



VRIJE  
UNIVERSITEIT  
BRUSSEL



Graduation thesis submitted in partial fulfillment of the requirements for the  
degree of Master of Science in Mathematics

# EXPLORATION OF VARIOUS MACHINE LEARNING TECHNIQUES IN THE CONTEXT OF NON-LIFE INSURANCE

Othman El Hammouchi

June 2023

Promotors: prof. dr. Robin Van Oirbeek    prof. dr. Tim Verdonck

**Sciences and Bioengineering Sciences**





VRIJE  
UNIVERSITEIT  
BRUSSEL



Proefschrift ingediend met het oog op het behalen van de graad van Master of  
Science in de Wiskunde

# VERKENNING VAN VERSCHILLENDE MACHINE LEARNING-TECHNIEKEN IN DE CONTEXT VAN NON-LIFE INSURANCE

Othman El Hammouchi

Juni 2023

Promotors: prof. dr. Robin Van Oirbeek    prof. dr. Tim Verdonck

**Wetenschappen en Bio-ingenieurswetenschappen**



# Abstract

Your abstract would go here.



# Contents

<b>Abstract</b>	<b>iii</b>
<b>1 Introduction</b>	<b>1</b>
<b>2 Pattern break detection</b>	<b>3</b>
2.1 Introduction . . . . .	3
2.2 The bootstrap method . . . . .	3
2.3 Mack's model . . . . .	4
2.3.1 Choosing residuals . . . . .	6
2.3.2 Bootstrapping the parameter error . . . . .	7
2.3.3 Bootstrapping the process error . . . . .	9
2.3.4 Implementation and numerical results . . . . .	10
<b>Conclusion</b>	<b>11</b>





# Chapter 1

## Introduction

The defining characteristic of the insurance industry is the inverted nature of its production cycle. In manufacturing, commerce, transportation, etc. payment is usually received upon delivery of goods or services. By contrast, insurance products are paid up-front, long before the adverse events they provide protection against occur, if at all. Insurance contracts exchange fixed payments today (premiums) for contingent claims in the future, and the insurer must set aside a sufficient portion of the former in order to cover the latter. The *claims reserving problem* involves forecasting the funds which will be needed to settle outstanding contracts as well as their uncertainty. This poses clear challenges from a risk-management perspective, especially in long-tailed lines of business, and makes careful and rigorous statistical modelling indispensable to the actuary. Failure to do so not only invites potentially burdensome regulatory scrutiny, but can also lead to solvency problems.



## Chapter 2

# Pattern break detection

### 2.1 Introduction

The chain ladder method ranks among the most frequently applied loss reserving techniques in insurance. Originally conceived as a purely computational algorithm, various models have since been proposed to cast it in a stochastic framework. Regardless of which one the reserving actuary chooses to employ, the central assumption he is obliged to make is that the development pattern observed in earlier cohorts is applicable to later ones. While this requirement seems eminently reasonable - all models ultimately rely on the past serving as a sufficiently reliable guide to the future - it turns out to be difficult to verify quantitatively in practice.

A special difficulty which arises in the actuarial context is the relatively small quantity of data which is typically available. This dearth of observations sharply constrains the efficacy of classical statistical tests. Historical data is most often presented in the form of a *loss* or *run-off triangle*  $\mathcal{D}_I$ , which consists either of cumulative or incremental amounts of some actuarial variable (payments, number of claims, etc.), respectively denoted by  $C_{ij}$  and  $X_{ij}$ . Here  $1 \leq i \leq I$  denotes the *cohort*, *origin year* or *accident year* and  $1 \leq j \leq J$  the *development year*, so that

$$\mathcal{D}_I = \{C_{ij} \mid 1 \leq j \leq J, i + j \leq I + 1\} \quad \text{or} \quad \mathcal{D}_I = \{X_{ij} \mid 1 \leq j \leq J, i + j \leq I + 1\}. \quad (2.1)$$

To simplify the formulas, we assume throughout this exposition that  $I = J$ . Embedding  $\mathcal{D}_I$  into a matrix as the triangle on and above the anti-diagonal, the reserving actuary then seeks to estimate the *total outstanding loss liabilities*

$$R = \sum_{i=1}^I (C_{i,I} - C_{i,I+1-i}) \quad (2.2)$$

by forecasting the values in the lower triangle  $\mathcal{D}_I^c$ .

