Individual claims reserving for RBNS claims using state of the art data science

Łukasz Delong, EY Poland

Presenter

- Łukasz Delong
- Senior Manager at EY Poland
- Full Professor at University of Warsaw, PhD, Habilitation Degree
- Specialize in mathematical and statistical methods in insurance

Agenda

1

Why individual claims reserving?

2

General model for individual claims reserving for IBNER and RBNS claims

3

Case study for RBNS claims with a particular model

4

Conclusions

Why individual claims reserving?

Chain-Ladder algorithm

- Chain-Ladder has a long history in the actuarial science and is still the most popular algorithm in claims reserving,
- The idea of the C-L algorithm:
  - We aggregate claim observations in a loss triangle with cumulative payments based on accident periods and development periods,
  - We calculate individual development factors between consecutive development periods for each accident period and estimate average development factors.

Accident year	Development year						
	1	2	3	4	5	6	7
2017	1.02	1.75	1.04	1.02	1.01	1.00	1.00
2018	1.11	1.05	1.03	0.99	1.01	1.01	
2019	1.14	1.11	1.02	1.03	1.02		
2020	1.06	1.43	1.05	1.02			
2021	1.02	1.06	1.04				
2022	1.31	1.03					
2023	1.15						
2024							

Why individual claims reserving?

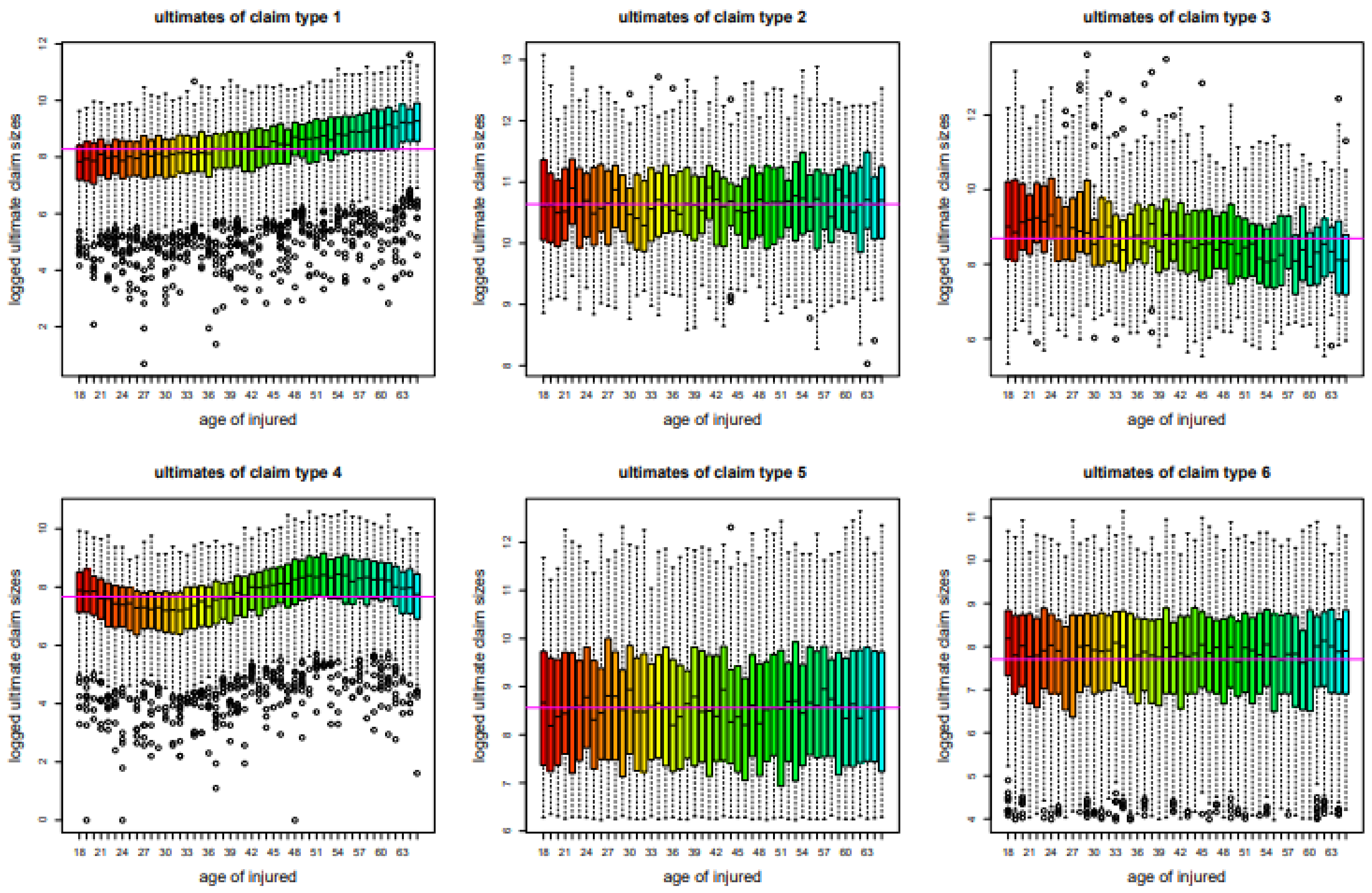
- We often observe problems with the Chain-Ladder algorithm when loss triangles are inhomogeneous and non-stationary across accident years, development years and calendar years,

