



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 . . . . .	4
2.2.1 Bootstrapping a classical estimator . . . . .	4
2.2.2 Bootstrapping a regression model . . . . .	5
2.3 Mack's model . . . . .	6
2.3.1 A challenging simulation . . . . .	8
2.3.2 Parameter error . . . . .	8
2.3.3 Process error . . . . .	11
2.3.4 Numerical implementation and results . . . . .	14
2.4 Poisson GLM . . . . .	14
2.4.1 Bootstrap methodology . . . . .	14
2.4.2 Numerical implementation and results . . . . .	14
<b>Conclusion</b>	<b>15</b>





## Chapter 1

# Introduction



## Chapter 2

# Pattern break detection

### 2.1 Introduction

The most defining characteristic of the insurance industry is the inverted nature of its production cycle. In manufacturing, commerce, transport, etc., payment is usually only received upon delivery of goods or services. By contrast, insurance products are usually purchased long before the adverse events which they protect against have occurred, if they ever do. Insurers therefore face the challenge of forecasting the amount and variability of funds needed to settle outstanding contracts, a process known as *claims reserving*. In this the reserving actuary relies historical data which 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

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

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

$$R = \sum_{i=2}^I (C_{i,I} - C_{i,I+1-i})$$

by forecasting the values in the lower triangle  $\mathcal{D}_I^c$ . A special difficulty arising in the actuarial context is the relatively small number of observations which is usually available.

One of the most frequently used loss reserving techniques in practice is the *chain ladder*, a recursive method which estimates cumulative claim in development year  $j$  by multiplying the previous year's claim by a so-called *age-to-age factor* or *link ratio*

$$f_j := \frac{\sum_{i=1}^{I-j} C_{i,j+1}}{\sum_{i=1}^{I-j} C_{i,j}}.$$

It was originally conceived as a purely computational algorithm, but there have since been various attempts to frame it in terms of a stochastic model. The central assumption it makes is that the pattern observed in earlier cohorts is applicable to later ones. In one sense, this is of course perfectly reasonable: all models ultimately use the past as a guide to the future. The dearth of data typically available to the actuary makes it challenging to verify its validity, as it limits the

efficacy of classical statistical techniques. In particular, this makes it difficult to detect structural breaks in the claims development pattern.

Our aim in this chapter is to investigate whether it is possible to use bootstrap simulations to remedy this problem. Specifically, we start from two widely-used chain ladder models (Mack's model and the Poisson GLM) and simulate run-off triangles which perfectly follow their assumptions. We then perturb the triangles in a myriad of ways and calculate a bootstrap reserve for the resulting data, allowing us to investigate how the simulated reserve is impacted by deviations from the model assumptions.

The chapter is divided as follows. The next section introduces the bootstrap method by means of a simple example from classical statistics, and explains how it can be applied to regression models, which will

## 2.2 The bootstrap method

When using a statistical model to describe a dataset in terms of a reduced number of parameters, we are not only interested in producing *estimates* of these parameters, but also in quantifying their *uncertainty*. In the classical theory, the usual approach to achieve this is to start from the model assumptions and derive from them analytically the sampling distribution of the estimators. In most cases (the Gaussian model being a notable exception) this leads to intractable calculations, so that one is either forced to rely on approximations and asymptotic results, or make unrealistic simplifying assumptions. Moreover, estimates obtained in this way often heavily depend on their underlying assumptions, which can potentially lead to gross errors if these are violated.

The bootstrap method aims to remedy this problem by using numerical simulations to compute estimates of model uncertainty. At its core, it is premised on the idea that the empirical distribution of the sample forms a good proxy for that the population distribution. Consequently, we can approximate sampling from the population by *resampling our data*, which, to the uncaring observer, can give the impression that we're 'magically' producing new information, using our single sample to 'pull ourselves up by our own bootstraps', which is where the procedure derives its name from. The next section shows how this can be applied concretely for a very simple estimation problem. After that, we discuss the special case which will be of primary interest to us, namely bootstrapping regression models.

