Bridging the gap between pricing and reserving with an occurrence and development model for non-life insurance claims

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

November 6, 2023

Presenter + Authors

Jonas Crevecoeur

Katrien Antonio

Stijn Desmedt

Alexandre Masquelein

Paper appeared in ASTIN Bulletin: The Journal of the IAA, 53(2), 2023.

Contributions

Contributions of the paper

1. Reflect on inconsistencies between using **actual observations** next to **best estimates** in insurance pricing data sets.

2. Model both occurrence + reporting and development of claims and use the **combined model for pricing and reserving**, hence: attempt to **bridge** two key actuarial tasks.

3. Demonstrate the approach on a **portfolio from insurance** as well as **reinsurance**, where delays (in reporting and settlement) are significant.

Related literature

Our work is related to contributions:

- in non-life **insurance pricing** with machine learning methods (cfr. infra)
- in non-life **claims reserving** using the development history of individual claims, e.g., Larsen (2007, ASTIN), Wüthrich (2018, SAJ), Delong et al (2022, SAJ) and infra
- in **reinsurance**, with Albrecher et al. (2017, Wiley) and Albrecher & Bladt (2022, preprint).

Non-life insurance pricing

Non-life insurance pricing 101

- Denote for policy i in a given policy period:
  - e_i: exposure-to-risk
  - N_i: number of claims filed during the exposure period
  - L_i: total loss amount reported during the exposure period.
- The **technical, pure premium** π_i:

$$\pi_i = \mathbb{E}\left[\frac{L_i}{e_i}\right] \overset{indep.}{=} \mathbb{E}\left[\frac{N_i}{e_i}\right] \times \mathbb{E}\left[\frac{L_i}{N_i} \mid N_i > 0\right] = \underbrace{\widehat{\text{Freq}_i}}_{\textbf{frequency}} \times \underbrace{\widehat{\text{Sev}_i}}_{\textbf{severity}}$$

- Build predictive models f(**risk factors**) for frequency and severity, respectively.

Our lab's recent work on insurance pricing analytics

[Henckaerts et al., 2018]

SAJ

[Henckaerts et al., 2021]

NAAJ

[Henckaerts et al., 2022]

Expert Syst. Appl.

[Henckaerts & Antonio, 2022]

IME

github/henckr/distRforest

github/henckr/maidrr

These contributions assume a **complete**, historical data set, with observations on:

- total number of claims N_i reported per policy i, during given exposure e_i, with characteristics $\mathbf{x}_i$
- ultimate claim size $L_i = Y_{i1} + \ldots + Y_{in_i}$, with the Y_{ij} the ultimate individual claim sizes.

However, pricing data are often **incomplete** and **preprocessing steps** are put into place!

First, examples of preprocessing steps to put a **complete** pricing data set together:

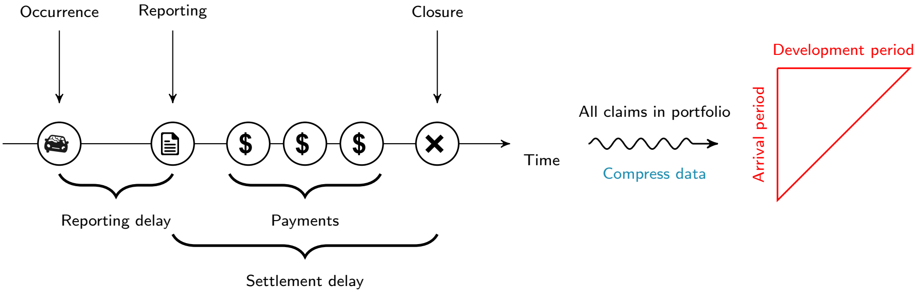
- (frequency) ignore unreported claims
- (severity) only consider settled claims, hence: ignore right-censored, open claims
- (severity) replace the future development of open claim with zero or with a best estimate constructed based on expert opinion or via data-driven methods.

Second, predictive models calibrated for severity often treat these best estimates as actual observations.