- The inhomogeneity and the non-stationarity might be caused:
  - Large claims from extreme events,
  - Changes in a composition of claims in a portfolio (types of claims, open/claims),
  - Changes in a settlement process (due to a change in a regulation),
  - Claim inflation rates linked to external variables (e.g. superimposed inflation depending on claim sizes),

- C-L has more disadvantages:
  - It uses a single pattern for all claims that occur in an accident year, independently of accident years,
  - It neglects information about the business environment and individual claims characteristics,
  - It assumes a fixed structural form for accident years and development years,

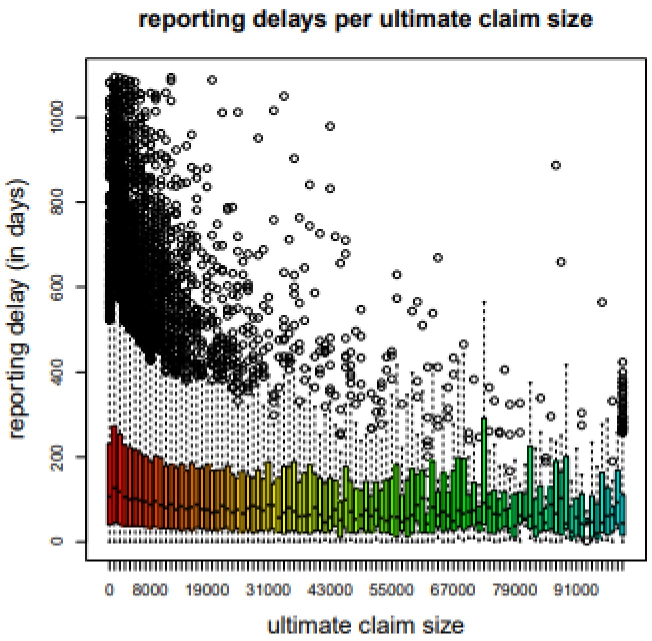
- To mitigate these problems and improve predictions, individual claims reserving models have been developed.

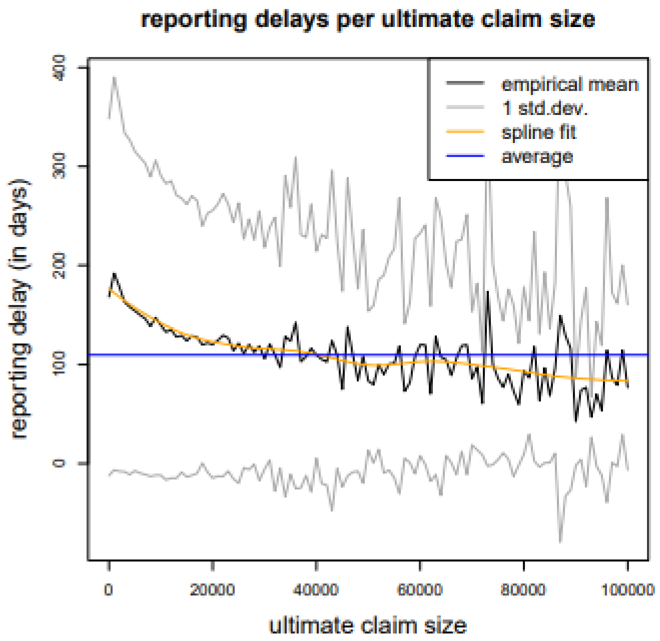
Why individual claims reserving?



Source: Wüthrich, M., Wang, M., 2022, Individual Claims Generator for Claims Reserving Studies: Data Simulation.R, available ar SSRN

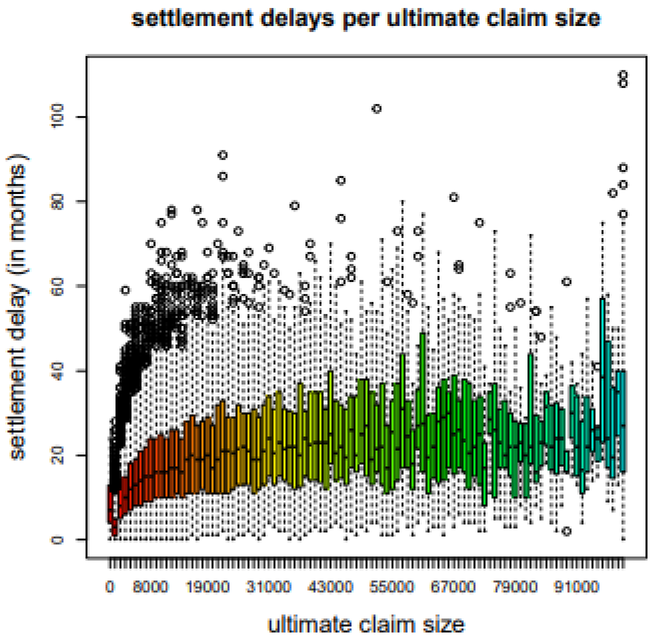
Why individual claims reserving?

