Recent developments in micro-level reserving

R in Insurance, Paris

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

June 2017

Telematics insurance

Actuarial pricing models

Stochastic mortality models

Micro-level claims reserving!

Mission statement

Launch a discussion of micro-level or granular data for claims reserving, and their features.

Sketch ongoing research on the modeling of IBNR claim counts.

Discuss recent developments in literature.

Mission statement

The talk is based on two papers (in progress) with:

Roel Verbelen, Jonas Crèvecoeur (present!) and Gerda Claeskens.

Mission statement

Mr M. H. Tripp, F.I.A.: Why do we throw away information? This question has already been hinted at, and needs reinforcing in the domain for thinking about in the future. I have never been keen on silos, and it is important to learn between disciplines. Looking at the life side of our profession, you realise that work like this takes place at policy level detail. If you look within the general insurance part of the actuarial profession, there is a body of thinking that has grown up around premium rating and a body of thinking that has grown up around reserving. Are we getting 'over-siloed'? Could aspects of the methodology and the thinking that has gone into using GLMs for premium rating be brought more into play when it comes to reserving, where, at present, we tend to use aggregated claims data? I wonder whether we are missing out on using information that is available from exposure descriptions and from the circumstances of individual claims. I know that the traditional response to this is that there is all too much variability, but, in attempts to remove heterogeneity from data and to try to find better for the future, I look for support in thinking this through.

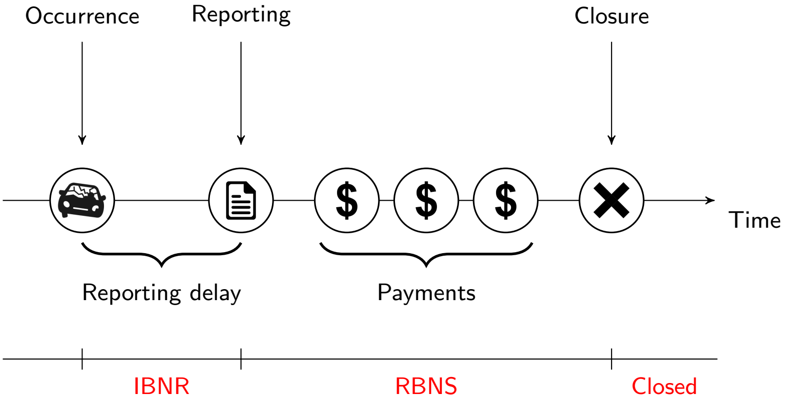
M. Tripp, F.I.A., Stochastic claims reserving - Abstract of the discussion, British Actuarial Journal, 2002.

Mission statement

Katrien's LinkedIn timeline, two days ago.

Introduction

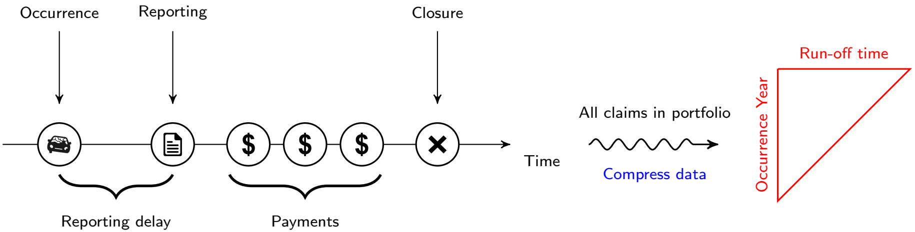
Development of a single claim



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:
    - robust (law of large numbers);
    - useful for accounting figures (audit);
    - established over years;
    - low data requirements and computational power.

