LRISK

Individual Reserving with Claim Specific Covariates

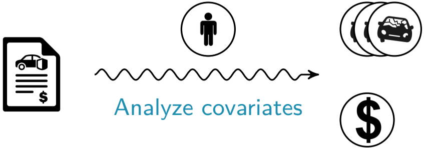
Jonas Crevecoeur
Joint work with Katrien Antonio

Insurance Data Science, Zürich
June, 2019

Introduction

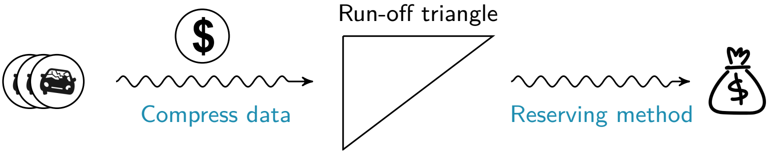
Similarities in pricing and reserving

Pricing



Analyze covariates to price individual contracts

Classical reserving



Aggregate data into a runoff triangle to calculate the total reserve

Advantages of compressing the data

Advantages of the aggregated approach:

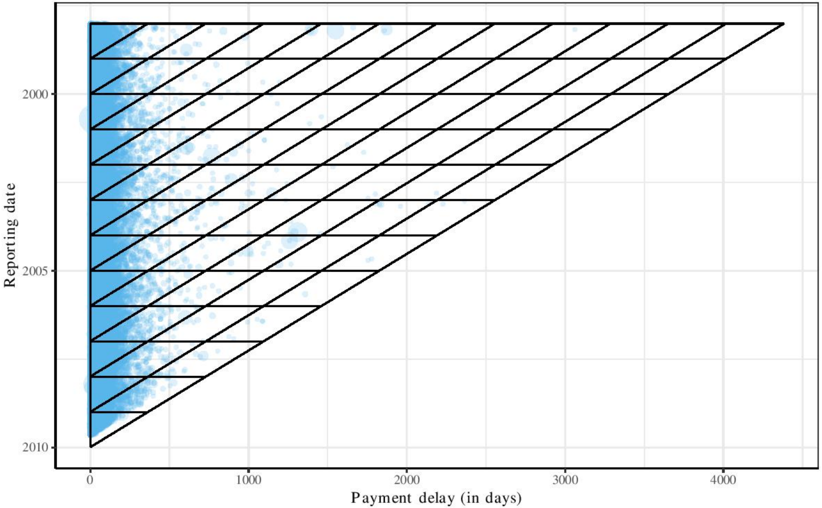
- low data requirement and computational power;
- simple to implement;
- easy to interpret;
- . . . .

Goal: Include data insights,
while preserving the advantages of the aggregate approach!

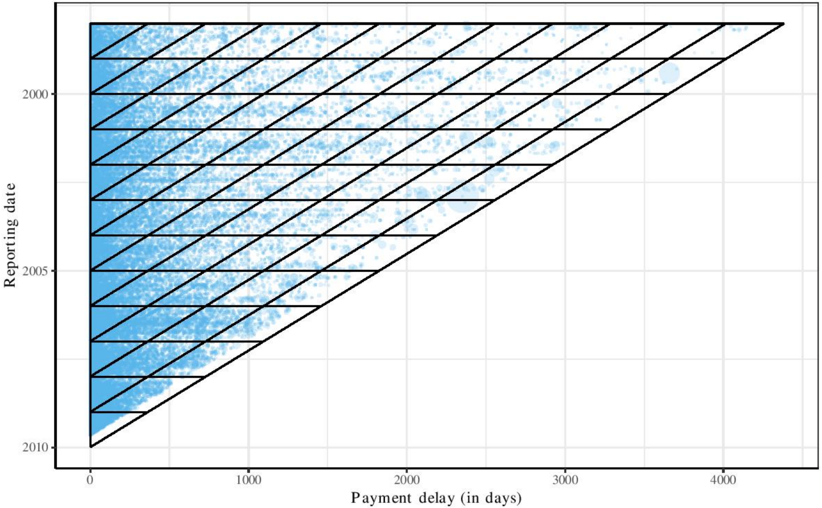
Hybrid approach: Combine aggregate and micro level methodology.

