Poisson distributed with intensity λ_t

$$N_t \sim \text{POI}(\lambda_t).$$

(A2) Conditional on N_t, the claim counts $N_{t,s}$ for $s = t, t+1, \ldots$ are multinomially distributed with probabilities $p_{t,s}$.

Combining $(A1)$ and $(A2)$, the $N_{t,s}$ are independent and

$$N_{t,s} \sim \text{POI}(\lambda_t \cdot p_{t,s}).$$

The statistical model

The likelihood

- We observe

$$\boldsymbol{N}^R = \{N_{t,s} \mid t \leq s \leq \tau\}$$

where $t \leq \tau$ indicates claim occurrence and $t \leq s \leq \tau$ reporting of the claim.

- Log-likelihood of observed data:

$$\ell(\boldsymbol{\lambda}, \boldsymbol{p}; \boldsymbol{N}^R) = \sum_{t=1}^{\tau} \sum_{s=t}^{\tau} \Big(\underbrace{-\lambda_t \cdot p_{t,s}}_{(\star)} + \log(\lambda_t) \cdot N_{t,s} + N_{t,s} \cdot \log(p_{t,s}) - \log(N_{t,s}!) \Big)$$

'difficult' to optimize (due to $\star$).

Expectation maximization strategy

The complete data likelihood

- Key idea: likelihood is difficult to optimize, because of unobserved data.
- Assume there is no unobserved data:

$$\boldsymbol{N} = \{N_{t,s} \mid t \leq \tau, t \leq s \leq \infty\}.$$

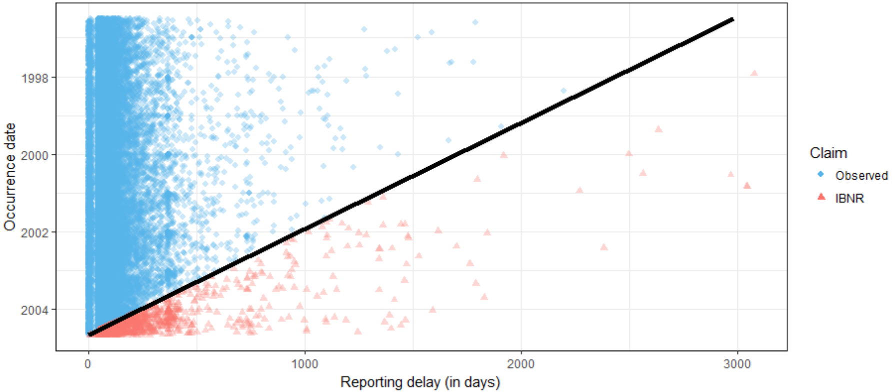
Then the likelihood of the complete data becomes:

$$\ell(\boldsymbol{\lambda}, \boldsymbol{p}; \boldsymbol{N}) = \sum_{t=1}^{\tau} \Big(-\lambda_t \cdot \underbrace{\sum_{s=t}^{\infty} p_{t,s}}_{=1} + \log(\lambda_t) \cdot \sum_{s=t}^{\infty} N_{t,s} + \sum_{s=t}^{\infty} N_{t,s} \cdot \log(p_{t,s}) - \sum_{s=t}^{\infty} \log(N_{t,s}!) \Big).$$

which splits into occurrence process and reporting delay likelihoods!

Expectation maximization strategy

Key idea



- Iteratively "guess" the complete loglikelihood (E-step) and maximize this loglikelihood (M-step).

Joint estimation of occurrence and reporting delay

A model for occurrences

- We propose a Poisson regression model:

$$
\begin{aligned}
N_t &\sim \text{POI}(\lambda_t) \\
\lambda_t &= e_t \cdot \exp\left(\boldsymbol{x}_t^{'}\boldsymbol{\alpha}\right),
\end{aligned}
$$

where e_t is the exposure on day t.

Joint estimation of occurrence and reporting delay

A model for reporting delay

- Probability of reporting after d days:

$$p_{t,d} = \begin{cases} p^W_{t,0} \cdot p^1_{t,d} & \text{for } d < 7 \\ p^W_{t,\lfloor \frac{d}{7} \rfloor} \cdot p^2_{t,d} & \text{otherwise} \end{cases}.$$

- Here:
  - $p^W_{t,w}$ probability of reporting in week w when the claim has occurred at t
  - $p^i_{t,d}$ probability of having a reporting delay d,

    given that the claim is reported in first week ($i = 1$) or later ($i = 2$), and has occurred at time t.