Our aim in this chapter is to use bootstrap simulation methods to investigate whether it is possible to detect structural breaks in the claims development pattern. We do this by examining the sensitivity of widely-used actuarial models to the deviations from their assumptions.

### 2.2 The bootstrap method

In its simplest form, the bootstrap is a statistical technique which employs simulation to compute estimates for the uncertainty of a given model. When applied correctly, it is an enormously

powerful method which can free us from the dilemma of either having to navigate convoluted analytical expressions which are most likely intractable, or being forced to make debilitating oversimplifications in order to make the calculations feasible. Fundamentally, bootstrapping is premised on the idea that the distribution of the observed samples serves as a good proxy for the population, and that we may therefore approximate i.i.d. sampling from the latter by resampling the former. Depending on their assumptions, we can broadly distinguish<sup>1</sup> between two classes of bootstrapping: *parametric* and *non-parametric*. The following example will illustrate both.

Suppose we are given a sample  $X_1, \dots, X_n$  of independent and identical random variables drawn from a distribution  $F$  and wish to estimate a quantity  $h(F)$  which is a function of it. Assume moreover that we have already fixed an estimator  $\widehat{h(F)} = g(X_1, \dots, X_n)$ . How can we quantify the uncertainty of our result? The non-parametric bootstrap does this by approximating  $F$  with the *empirical cumulative distribution function* (ECDF) defined by

$$\hat{F}(x) := \sum_{k=1}^n I_{\{X_k \leq x\}}. \quad (2.3)$$

We then use  $\hat{F}$  to simulate new samples  $X_1^{(b)}, \dots, X_n^{(b)}$ , which simply corresponds to drawing with replacement from  $X_1, \dots, X_n^*$ . For each of these  $B$  bootstrap samples, we can then compute  $g^{(b)} = g(X_1^{(b)}, \dots, X_n^{(b)})$ , which we think of as an approximate sample from the true distribution of  $\widehat{h(F)}$ . The uncertainty of our estimate could then for instance be quantified by

$$\frac{1}{B-1} \sum_{b=1}^B (g^{(b)} - \widehat{h(F)})^2. \quad (2.4)$$

The parametric bootstrap follows the same logic, but makes the additional assumption that  $F$  is of a known form fully determined by a set of parameters  $\theta$ . It then estimates  $\theta$  with the available data and uses the fitted model  $F_\theta$  to simulate new samples  $X_1^{(b)}, \dots, X_n^{(b)}$  and  $g^{(b)}$  with the help of a random number generator.

Although we introduced the bootstrap in this section as a device for estimating the error of our models, its scope does not limit itself to this. The output of the bootstrap gives us a complete *simulated distribution* which can be used for all kinds of inference, e.g. estimating higher moments, computing confidence intervals.

## 2.3 Mack's model

We begin by considering the distribution-free model due to Mack<sup>2</sup>, which applies to the cumulative claims triangle  $\mathcal{D}_I = (C_{ij})$ , and makes the following assumptions:

**Model 1 (Mack Chain Ladder).**

(i) *There exist development factors  $f_1, \dots, f_{I-1}$  such that*

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

*for  $1 \leq i \leq I$ .*

---

<sup>1</sup>A. C. Davison and D. V. Hinkley. *Bootstrap methods and their application*. Cambridge University Press, 1997, p. 11.

<sup>2</sup>Thomas Mack. "Distribution-free Calculation of the Standard Error of Chain Ladder Reserve Estimates". In: *Astin Bulletin* 23.2 (1993).

(ii) There exist parameters  $\sigma_1, \dots, \sigma_{I-1}$  such that

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

for  $1 \leq i \leq I$ .

(iii) Cumulative claims processes  $(C_{ij})_j, (C_{i'j})_j$  are independent for  $i \neq i'$ .

Assumption (i) states that the conditional expectation of the cumulative claim amount in any period depends only on that of the previous period, and that this dependence is moreover linear. Assumption (ii) makes a similar assertion about the conditional variance, and assumption (iii) states that the rows of our data triangle represent randomly drawn sample paths.