Introduction

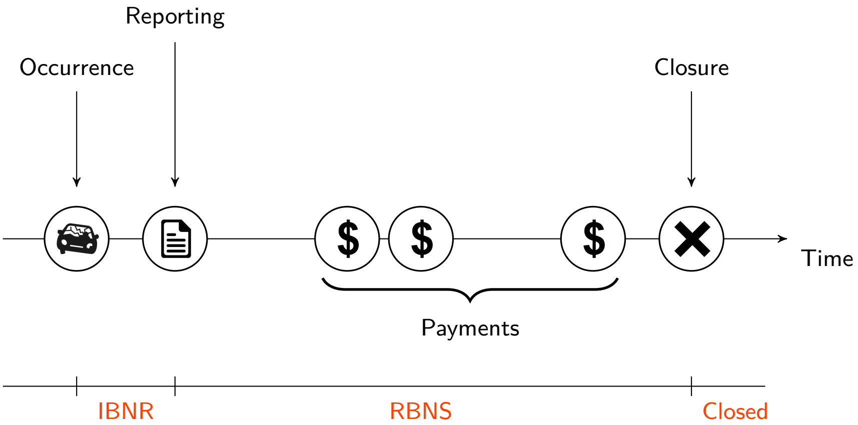
Aggregated and individual data



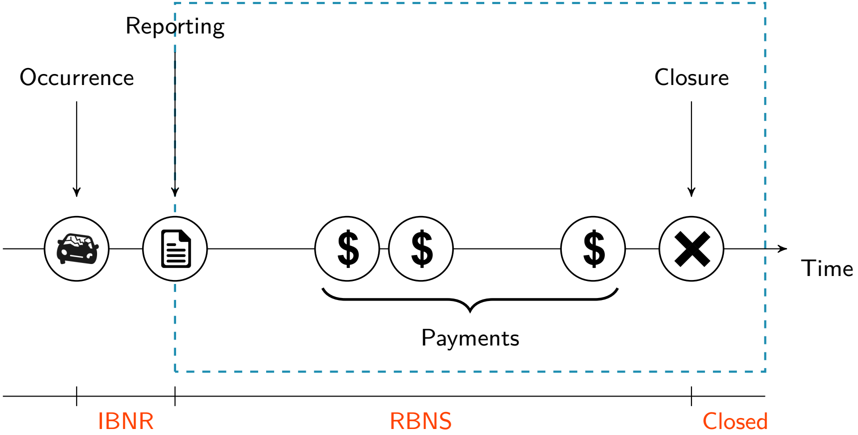
Aggregated and individual data



Development of a single claim

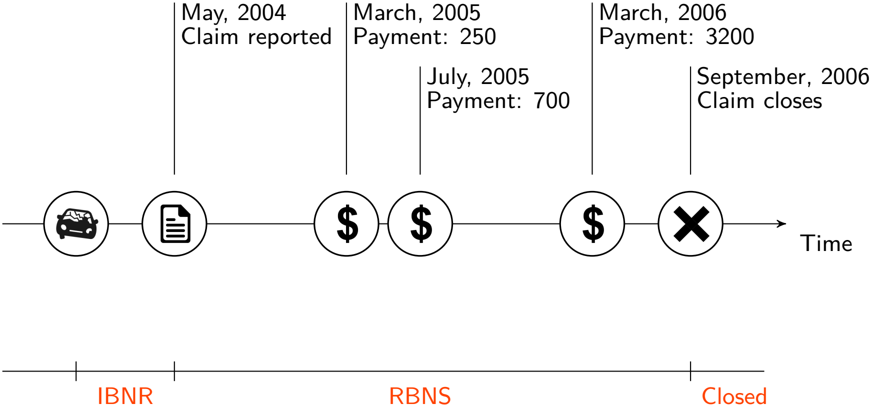


Development of a single claim



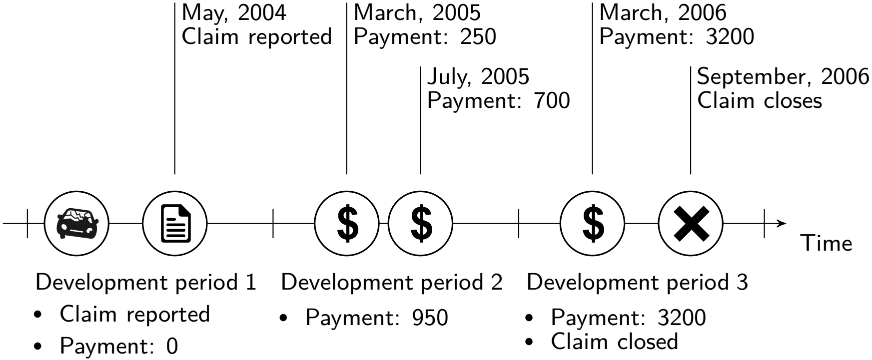
We focus on the modeling of the RBNS reserve
After the reporting date claim covariates are available.

Development of a single claim



The insurer registers for each claim the event dates and the payment sizes.

Development of a single claim



We discretize the data by development year.

We index the individual claims by k and the development years by j.

In each development year, we observe:

- C_{kj}: Closure indicator
- P_{kj}: Payment indicator
- Y_{kj}: Payment size

For each claim, we observe:

- Policyholder information
- Policy characteristics
- Claim covariates

Data management

We construct our reserving data set by combining policy and claims data.

A quick glance at both data sets:

Policy data

```
## # A tibble: 432080 x 2
##     policy_nr policy_covariates
##         <dbl> <list>
## 1 30000037738 <tibble [1 x 106]>
## 2 30000124129 <tibble [1 x 106]>
## 3 30000125846 <tibble [1 x 106]>
## 4 30000194251 <tibble [1 x 106]>
## 5 30000265383 <tibble [1 x 106]>
```

Claims data

```
## # A tibble: 56698 x 2
##     policy_nr  accident_nr claim_covariates
##         <dbl>        <dbl> <list>
## 1 30000037738 898000390380 <tibble [1 x 15]>
## 2 30000124129 898001131523 <tibble [1 x 15]>
## 3 30000125846 898001053014 <tibble [1 x 15]>
## 4 30000194251 898000308942 <tibble [2 x 15]>
## 5 30000194251 898000446055 <tibble [1 x 15]>
```