However, many other properties of the loss r.v. (e.g., the variance) are not preserved when treating best estimates as actual observations (cfr. Section 1 in our paper).

Non-life insurance reserving

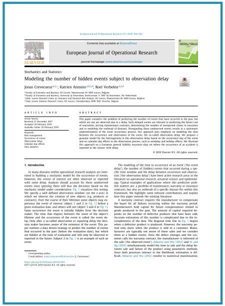
Non-life insurance reserving 101



We typically **aggregate** the data from the time line into a run-off triangle.

Our lab's recent work on non-life reserving analytics

[Crevecoeur et al., 2019]

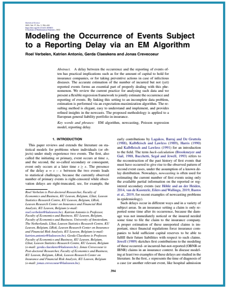

European Journal of Operational Research

Modeling the number of hidden events subject to observation delay

Jonas Crevecoeur, Katrien Antonio, Roel Verbelen

EJOR

[Verbelen et al., 2022]


Modeling the Occurrence of Events Subject to a Reporting Delay via an EM Algorithm

Roel Verbelen, Katrien Antonio, Gerda Claeskens and Jonas Crevecoeur

Stat Science

[Crevecoeur et al., 2022]


Insurance: Mathematics and Economics

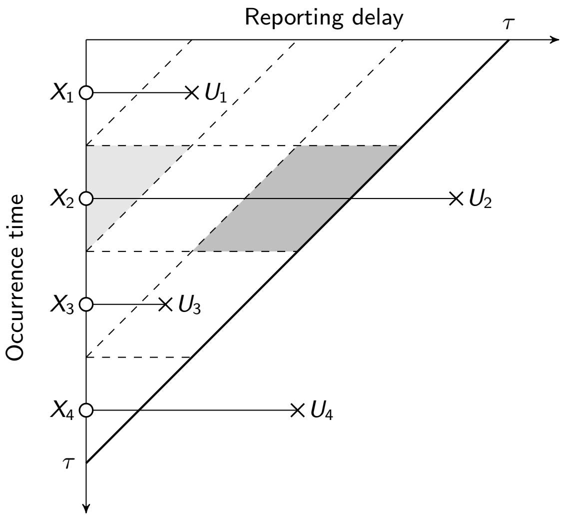
A hierarchical reserving model for reported non-life insurance claims

Jonas Crevecoeur, Jens Robben, Katrien Antonio

IME

IBNR reserving

From continuous time setting ...



IBNR reserving

. . . to granular runoff triangles

Occurrence period	Reporting delay 0	$\cdots$	$\tau - t$	$\cdots$	$\tau - 1$
1	N_{10}	$\cdots$	$N_{1,\tau-t}$	$\cdots$	$N_{1,\tau-1}$
$\vdots$					
t	N_{t0}	$\cdots$	$N_{t,\tau-t}$		
$\vdots$					
τ	$N_{\tau 0}$				

An **incomplete two-way contingency table**: the run-off triangle in actuarial science or reporting triangle in epidemiology.

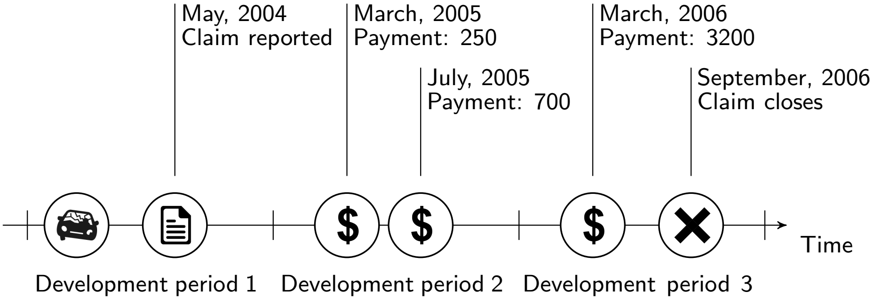
The dimension of the triangle depends on the **granularity of the discretization**!