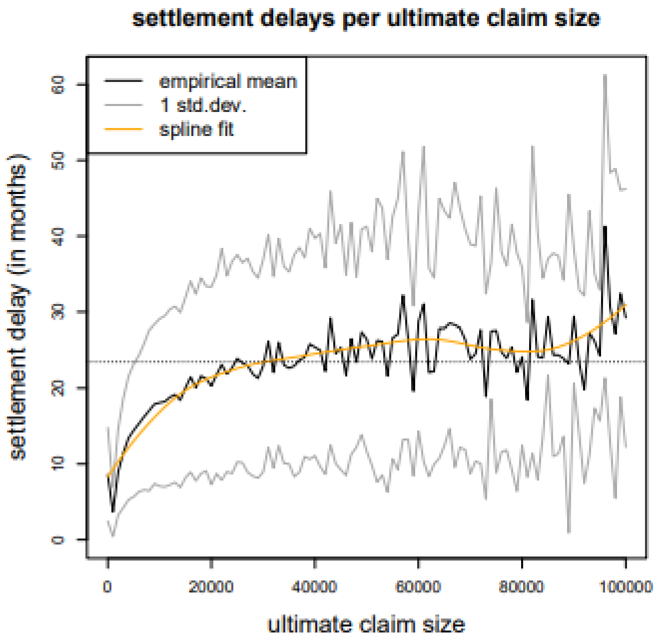




Source: Wüthrich, M., Wang, M., 2022, Individual Claims Generator for Claims Reserving Studies: Data Simulation.R, available ar SSRN

Why individual claims reserving?





Source: Wüthrich, M., Wang, M., 2022, Individual Claims Generator for Claims Reserving Studies: Data Simulation.R, available ar SSRN

Predictive models and covariates

We should build predictive models – we should predict ultimate claim sizes based on covariates

- Policy covariates:
  - The age of policyholder,
  - The area of residence,
  - The value of the property/car,
  - The line of business (risk group),
- Claim covariates:
  - The date on which the accident occur,
  - The date on which the claim was reported,
  - The cause of claim,
  - The claim type,
  - The initial expert case reserve
- Development covariates:
  - The number of periods elapsed since reporting,
  - The historical payments,
  - The historical case reserves (incurred values),
  - The settlement status (open or closed).

General model for individual claims reserving

Individual claims reserving

- In general, we need two separate models, for individual claims reserving:
- Incurred But Not Yet Reported (IBNYR):
    - We model the number of reported claims, i.e. we model the time of occurrence of a claim and the time of reporting,
    - Since policy specific information is available, we can also use it to predict the number of claims,
    - We estimate the ultimate claim size of a reported claim depending on its accident year and reporting delay, and policy features,
    - Claim and development specific information is not available for differentiating the cost per claim.
- Reported But Not Settled (RBNS):
    - We model the ultimate claim size for both open and closed claims on individual basis,
    - Policy specific together with claim and development specific information is available for differentiating the cost per claim,
    - The goal is to predict the future cost on a per claim basis as more and more individual claims specific information becomes available in the course of claim's development.

We focus on individual claims reserving for RBNS claims: Claims reported with different features should generate different cash flows in time and amount

EY

RBNS model

- We present a hierarchical model with layers introduced by
    - *Crevecoeur, J., Robben, J., Antonio, K., 2022, A hierarchical reserving model for reported non-life insurance claims, Insurance: Mathematics and Economics 104, 158-184.*

RBNS model

- Let ℓ denote a reported claim, $\ell = 1, \ldots, n$,
- Let the vector $\mathbf{U}_k^\ell$ describe the claim development information of claim ℓ after k development periods since reporting, $k \geq 0$,
- The vector $\mathbf{U}_0^\ell = \boldsymbol{z}_k^\ell$ describes the claim information registered at the reporting date,
- The vector $\mathbf{U}_k^\ell$ has m components (covariates) $\mathbf{U}_k^\ell = \left(U_{k,1}^\ell, \ldots, U_{k,m}^\ell\right)$,
- For each reported claim, the goal is to model the vector which describes the development of a claim:

$$\mathbf{U}_1^\ell, \ldots, \mathbf{U}_T^\ell | \mathbf{U}_0^\ell,$$

- We predict the components:

$$\mathbf{U}_{\mathrm{t}_\ell+1}^\ell, \ldots, \mathbf{U}_T^\ell | \mathbf{U}_0^\ell, \mathbf{U}_1^\ell, \ldots, \mathbf{U}_{\mathrm{t}_\ell}^\ell$$

in future development periods after the evaluation date.

RBNS model

- Since the future development depends on its development in previous years, we split the vector by development periods in chronological order:

$$\begin{gathered}\mathbf{U}_1^\ell | \mathbf{U}_0^\ell, \\ \mathbf{U}_2^\ell | \mathbf{U}_0^\ell, \mathbf{U}_1^\ell, \\ \dots \\ \mathbf{U}_k^\ell | \mathbf{U}_0^\ell, \mathbf{U}_1^\ell, \dots, \mathbf{U}_{k-1}^\ell, \\ \dots\end{gathered}$$

- We split the vector $\mathbf{U}_k^\ell$ by iterating over all its elements (layers):

$$\begin{gathered}U_{k,1}^\ell | \mathbf{U}_0^\ell, \mathbf{U}_1^\ell, \dots, \mathbf{U}_{k-1}^\ell, \\ U_{k,2}^\ell | \mathbf{U}_0^\ell, \mathbf{U}_1^\ell, \dots, \mathbf{U}_{k-1}^\ell, U_{k,1}^\ell, \\ \dots \\ U_{k,m}^\ell | \mathbf{U}_0^\ell, \mathbf{U}_1^\ell, \dots, \mathbf{U}_{k-1}^\ell, U_{k,1}^\ell, \dots, U_{k,m-1}^\ell,\end{gathered}$$

- We can now use regression models to model one-dimensional responses based on covariates.