### 2.2.1 Bootstrapping a classical estimator

Let  $X_1, \dots, X_n$  be an i.i.d. sample drawn from some distribution  $F$ . Suppose we have estimated the quantity  $h(F)$  with  $\widehat{h(F)} = g(X_1, \dots, X_n)$  and want to bootstrap an estimate of the uncertainty of our result, e.g. the variance of  $\widehat{h(F)}$ . In general, we can distinguish between *parametric* and *non-parametric* bootstrap procedures (see for example [1, Chapter 2]), which differ in the assumptions they make.

In the non-parametric bootstrap, we use the data directly, drawing with replacement from  $X_1, \dots, X_n$  to simulate new samples  $X_1^{(b)}, \dots, X_n^{(b)}$ . This can be viewed equivalently as approximating  $F$  with the *empirical cumulative distribution function*

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

In our example, we would then compute  $g^{(b)} = g(X_1^{(b)}, \dots, X_n^{(b)})$  for every bootstrap sample,

yielding an approximate sample from the sampling distribution of  $\widehat{h(F)}$ , provided  $F \approx \hat{F}_n$  with sufficient accuracy. Denoting the total number of bootstrap samples by  $B$ , we can estimate the variance of  $\widehat{h(F)}$  by

$$\frac{1}{B-1} \sum_{b=1}^B (g^{(b)} - \bar{g})^2,$$

with  $\bar{g} = \frac{1}{B} \sum_{b=1}^B g^{(b)}$ .

The parametric bootstrap, by contrast, first uses the data to fit a model, from which it then produces simulated samples with the help of a random number generator. In our example, this means that we would have to assume that  $F$  belongs to some parametric family  $\{F_\theta \mid \theta \in \Theta\}$ . The sample  $X_1, \dots, X_n$  is then used to produce an estimate  $\hat{\theta}$ , and plugging this in gives us  $F_{\hat{\theta}}$  which we can use to simulate  $X_1^{(b)}, \dots, X_n^{(b)}$  and  $g^{(b)}$  as before.

As always, the parametric approach has the advantage of being more powerful when its assumptions are met, at the cost of making gross errors if they are violated. It is also instructive to observe that, although our example used the bootstrap to calculate a single statistic, its application naturally yields a whole *simulated distribution*, which can be used for any kind of inference. This shows the tremendous potential of bootstrap as a tool for statistical analysis, if properly applied and interpreted.

### 2.2.2 Bootstrapping a regression model

We now turn our attention towards bootstrap methods for regression models, as these will be our primary focus in the next sections. Let  $X_1, \dots, X_p$  and  $Y$  be a set of random variables whose relationship is modelled by means of a parametrised mapping  $f(X_1, \dots, X_p; \beta)$ . Given a sample  $(X_1, Y_1), \dots, (X_n, Y_n)$ , we can apply a fitting procedure to obtain an estimate  $\hat{\beta}$  of  $\beta$ . For new values  $x_1^+, \dots, x_n^+$  of the regressors, we can then predict the response  $y^+$  by  $f(x_1^+, \dots, x_n^+; \hat{\beta})$ . It is important to understand that these are two distinct problems, which lead to different bootstrap procedures (see [1, Sections 6.3.3 and 7.2.4]).

In *estimation*, we seek to identify the value of a fixed but unknown quantity. With *prediction*, our aim is to forecast the value of a random quantity, usually by means of some measure of location. If we're using the least squares criterion, for example, then we know that the optimal predictor of  $Y$  is the conditional expectation  $\mathbb{E}[Y \mid X = x]$ ; this is an ordinary function which returns a fixed number for every value of  $x$ , and which can therefore be estimated. We do this by assuming that it can be approximated by some well-known class of functions which depend on some parameters  $\beta$ . We

we additionally have to incorporate the intrinsic randomness of the response variable.

To illustrate this more clearly, we consider the classic example of the normal linear model

$$Y = \sum_{k=1}^p \beta_k X_k + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, 1).$$

Reasoning as in the previous section, we need to identify a fundamental unit of the resampling, taking care to ensure that this yields quantities which are i.i.d. Considering the non-parametric case first, one option would be to resample the pairs  $(X_i, Y_i)$  (see [1, Algorithm 6.2]); this is particularly appropriate when the regressors are stochastic (i.e. nondegenerate), as is the case for socio-economic data, for instance.