Joint estimation of occurrence and reporting delay

A model for reporting delay

- Probability of reporting after d days:

$$p_{t,d} = \begin{cases} p^W_{t,0} \cdot p^1_{t,d} & \text{for } d < 7 \\ p^W_{t,\lfloor \frac{d}{7} \rfloor} \cdot p^2_{t,d} & \text{otherwise} \end{cases}.$$

- Here:
  - $p^W_{t,w}$ probability of reporting in week w when the claim has occurred at t.
  - $p^i_{t,d}$ probability of having a reporting delay d,

    given that the claim is reported in first week $(i = 1)$ or later $(i = 2)$, and has occurred at time t.

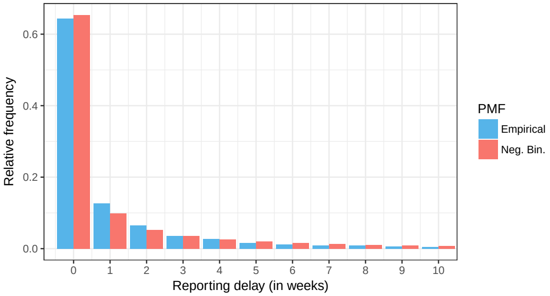
Joint estimation of occurrence and reporting delay

A model for reporting delay

- Use a Negative Binomial distribution for $(p^W_{t,w})_{w \geq 0}$:

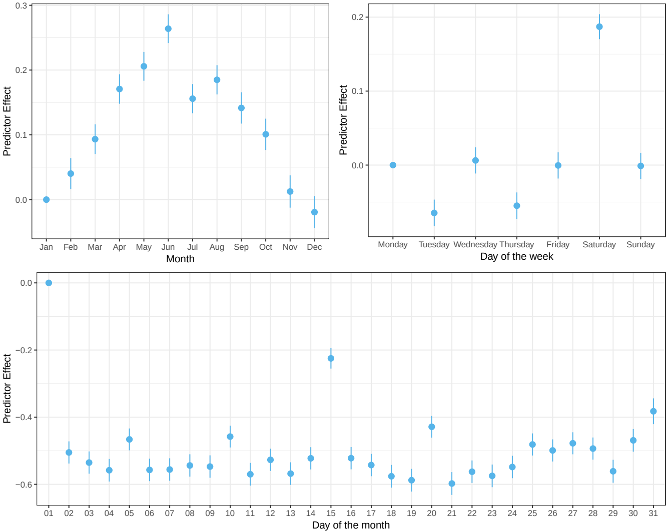
$$p^W_{t,w} = \frac{\Gamma(\phi + w)}{w!\Gamma(\phi)} \cdot \frac{\phi^\phi \mu_t^w}{(\phi + \mu_t)^{\phi + w}},$$

with $\mu_t = \exp(\boldsymbol{z}_t' \boldsymbol{\beta})$ incorporating covariate information.



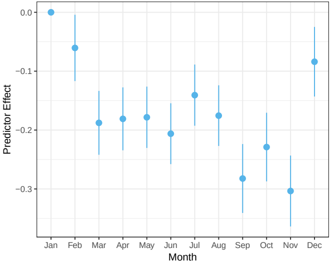
Joint estimation results

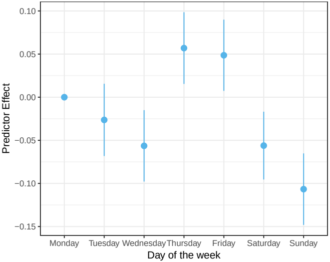
Covariate effects for the occurrence model ($\boldsymbol{\alpha}$)

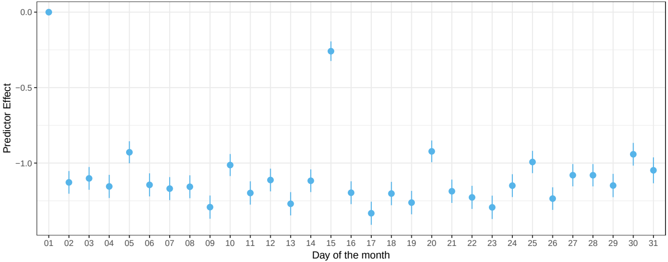


Joint estimation results

Covariate effects for reporting delay (β)