RBNS model

- Remarks on the model's structure:
  - The outcome from one step becomes a covariate in the next step (multi-period and sequential predictions),
  - The layers and their order are important part of the model,
  - We can use any predictive modelling technique, e.g. GLM, GAMLSS, tree, neural network to predict the claims development and model the conditional distributions,
  - Not only the conditional mean values are important, as in regression modelling, but

    we need the whole conditional distributions at future development periods,
  - Compared to C-L models, we allow for a non-Markovian model since the whole history of a claim is included.

RBNS model

- Predictions and simulations from the model:
  - We simulate the future development for each reported claim,
  - We perform conditional simulations in line with the hierarchical structure and the order of the layers,
  - Note that the covariates in the regression models are realizations of random variables.

    If the mean values of these covariates are used in simulations,

    then the inputs for predictions are smoothed, whereas the models have been trained on unsmoothed inputs –

    we must have a stochastic model of good quality,
  - We obtain the distribution of an individual ultimate loss for each claim ℓ adapted to its features,
  - Calculating the mean values gives us individual claims reserves.

Case study for RBNS claims

Case study

- We present the results of two papers:
  - Ł. Delong, M.V. Wüthrich (2020) *Neural networks for the joint development of individual payments and claim incurred*, Risks 8, 1-33,
  - Delong, M.M. Lindholm, M.V. Wüthrich (2022) *Collective reserving using individual claims data*, Scandinavian Actuarial Journal 1, 1-28.

Data set

- We had a data set consisting of 1, 331, 856 individual claims. The data set describes the development processes of claims with accident dates and reporting dates both between January 2005 and December 2018,
- In order to anonymize the results, incremental payments and case reserves are scaled with a constant,
- We have claim and development information about:
  - Accident dates and reporting dates,
  - Claim type (property or bodily injured),
  - Claim segment (5 segments)
  - Claim origin (Poland and abroad),
  - Payments and case reserves on monthly basis,
- The goal is to predict ultimate claim size for each reported claim with regression models using the claim and development covariates.

RBNS model

- Let $i \in \{1,2,\dots\}$ denote the accident period of the occurrence date of an insurance claim, $j \in \{0,1,2,\dots\}$ denote the reporting delay after the claim occurrence date, $k \in \{0,1,2,\dots\}$ measure the development period of a reported claim, initialized to the respective reporting date $i+j$,
- Let $P_k^{i,j}$ denote the incremental payment in development period k, $I_k^{i,j}$ denote the claim incurred at the end of development period k, $R_k^{i,j}$ denote the case reserve at the end of development period k, for a claim from accident period i reported with delay j,
- At the reporting date of a claim, we observe the first payment and the first evaluation of the claim incurred, i.e. we have information $\left(P_0^{i,j}, I_0^{i,j}\right)$ .Next, we observe a process $\left(P_k^{i,j}, I_k^{i,j}\right)_{k\geq 1}$,
- To each individual claim, we associate a vector of features, which we denote by $\mathbf{z}_k^{i,j}$,
- We also define a filtration $\left(C_k^{i,j}\right)_{k=0,1,\dots}$ which describes the history of payments and claim incurred on an individual claim:

$$C_k^{i,j} = \sigma\left(P_s^{i,j}, I_s^{i,j}, s = 0,1,\dots,k\right), \qquad k = 0,1,\dots.$$

RBNS model

We aim at modeling the development $\left(P_k^{i,j}, I_k^{i,j}\right)_{k \geq 1}$ for each individual claim for all later time points $k = 1, 2, \ldots,$ given the individual claim history and the claim features

- Overview of the modelling and prediction process:

- For each RBNS claim,
  - In each development period k of an individual RBNS claim, we have to model:
    - Model 1: the event that there is a new payment and a change in a claim incurred,
    - Model 2-3: the payment (if it occurs), positive or negative,
    - Model 4: the event that the claim is closed,
    - Model 5: the new claim incurred (if the claim is still open and the claim incurred is changed),
  - And we can go to the next development period $k + 1$,
  - In each development period we use regression models to make predictions.

