Sensitivity analysis of stochastic reserving models using bootstrap simulations

Master's thesis defence

Othman El Hammouchi

June 28, 2023

VRIJE
UNIVERSITEIT
BRUSSEL

OVERVIEW

Introduction

INSURANCE INDUSTRY

- Inverted production cycle
- Future liabilities not known today
- Prudential and regulatory requirement to make provisions

THE ACTUARIAL RESERVING PROBLEM

- Claims reserving: forecast future funds needed to settle outstanding contracts
- Not just point estimate, but also variability and shape of distribution
- Traditional approach based on claims, loss or run-off triangles

CLAIMS TRIANGLE EXAMPLE

Origin	Dev						
	1	2	3	4	5	6	7
2007	3511	6726	8992	10704	11763	12350	12690
2008	4001	7703	9981	11161	12117	12746	
2009	4355	8287	10233	11755	12993		
2010	4295	7750	9773	11093			
2011	4150	7897	10217				
2012	5102	9650					
2013	6283						

Table: Cumulative payments triangle for a motor insurance account from the UK

CLAIMS TRIANGLES IN GENERAL

Cumulative					Incremental				
C_{11}	C_{12}	C_{13}	C_{14}	C_{15}	X_{11}	X_{12}	X_{13}	X_{14}	X_{15}
C_{21}	C_{22}	C_{23}	C_{24}		X_{21}	X_{22}	X_{23}	X_{24}	
C_{31}	C_{32}	C_{33}			X_{31}	X_{32}	X_{33}		
C_{41}	C_{42}				X_{41}	X_{42}			
C_{51}					X_{51}				

Origin period ↓

Development period →

CLAIMS TRIANGLES IN GENERAL

Calendar period (current)

Cumulative

$$\begin{array}{ccccc} C_{11} & C_{12} & C_{13} & C_{14} & C_{15} \\ C_{21} & C_{22} & C_{23} & C_{24} & \\ C_{31} & C_{32} & C_{33} & & \\ C_{41} & C_{42} & & & \\ C_{51} & & & & \end{array}$$

Incremental

$$\begin{array}{ccccc} X_{11} & X_{12} & X_{13} & X_{14} & X_{15} \\ X_{21} & X_{22} & X_{23} & X_{24} & \\ X_{31} & X_{32} & X_{33} & & \\ X_{41} & X_{42} & & & \\ X_{51} & & & & \end{array}$$

Origin period

Development period

FORECASTING USING CLAIMS TRIANGLES

$$
\begin{array}{ccccc}
C_{11} & C_{12} & C_{13} & C_{14} & C_{15} \\
C_{21} & C_{22} & C_{23} & C_{24} & \color{red}{\widehat{C}_{25}} \\
C_{31} & C_{32} & C_{33} & \color{red}{\widehat{C}_{34}} & \color{red}{\widehat{C}_{35}} \\
C_{41} & C_{42} & \color{red}{\widehat{C}_{43}} & \color{red}{\widehat{C}_{44}} & \color{red}{\widehat{C}_{45}} \\
C_{51} & \color{red}{\widehat{C}_{52}} & \color{red}{\widehat{C}_{53}} & \color{red}{\widehat{C}_{54}} & \color{red}{\widehat{C}_{55}}
\end{array}
\qquad
\begin{array}{ccccc}
X_{11} & X_{12} & X_{13} & X_{14} & X_{15} \\
X_{21} & X_{22} & X_{23} & X_{24} & \color{red}{\widehat{X}_{25}} \\
X_{31} & X_{32} & X_{33} & \color{red}{\widehat{X}_{34}} & \color{red}{\widehat{X}_{35}} \\
X_{41} & X_{42} & \color{red}{\widehat{X}_{43}} & \color{red}{\widehat{X}_{44}} & \color{red}{\widehat{X}_{45}} \\
X_{51} & \color{red}{\widehat{X}_{52}} & \color{red}{\widehat{X}_{53}} & \color{red}{\widehat{X}_{54}} & \color{red}{\widehat{X}_{55}}
\end{array}
$$