Estimates for the ultimate claim amounts  $C_{iI}$  are obtained by substituting the chain ladder estimates

$$\hat{f}_j(\mathcal{D}_I) = \hat{f}_j(C_{1j}, \dots, C_{I-j,j}, \dots, C_{1,j+1}, \dots, C_{I-j,j+1}) := \frac{\sum_{i=1}^{I-j} C_{i,j+1}}{\sum_{i=1}^{I-j} C_{i,j}} \quad (2.5)$$

for the unknown development factors  $f_j$  in the expression for the conditional expectation, yielding

$$\hat{C}_{i,I} := \hat{\mathbb{E}}[C_{i,I} \parallel C_{i,I-i}] = C_{i,I+1-i} \prod_{j=I-i}^{I-1} \hat{f}_j, \quad (2.6)$$

from which we obtain the reserve estimate

$$\hat{R} = g(\mathcal{D}_I) := \sum_{i=1}^I (\hat{C}_{i,I} - C_{i,I+1-i}), \quad (2.7)$$

which should, strictly speaking, more properly be called a *predictor*, as the total outstanding loss, being a function of  $(C_{iI})_{1 \leq i \leq I}$ , is a random variable. Although we shall generally follow the common practice of using both terms interchangeably, as this is rarely cause for confusion, it will sometimes prove fruitful to make this distinction, which is why we emphasize it here.

Predicting the reserve, then, can be viewed as a two-step process, the conditional mean being estimated by plugging in  $\hat{f}_j$ , and the former in turn serving to predict the ultimate loss. The error thus made can be quantified using the (conditional) *mean square error of prediction*

$$\text{MSEP}_{R|\mathcal{D}_I}(\hat{R}) := \mathbb{E}[(\hat{R} - R)^2 \parallel \mathcal{D}_I], \quad (2.8)$$

which we can divide into an *estimation error* and a *process error* corresponding to both steps of the procedure:

$$\mathbb{E}[(\hat{R} - R)^2 \parallel \mathcal{D}_I] = \mathbb{E}[(R - \mathbb{E}[R \parallel \mathcal{D}_I])^2 \parallel \mathcal{D}_I] + \mathbb{E}[(\mathbb{E}[R \parallel \mathcal{D}_I] - \hat{R})^2 \parallel \mathcal{D}_I] \quad (2.9)$$

$$\begin{aligned} & - 2\mathbb{E}[(R - \mathbb{E}[R \parallel \mathcal{D}_I])(\mathbb{E}[R \parallel \mathcal{D}_I] - \hat{R}) \parallel \mathcal{D}_I] \\ & = \text{Var}(R \parallel \mathcal{D}_I) + (\mathbb{E}[R \parallel \mathcal{D}_I] - \hat{R})^2 \\ & \quad - 2(\mathbb{E}[R \parallel \mathcal{D}_I] - \hat{R})(\mathbb{E}[R - \mathbb{E}[R \parallel \mathcal{D}_I] \parallel \mathcal{D}_I]) \end{aligned} \quad (2.10)$$

$$= \underbrace{\text{Var}(R \parallel \mathcal{D}_I)}_{\text{process error}} + \underbrace{(\mathbb{E}[R \parallel \mathcal{D}_I] - \hat{R})^2}_{\text{estimation error}}. \quad (2.11)$$

It is not difficult to see the analogy between this expression and the familiar bias-variance decomposition from classical statistics.

### 2.3.1 Choosing residuals

If we attempt to apply the methodology of Section 2.2 to (2.7), we are faced with the difficulty that the  $C_{ij}$ 's are neither independent nor share the same distribution. This problem is encountered when bootstrapping any statistical model, and is most commonly solved by resorting to the residuals obtained after fitting it. Here, the limitations of ?? with its distribution-free paradigm become apparent, and we will need to supplement it with some stronger assumptions. For now, we will use the following autoregressive Gaussian time series model:

$$C_{i,j+1} = f_j C_{i,j} + \sigma_j \sqrt{C_{i,j}} \varepsilon_{i,j} \sim \mathcal{N}(f_j C_{i,j}, \sigma_j^2 C_{i,j}). \quad (2.12)$$

It can easily be verified that this is compatible with ?. Its error terms are given by

$$\hat{\varepsilon}_{ij} := \frac{(F_{i,j} - f_j) \sqrt{C_{i,j}}}{\sigma_j}, \quad (2.13)$$

and we can obtain residuals by substituting appropriate estimators for the unknown parameters in this expression. For  $f_j$ , these will be the chain-ladder estimators 2.5, but we haven't discussed how to estimate the  $\sigma_j$ 's yet. The next theorem, which we state without proof<sup>3</sup>, remedies this, as well as giving some useful properties of these estimators.