An alternative is to first fit the model and use it to compute some kind of standardised residuals  $r_i$ , which are then resampled by the bootstrap procedure (see [1, Algorithm 6.1]).

This approach is sometimes called "semiparametric", as it makes use of parameter estimates to produce the residuals, but makes no assumptions regarding their distribution (aside from them having common one). Choosing the (externally) studentised residuals

$$r_i^S = \frac{Y_i - X_i^T \hat{\beta}}{\hat{\sigma}_{(i)} \sqrt{1 - h_{ii}}},$$

for example, we can resample these to obtain bootstrap samples  $r_1^{S,(b)}, \dots, r_n^{S,(b)}$ . These will then yield bootstrap responses

$$Y^{(b)} := \sum_{k=1}^p \beta_k X_k + \hat{\sigma} r_k^{S,(b)},$$

which in turn can be used to obtain bootstrap samples  $\hat{\beta}^{(b)}$  by fitting the model to the new data. Note that only the responses are randomised in this procedure; it would therefore be ill-suited when the independent variables are not fixed quantities.

If we are mainly concerned with estimation, then our task is done at this point, as we have obtained an approximation for the sampling distribution of  $\hat{\beta}$ . If, on the other hand, our aim is prediction, then we still need to bring in the variability inherent in the response. For a new point  $x^+ = (x_1^+, \dots, x_n^+)$ , this can be done by resampling the residuals a second time to produce bootstrap responses

$$Y_*^{(b)} := \sum_{k=1}^p \beta_k x_k^+ + \hat{\sigma} r_*^{(S,(b))},$$

which gives us an approximation to the distribution of  $Y_{x^+}$ .

The development of the parametric case follows the same logic, the only difference being that we generate bootstrap residuals directly from some distribution derived from the model with the help of a random number generator. Our example exhibits the most favourable circumstances for this approach, as in this case we have theoretical results stating that they follow a  $t$ -distribution with  $n - p - 1$  degrees of freedom. In most other circumstances, notably with generalised linear models, we will not be so lucky.

Having introduced the bootstrap method, we can now begin our discussion of the first model.

## 2.3 Mack's model

Mack's model (originally proposed in [5]) is a distribution-free model applying to the cumulative claims triangle  $\mathcal{D}_I = \{C_{ij} \mid 1 \leq i \leq I, i + j \leq I + 1\}$ . It 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  $1 \leq i \leq I$ .*

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

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

*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 the data triangle represent randomly drawn sample paths.

The chain ladder link ratios are now viewed as estimators of the development factors:

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

As for the  $\sigma_j$ 's, if we define the individual development factors as

$$F_{i,j+1} := \frac{C_{i,j+1}}{C_{ij}}$$

then these are estimated by

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

Under the assumptions of Model 1,  $\hat{f}_j$  and  $\hat{\sigma}_j$  are (conditionally) unbiased, and the  $\hat{f}_j$ 's are moreover uncorrelated (the proofs can be found in [10, pp. 17 sqq.]). Estimates for the ultimate claim amounts  $C_{iI}$  are obtained by substituting the chain ladder estimates for the unknown development factors  $f_j$  in the expression for the conditional expectation, yielding

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

from which we obtain the reserve estimate

$$\hat{R} = g(\mathcal{D}_I) := \sum_{i=2}^I (\hat{C}_{i,I} - C_{i,I+1-i}). \quad (2.1)$$

Strictly speaking,  $\hat{R}$  should more properly be referred to as a predictor, as the total outstanding loss is a random variable, being a function of  $(C_{iI})_{2 \leq i \leq I}$ . While it is typically harmless to use the terms interchangeably, the distinction between them becomes important in the context of the bootstrap, as we saw in Section 2.2, and we therefore emphasize it here. We predict the ultimate loss using the conditional mean, and we estimate the conditional mean by plugging in the  $\hat{f}_j$ 's; these steps represent the two sources of variability in our model. The total uncertainty can be quantified using the (conditional) *mean square error of prediction*

$$\text{MSEP}(\hat{R}) := \mathbb{E} \left[ (\hat{R} - R)^2 \mid \mathcal{D}_I \right],$$

and this admits a corresponding decomposition into *parameter* or *estimation error* and a *process*

error:

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

It is not difficult to see the analogy between this expression and the familiar bias-variance decomposition from classical statistics. It will prove to be immensely valuable for our development of a bootstrap procedure in the sequel.