FORECASTING USING CLAIMS TRIANGLES

$$
\begin{array}{ccccc}
C_{11} & C_{12} & C_{13} & C_{14} & C_{15} \\
C_{21} & C_{22} & C_{23} & C_{24} & \widehat{C}_{25} \\
C_{31} & C_{32} & C_{33} & \widehat{C}_{34} & \widehat{C}_{35} \\
C_{41} & C_{42} & \widehat{C}_{43} & \widehat{C}_{44} & \widehat{C}_{45} \\
C_{51} & \widehat{C}_{52} & \widehat{C}_{53} & \widehat{C}_{54} & \widehat{C}_{55}
\end{array}
\qquad
\begin{array}{ccccc}
X_{11} & X_{12} & X_{13} & X_{14} & X_{15} \\
X_{21} & X_{22} & X_{23} & X_{24} & \widehat{X}_{25} \\
X_{31} & X_{32} & X_{33} & \widehat{X}_{34} & \widehat{X}_{35} \\
X_{41} & X_{42} & \widehat{X}_{43} & \widehat{X}_{44} & \widehat{X}_{45} \\
X_{51} & \widehat{X}_{52} & \widehat{X}_{53} & \widehat{X}_{54} & \widehat{X}_{55}
\end{array}
$$

Ultimates

FORECASTING USING CLAIMS TRIANGLES

Latest

$$
\begin{array}{ccccc}
C_{11} & C_{12} & C_{13} & C_{14} & C_{15} \\
C_{21} & C_{22} & C_{23} & C_{24} & \widehat{C}_{25} \\
C_{31} & C_{32} & C_{33} & \widehat{C}_{34} & \widehat{C}_{35} \\
C_{41} & C_{42} & \widehat{C}_{43} & \widehat{C}_{44} & \widehat{C}_{45} \\
C_{51} & \widehat{C}_{52} & \widehat{C}_{53} & \widehat{C}_{54} & \widehat{C}_{55}
\end{array}
\qquad
\begin{array}{ccccc}
X_{11} & X_{12} & X_{13} & X_{14} & X_{15} \\
X_{21} & X_{22} & X_{23} & X_{24} & \widehat{X}_{25} \\
X_{31} & X_{32} & X_{33} & \widehat{X}_{34} & \widehat{X}_{35} \\
X_{41} & X_{42} & \widehat{X}_{43} & \widehat{X}_{44} & \widehat{X}_{45} \\
X_{51} & \widehat{X}_{52} & \widehat{X}_{53} & \widehat{X}_{54} & \widehat{X}_{55}
\end{array}
$$

MORE NOMENCLATURE

- I origin periods, J development periods
- We assume $I = J$ (square triangles)
- Reserve $R = \sum_{i=2}^{I}(C_{i,I} - C_{i,I+1-i}) = \sum_{j=2}^{I}\sum_{i=I+2-j}^{I} X_{ij}$

THE CHAIN LADDER

- Most popular reserving method [1]
- Originally deterministic algorithm
- Various attempts to frame it as a stochastic model
- Main assumption: there exist development factors $f_1, \ldots, f_{I-1}$ such that

$$\mathbb{E}\left[C_{ij} \middle| C_{i,j-1}, \ldots, C_{i1}\right] = f_{j-1} C_{i,j-1}$$

[1] According to the ASTIN 2016 Non-Life Reserving Practices Report

VISUALISATION OF THE CHAIN LADDER

$$\begin{array}{ccccc} C_{11} & C_{12} & C_{13} & C_{14} & C_{15} \\ C_{21} & C_{22} & C_{23} & C_{24} & \\ C_{31} & C_{32} & C_{33} & & \\ C_{41} & C_{42} & & & \\ C_{51} & & & & \end{array}$$

$$f_1 \quad f_2 \quad f_3 \quad f_4$$

Column sum average

$$\widehat{f}_j = \frac{\sum_{i=1}^{I-j} C_{i,j+1}}{\sum_{i=1}^{I-j} C_{ij}}$$

Chain ladder prediction

$$\widehat{C}_{ij} = C_{i,I+1-i} \prod_{k=I+1-i}^{j-1} \widehat{f}_k$$

VISUALISATION OF THE CHAIN LADDER

$$
\begin{array}{ccccc}
C_{11} & C_{12} & C_{13} & C_{14} & C_{15} \\
C_{21} & C_{22} & C_{23} & C_{24} & \\
C_{31} & C_{32} & C_{33} & & \\
C_{41} & C_{42} & & & \\
C_{51} & \color{red}{\widehat{C}_{52}} & & &
\end{array}
$$