In Verbelen et al. (2022, Stat Science) we propose:

- N_t for $t = 1, \ldots, \tau$ are **independently Poisson distributed** with intensity $\lambda_t = \exp(\boldsymbol{x}'_t \boldsymbol{\alpha})$, where $\boldsymbol{x}_t$ is a covariate vector corresponding to occurrence period t and $\boldsymbol{\alpha}$ is a parameter vector

- conditional on N_t, the N_{td} for $d = 0, 1, 2, \ldots$, are **multinomially distributed** with probabilities $p_{td} = p_{td}(\boldsymbol{\theta}, \boldsymbol{x}_{td})$, a well-defined reporting probability distribution

- use **EM** algorithm to optimize the likelihood in presence of missing data.



- Claim reported
- Payment: 0

- Payment: 950

- Payment: 3200
- Claim closed

A hierarchical reserving model for RBNS claims

Crevecoeur et al. (2022) - layers

- Index the individual claims by k and the development periods by j.
- Our approach is **modular** or **layered**:
  - $\boldsymbol{x}_k$ denotes the (observed, static) claim information available at the end of the first development period, i.e. the reporting period

    e.g. cause of claim, policy(holder) covariates, initial case estimate
  - $\boldsymbol{U}_k^j$ is the vector with **claim k's updated information in development period j**

    depends on portfolio at hand, e.g. $\boldsymbol{U}_k^j = (C_k^j, P_k^j, Y_k^j)$ with a settlement indicator C_k^j, a payment indicator P_k^j and payment size Y_k^j.

A hierarchical reserving model for RBNS claims

Crevecoeur et al. (2022) - predictive model per layer

- Fit **layer-specific predictive model** (e.g., GLM, Gradient Boosting Machine or a Neural Network):

$$f\left(U_{k,l}^{j} \mid \boldsymbol{U}_k^1, \ldots, \boldsymbol{U}_k^{j-1}, U_{k,1}^{j}, \ldots, U_{k,l-1}^{j}, \boldsymbol{x}_k\right),$$

  with

  - time dynamic, layered **hierarchical** structure for $\boldsymbol{U}_k^j$
  - **static** (via $\boldsymbol{x}_k$) as well as **dynamic** features (via the update vectors of previous periods 1 to $j-1$ or proceeding layers 1 to $l-1$).

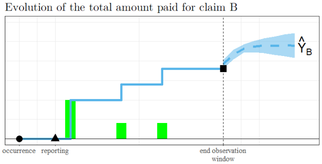
- Use the layer-specific predictive models to predict future development of reported claims.

An occurrence and development model for non-life insurance claims

- **Occurrence** model:
  - specify the occurrence + reporting model (cfr. IBNR reserving) **at level of individual policies i**
  - $N_i \sim \text{POI}(e_i \cdot \lambda_i)$ with λ_i a function of observed policy characteristics $\boldsymbol{x}_i$
  - from the N_i occurred claims, the reported claims N_{ij} are multinomially distributed with reporting probabilities $p_{ij}(\boldsymbol{x}_i)$.
- As such, we
  - transfer the ideas from Verbelen et al. (2022) to the **individual policy level**, and
  - can **estimate the number of unreported claims** at policy level in a data driven way, useful for pricing and reserving.

- A hierarchical **development model** for reported claims:
  - hierarchical reserving model for RBNS claims (cfr. RBNS reserving in Crevecoeur et al., 2022)
  - layers tailored to portfolio, e.g., in reinsurance case-study our development model distinguishes between I_k (in reporting period) and U_k^j (for development periods since reporting)
  - takes policy and claim characteristics (at reporting) as well as claim development history into account.
- This development model allows to
  - model the development of open claims in future development periods (reserving),
  - estimate the ultimate severity of claims (pricing).