**Theorem 1.** *The chain ladder estimators  $\hat{f}_j$  are (conditionally) unbiased and uncorrelated. Moreover, if we define the individual development factors as*

$$F_{i,j+1} := \frac{C_{i,j+1}}{C_{i,j}} \quad (2.14)$$

then

$$\hat{\sigma}_j := \frac{1}{I-j} \sum_{i=1}^{I-j} C_{i,j} (F_{i,j} - \hat{f}_j)^2 \quad (2.15)$$

are (conditionally) unbiased estimators for the parameters  $\sigma_1, \dots, \sigma_I$ .

For the residuals, a multitude of different choices are available. The most obvious is to simply use the *raw residuals*

$$\hat{\varepsilon}_{ij} := \frac{(F_{i,j} - \hat{f}_j) \sqrt{C_{i,j}}}{\hat{\sigma}_j}, \quad (2.16)$$

which have the advantage of being intuitive and simple to implement.

Note that the recursive dependence of loss amounts with the same origin year makes this a non-trivial task. Had we been dealing with a simple linear regression of pairs  $(X_1, Y_1), \dots, (X_n, Y_n)$ , for example, then we could use the fitted model and resampled residuals directly to generate simulated responses

$$Y_i^* := \hat{\beta}_0 + \hat{\beta}_1 X_i + \hat{\varepsilon}_i^*. \quad (2.17)$$

By contrast, it is not immediately clear how a resampled loss triangle can be obtained from  $\hat{f}_j$ ,  $\hat{\sigma}_j$  and  $\hat{\varepsilon}_{ij}^*$ . A number of different approaches have been proposed in the literature<sup>4</sup>, where they are framed in terms of producing a bootstrap estimate of (2.8), which is why we choose to adopt this point of view as well. Our exposition closely follows [5, pp. 45 sqq.], focusing on two methods in particular, which, for reasons that will become apparent shortly, we will refer to as the *conditional* and *unconditional* approach.

<sup>3</sup>The relevant proofs can be found in 5, pp. 17 sqq.

<sup>4</sup>See, for example, 4, Section D6; 6.

Our first task is to obtain a more manageable expression for the total MSEP by breaking it down into components which depend on the individual accident years:

$$\mathbb{E}[(\hat{R} - R)^2 \mid \mathcal{D}_I] = \sum_{i=1}^I \text{Var}(C_{iI} - C_{i,I+1-i} \mid \mathcal{D}_I) + \left( \sum_{i=2}^I \mathbb{E}[(\hat{C}_{iI} - C_{iI}) \mid \mathcal{D}_I] - \sum_{i=2}^I (\hat{C}_{iI} - C_{i,I+1-i}) \right)^2 \quad (2.18)$$

$$= \sum_{i=2}^I \text{Var}(C_{iI} \mid \mathcal{D}_I) + \left( \sum_{i=2}^I \mathbb{E}[C_{iI} \mid \mathcal{D}_I] - \hat{C}_{iI} \right)^2 \quad (2.19)$$

$$= \sum_{i=2}^I \underbrace{\left( \text{Var}(C_{iI} \mid \mathcal{D}_I) + (\mathbb{E}[C_{iI} \mid \mathcal{D}_I] - \hat{C}_{iI})^2 \right)}_{=: \text{MSEP}_{C_{iI} \mid \mathcal{D}_I}(\hat{C}_{iI})} + \sum_{\substack{2 \leq i, j \leq I \\ i \neq j}} (\mathbb{E}[C_{iI} \mid \mathcal{D}_I] - \hat{C}_{iI})(\mathbb{E}[C_{jI} \mid \mathcal{D}_I] - \hat{C}_{jI}). \quad (2.20)$$

The problem has thus been reduced to estimation of the MSEP for a single accident along with some additional cross terms. The latter are only relevant for obtaining a closed formula and we will therefore ignore them in the sequel. It then remains for us to deal with the parameter and process error of individual accident years.