$$f_1 \quad f_2 \quad f_3 \quad f_4$$

Column sum average

$$\widehat{f}_j = \frac{\sum_{i=1}^{I-j} C_{i,j+1}}{\sum_{i=1}^{I-j} C_{ij}}$$

Chain ladder prediction

$$\widehat{C}_{ij} = C_{i,I+1-i} \prod_{k=I+1-i}^{j-1} \widehat{f}_k$$

VISUALISATION OF THE CHAIN LADDER

$$
\begin{array}{ccccc}
C_{11} & C_{12} & C_{13} & C_{14} & C_{15} \\
C_{21} & C_{22} & C_{23} & C_{24} & \\
C_{31} & C_{32} & C_{33} & & \\
C_{41} & C_{42} & \color{red}{\widehat{C}_{43}} & & \\
C_{51} & \color{red}{\widehat{C}_{52}} & \color{red}{\widehat{C}_{53}} & & \\
f_1 & f_2 & f_3 & f_4 &
\end{array}
$$

Column sum average

$$\widehat{f}_j = \frac{\sum_{i=1}^{I-j} C_{i,j+1}}{\sum_{i=1}^{I-j} C_{ij}}$$

Chain ladder prediction

$$\widehat{C}_{ij} = C_{i,I+1-i} \prod_{k=I+1-i}^{j-1} \widehat{f}_k$$

VISUALISATION OF THE CHAIN LADDER

$$
\begin{array}{ccccc}
C_{11} & C_{12} & C_{13} & C_{14} & C_{15} \\
C_{21} & C_{22} & C_{23} & C_{24} & \\
C_{31} & C_{32} & C_{33} & \color{red}{\widehat{C}_{34}} & \\
C_{41} & C_{42} & \color{red}{\widehat{C}_{43}} & \color{red}{\widehat{C}_{44}} & \\
C_{51} & \color{red}{\widehat{C}_{52}} & \color{red}{\widehat{C}_{53}} & \color{red}{\widehat{C}_{54}} & \\
\end{array}
$$

$$f_1 \quad f_2 \quad f_3 \quad f_4$$

Column sum average

$$\widehat{f}_j = \frac{\sum_{i=1}^{I-j} C_{i,j+1}}{\sum_{i=1}^{I-j} C_{ij}}$$

Chain ladder prediction

$$\widehat{C}_{ij} = C_{i,I+1-i} \prod_{k=I+1-i}^{j-1} \widehat{f}_k$$

VISUALISATION OF THE CHAIN LADDER

$$
\begin{array}{ccccc}
C_{11} & C_{12} & C_{13} & C_{14} & C_{15} \\
C_{21} & C_{22} & C_{23} & C_{24} & \widehat{C}_{25} \\
C_{31} & C_{32} & C_{33} & \widehat{C}_{34} & \widehat{C}_{35} \\
C_{41} & C_{42} & \widehat{C}_{43} & \widehat{C}_{44} & \widehat{C}_{45} \\
C_{51} & \widehat{C}_{52} & \widehat{C}_{53} & \widehat{C}_{54} & \widehat{C}_{55} \\
\multicolumn{5}{c}{\quad f_1 \qquad\quad f_2 \qquad\quad f_3 \qquad\quad f_4 \qquad\quad}
\end{array}
$$

Column sum average

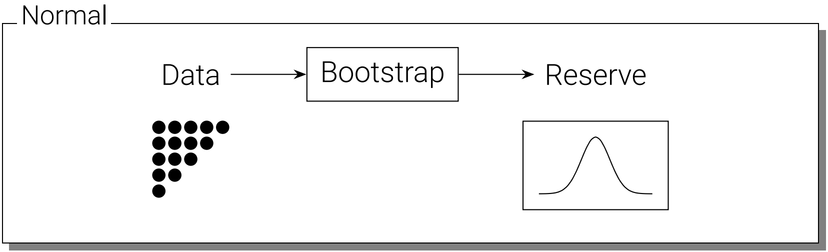
$$\widehat{f}_j = \frac{\sum_{i=1}^{I-j} C_{i,j+1}}{\sum_{i=1}^{I-j} C_{ij}}$$