Joint estimation results

Covariate effects for reporting delay

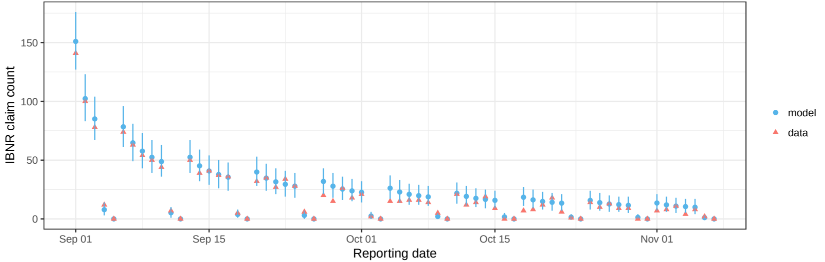
Reporting day probabilities in first week: (dow of occurrence date)

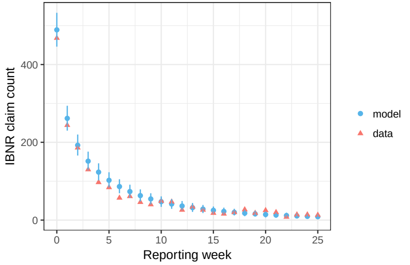
	wday						
dow	wday1	wday2	wday3	wday4	wday5	Saturday	Sunday
Monday	0.2600	0.4006	0.1638	0.0957	0.0744	0.0055	0.0000
Tuesday	0.2722	0.4131	0.1486	0.0900	0.0689	0.0072	0.0000
Wednesday	0.2699	0.3802	0.1739	0.0972	0.0700	0.0088	0.0000
Thursday	0.2639	0.4106	0.1464	0.0925	0.0695	0.0170	0.0000
Friday	0.2985	0.3003	0.1527	0.1006	0.0712	0.0767	0.0000
Saturday	0.4575	0.2045	0.1284	0.0843	0.0722	0.0531	0.0000
Sunday	0.4778	0.2232	0.1375	0.0890	0.0673	0.0051	0.0001

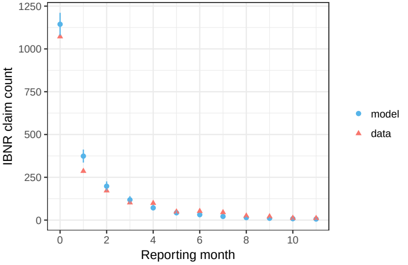
Reporting day probabilities in later weeks:

wday1	wday2	wday3	wday4	wday5	Saturday	Sunday
0.2886	0.2117	0.1829	0.1542	0.1429	0.0196	0.0000

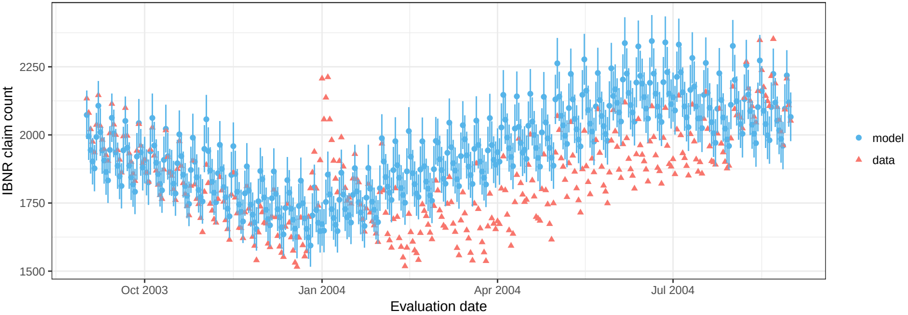
Joint estimation results







Joint estimation results



Predictions of the total IBNR claim counts for varying evaluation dates τ in between September 1, 2003, and August 31, 2004. Prediction intervals are constructed with 95% confidence level. The actual total IBNR claim counts are derived based on the full data set until August 2009.

Time change strategy to model reporting delay dynamics

Crèvecoeur et al., 2019

Time change strategy

Non parametric occurrence process

- Estimate the occurrence process non-parametrically.
- Likelihood is maximal when

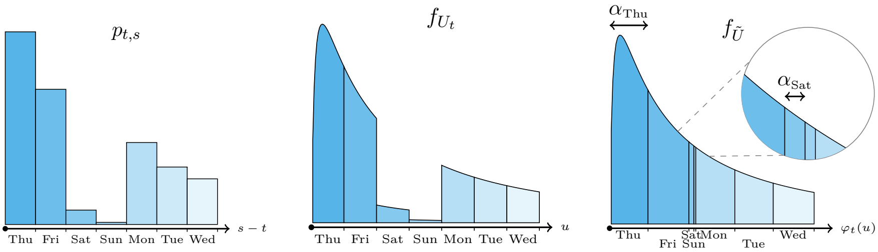
$$\lambda_t = \frac{\sum_{s=t}^{\tau} N_{t,s}}{\sum_{s=t}^{\tau} p_{t,s}} = \frac{N_t^R(\tau)}{p_t^R(\tau)}.$$