Pricing and reserving with the ODM

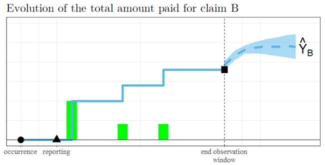


Let's focus on **pricing**:

- claim **frequency** estimates adjusted for unreported claims follow from ODM
- claim **severity**:
  - simulate ultimate claim sizes **from ground-up** for a given policy with characteristics $\boldsymbol{x}$
  - simulate n_{path} paths of the future development of **open claims**, then **fit a severity distribution** $f_Y(.)$ by maximizing

$$\mathcal{L}^{\texttt{ODM}}(f_Y) = \sum_{k=1}^{m} \left\{ \texttt{settled}_k \cdot \log(f_Y(Y_k)) + (1 - \texttt{settled}_k) \cdot \frac{1}{n_{\texttt{path}}} \cdot \sum_{p=1}^{n_{\texttt{path}}} \log(f_Y(Y_{k,p})) \right\}.$$

Pricing and reserving with the ODM



Let's focus on **reserving**: $\mathcal{R} = \mathcal{R}^{\text{IBNR}} + \mathcal{R}^{\text{RBNS}}$

- we estimate the IBNR reserve via

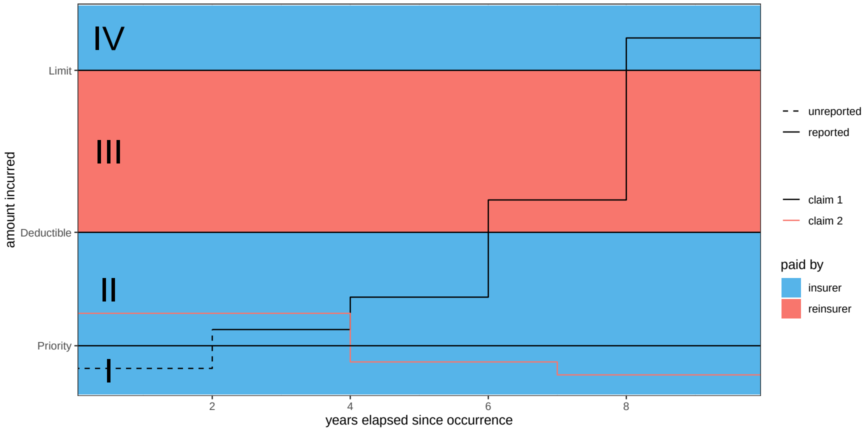
$$E(\mathcal{R}^{\text{IBNR}}) \quad = \quad \sum_{i} \sum_{j=\tau_i+1}^{d} E(N_{ij}) \cdot E(Y_i | \text{rep.delay} = j)$$

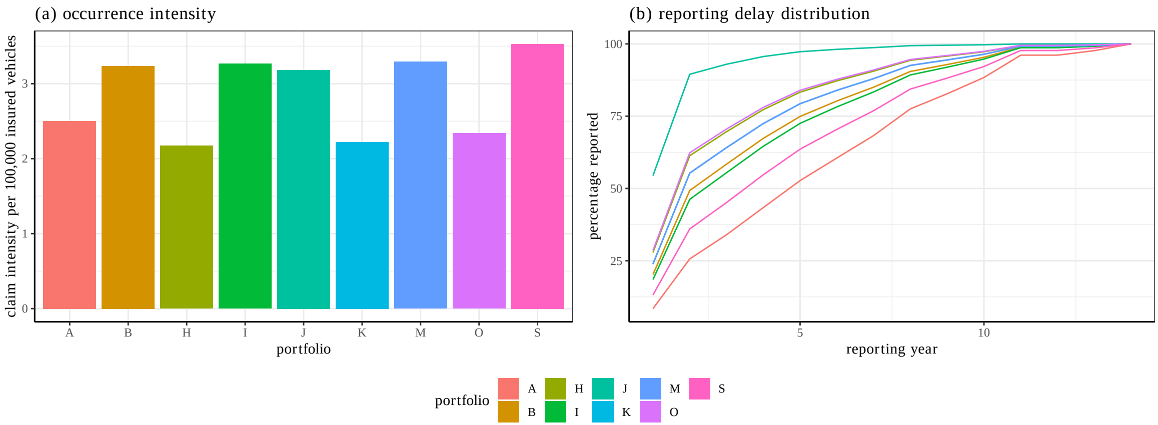
- for $\mathcal{R}^{\text{RBNS}}$ we use the hierarchical reserving model and simulate the joint evolution of all open claims.