### 2.3.2 Bootstrapping the parameter error

Using the definitions from the previous section, we can write the process error for a single accident year  $i \in \{1, \dots, I\}$  as

$$(\mathbb{E}[C_{iI} \mid \mathcal{D}_I] - \hat{C}_{iI})^2 = C_{i,I+1-i}^2 \left( \prod_{j=I+1-i}^{I-1} f_j - \prod_{j=I+1-i}^{I-1} \hat{f}_j \right)^2 \quad (2.21)$$

$$= C_{i,I+1-i}^2 \left( \prod_{j=I+1-i}^{I-1} f_j^2 + \prod_{j=I+1-i}^{I-1} \hat{f}_j^2 - 2 \prod_{j=I+1-i}^{I-1} f_j \hat{f}_j \right). \quad (2.22)$$

Notice that this expression contains the unknown development factors  $f_j$  and so cannot be computed directly. If we tried to estimate it by substituting the  $\hat{f}_j$ 's for them, however, the result would be a constant 0, which is clearly not accurate. We must therefore find a way to express the variability of  $\hat{f}_j$  around  $f_j$ . Until now, we have worked under the assumption that  $\mathcal{D}_I$  is held fixed, which is problematic because, conditional on  $\mathcal{D}_I$ ,  $\hat{f}_j$  is a scalar. Relaxing this assumption therefore means reducing the dataset we condition on.

It's instructive to reflect upon the relation between

They can be most easily understood if we think of conditioning as providing a 'recipe' for a particular kind of resampling, e.g.  $\mathbb{E}[X \mid Y]$  corresponds to sampling observations from  $X$  for a fixed value of  $Y$  and taking their average. Unsurprisingly, this point of view will allow us to transition easily from analytic formulae to a bootstrapping procedure.

If we define the subset of observation up to (and including) development year  $k$  as

$$\mathcal{B}_k := \{C_{ij} \in \mathcal{D}_I \mid j \leq k\}, \quad (2.23)$$

then the unconditional approach fixes  $\mathcal{B}_{I+1-i}$  in accident year  $i$ , resampling the observations in

$$\mathcal{D}_{I,i}^O := \{C_{ij} \in \mathcal{D}_I \mid j > I + 1 - i\} . \quad (2.24)$$

completely, which leads to the estimate

$$\mathbb{E}[(\mathbb{E}[C_{iI} \mid \mathcal{D}_I] - \hat{C}_{iI})^2 \mid \mathcal{B}_{I+1-i}] = C_{i,I+1-i}^2 \mathbb{E} \left[ \left( \prod_{j=I+1-i}^{I-1} f_j - \prod_{j=I+1-i}^{I-1} \hat{f}_j \right)^2 \mid \mathcal{B}_{I+1-i} \right] \quad (2.25)$$

$$= C_{i,I+1-i}^2 \left( \mathbb{E} \left[ \prod_{j=I+1-i}^{I-1} \hat{f}_j^2 \mid \mathcal{B}_{I+1-i} \right] - \prod_{j=I+1-i}^{I-1} \hat{f}_j^2 \right) , \quad (2.26)$$

where we used the fact that the  $\hat{f}_j$ 's are uncorrelated.

then we can view (2.25) as resampling  $\mathcal{D}_{I,i}^O$  completely, computing  $\prod_{j=I+1-i}^{I-1} \hat{f}_j^2$  and taking the average.