### 2.3.1 A challenging simulation

The nonparametric and recursive nature of Mack's model makes applying the theory from Section 2.2.2 a non-trivial task. A first difficulty is how we should choose the residuals. Our definition has to ensure that they are identically distributed, but it is impossible to guarantee when our assumptions only deal with the first two moments. Even if we find it reasonable to extrapolate from second order statements to homogeneity of the distributions, it still remains unclear how we can obtain a resampled loss triangle from  $\hat{f}_j$ ,  $\hat{\sigma}_j$  and  $r_{ij}$ .

Although the literature does not address this problem directly, discussions of resampling methods have come up in the context of deriving a closed-form expression for the MSE (see [11, 7, 4, 9]). To understand this connection, it's instructive to reflect upon the relationship which exists between certain resampling schemes and analytical probability statements. For example, taking the conditional expectation  $\mathbb{E}[X \mid Y]$  would correspond to sampling a large number of observations from  $X$  while keeping  $Y$  fixed and computing their average. We will use this point of view in the sequel to derive a bootstrap method for Mack's model. Our exposition follows the one in [10, pp. 45 sqq.], albeit with a focus on simulation.

For any  $i \in \{1, \dots, I\}$ , consider the decomposition of the MSE for the ultimate claim amount  $C_{iI}$  into parameter and process error:

$$\text{MSE}_{C_{iI} \mid \mathcal{D}_I}(\hat{C}_{iI}) = (\mathbb{E}[C_{iI} \mid \mathcal{D}_I] - \hat{C}_{iI})^2 + \text{Var}(C_{iI}) \quad (2.2)$$

This gives us the two components of the model uncertainty for accident year  $i$  which our bootstrap method will have to account for. We will consider each of them in turn.

### 2.3.2 Parameter error

Starting with the parameter error, we can use the definitions from the previous section to rewrite this as

$$\begin{aligned}
(\mathbb{E}[C_{iI} \mid \mathcal{D}_I] - \hat{C}_{iI})^2 &= C_{i,I+1-i}^2 \left( \prod_{j=I+1-i}^{I-1} f_j - \prod_{j=I+1-i}^{I-1} \hat{f}_j \right)^2 \\
&= C_{i,I+1-i}^2 \left( \prod_{j=I+1-i}^{I-1} f_j^2 + \prod_{j=I+1-i}^{I-1} \hat{f}_j^2 - 2 \prod_{j=I+1-i}^{I-1} f_j \hat{f}_j \right).
\end{aligned} \quad (2.3)$$



Notice that this expression depends on the unknown development factors  $f_j$  and so cannot be computed directly. If we substituted  $\hat{f}_j$  in place of these, however, the result would be a constant 0, which is clearly not an accurate estimate. This problem was recognised by Mack himself in [6, Appendix D]. It is caused by the fact that, conditional on  $\mathcal{D}_I$ , the development factors  $\hat{f}_j$  are scalars. We must therefore find a way to allow  $\hat{f}_{I+1-i}, \dots, \hat{f}_{I-1}$  to vary around  $f_{I+1-i}, \dots, f_{I-1}$ , which means reducing the set we condition on. Depending on our choice, this will yield different estimates of eq. (2.3) with corresponding resampling schemes. We will discuss two in particular here.

It will be useful to introduce the notation

$$\mathcal{B}_k := \{C_{ij} \in \mathcal{D}_I \mid j \leq k\}$$

for the subset of observation up to and including development year  $k$ , and

$$\mathcal{D}_{I,k}^O := \{C_{ij} \in \mathcal{D}_I \mid j > I + 1 - k\}$$

for its complement in  $\mathcal{D}_I$ .