Model 1: New payments and changes in claim incurred

- We define the indicator process $(\mathcal{I}_k^{i,j}, \mathcal{P}_k^{i,j})_{k=1,2,\ldots}$:

$$\mathcal{I}_k^{i,j} = \mathbb{1}_{\{I_k^{i,j} - I_{k-1}^{i,j} \neq 0\}} \quad \text{and} \quad \mathcal{P}_k^{i,j} = \mathbb{1}_{\{P_k^{i,j} \neq 0\}},$$

  and we introduce the stochastic process $(Y_k^{i,j})_{k=1,2,\ldots}$:

$$Y_k^{i,j} = 2\mathcal{I}_k^{i,j} + \mathcal{P}_k^{i,j} = \begin{cases} 0 & \text{if} \quad \mathcal{P}_k^{i,j} = 0 \text{ and } \mathcal{I}_k^{i,j} = 0, \\ 1 & \text{if} \quad \mathcal{P}_k^{i,j} = 1 \text{ and } \mathcal{I}_k^{i,j} = 0, \\ 2 & \text{if} \quad \mathcal{P}_k^{i,j} = 0 \text{ and } \mathcal{I}_k^{i,j} = 1, \\ 3 & \text{if} \quad \mathcal{P}_k^{i,j} = 1 \text{ and } \mathcal{I}_k^{i,j} = 1. \end{cases}$$

- We use a multinomial logistic regression to model the categorical conditional probabilities for the two-dimensional indicator process $(\mathcal{I}_k^{i,j}, \mathcal{P}_k^{i,j})$:

$$\log\left(\frac{\mathbb{P}\left(Y_k^{i,j} = y \middle| \mathcal{C}_{k-1}^{i,j}, \boldsymbol{z}_{k-1}^{i,j}\right)}{\mathbb{P}\left(Y_k^{i,j} = 0 \middle| \mathcal{C}_{k-1}^{i,j}, \boldsymbol{z}_{k-1}^{i,j}\right)}\right) = f^y\left(\mathcal{C}_{k-1}^{i,j}, \boldsymbol{z}_{k-1}^{i,j}\right), \quad y = 1, 2, 3, \quad k \geq 1,$$

  where f_y denotes a regression function,

- If $Y_k^{i,j} = 0$,then we immediately know the values of the process $(P_k^{i,j}, I_k^{i,j})$ in the next development period k.

Model 2 and 3: Claim severities of incremental payments

- In practice, we observe both positive and negative incremental payments (salvages and subrogations). We use a spliced distribution to model non-zero incremental payments,

- We introduce the sequences of random variables $\left(P_k^{i,j,(+)}, P_k^{i,j,(-)}\right)_{k=1,2,\ldots}$:

$$
\begin{aligned}
P_k^{i,j,(+)} &= P_k^{i,j} \middle| Y_k^{i,j} \in \{1,3\}, P_k^{i,j} > 0, \quad k = 1,2,\ldots, \\
P_k^{i,j,(-)} &= -P_k^{i,j} \middle| Y_k^{i,j} \in \{1,3\}, P_k^{i,j} < 0, \quad k = 1,2,\ldots,
\end{aligned}
$$

- We use a binomial logistic regression to model the conditional probabilities of a positive or a negative incremental payment:

$$
\log\left(\frac{\mathbb{P}\left(P_k^{i,j} > 0 \middle| Y_k^{i,j} \in \{1,3\}, \mathcal{I}_k^{i,j}, \mathcal{C}_{k-1}^{i,j}, \boldsymbol{z}_{k-1}^{i,j}\right)}{\mathbb{P}\left(P_k^{i,j} < 0 \middle| Y_k^{i,j} \in \{1,3\}, \mathcal{I}_k^{i,j}, \mathcal{C}_{k-1}^{i,j}, \boldsymbol{z}_{k-1}^{i,j}\right)}\right) = f\left(\mathcal{I}_k^{i,j}, \mathcal{C}_{k-1}^{i,j}, \boldsymbol{z}_{k-1}^{i,j}\right), \quad k \geq 1.
$$

where f denotes a regression function

Model 2 and 3: Claim severities of incremental payments

- We model claim severities with a double Gamma regression,

- We assume that $P_k^{i,j,(+)}$ and $P_k^{i,j,(-)}$ have Gamma distributions with the mean value:

$$\log\left(\mathbb{E}\left[P_k^{i,j,(+)}\middle|\mathcal{I}_k^{i,j},\mathcal{C}_{k-1}^{i,j},\boldsymbol{z}_{k-1}^{i,j}\right]\right)=f\left(\mathcal{I}_k^{i,j},\mathcal{C}_{k-1}^{i,j},\boldsymbol{z}_{k-1}^{i,j}\right),\quad k\geq 1,$$

  for a regression function f, and the second moment given by

$$\begin{aligned}\operatorname{Var}\left[P_k^{i,j,(+)}\middle|\mathcal{I}_k^{i,j},\mathcal{C}_{k-1}^{i,j},\boldsymbol{z}_{k-1}^{i,j}\right] &= e^{\phi\left(\mathcal{I}_k^{i,j},\mathcal{C}_{k-1}^{i,j},\boldsymbol{z}_{k-1}^{i,j}\right)}\\ &\quad\cdot\left(\mathbb{E}\left[P_k^{i,j,(+)}\middle|\mathcal{I}_k^{i,j},\mathcal{C}_{k-1}^{i,j},\boldsymbol{z}_{k-1}^{i,j}\right]\right)^2,\quad k\geq 1,\end{aligned}$$

  for another regression function ϕ,

- If $Y_k^{i,j}=1$,we can derive the values of the process $(P_k^{i,j}, I_k^{i,j})$ in the next development period k. If $Y_k^{i,j}=3$, we have to model the change in claim incurred at the end of the development period k. If $Y_k^{i,j}=2$,the payment $P_k^{i,j}=0$ is zero but we need to consider a change in claim incurred.