---

**Algorithm 1** Unconditional resampling

---

**Input:** Loss triangle  $\mathcal{D}_I = (C_{ij})_{1 \leq i, j \leq I}$ , number of iterations  $N$ , residuals  $(\varepsilon_{ij})_{1 \leq i, j \leq I-1}$

**Output:** Simulated reserve sample  $\hat{R}_1^*, \dots, \hat{R}_N^*$

```

for  $l \leftarrow 1, N$  do
  for  $k \leftarrow 2, I$  do
    for  $i \leftarrow 1, I + 1 - k$  do
       $C_{i,I+1-k}^* \leftarrow C_{i,I+1-k}$ 
    end for
    for  $j \leftarrow I + 1 - k, I - 1$  do
      for  $i \leftarrow 1, I + 1 - j$  do
         $F_{ij}^* \leftarrow \hat{f}_j + \frac{\hat{\sigma}_j}{\sqrt{C_{ij}^*}} \varepsilon_{ij}^*$ 
         $C_{i,j+1}^* \leftarrow F_{ij}^* C_{ij}^*$ 
      end for
       $\hat{f}_j^* \leftarrow \frac{\sum_{i=1}^{I-j} F_{ij}^* C_{ij}^*}{\sum_{i=1}^{I-j} C_{ij}^*}$ 
    end for
     $\hat{C}_{i,I} \leftarrow \prod_{j=I-i}^{I-1} \hat{f}_j^* C_{i,I+1-i}$ 
     $\hat{R}_{i,l}^* \leftarrow \hat{C}_{i,I} - C_{i,I+1-i}$ 
  end for
   $\hat{R}_l^* \leftarrow \sum_{i=2}^I \hat{R}_{i,l}^*$ 
end for
```

---

By contrast, the conditional approach only allows us to vary a given factor  $\hat{f}_j$  over the observations after  $j$ , so that every point in  $\mathcal{D}_{I,i}^O$  (except the upper right corner) will be conditioned on for at least one  $\hat{f}_j$  when resampling. Phrased in terms of (2.12), we keep the original  $C_{ij}$



fixed at every step of the time series and only resample the next value  $C_{i,j+1}$ . Writing  $\mathbb{P}_{\mathcal{D}_I}^*$  for the probability measure induced on  $\mathcal{D}_{I,i}^O$  by this process, our estimate for the parameter error in the conditional approach then becomes

$$\mathbb{E}_{\mathbb{P}_{\mathcal{D}_I}^*} \left[ (\mathbb{E}[C_{iI} \parallel \mathcal{D}_I] - \hat{C}_{iI})^2 \right] = C_{iI}^2 \mathbb{E}_{\mathbb{P}_{\mathcal{D}_I}^*} \left[ \left( \prod_{j=I+1-i}^{I-1} f_j - \prod_{j=I+1-i}^{I-1} \hat{f}_j \right)^2 \right] \quad (2.27)$$

$$= C_{i,I+1-i}^2 \left( \mathbb{E}_{\mathbb{P}_{\mathcal{D}_I}^*} \left[ \prod_{j=I+1-i}^{I-1} \hat{f}_j^2 \right] - \prod_{j=I+1-i}^{I-1} f_j^2 \right) \quad (2.28)$$

$$= C_{i,I+1-i}^2 \left( \prod_{j=I+1-i}^{I-1} \mathbb{E}[\hat{f}_j^2 \parallel \mathcal{B}_j] - \prod_{j=I+1-i}^{I-1} f_j^2 \right). \quad (2.29)$$

---

**Algorithm 2** Conditional resampling

---

**Input:** Loss triangle  $\mathcal{D}_I = (C_{ij})_{1 \leq i,j \leq I}$ , number of iterations  $N$ , residuals  $(\varepsilon_{ij})_{1 \leq i,j \leq I-1}$

**Output:** Simulated reserve sample  $\hat{R}_1^*, \dots, \hat{R}_N^*$

```

for  $l \leftarrow 1, N$  do
  for  $k \leftarrow 2, I$  do
    for  $j \leftarrow I+1-k, I-1$  do
      for  $i \leftarrow 1, I+1-j$  do
         $F_{ij}^* \leftarrow \hat{f}_j + \frac{\hat{\sigma}_j}{\sqrt{C_{ij}}} \varepsilon_{ij}^*$ 
      end for
       $\hat{f}_j^* \leftarrow \frac{\sum_{i=1}^{I-j} F_{ij}^* C_{ij}}{\sum_{i=1}^{I-j} C_{ij}}$ 
    end for
     $\hat{C}_{i,I} \leftarrow \prod_{j=I-i}^{I-1} \hat{f}_j^* C_{i,I+1-i}$ 
     $\hat{R}_{i,l}^* \leftarrow \hat{C}_{i,I} - C_{i,I+1-i}$ 
  end for
   $\hat{R}_l^* \leftarrow \sum_{i=2}^I \hat{R}_{i,l}^*$ 
end for

```

---

We will implement both approaches in Section 2.3.4.