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

Introduction

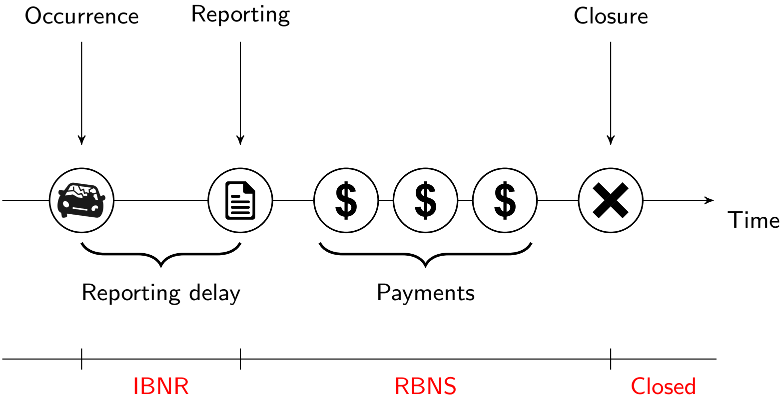
Pros and cons of aggregated approach

- Disadvantages of aggregating, pros of micro-level:
  - a lot of (detailed) data gets lost;
  - individual claims (types) prediction is not available (viz. pricing of products);
  - case management (and early warning) is not possible;
  - non-stationarity is difficult to detect.

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

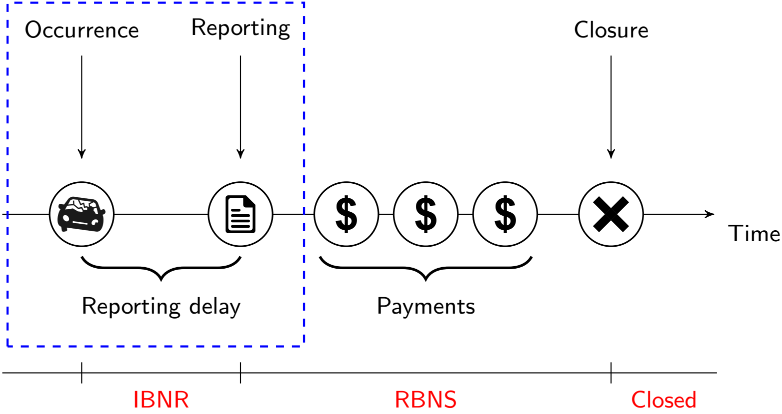
Research focus

IBNR claim counts



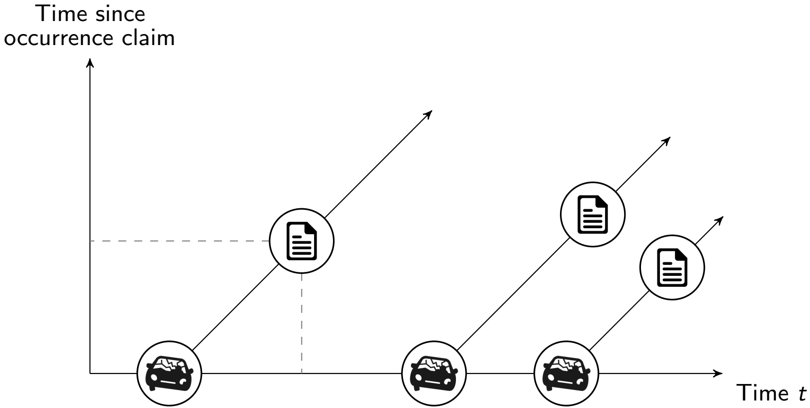
Research focus

IBNR claim counts



Research focus

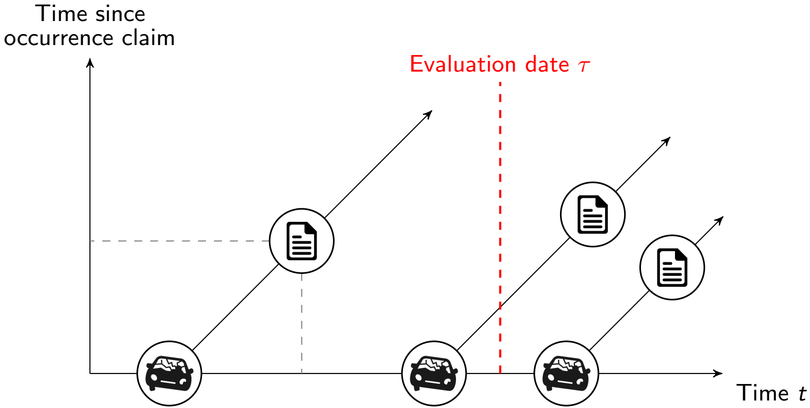
IBNR claim counts



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

Research 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 their reporting delay is right truncated
    (i.e. larger than $\tau - t$, with t occurrence date of accident)?
  - When will these IBNR claims be reported?
  - Study claim occurrences and reporting delay at daily level (=natural time unit).
  - Incorporate covariate information.