Chain ladder prediction

$$\widehat{C}_{ij} = C_{i,I+1-i} \prod_{k=I+1-i}^{j-1} \widehat{f}_k$$

STOCHASTIC CHAIN LADDER

- Many different variants
- Reproduce chain ladder point estimates
- Make different assumptions
- Difficult to verify with small data sizes
- Idea: detect violations by excluding points and gauging effect on bootstrapped reserve

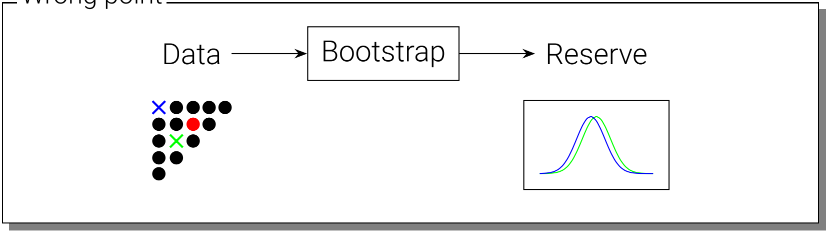


DETECTING ASSUMPTION VIOLATIONS

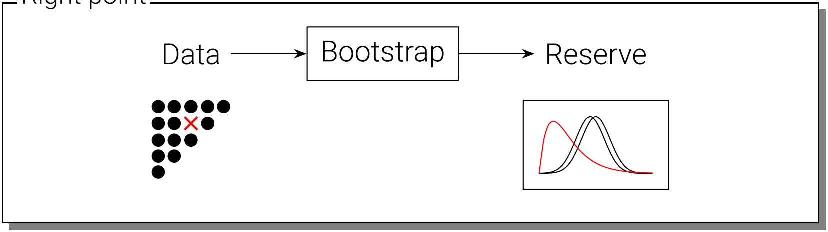
- Generate triangles which follow assumptions perfectly
- Apply perturbation
- Remove one point at a time and study impact on reserve
- Significant impact $\Rightarrow$ reverse-engineer

DETECTING ASSUMPTION VIOLATIONS

Wrong point

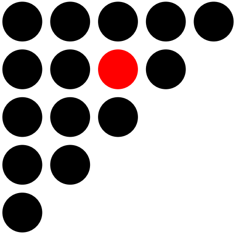


Right point



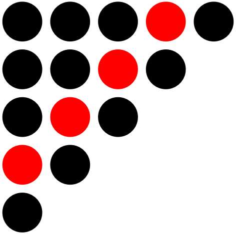
PERTURBATIONS

Single observation



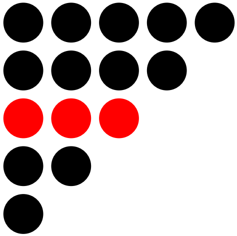
PERTURBATIONS

Calendar period



PERTURBATIONS

Origin period



The bootstrap method

MAIN IDEA

- Classical inference often intractible
- Relies on approximations and asymptotics
- Solution: resampling to produce pseudo-replicates

MAIN IDEA

- Classical inference often intractible
- Relies on approximations and asymptotics
- Solution: resampling to produce pseudo-replicates

CLASSICAL ESTIMATOR

- Independent identically distributed sample $X_1, \dots, X_n$
- Parameter θ estimated by $\widehat{\theta} := g(X_1, \dots, X_n)$
- For $b = 1, \dots, B$
  - Resample to obtain $X_1^{(b)}, \dots, X_n^{(b)}$
  - Compute $\widehat{\theta}^{(b)} := g(X_1^{(b)}, \dots, X_n^{(b)})$
  - $\{\widehat{\theta}^{(b)} \mid b = 1, \dots, B\}$ used for inference, e.g. variance estimation:

$$\widehat{\mathrm{Var}}(\theta) := \frac{1}{B-1} \sum_{b=1}^{B} (\widehat{\theta}^{(b)} - \overline{\theta}^{B})^2$$

with $\overline{\theta}^{B} := \frac{1}{B} \sum_{b=1}^{B} \widehat{\theta}^{(b)}$