One possible approach would be to condition on  $\mathcal{B}_{I+1-i}$ , leading to the estimate

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

for the parameter error, where we used the fact that the  $\hat{f}_j$ 's are uncorrelated. This corresponds to averaging over the distribution of  $\mathcal{D}_{I,k}^O$ . Expressed in terms of resampling, it means that we generate new observations in the entire upper right triangle in order to produce new samples  $\hat{f}_{I+1-i}, \dots, \hat{f}_{I-1}$ . Borrowing the nomenclature from [11], we call this the *unconditional approach* to resampling.

An alternative is to average each  $\hat{f}_j$  only over the observations after  $j$ , which is equivalent to averaging over  $C_{1,j+1}, \dots, C_{I-j,j+1}$ , since these are the only relevant such points for the chain ladder estimator, and keeping the denominator  $\sum_{i=1}^{I-j} C_{ij}$ . Formally, this means that we estimate eq. (2.3) by taking the expectation with respect to the probability measure  $\mathbb{P}_{\mathcal{D}_I}^{(i)}$  on  $\mathcal{D}_{I,i}^O$  by

$$\mathbb{P}_{\mathcal{D}_I}^{(i)}(dx_{i,I+1-j}, \dots, dx_{i,I-1}) := \prod_{j=1} \mathbb{P}_{C_{i,j+1}}(dx_{i,j+1} \mid C_{ij} = c_{ij}),$$

where  $c_{ij}$  denotes the observed value of  $C_{ij}$ , yielding

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

We refer to this as the *conditional approach*. It corresponds to a resampling scheme in which, at every step  $j$ , only the observations from the next period are resampled in order to produce a new realisation of  $\hat{f}_j$ .

Note that our statements about resampling have so far been with respect to theoretical distributions; we have not yet specified the mechanism by which to implement this in the bootstrap. This will require us to define suitable residuals, i.e. ones which are identically distributed; however, as noted previously, Model 1 provides no way to do this. Starting from the observation that the *raw residuals*

$$e_i := C_{i,j+1} - \hat{C}_{i,j+1} = C_{i,j+1} - \hat{f}_j C_{ij}$$

suffer from heteroscedasticity,

$$\text{Var}(e_i \parallel C_{ij}) = \sigma_j^2 \left( C_{ij} - \frac{C_{ij}}{\sum_{i=1}^{I-j} C_{ij}} \right),$$

an intuitive idea would be to consider

$$r_{ij} := \frac{e_i}{\sqrt{C_{ij} - \frac{C_{ij}^2}{\sum_{i=1}^{I-j} C_{ij}}}} = \frac{(F_{i,j+1} - \hat{f}_j) \sqrt{C_{i,j}}}{\sqrt{1 - \frac{C_{ij}}{\sum_{i=1}^{I-j} C_{ij}}}}, \quad (2.4)$$

which satisfy

$$\mathbb{E}[r_{ij} \parallel C_{i,j-1}] = 0, \quad \text{Var}(r_{ij} \parallel C_{i,j-1}) = \sigma_j^2.$$

The problem is that our model don't allow us to say anything more than this, because it only makes second-order statements. Of course, we could choose to extrapolate from this to say that the  $r_{ij}$ 's are identically distributed, but this is an additional assumption which should not be made lightly; see [2] for an example where the percentiles of error distribution depend on the regressor.

Alternatively, we could deliberately choose to augment our assumptions so that we can make more precise statements. In particular, we will consider the autoregressive Gaussian time series model:

$$C_{i,j+1} = f_j C_{ij} + \sigma_j \sqrt{C_{ij}} \varepsilon_{i,j}, \quad \varepsilon \sim \mathcal{N}(0, 1), \quad (2.5)$$

which can easily be seen to be compatible with Model 1. This will allow us to choose residuals whose distribution can be derived analytically. To see how, observe that the chain ladder can be viewed as a collection of weighted linear regressions through the origin (i.e. without intercept term). Indeed, for any development year  $j \in \{1, \dots, I-1\}$ , the assumptions of Model 1 can be equivalently stated as