### 2.3.3 Bootstrapping the process error

Moving on to the remaining component in (2.20), which quantifies the intrinsic variation of the ultimate loss amount, conditional on the observations in  $\mathcal{D}_I$ . Observe that this term only arises because we are dealing with a prediction problem, so that our target is a random variable rather than some fixed parameter. This suggests the following method for incorporating the process error into our bootstrap procedure. At each iteration of the innermost loop in algorithm 1 and

section 2.3.2, we additionally compute the resampled parameters, respectively

$$\hat{\sigma}_j^* = \frac{1}{I-j} \sum_{i=1}^{I-j} C_{ij}^* \left( F_{ij}^* - \hat{f}_j^* \right)^2 \quad (2.30)$$

and

$$\hat{\sigma}_j^* = \frac{1}{I-j} \sum_{i=1}^{I-j} C_{ij} \left( F_{ij}^* - \hat{f}_j^* \right)^2. \quad (2.31)$$

We then use these to simulate lower pseudo-triangles  $(\mathcal{D}_I^c)^* = (C_{ij}^*)$ , which in turn yield simulated reserve samples

$$R^* := \sum_{i=2}^I (C_{iI}^* - C_{i,I+1-i}). \quad (2.32)$$

Notice that this allows us to bypass (2.7) completely, although we still expect that

$$\frac{1}{N} \sum_{k=1}^N R_k^* \approx \hat{R}. \quad (2.33)$$

In view of (2.12), the most evident way to generate  $(\mathcal{D}_I^c)^*$  is to start from the antidiagonal in  $\mathcal{D}_I$  and successively sample

$$C_{i,j+1}^* \sim \mathcal{N}(\hat{f}_j^* C_{ij}, \hat{\sigma}_j^* C_{ij}). \quad (2.34)$$

This approach has a major drawback, however: it makes it possible to draw negative samples for the cumulative loss amounts. One way to remedy this would obviously be to simply discard a simulated triangle as soon as it contains a negative value.

An alternative is to follow the suggestion given in [2, p. 238] and substitute in place of (2.34) a gamma distribution with the same mean and variance. If we write  $C_{ij} \sim \Gamma(\alpha, \beta)$ , this means that  $\alpha, \beta$  must satisfy

$$\frac{\alpha}{\beta} = f_{j-1} C_{i,j-1} \quad \text{and} \quad \frac{\alpha}{\beta^2} = \sigma_{j-1}^2 C_{i,j-1}, \quad (2.35)$$

from which it follows that

$$\alpha = \frac{f_{j-1}^2 C_{i,j-1}}{\sigma_{j-1}^2} \quad \text{and} \quad \beta = \frac{f_{j-1}}{\sigma_{j-1}^2}. \quad (2.36)$$

This ensures that Mack's assumptions are still valid while avoiding nonsensical outcomes.

### 2.3.4 Implementation and numerical results

# Conclusion



# Bibliography

- [1] A. C. Davison and D. V. Hinkley. *Bootstrap methods and their application*. Cambridge University Press, 1997.
- [2] P. D. England and R. J. Verrall. “Predictive Distributions of Outstanding Liabilities in General Insurance”. In: *Annals of Actuarial Science* 1.1 (2006).
- [3] Thomas Mack. “Distribution-free Calculation of the Standard Error of Chain Ladder Reserve Estimates”. In: *Astin Bulletin* 23.2 (1993).
- [4] Thomas Mack. “Measuring the Variability of Chain Ladder Reserve Estimates”. In: *Claims Reserving Manual*. v. 2. Institute of Actuaries, 1997.
- [5] Mario V. Wüthrich and Michael Merz. *Stochastic Claims Reserving Methods in Insurance*. John Wiley & Sons, 2008.
- [6] Mario V. Wüthrich et al. “The Mean Square Error of Prediction in the Chain Ladder Reserving Method (Mack and Murphy Revisited)”. In: *Astin Bulletin* 36.2 (2006).