- Replacing λ_t

$$\ell(p; \boldsymbol{N}^R) = \sum_{t=1}^{\tau} \sum_{s=t}^{\tau} N_{t,s} \cdot \log(p_{t,s}) - \sum_{t=1}^{\tau} N_t^R(\tau) \cdot \log(p_t^R(\tau)) + \text{constants}.$$

Time change strategy

The idea pictured!



$$
\begin{aligned}
p_{t,s} &= \int_{s-t}^{s-t+1} f_{U_t}(u)du \\
&= F_{U_t}(s-t+1) - F_{U_t}(s-t).
\end{aligned}
$$

Time change strategy

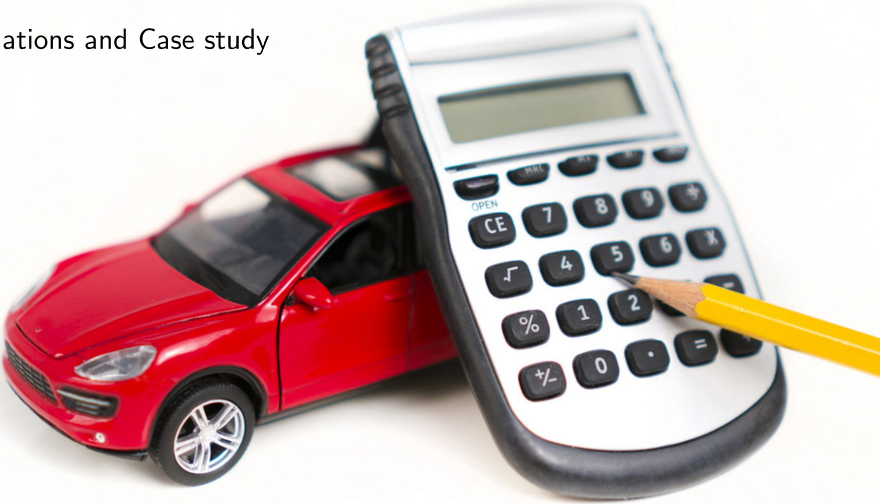
Structuring the reporting exposures

- Use a standard distribution for $\tilde{U}$.
- Explain the daily reporting exposures as a function of covariates:

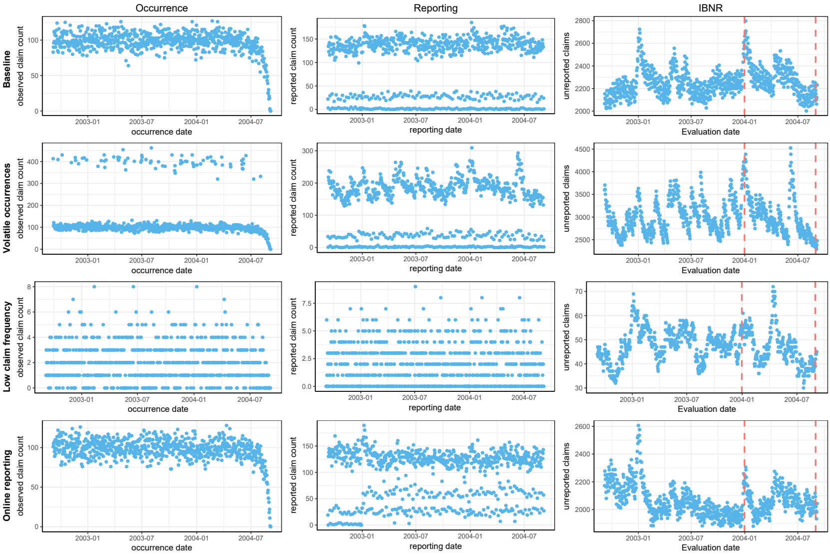
$$\alpha_{t,s} = \exp(\boldsymbol{x}_{t,s}^{'} \cdot \boldsymbol{\gamma}).$$

- Joint estimation of distribution $\tilde{U}$ and regression parameters to structure $\alpha_{t,s}$.