Basic notations

- N_t: the (total) number of claims that occurred on day t.
- $N_{t,d}$: the number of claims from day t that are reported after d days.
- Each claim has a reporting delay, thus

$$N_t = \sum_{d=0}^{\infty} N_{t,d},$$

where $d = 0$ when the claim is reported on the occurrence date.

Basic notations

A daily run-off triangle with reported claims

occurrence day	reporting delay (in days)				
	0	$\cdots$	$\tau - t$	$\cdots$	$\tau - 1$
1	N_{10}	$\cdots$	$N_{1,\tau-t}$	$\cdots$	$N_{1,\tau-1}$
$\vdots$					
t	N_{t0}	$\cdots$	$N_{t,\tau-t}$		
$\vdots$				IBNR	
τ	$N_{\tau 0}$				

A closer look at the micro-level data!

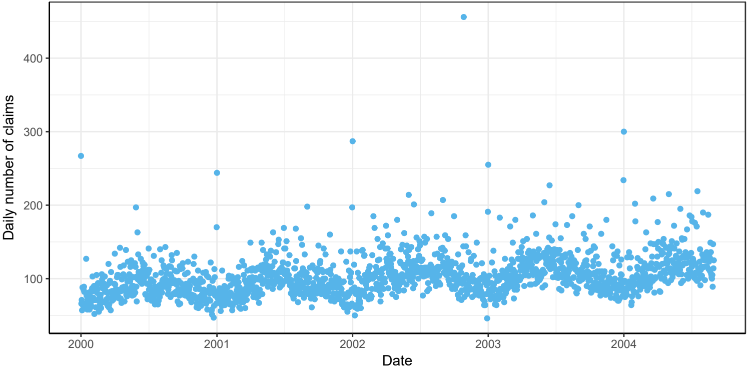
Case-study

Structure of the data

- Large European dataset of liability claims (from private individuals).
- Three essential variables *(for work on IBNR claim counts)*:
  - Occurrence date;
  - Reporting date;
  - Monthly earned exposure.
- Restrict our analysis to claims that have occurred between January 1, 2000 and August 31, 2004 ($= \tau$, the evaluation date).
- Remaining data until until August 2009 used for out-of-sample prediction.

Case-study

Exploratory analysis: occurrence



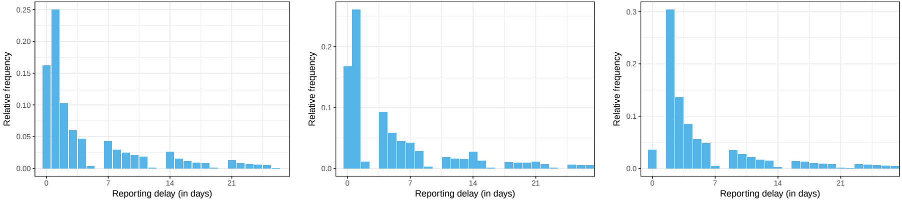
Number of claims from a specific occurrence date, reported before or at August 31, 2009; $\sum_{d=0}^{\tau-t} N_{t,d}$ with t from to January 1, 2000 to August 31, 2004 and τ is 31 Aug 2004.

Case-study

Exploratory analysis: reporting delay distribution

Weekly declining pattern in reporting delay + daily pattern within each week, depending on occurrence day of week.

(a) Monday **(b)** Thursday **(c)** Saturday

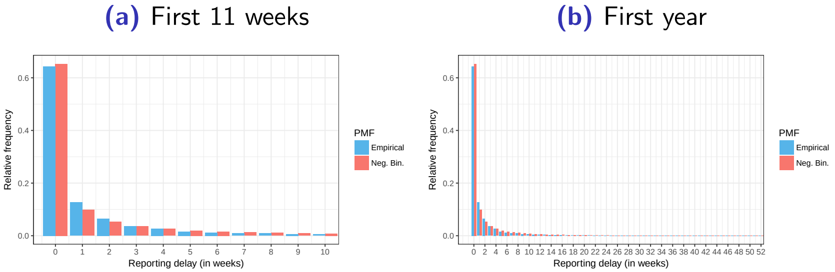


Empirical reporting delay distribution in the first 4 weeks
for claims that occurred on (a) Monday, (b) Thursday and (c) Saturday
between January 1, 2000 and August 31, 2004 and have been reported before August 31, 2009.