PARAMETRIC VS. NONPARAMETRIC

- How to do bootstrap resampling?
- Nonparametric: resample with replacement directly from data
- Parametric: fit model first, use this to simulate from RNG
- Can be extended to regression models

Mack’s model

FORMULATION

Model 1 (Mack chain ladder)

1. *There exist development factors* $f_1, \ldots, f_{I-1}$ *such that*

$$\mathbb{E}[C_{ij} \parallel C_{i,j-1}, \ldots, C_{i1}] = \mathbb{E}[C_{ij} \parallel C_{i,j-1}] = f_{j-1} C_{i,j-1}$$

   for $1 \leq i \leq I$

2. *There exist variance parameters* $\sigma_1, \ldots, \sigma_{I-1}$ *such that*

$$\mathrm{Var}[C_{ij} \parallel C_{i,j-1}, \ldots, C_{i1}] = \mathrm{Var}[C_{ij} \parallel C_{i,j-1}] = \sigma_{j-1}^2 C_{i,j-1}$$

   for $1 \leq i \leq I$

3. *The cumulative claims processes* $(C_{ij})_j, (C_{i'j})_j$ *are independent for* $i \neq i'$

PROPERTIES

- Cumulative triangle
- Distribution-free
- Recursive
- For any pair of consecutive columns: equivalent to

$$\mathbf{c}_{\boldsymbol{j+1}} = f_j \mathbf{c}_{\boldsymbol{j}} + \boldsymbol{\varepsilon}$$

with

$$\mathbb{E}\left[\boldsymbol{\varepsilon} \middle| C_{1,j}, \ldots, C_{I-j,j}\right] = \mathbf{0}$$

$$\operatorname{Var}\left[\boldsymbol{\varepsilon} \middle| C_{1,j}, \ldots, C_{I-j,j}\right] = \sigma_j^2 \begin{bmatrix} C_{1j} & & \\ & \ddots & \\ & & C_{I-j,j} \end{bmatrix}$$

PROPERTIES

- Model assumptions correspond to Gauss-Markov
- Optimal estimator: weighted least squares with

$$\mathbf{W} = \begin{bmatrix} 1/C_{1j} & & \\ & \ddots & \\ & & 1/C_{I-j,j} \end{bmatrix}$$

- Same as column sum estimator!

$$\widehat{f}_j^{\mathrm{WLS}} = (\mathbf{c}_{\boldsymbol{j}}^T \mathbf{W} \mathbf{c}_{\boldsymbol{j}})^{-1} \mathbf{c}_{\boldsymbol{j}}^T \mathbf{W} = \frac{\sum_{i=1}^{I-j} C_{i,j+1}}{\sum_{i=1}^{I-j} C_{i,j}}$$

- We can adapt the regression bootstrap!

CONDITIONAL VS. UNCONDITIONAL

- Recursivity leads to different bootstrap types
- Simulate next development year based on original data vs. generated bootstrap replicate
- Parametric example:

$$C_{i,j+1}^{(b)} \sim \mathcal{N}(\widehat{f}_j C_{ij}, \widehat{\sigma}_j^2) \quad \text{vs.} \quad C_{i,j+1}^{(b)} \sim \mathcal{N}(\widehat{f}_j C_{ij}^{(b)}, \widehat{\sigma}_j^2)$$

WEALTH OF CONFIGURATIONS

- Conditional vs. unconditional
- Nonparametric: only conditional is possible!
- Parametric: which distribution?
  - Normal
  - Gamma
- Semiparametric: which residuals?
  - Standardised
  - Studentised
  - Log-normal
- Computationally very intensive!