Data management

First discretize the claims data

```
claim_data <- claim_data %>%
  group_by(accident_nr, dev_year) %>%
  mutate(close = sum(close) >= 1,
         payment = sum(payment) >= 1,
         size = sum(size))
```

```
## # A tibble: 32051 x 7
##     policy_nr  accident_nr dev_year close payment   size claim_covariates
##         <dbl>        <dbl>    <dbl> <lgl> <lgl>    <dbl> <list>
## 1 30000037738 898000390380        1 TRUE  FALSE      0   <tibble [1 x 11]>
## 2 30000124129 898001131523        1 TRUE  TRUE      57.2 <tibble [1 x 11]>
## 3 30000125846 898001053014        1 TRUE  FALSE      0   <tibble [1 x 11]>
## 4 30000194251 898000308942        1 FALSE  TRUE     120   <tibble [1 x 11]>
## 5 30000194251 898000308942        2 TRUE TRUE    2031.  <tibble [1 x 11]>
```

Then merge the policy and discretized claims data set.

```
reserving_data <-
  left_join(claim_data,
            policy_data,
            by = "policy_nr")

reserving_data
```

```
## # A tibble: 32051 x 4
##     policy_nr  accident_nr claim_covariates  policy_covariates
##         <dbl>        <dbl> <list>            <list>
## 1 30000037738 898000390380 <tibble [1 x 15]> <tibble [1 x 106]>
## 2 30000124129 898001131523 <tibble [1 x 15]> <tibble [1 x 106]>
## 3 30000125846 898001053014 <tibble [1 x 15]> <tibble [1 x 106]>
## 4 30000194251 898000308942 <tibble [1 x 15]> <tibble [1 x 106]>
## 5 30000194251 898000446055 <tibble [1 x 15]> <tibble [1 x 106]>
```