Case-study

- 4 277 large motor insurance claims with occurrence in 2000-2017 and their detailed development.
- Reported by 21 insurance companies (`A - U`), indexed with i:
  - exposure $e_{i,t}$ is number of vehicles covered by company i in year t
  - **reporting priority** P_i of company i.
- For each claim, indexed with k:
  - occurrence year, year of reporting to reinsurer, settlement year
  - **paid and incurred** amount in every development year since reporting.

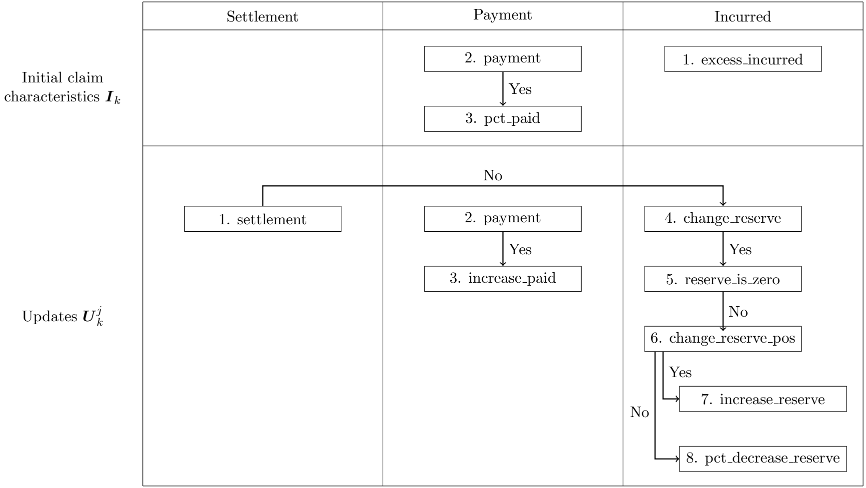
Portfolio of excess-of-loss contracts





(a) Estimated number of claims exceeding the priority of 750 000 per 100 000 insured vehicles in 9 portfolios and (b) fitted reporting delay distribution per portfolio, where reporting of a claim captures the first exceedance of the incurred claim amount above the priority of 750 000.