Model 4: Closing times

- We introduce the sequences of random variables $\left(\mathcal{R}_k^{i,j}\right)_{k=1,2,\dots}$:

$$\mathcal{R}_k^{i,j} = R_k^{i,j} \middle| Y_k^{i,j} \in \{2,3\}, \quad k = 1,2,\dots.$$

- The event $\{\mathcal{R}_k^{i,j} = 0\}$ is interpreted as claim closing in development period k,

- We a binomial logistic regression to model the event that a claim is closed:

$$\log\left(\frac{\mathbb{P}\left(\mathcal{R}_k^{i,j} = 0 \middle| \mathcal{P}_k^{i,j}, P_k^{i,j}, \mathcal{C}_{k-1}^{i,j}, \boldsymbol{z}_{k-1}^{i,j}\right)}{\mathbb{P}\left(\mathcal{R}_k^{i,j} \neq 0 \middle| \mathcal{P}_k^{i,j}, P_k^{i,j}, \mathcal{C}_{k-1}^{i,j}, \boldsymbol{z}_{k-1}^{i,j}\right)}\right) = f\left(\mathcal{P}_k^{i,j}, P_k^{i,j}, \mathcal{C}_{k-1}^{i,j}, \boldsymbol{z}_{k-1}^{i,j}\right), \quad k \geq 1.$$

- For claims that have zero case reserves at the end of development period k, we exactly know the change in claim incurred.

Model 5: Severities for claim incurred for open claims

- We define the sequence of random variables $\left(I_k^{i,j,(open)}\right)_{k=1,2,\dots}$ which describe claims incurred:

$$I_k^{i,j,(open)} = I_k^{i,j} | Y_k^{i,j} \in \{2,3\}, R_k^{i,j} \neq 0, \quad k = 1,2, \dots,$$

- As for payments, we assume a double Gamma regression. The claim incurred $I_k^{i,j,(open)}$ has Gamma distribution with the mean value:

$$\log\left(\mathbb{E}\left[I_k^{i,j,(open)} \middle| \mathcal{P}_k^{i,j}, P_k^{i,j}, \mathcal{C}_{k-1}^{i,j}, \boldsymbol{z}_{k-1}^{i,j}\right]\right) = f\left(\mathcal{P}_k^{i,j}, P_k^{i,j}, \mathcal{C}_{k-1}^{i,j}, \boldsymbol{z}_{k-1}^{i,j}\right), \quad k \geq 1,$$

  for a regression function f, and choosing another regression function ϕ we specify the second moment of the distribution:

$$\begin{aligned} \mathrm{Var}\left[I_k^{i,j,(open)} \middle| \mathcal{P}_k^{i,j}, P_k^{i,j}, \mathcal{C}_{k-1}^{i,j}, \boldsymbol{z}_{k-1}^{i,j}\right] &= e^{\phi\left(\mathcal{P}_k^{i,j}, P_k^{i,j}, \mathcal{C}_{k-1}^{i,j}, \boldsymbol{z}_{k-1}^{i,j}\right)} \\ &\cdot \left(\mathbb{E}\left[I_k^{i,j,(open)} \middle| \mathcal{P}_k^{i,j}, P_k^{i,j}, \mathcal{C}_{k-1}^{i,j}, \boldsymbol{z}_{k-1}^{i,j}\right]\right)^2, \quad k \geq 1. \end{aligned}$$

- We model claim incurred $I_k^{i,j} | Y_k^{i,j} \in \{2,3\}$ at the end of development period k and we can derive the value of the process $\left(P_k^{i,j}, I_k^{i,j}\right)$ in the next development period k.

Modeling approach

- As predictive models, we calibrate deep neural networks with two hidden layers,
- We fit the neural networks on quarterly data. Hence, we deal with development period $k = 1, 2, \ldots, 55$,
- We only model positive payments and resign from modeling negative payments (salvages and subrogations),
- We fit separate neural networks for each development period $k = 1, \ldots, 16$, and one neural network for all development periods $k = 17, \ldots, 55$, for Models 1, 3 positive, 4 and 5.

Results

Accident year	RBNS reserve Mean	25th quantile	75th quantile	Chain-Ladder	Case reserve
2009	1.26	1.13	1.39	1.35	0.65
2010	1.51	1.32	1.68	1.97	1.56
2011	2.29	2.03	2.39	2.66	3.14
2012	3.65	3.26	3.78	3.85	3.63
2013	5.10	4.69	5.18	5.11	5.07
2014	5.99	5.43	6.35	6.61	7.00
2015	8.27	7.57	8.36	9.36	6.90
2016	11.91	11.48	12.27	13.43	10.81
2017	17.20	16.67	17.80	19.62	11.65
2018	35.09	34.64	35.59	39.77	28.02
All	92.27	88.21	94.78	103.74	78.43

Table: Simulations results and Chain-Ladder (CL) estimates (in MM).