The likelihood of the observed development process for a single claim is:

$$f(C_{1,\ldots,T}, P_{1,\ldots,T}, Y_{1,\ldots,T}) = \prod_{j=1}^{T} f(C_j \mid C_{1,\ldots,j-1}, P_{1,\ldots,j-1}, Y_{1,\ldots,j-1}) \times \prod_{j=1}^{T} f(P_j \mid C_{1,\ldots,j}, P_{1,\ldots,j-1}, Y_{1,\ldots,j-1}) \times \prod_{j=1}^{T} f(Y_j \mid C_{1,\ldots,j}, P_{1,\ldots,j}, Y_{1,\ldots,j-1}),$$

where T is the number of observed development years.
We model the building blocks C, P and Y in this likelihood with a Generalized Linear Model (GLM).

- Closure indicator:

  Binomial GLM with complementary log-log link.

- Payment indicator:

  Binomial GLM with logit link.

- Payment size:

  Gamma GLM.

Model

Building blocks

- Closure indicator:

  Binomial GLM with complementary log-log link.

- Payment indicator:

  Binomial GLM with logit link.

- Payment size:

  Gamma GLM.

- Closure indicator:

  Binomial GLM with complementary log-log link.

- Payment indicator:

  Binomial GLM with logit link.

- Payment size:

  Gamma GLM.

Literature

Recent developments in individual reserving

Combining multiple triangles

Martínez Miranda et al. (2012); Denuit and Trufin (2018)

Machine learning methods

Lopez et al. (2016); Wüthrich (2018); Jamal et al. (2018)

Generalized linear models

Larsen (2007)

Focus on the individual reserving framework:

- Choosing between individual or aggregate reserving.
- Model selection techniques.
- Model evaluation.

Closure indicator

Only development year selected in the closure GLM model:

$$f_j := \mathcal{P}(C_j = 1 \mid C_{1,\ldots,j-1} = 0) = 1 - \exp(-\exp(\beta_j)).$$

The closure probability is estimated as:

$$\hat{f}_j = \frac{d_j}{n_j},$$

where

- d_j is the number of claims that close in development period j
- n_j is the number of claims that were open in development period j

This is the Kaplan-Meier estimator for the time to settlement.

Closure indicator

Use model selection techniques (AIC, BIC, . . .) to choose between:

- Aggregate approach, (only development period - KM estimator):

$$\mathcal{P}(C_j = 1 \mid C_{1,\ldots,j-1} = 0, P_{1,\ldots,j-1}, Y_{1,\ldots,j-1}) = 1 - \exp(-\exp(\beta_j)).$$

- Individual approach:

$$\mathcal{P}(C_j = 1 \mid C_{1,\ldots,j-1} = 0, P_{1,\ldots,j-1}, Y_{1,\ldots,j-1}) = 1 - \exp(-\exp(\boldsymbol{y}^{'} \cdot \boldsymbol{\beta})).$$

Hybrid approach: Use the aggregate approach when possible and the individual approach when needed.

Model selection

Imbalance in-sample and out-of-sample data

Runoff triangle of the number of open claims in each reporting year and development year

reporting year	development year 1	2	3	4	5	6
1998	14 507	2256	51	11	6	5
1999	15 936	2325	75	24	11	4
2000	15 818	2224	73	18	6	3
2001	17 079	2895	103	29	14	3
2002	19 656	2929	112	31	12	6
2003	18 342	2713	137	25	12	3

In-sample and out-of-sample distribution of the development year

	1	2	3	4	5	6
in-sample (%)	88.626	11.045	0.264	0.046	0.015	0.004
out-of-sample (%)	0	87.151	7.999	2.730	1.413	0.707