The hierarchical development process: layers

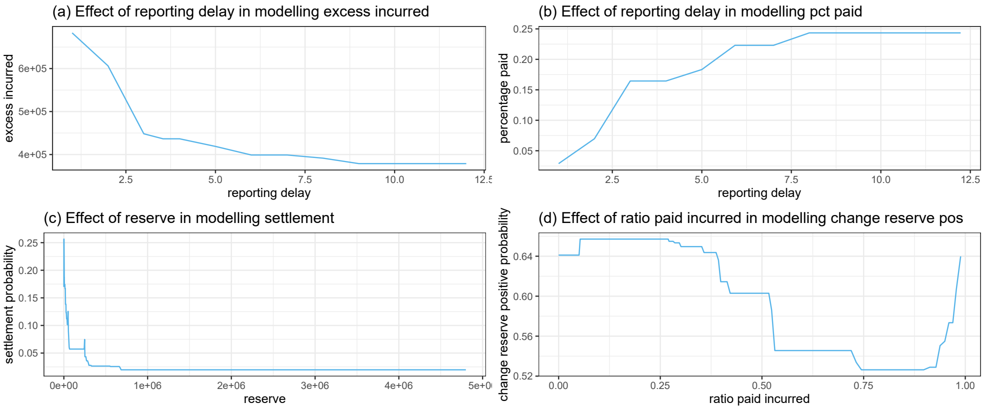


The hierarchical development process: covariates

covariates \ layers	initial status I_k: excess incurred	initial status I_k: payment	initial status I_k: percentage paid	updates U_k^j: settlement	updates U_k^j: payment	updates U_k^j: increase paid	updates U_k^j: change reserve	updates U_k^j: change reserve pos	updates U_k^j: reserve is zero	updates U_k^j: increase reserve	updates U_k^j: pct decrease reserve
portfolio	62	44	37	20	18	19	56	26	2	25	27
years elapsed since reporting	×	×	×	4	5	4	2	5	0	4	5
reporting delay	38	18	32	3	3	3	2	3	0	2	3
increase paid	×	×	×	10	29	14	13	16	2	18	15
total amount paid	×	×	×	9	11	15	7	14	1	10	15
reserve	×	×	×	34	22	29	11	16	53	32	18
ratio paid incurred	×	×	×	20	11	13	9	20	42	9	17
incurred	×	38	31	×	×	×	×	×	×	×	×
settlement	×	×	×	×	1	3	×	×	×	×	×

× covariate not included

Tree-based Gradient Boosting Machine (GBM) for each layer.



MTPL reinsurance data set: selected partial dependence plots in the hierarchical claim development model.

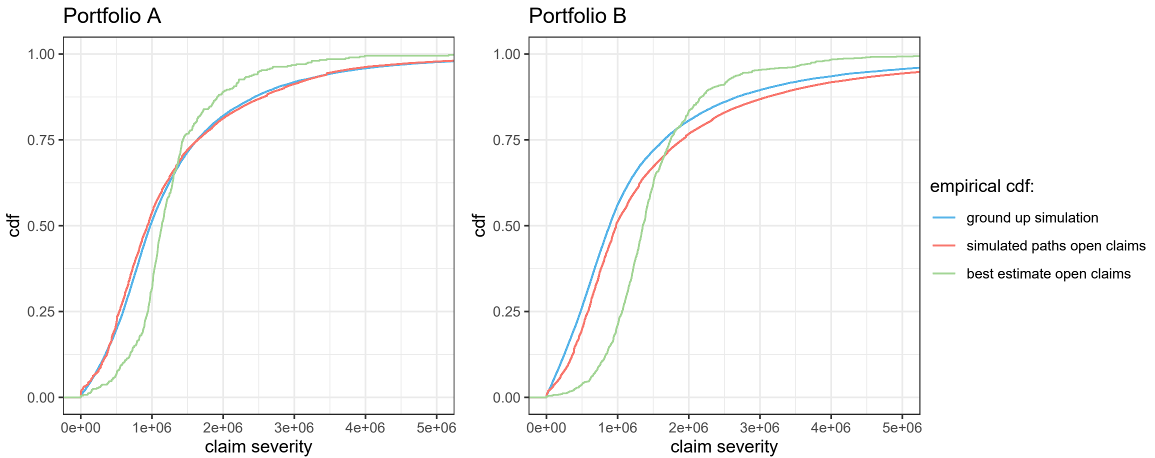
An **excess-of-loss reinsurance contract** covering loss from individual claim exceeding a **deductible** $D = 2\ 500\ 000$ up to a **limit** $L = 5\ 000\ 000$.

The pure premium π^P is

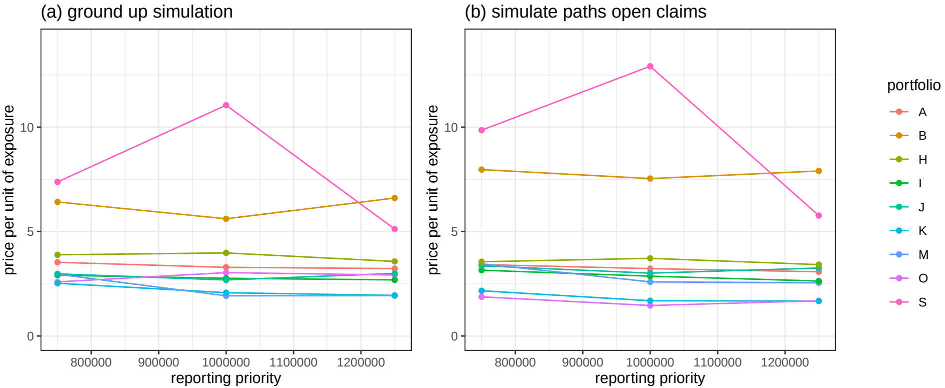
$$\pi^P = E(N^P) \cdot E(((Y^P \wedge L) - D)_+),$$