Case-study

Exploratory analysis: reporting delay distribution

Reporting delay in weeks: the number of weeks that elapses between occurrence and reporting of the claim.



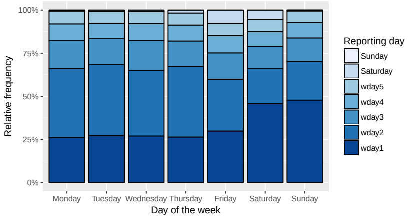
Empirical reporting delay distribution in weeks and its negative binomial approximation.
First 11 weeks in (a) and for the first year in (b).
Data on claims that occurred between January 1, 2000 and August 31, 2004 and have been reported before August 31, 2009.

Case-study

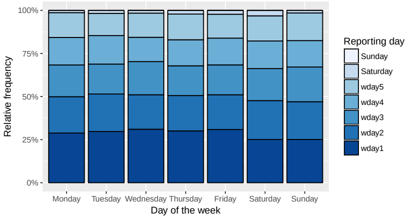
Exploratory analysis: reporting delay distribution

The reporting day probabilities model on which day a claim is reported within a given reporting week.

(a) First week



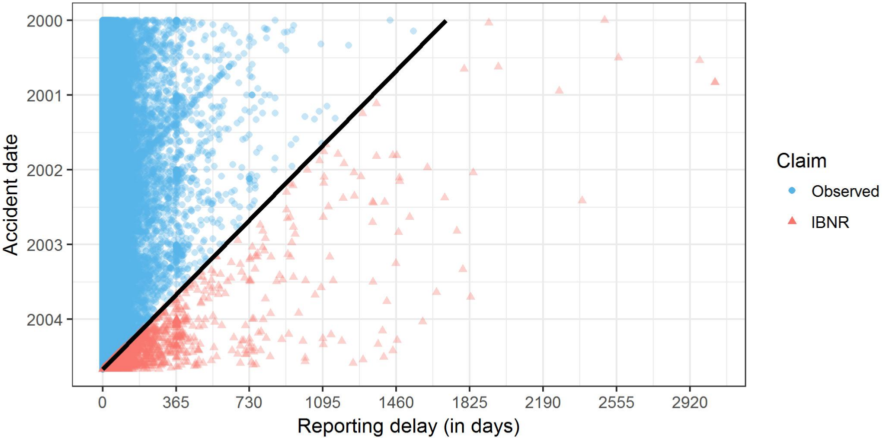
(b) From week 2 on



Empirical reporting delay day probabilities within a reporting week according to the day of the week of the occurrence date.
First reporting week in (a) and from the second reporting week onwards in (b).
Data on claims that occurred between January 1, 2000 and August 31, 2004 and have been reported before August 31, 2009.

Case-study

Exploratory analysis: reporting of claims



Daily run-off triangle of claims with occurrence dates between January 1, 2000 and August 31, 2004.
The black line indicates the evaluation date τ: August 31, 2004.

Statistical building blocks

The statistical model

Assumptions

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

$$N_t \sim \text{POI}(\lambda_t = e_t \cdot \exp(\boldsymbol{x}_t^{'}\boldsymbol{\alpha})),$$

where e_t is the exposure and $\boldsymbol{x}_t$ contains covariate information of day t.

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

$$N_t \sim \text{POI}(\lambda_t = e_t \cdot \exp(\boldsymbol{x}_t^{'}\boldsymbol{\alpha})),$$

where e_t is the exposure and $\boldsymbol{x}_t$ contains covariate information of day t.

(A2) Conditional on N_t, the claim counts N_{td} for $d = 0, 1, 2, \ldots$ are multinomially distributed with reporting delay probabilities p_{td}.

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

$$N_t \sim \text{POI}(\lambda_t = e_t \cdot \exp(\boldsymbol{x}_t^{'}\boldsymbol{\alpha})),$$

where e_t is the exposure and $\boldsymbol{x}_t$ contains covariate information of day t.