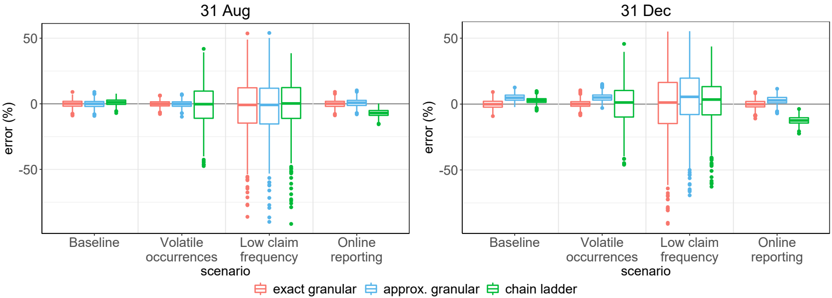
Simulations and Case study



Simulation study



Simulation study



Performance indicator

$$PE = 100 \cdot \frac{N^{IBNR} - \widehat{N^{IBNR}}}{N^{IBNR}}.$$

Case study

Results - first evaluation

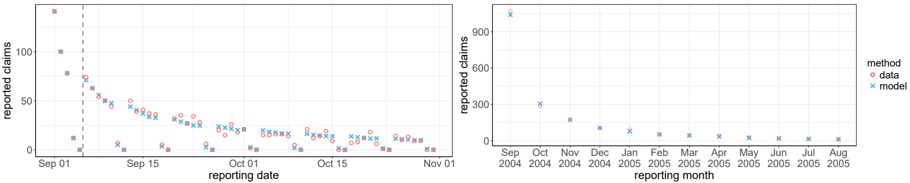
How many IBNR claims (on August 31, 2004) will be reported by August 31, 2009?

- Observed: 2049, claims
- Granular: 2012.7 claims
- Chain ladder: 2043.2 claims

Case study

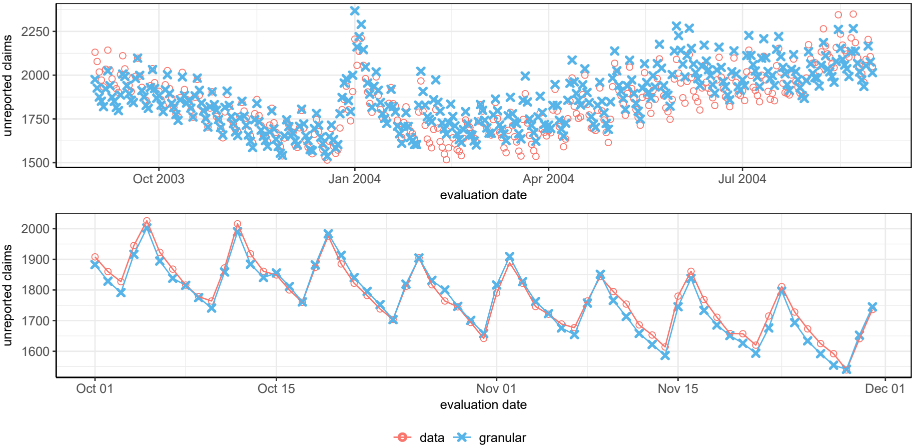
Results - second evaluation (only granular)

When are the claims that are IBNR (on August 31, 2004) reported?



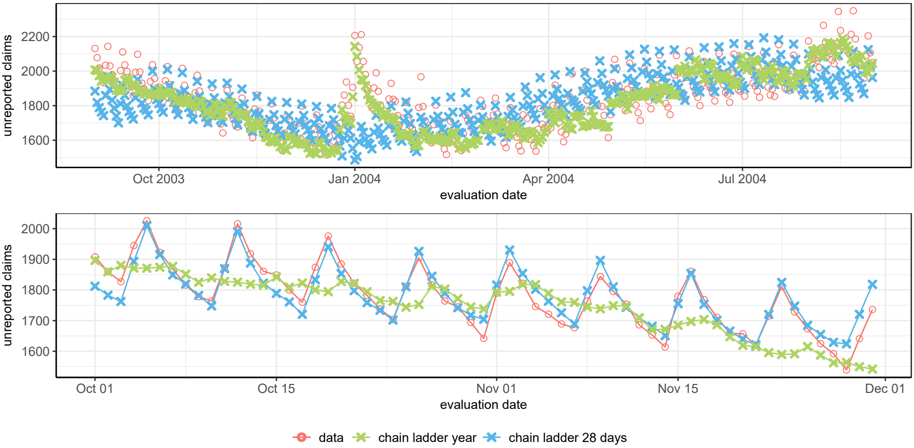
Case study

Results - third evaluation



Case study

Results - third evaluation



What else is there?

What else is there?
Recent developments

- Capture overdispersion and serial dependency in the occurrence process with a Cox process:
  - Avanzi, Wong & Yang (2016, IME) with a Shot Noise Cox Process.
  - Badescu, Lin & Tang (2016, IME) with a Hidden Markov Model.
- Focus on inhomogeneous marked Poisson process and reporting delay in continuous time, Verrall & Wüthrich (2016, Risks).