Model selection

Imbalance in-sample and out-of-sample data

Divide the runoff triangle in training, validation and evaluation cells

reporting year	development year 1	2	3	4	5	6
1998	14 507	2256	51	11	6	5
1999	15 936	2325	75	24	11	4
2000	15 818	2224	73	18	6	3
2001	17 079	2895	103	29	14	3
2002	19 656	2929	112	31	12	6
2003	18 342	2713	137	25	12	3

Calibrate the model on the training cells
Select covariates based on the validation cells
Recalibrate on the training and validation cells, predict the evaluation cells

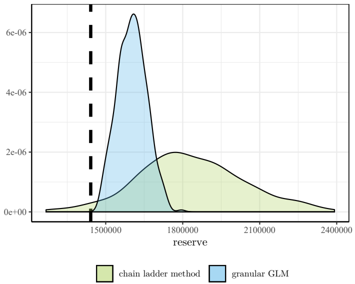
	1	2	3	4	5	6
training (%)	91.063	8.717	0.179	0.031	0.005	0.004
validation (%)	0	95.688	3.365	0.588	0.359	0
evaluation (%)	0	87.151	7.999	2.730	1.413	0.707

Model evaluation

Traditional one day view

Fit and evaluate the model on 31 December 2003

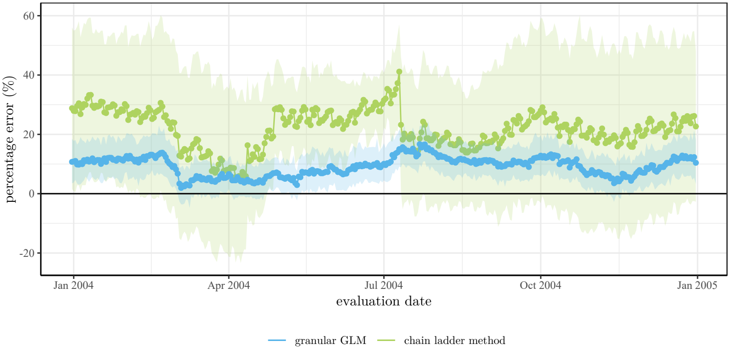
dev. year	actual	granular GLM	chain ladder
2	1 110 556	1 140 453	1 281 761
3	126 417	119 937	125 258
4	130 200	184 242	71 107
5	44 753	102 647	249 168
6	29 633	55 475	129 629
total	1 441 560	1 602 757	1 856 926



Model evaluation

Dynamic view

Moving window, fit and evaluate the reserve over an extended period of time.



Conclusions

Our ambitions for reserving

- Structure the scattered literature on claims reserving.
- Use multiple evaluation dates.
- Use multiple portfolios, no free lunch.
- Bridge pricing and reserving methodology, by using GLMs.
- Hybrid strategy, data driven approach to select position between individual and aggregated reserving.

More information

For more information, please visit:

LRisk website, www.lrisk.be

https://feb.kuleuven.be/jonas.crevecoeur

Thanks to

ARGENTA

fwo Research Foundation Flanders
Opening new horizons

Questions?

Denuit, M. and Trufin, J. (2018). Collective loss reserving with two types of claims in motor third party liability insurance. *Journal of Computational and Applied Mathematics*, 335:168 – 184.

Jamal, S., Canto, S., Fernwood, R., Giancaterino, C., Hiabu, M., Invernizzi, L., Korzhynska, T., Martin, Z., and Shen, H. (2018). Machine learning and traditional methods synergy in non-life reserving (mltms). *ASTIN WORKING PARTY*.

Larsen, C. R. (2007). An individual claims reserving model. *ASTIN Bulletin*, 37(1):113 – 132.

Lopez, O., Milhaud, X., and Thérond, P.-E. (2016). Tree-based censored regression with applications in insurance. *Electron. J. Statist.*, 10(2):2685–2716.

Martínez Miranda, M. D., Nielsen, J. P., and Verrall, R. J. (2012). Double Chain Ladder. *ASTIN Bulletin*, 42:59–76.

Wüthrich, M. V. (2018). Machine learning in individual claims reserving. *Scandinavian Actuarial Journal*, 2018(6):465–480.