(A2) Conditional on N_t, the claim counts N_{td} for $d = 0, 1, 2, \ldots$ are multinomially distributed with reporting delay probabilities p_{td}.

Combining $(A1)$ and $(A2)$

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

The statistical model

The likelihood

- We observe the upper triangle

$$\boldsymbol{N}^R = \{N_{td} \mid t \leq \tau, t + d \leq \tau\}$$

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

- Log-likelihood of observed data: difficult to optimize (due to $\star$)

$$\ell(\boldsymbol{\lambda}, \boldsymbol{p}; \boldsymbol{N}^R) = \sum_{t=1}^{\tau} \left(\underbrace{-\lambda_t \sum_{d=0}^{\tau-t} p_{t,d}}_{(\star)} + \log(\lambda_t) \sum_{d=0}^{\tau-t} N_{t,d} + \sum_{d=0}^{\tau-t} N_{t,d} \log(p_{t,d}) - \sum_{d=0}^{\tau-t} \log(N_{t,d}!) \right).$$

Parameter estimation

The complete data likelihood

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

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

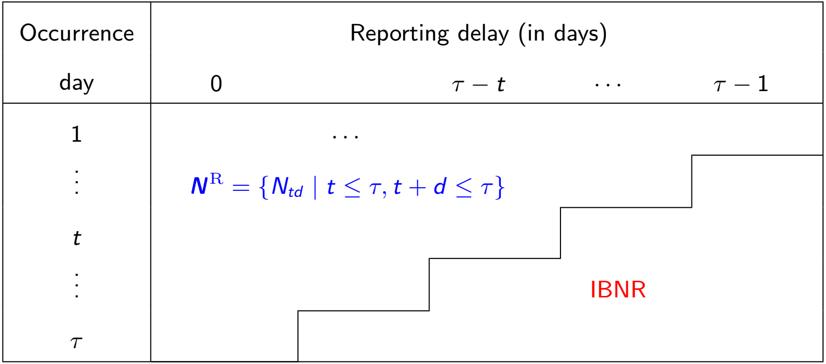
Then the likelihood of the complete data becomes:

$$\ell_c(\boldsymbol{\lambda}, \boldsymbol{p}; \boldsymbol{N}) = \sum_{t=1}^{\tau} \left(-\lambda_t \sum_{d=0}^{\infty} p_{t,d} + \log(\lambda_t) \sum_{d=0}^{\infty} N_{t,d} + \sum_{d=0}^{\infty} N_{t,d} \log(p_{t,d}) - \sum_{d=0}^{\infty} \log(N_{t,d}!) \right).$$

which splits into occurrence process and reporting delay likelihoods!

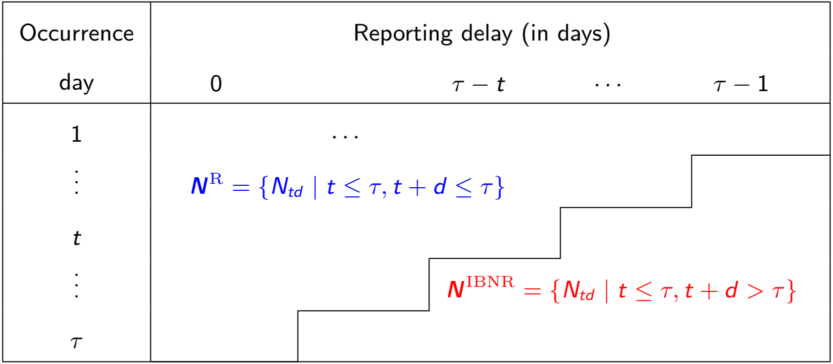
Parameter estimation

EM algorithm - key idea



Parameter estimation

EM algorithm - key idea



Complete data $\boldsymbol{N} = \boldsymbol{N}^{\mathrm{R}} \cup \boldsymbol{N}^{\mathrm{IBNR}} = \{N_{td} \mid t \leq \tau, d \geq 0\}$;

Iterate between an expectation step (E-step) and maximization step (M-step).

Joint estimation of occurrence and reporting delay

A model for occurrences

- We propose a Poisson regression model:

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

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.