with:

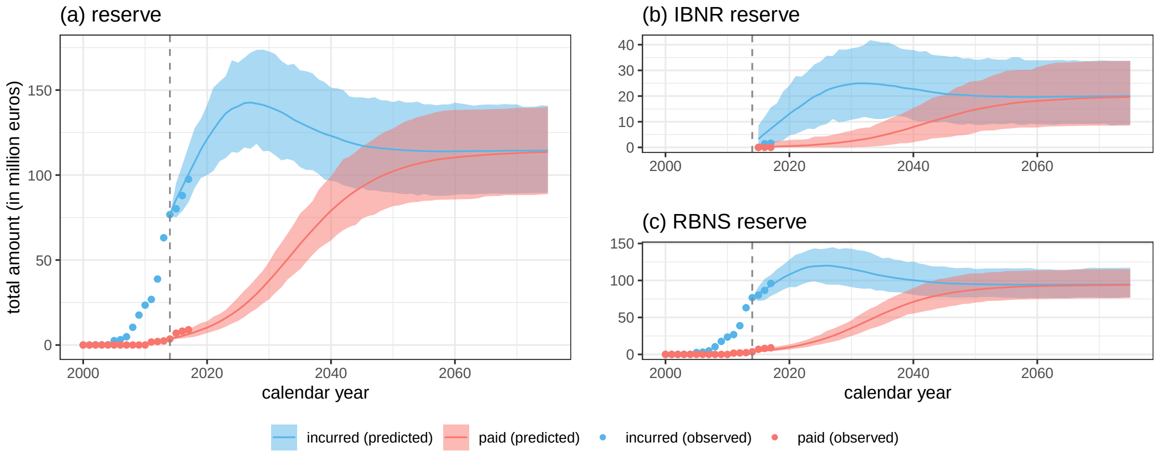
- N^P and Y^P the frequency and severity, respectively, of claims reported above a priority P
- $(Y^P \wedge L)$ the minimum of Y^P and L, and $(Y - D)_+$ is $Y - D$ if $Y \geq D$ and zero otherwise.



Simulated severity distribution of MTPL claims from portfolio A and B above a reporting priority of 750 000. For each portfolio, we show the severity distribution based on 20 000 from ground up simulated new claims (blue), observed claims complemented with 200 simulated paths per open claim (red) and observed claims where open claims have been replaced by best estimates (green).



Technical price per insured vehicle for an excess-of-loss contract with deductible $D = 2,500,000$ and limit $L = 5,000,000$. Claim severity is estimated based on (a) simulating 20 000 new claims from ground up and (b) observed claims complemented with 200 simulated paths per open claim. Prices are computed at reporting priorities: 750 000, 1 000 000 and 1 250 000.



Evolution of the aggregated amount incurred and paid between 2 500 000 and 5 000 000 for claims that occurred between 2000 and 2014. The (a) total reserve is split into the (b) IBNR and (c) RBNS reserve. 95% prediction intervals are shown for these amounts, with solid lines indicating expected values. Points indicate for calendar years 2015-2017 the actual out-of-time observations.

More information

For more information, please visit:

- journal website, and hirem package for R
- LRisk website, www.lrisk.be
- my homepage https://katrienantonio.github.io.

Special thanks to

- the organizers of the seminar
- the collaboration with Argenta and QBE Re on reserving analytics.