$$\begin{cases} \mathbb{E} \left[ \begin{bmatrix} C_{1,j+1} \\ \vdots \\ C_{I-j,j+1} \end{bmatrix} \parallel \begin{bmatrix} C_{1,j} \\ \vdots \\ C_{I-j,j} \end{bmatrix} \right] = f_j \begin{bmatrix} C_{1,j+1} \\ \vdots \\ C_{I-j,j+1} \end{bmatrix} \\ \text{Var} \left( \begin{bmatrix} C_{1,j+1} \\ \vdots \\ C_{I-j,j+1} \end{bmatrix} \parallel \begin{bmatrix} C_{1,j} \\ \vdots \\ C_{I-j,j} \end{bmatrix} \right) = \sigma_j^2 \begin{bmatrix} C_{1j} & & \\ & \ddots & \\ & & C_{I-j,j} \end{bmatrix}, \end{cases}$$

to which the weighted least squares method with weights matrix

$$\mathbf{W} = \frac{1}{\sigma_j^2} \begin{bmatrix} 1/C_{1j} & & \\ & \ddots & \\ & & 1/C_{I-j,j} \end{bmatrix},$$

can be applied. Writing

$$\mathbf{c}_j \stackrel{\text{not.}}{=} \begin{bmatrix} C_{1,j} \\ \vdots \\ C_{I-j,j} \end{bmatrix},$$

we then estimate the development factor  $f_j$  using

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

which is precisely the chain ladder estimator. The corresponding hat matrix is given by

$$\begin{aligned} \mathbf{H} &= \mathbf{c}_j (\mathbf{c}_j^T \mathbf{W} \mathbf{c}_j)^{-1} \mathbf{c}_j^T \mathbf{W} \\ &= \frac{1}{\sum_{i=1}^{I-j} C_{ij}} \begin{bmatrix} C_{1j} & \cdots & C_{1j} \\ \vdots & \ddots & \vdots \\ C_{I-j,j} & \cdots & C_{I-j,j} \end{bmatrix}, \end{aligned}$$

which yields the *modified residuals*

$$r_{i,j+1} := \frac{e_i}{\sqrt{1 - \mathbf{H}_{ii}}} \sqrt{\mathbf{W}_{ii}} = \frac{(F_{i,j+1} - \hat{f}_j) \sqrt{C_{i,j}}}{\sqrt{1 - \frac{C_{ij}}{\sum_{i=1}^{I-j} C_{ij}}}},$$

which is the same expression we obtained earlier in eq. (2.4). The advantage of this new point of view is that it gives us access to the classical regression theory; in particular, we know (see [8, pp. 267 sqq.]) that the *externally studentised residuals*

$$r_{i,j+1} := \frac{e_i}{\hat{\sigma}_{j(i)} \sqrt{1 - \mathbf{H}_{ii}}} \sqrt{\mathbf{W}_{ii}} = \frac{(F_{i,j+1} - \hat{f}_j) \sqrt{C_{i,j}}}{\hat{\sigma}_{j(i)} \sqrt{1 - \frac{C_{ij}}{\sum_{i=1}^{I-j} C_{ij}}}},$$

where  $\hat{\sigma}_{j(i)}$  is the leave- $i$ -out estimator for  $\sigma_j$ , follow a  $t$  distribution with  $I - j - 1$  degrees of freedom.

Writing  $\delta(C_{ij}, \hat{C}_{ij})$  for any one of these choices (with  $\hat{C}_{ij}$  denoting the fitted value at position  $(i, j)$ ), we outline both resampling procedures in Algorithm 1.

### 2.3.3 Process error

Finally, we deal with the remaining component of eq. (2.2), i.e. the process error.