Results

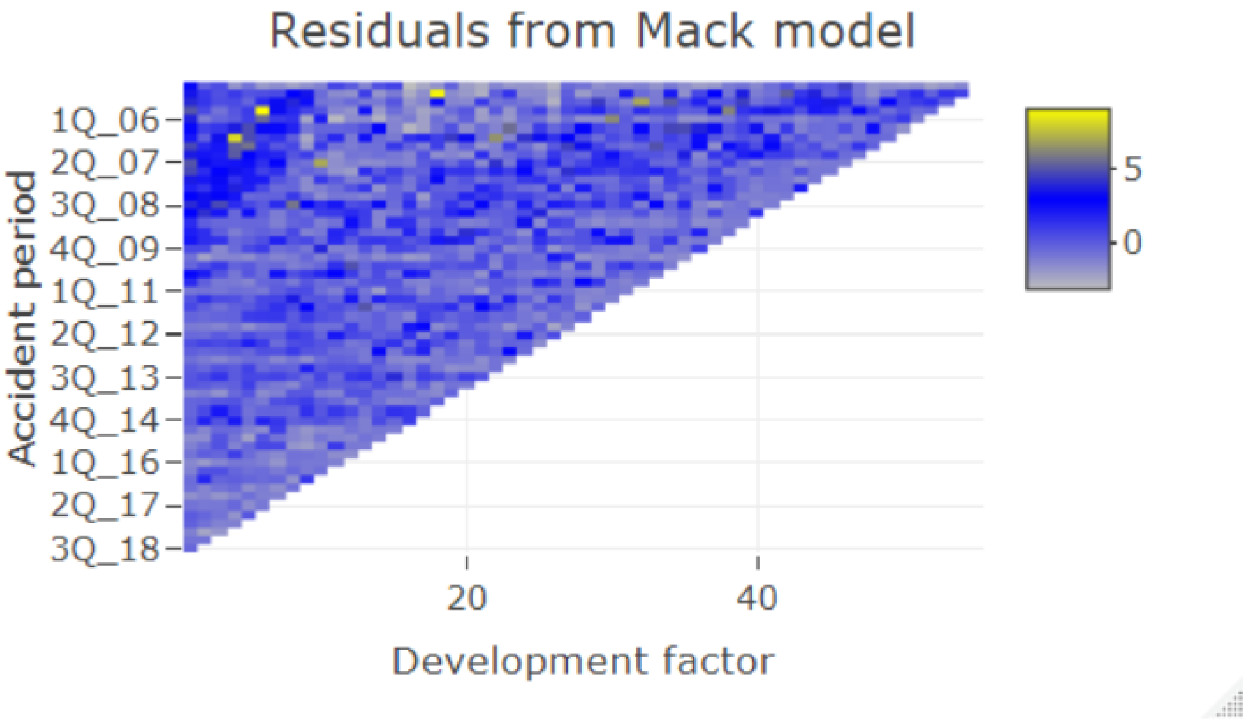


Figure: Residuals from Mack model fitted to development factors in a run-off triangle.