IMPLEMENTATION

- R package `claimsBoot`
- Front-end in R
- Heavy-duty numerical code in Fortran
- Parallelised using OpenMP
- Glued together with `Rcpp`
- Available on Github

RESULTS

- Parametric
  - Good performance
  - Unconditional better than conditional
- Semiparametric
  - Standardised & log-normal residuals yield bad results
  - Studentised residuals do reasonably well
- Nonparametric
  - Performance in-between parametric/studentised and standardised/log-normal
- Differences more noticeable closer to current calendar period
- For calendar & origin outliers: same trends, more pronounced
- Visualisation: Shiny app

The ODP model

FORMULATION

Model 2 (overdispersed Poisson GLM)

1. *The incremental claims are independent from each other*
2. *There exist parameters* c, $a_1, \ldots, a_I$ *and* $b_1, \ldots, b_I$ *such that*

$$\log(\mu_{ij}) = c + a_i + b_j$$

   with $\mu_{ij} := \mathbb{E}\left[X_{ij}\right]$ *and* $a_1 = b_1 = 0$

3. *There exists a parameter* ϕ *such that*

$$\mathrm{Var}\left[X_{ij}\right] = \phi\mu_{ij}$$

PROPERTIES

- Incremental triangle
- Belongs to family of generalised linear models
  - Extend normal linear model
  - Response can follow any distribution from the EDM family
  - Covariates related to response via link function
- Dispersion parameter allowing mean to differ from variance (cfr. Poisson)
- Fitted using quasi-maximum likelihood
- Equations solved iteratively using Fisher scoring

TRIANGLE TO REGRESSION

- Flatten triangle to obtain regression model
- Development and origin year become the covariates
- Cfr. two-way ANOVA (without interaction)
- We can adapt the regression bootstrap!

TRIANGLE TO REGRESSION

- Flatten triangle to obtain regression model
- Development and origin year become the covariates
- Cfr. two-way ANOVA (without interaction)
- We can adapt the regression bootstrap!

Origin	Dev	Value
2007	1	3511
2008	1	4001
2009	1	4355
2010	1	4295
2011	1	4150
2012	1	5102
2013	1	6283
2007	2	3215
2008	2	3702
2009	2	3932
2010	2	3455
2011	2	3747
2012	2	4548
2007	3	2266
2008	3	2278

TRIANGLE TO REGRESSION

$$\begin{array}{ccccccc}
c & & & b_2 & b_3 & b_4 & b_5 \\
 & \nwarrow & & \uparrow & \uparrow & \uparrow & \uparrow \\
 & & X_{11} & X_{12} & X_{13} & X_{14} & X_{15} \\
a_2 & \longleftarrow & X_{21} & X_{22} & X_{23} & X_{24} & \\
a_3 & \longleftarrow & X_{31} & X_{32} & X_{33} & & \\
a_4 & \longleftarrow & X_{41} & X_{42} & & & \\
a_5 & \longleftarrow & X_{51} & & & &
\end{array}$$

$$\begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ 1 & 1 & 0 & \dots & 0 \\ 1 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & 0 & \dots & 1 \end{bmatrix} \begin{bmatrix} c \\ a_2 \\ a_3 \\ \vdots \\ b_5 \end{bmatrix} = \begin{bmatrix} \log(\mu_{11}) \\ \log(\mu_{21}) \\ \log(\mu_{31}) \\ \vdots \\ \log(\mu_{15}) \end{bmatrix}$$

CONFIGURATIONS

- Parametric: which distribution?
  - Normal
  - Gamma
  - Poisson
- Semiparametric: which residuals?
  - Most popular ones for GLM: Pearson and deviance
  - Deviance suffer technical shortcoming which inhibits resampling
- Nonparametric: impossible
- Computationally very intensive!

IMPLEMENTATION

- R package `claimsBoot`
- Front-end in R
- Heavy-duty numerical code in Fortran
- Parallelised using OpenMP
- Glued together with `Rcpp`
- Available on Github

RESULTS

- Parametric outperforms semiparametric
- Differences more noticeable closer to current calendar period
- For calendar & origin outliers: same trends, more pronounced
- Visualisation: Shiny app

Conclusion

KEY TAKEAWAYS

- Parametric bootstraps perform very well
- For semiparametric bootstraps, result depends on residuals
- For Mack's model: nonparametric bootstrap performs reasonably well, but outclassed by parametric variant
- Flag suspicious datapoints by reverse-engineering the simulation process

FUTURE RESEARCH

- Other types of deviations
- Robust methods in semiparametric bootstrap