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

---

**Algorithm 1** Two resampling methods
 

---

```

procedure UNCONDITIONALRESAMPLE( $C, r, \hat{f}$ )
  for  $i \in \{1, \dots, I\}$  do
     $C^*[i, 1] \leftarrow C[i, 1]$ 
  end for
  for  $j \in \{2, \dots, I\}$  do
    for  $i \in \{1, \dots, I + 1 - j\}$  do
       $\hat{C}[i, j] \leftarrow \hat{f}[j - 1] \cdot C^*[i, j - 1]$ 
      Solve  $\delta(y, \hat{C}[i, j]) = r^*[i, j]$  for  $y$ .
       $C^*[i, j] \leftarrow y$ 
    end for
  end for
  return  $C^*$ 
end procedure

procedure CONDITIONALRESAMPLE( $C, r, \hat{f}$ )
  for  $i \in \{1, \dots, I\}$  do
     $C^*[i, 1] \leftarrow C[i, 1]$ 
  end for
  for  $j \in \{2, \dots, I\}$  do
    for  $i \in \{1, \dots, I + 1 - j\}$  do
       $\hat{C}[i, j] \leftarrow \hat{f}[j - 1] \cdot C[i, j - 1]$ 
      Solve  $\delta(y, \hat{C}[i, j]) = r^*[i, j]$  for  $y$ .
       $C^*[i, j] \leftarrow y$ 
    end for
  end for
  return  $C^*$ 
end procedure

```

---

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

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

Notice that this allows us to bypass (2.1) completely, although we still expect that

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

In view of (2.5), the most evident way to generate  $(\mathcal{D}_I^c)^*$  is to start from the antidiagonal in  $\mathcal{D}_I$  and successively sample

$$C_{i,j+1}^* \sim \mathcal{N}(\hat{f}_j^* C_{ij}, \hat{\sigma}_j^* C_{ij}). \quad (2.6)$$

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 [3, p. 238] and substitute in place of (2.6) 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

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

from which it follows that

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

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

---

**Algorithm 2** Mack bootstrap

---

**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

```

---

### 2.3.4 Numerical implementation and results

## 2.4 Poisson GLM

**Model 2 (Poisson GLM).**

1. *The incremental claims are independent from each other.*
2. *There exist parameters  $c, \alpha_2, \dots, \alpha_I$  and  $\beta_2, \dots, \beta_I$  such that*

$$\log(\mathbb{E}[X_{ij}]) = c + \alpha_i + \beta_j.$$

3. *The incremental claims follow a Poisson distribution with  $\mu_{ij} = \mathbb{E}[X_{ij}]$ :*

$$X_{ij} \sim \text{Pois}(e^{c+\alpha_i+\beta_j}).$$

### 2.4.1 Bootstrap methodology

### 2.4.2 Numerical implementation and results

# Conclusion





# Bibliography

- [1] A. C. Davison and D. V. Hinkley. *Bootstrap methods and their application*. Cambridge University Press, 1997.
- [2] Bradley Efron and Robert Tibshirani. *An introduction to the bootstrap*. Boca Raton, Fla Chapman & Hall/Crc, 1998. ISBN: 9780412042317.
- [3] P. D. England and R. J. Verrall. “Predictive Distributions of Outstanding Liabilities in General Insurance”. In: *Annals of Actuarial Science* 1.1 (2006).
- [4] Alois Gisler. “The Estimation Error in the Chain-Ladder Reserving Method: A Bayesian Approach”. In: *ASTIN Bulletin: The Journal of the IAA* 36.2 (2006), 554–565. DOI: 10.1017/S0515036100014653.
- [5] Thomas Mack. “Distribution-free Calculation of the Standard Error of Chain Ladder Reserve Estimates”. In: *Astin Bulletin* 23.2 (1993).
- [6] Thomas Mack. “Measuring the Variability of Chain Ladder Reserve Estimates”. In: 1999.
- [7] Thomas Mack, Gerhard Quarg, and Christian Braun. “The Mean Square Error of Prediction in the Chain Ladder Reserving Method – A Comment”. In: *ASTIN Bulletin: The Journal of the IAA* 36.2 (2006), 543–552. DOI: 10.1017/S051503610001463X.
- [8] G A F Seber and A J Lee. *Linear regression analysis*. Wiley-Interscience, 2003. ISBN: 9780471415404.
- [9] Gary G. Venter. “Discussion of the Mean Square Error of Prediction in the Chain Ladder Reserving Method”. In: *ASTIN Bulletin: The Journal of the IAA* 36.2 (2006), 566–571. DOI: 10.1017/S0515036100014665.
- [10] Mario V. Wüthrich and Michael Merz. *Stochastic Claims Reserving Methods in Insurance*. John Wiley & Sons, 2008.
- [11] 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).