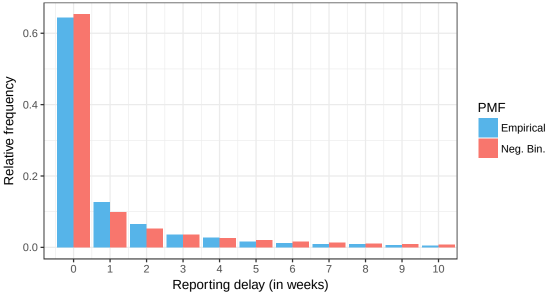
Joint estimation of occurrence and reporting delay

A model for reporting delay

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

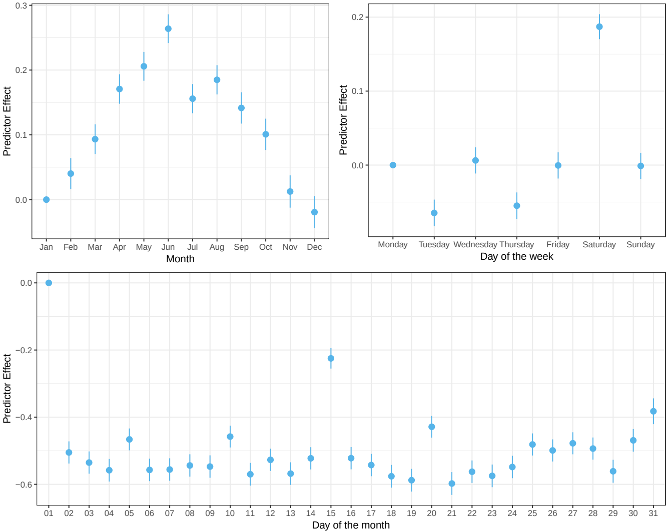
$$p^W_{t,w} = \frac{\Gamma(\phi + w)}{w!\Gamma(\phi)} \cdot \frac{\phi^\phi \mu_t^w}{(\phi + \mu_t)^{\phi + w}},$$

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



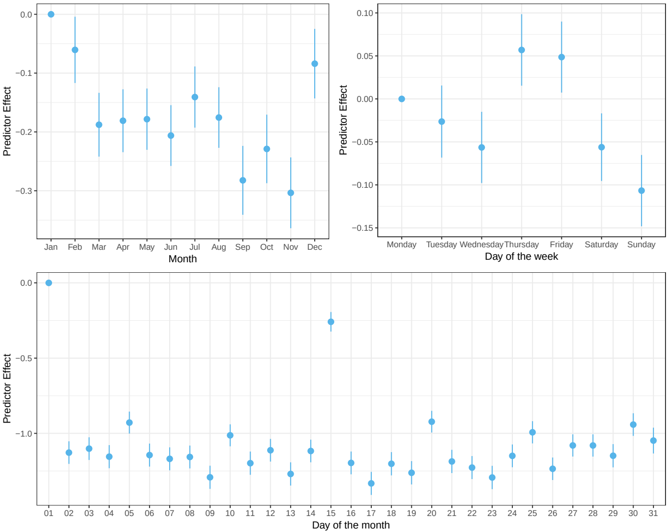
Results

Covariate effects for the occurrence model



Results

Covariate effects for reporting delay



Results

Covariate effects for reporting delay

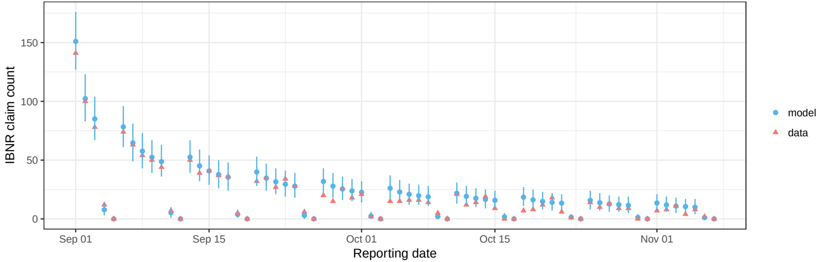
Reporting day probabilities in first week:

	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

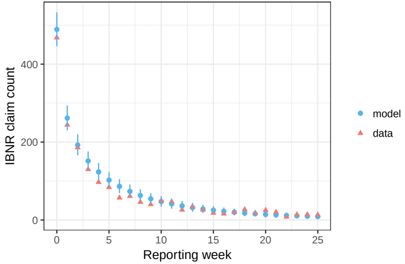
Reporting day probabilities in later weeks:

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

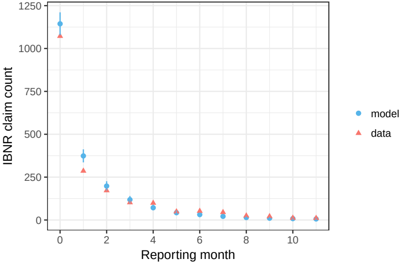
Results



IBNR claim count
150
100
50
0
Sep 01
Sep 15
Oct 01
Oct 15
Nov 01
Reporting date
model
data



IBNR claim count
400
200
0
0
5
10
15
20
25
Reporting week
model
data



IBNR claim count
1250
1000
750
500
250
0
0
2
4
6
8
10
Reporting month
model
data

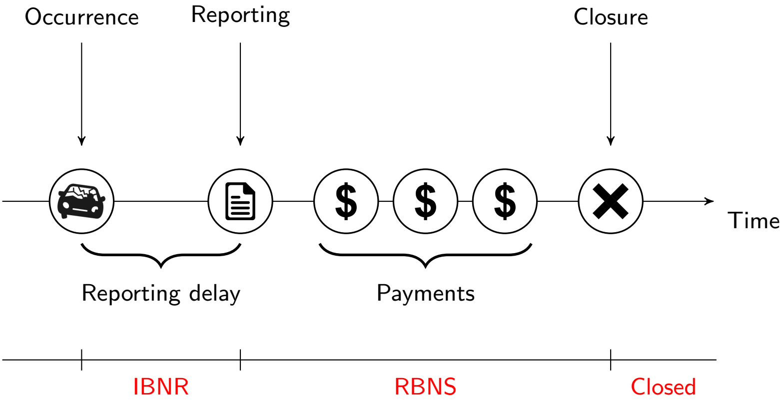
What else is there?

What else is there?

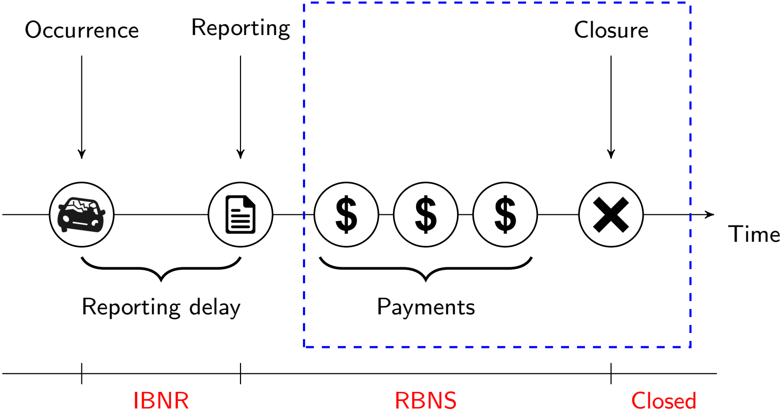
Recent developments

- 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).
- Crèvecoeur, Antonio & Verbelen on calendar effects in reporting of claims.

What else is there?



What else is there?



What else is there?

Research questions

- More research questions with focus on micro-level data?
  - What is the number of payments for an RBNS claim?
  - What is the size of these future payments?
  - When do we make these payments?
  - When will the claim settle or close?

What else is there?

Recent developments

- Wüthrich (2017, SSRN) on machine learning in individual claims reserving.
- A multi-state approach in Antonio, Godecharle & Van Oirbeek (2016, SSRN) and Gerards, N. & Antonio, K. (2017).


Conclusions

Wrap-up

- The message is not that chain-ladder should disappear!
- Take home messages:
  - the presented methods increase insight in the available data and the dynamics in claim development patterns;

    (fits within the increasing interest in data analytics);

    (claim and policy characteristics can be taken into account).
  - caution: many choices involved, should be done with care!

More information

For more information, please visit:

- LRisk website, www.lrisk.be;
- my homepage, www.econ.kuleuven.be/katrien.antonio.

Thanks to

Ageas Continental Europe

Argenta