Results

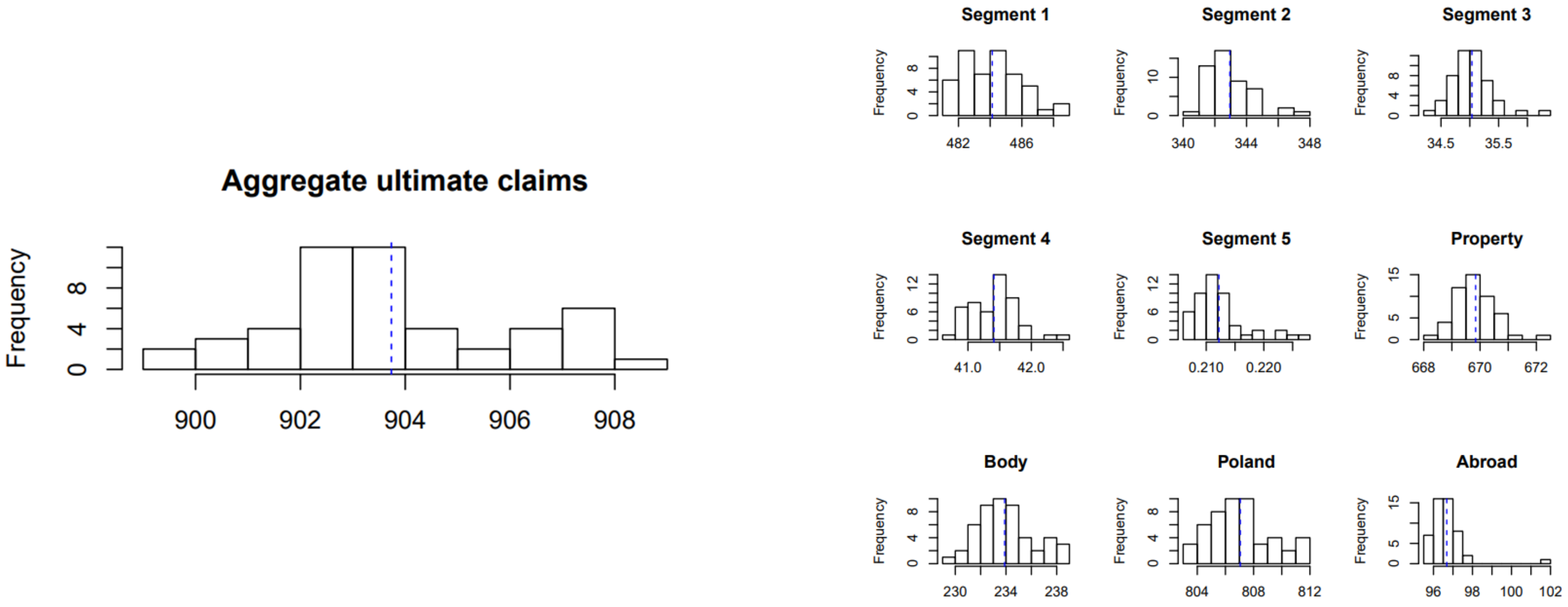
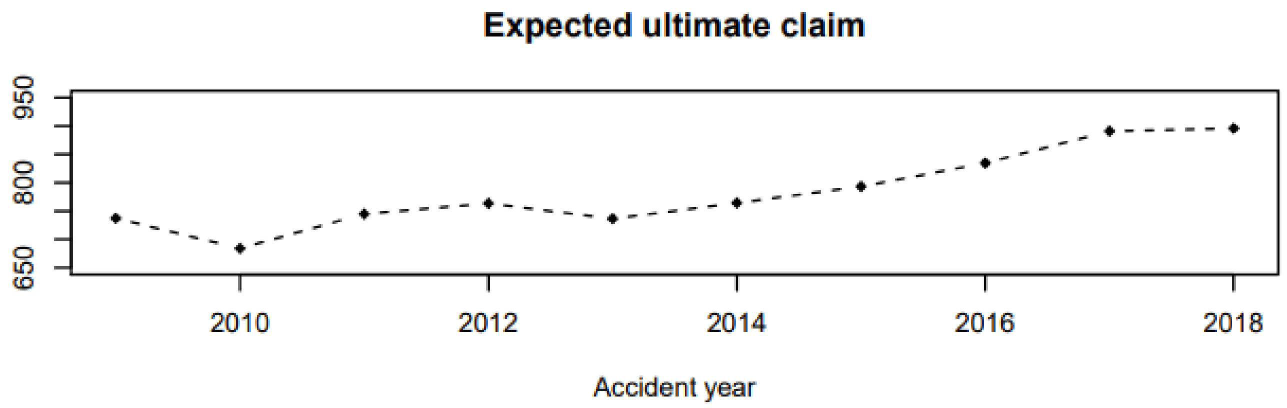


Figure: Histograms of the aggregate ultimate payments from the RBNS claims (in MM). The dashed lines indicate the mean value from the simulations.

Results



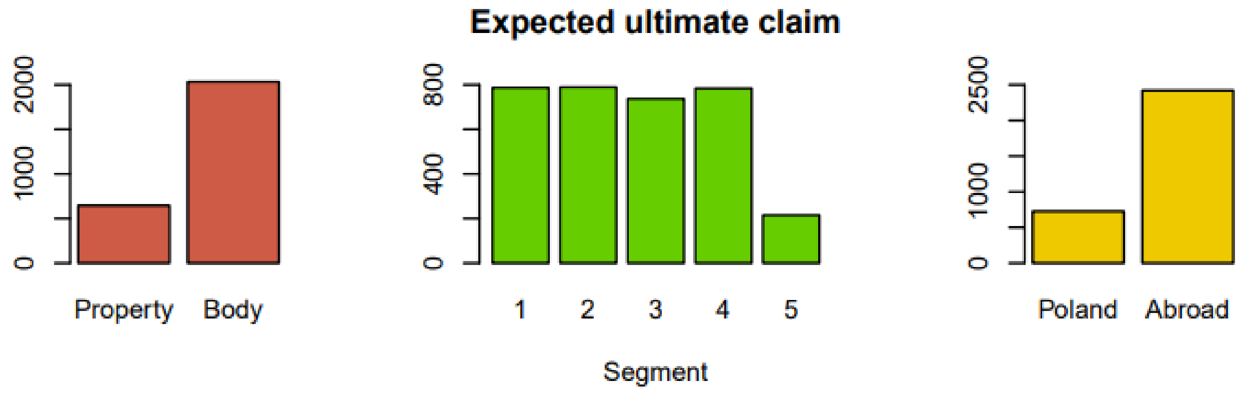


Figure: Expected aggregate ultimate payments from an individual claim.

Conclusions

Pros and cons of individual claims reserving

- Advantages:
  - Individual claims reserving models give us more granular view on the risk drivers behind claims development and provides more insight into claims development,
  - Predictions are more accurate since unique characteristics of claims are taken into account

    and claim characteristics have predictive power of the ultimate claim sizes (if machine learning models are implemented),
  - Individual claims reserving models can be tailored to specific needs by incorporating important factors and adapting to new claims enviroments,
  - They allow for early detection of trends or anomalies in claims developments and dynamic management of claims reserve risk
  - We can link claims reserving to actuarial pricing,

- Disadvantages:
  - At the aggregate level and in stationary environment, individual claims reserving are in line with traditional methods,
  - The individual claims reserving models may be complex, sensitive to assumptions and time-consuming in estimation and simulations compared to C-L type algorithms.

Questions