VRIJE
UNIVERSITEIT
BRUSSEL

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

SENSITIVITY ANALYSIS OF STOCHASTIC RESERVING MODELS USING BOOTSTRAP SIMULATIONS

Othman El Hammouchi

June 2023

Sciences and Bioengineering Sciences

VRIJE
UNIVERSITEIT
BRUSSEL

SENSITIVITY ANALYSIS OF STOCHASTIC RESERVING MODELS USING BOOTSTRAP SIMULATIONS

Othman El Hammouchi

June 2023

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

Sciences and Bioengineering Sciences

Acknowledgements

It has become almost obligatory to quote the immortal words of John Donne at the start of any paragraph of acknowledgements: "No man is an island." A shibboleth though it may be, yet it does have the considerable virtue of being true, which is why I have allowed myself to indulge in it here. The list of those to whom I have become indebted in the course of writing this thesis could fill a seperate bibliography, so that I can perforce only mention its most significant entries. Chief among these are my parents and sister, who have been my support and reliance throughout my life; words cannot express the gratitude I feel towards them. I would also like to extend a warm thanks to my active supervisor Dr. Robin Van Oirbeek, whose direction, encouragement and generosity have been absolutely wonderful; it was truly a joy to work with him.

Abstract

The problem of estimating the reserve needed to cover outstanding claims liabilities, and of quantifying the uncertainty of such estimates, has been the subject of much research within actuarial science. The insurance sector distinguishes itself from other model-heavy fields in the high degree of deference it accords to expert judgement. Reserving actuaries are reluctant to move away from established techniques, which are based on macro-level models that aggregate claims data over certain periods. The resulting datasets contain relatively few observations, limiting the efficacy of classical diagnostic methods, and making it difficult to verify model assumptions. We investigate the use of the statistical bootstrap to remedy this, studying how the simulatied predictive distribution of the reserve is impacted by the presence of observations which deviate from the assumptions of two major stochastic reserving models: the Mack chain ladder and the Overdispersed Poisson GLM. Our findings indicate that certain types of bootstraps are very promising for the purpose of distinguishing contaminated points from normal observations.

Samenvatting

Het schatten van de reserve die nodig is om openstaande schadeverplichtingen te dekken, even-als het kwantificeren van de onzekerheid van dergelijke schattingen, is het voorwerp van veel onderzoek binnen de actuariële wetenschap. De verzekeringssector onderscheidt zich van andere domeinen waarin wiskundige modellen een belangrijke rol spelen door de hoge achting die het toekent aan het oordeel van deskundigen. Actuarissen die verantwoordelijk zijn voor het schade-reserveringsproces willen typisch niet afwijken van gevestigde technieken, en deze zijn gebaseerd op macro-modellen die schadegegevens aggregeren over bepaalde periodes. De datasets die men hieruit verkrijgt bevatten relatief weinig observaties, hetgeen de doeltreffendheid van klassieke diagnostische methodes beperkt en de verificatie van modelaannames beperkt. We onderzoeken het gebruik van de statistische bootstrap om deze moeilijkheid te remediëren. Dit doen we door te bestuderen hoe de gesimuleerde voorspellende verdeling van de reserve beïnvloed wordt door de aanwezigheid van observaties die afwijken van de aannames van twee belangrijke stochasti-sche schadereserveringsmodellen: de Mack chain ladder en de Overdispersed Poisson GLM. Onze bevindingen geven aan dat bepaalde types bootstrap beloftevol zijn voor het onderscheiden van gecontamineerde datapunten van gewone waarnemingen.

Contents

List of Tables

List of Figures

List of Symbols

The following list is intended to aid in cross-referencing the most commonly used symbols in the text. As a matter of general notation, an uppercase letter like X denotes a random variable and a lowercase letter such as c is used for a deterministic one. A bold lowercase letter like $\boldsymbol{\beta}$ is used for vectors and a bold uppercase one such as $\mathbf{W}$ denotes a matrix; both may or may not be random, which should be clear from the context. Any letter adorned with a hat denotes an estimate.

Probability & statistics

$\varepsilon_{ij}, \varepsilon_i$ Regression errors

r_{ij}, r_i Regression residuals

$\mathcal{N}(\mu, \sigma^2)$ Normal distribution with mean μ and variance σ^2

$\Gamma(\alpha, \beta)$ Gamma distribution with shape parameter α and rate parameter β

$\text{Pois}(\lambda)$ Poisson distribution with rate parameter λ

I Fisher information matrix

l Log-likelihood function

Mack chain ladder

f_j Development factor

σ_j Variance parameter

ξ_{ij} Log-normal residuals shift parameter

m_{ij} Log-normal residuals mean parameter

s_{ij}^2 Log-normal residuals variance parameter

ODP model

μ_{ij} Incremental triangle means

c Intercept parameter

a_i Origin period parameters

b_j Development period parameters

ϕ Dispersion parameter

V Variance function

Q Quasi-likelihood function

Other symbols

C_{ij} Cumulative claim amount

X_{ij} Incremental claim amount

$\mathcal{D}_I$ Claims triangle

$\mathcal{D}_I^c$ Lower triangle of future claims

Chapter 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 received only upon delivery of goods or services. By contrast, insurance products are purchased long before the adverse events which they protect against have occured, 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*. The reserving actuary must rely here on historical data, which is most often presented in the form of a *loss* or *run-off triangle* $\mathcal{D}_I$ consisting of either cumulative or incremental amounts of some actuarial variable (payments, number of claims, etc.), respectively denoted by C_{ij} and X_{ij}. The row index $1 \leq i \leq I$ denotes the *cohort*, *origin year* or *accident year*, and the column index $1 \leq j \leq J$ gives the *development year*, so that

$$\mathcal{D}_I = \{C_{ij} \mid 1 \leq j \leq J, i + j \leq I + 1\} \quad \text{or} \quad \mathcal{D}_I = \{X_{ij} \mid 1 \leq j \leq J, i + j \leq I + 1\} . \tag{1.1}$$

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 predict the *total outstanding loss liabilities*

$$R = \sum_{i=2}^{I} (C_{i,I} - C_{i,I+1-i}) \tag{1.2}$$

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

One of the most frequently used loss reserving techniques in practice is the so-called *chain ladder* (CL), which predicts the cumulative claim in development year j by multiplying the previous year's amount by a so-called *age-to-age factor*, *link ratio* or *development factor*. It

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}				

(a) Cumulative

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}				

(b) Incremental

Table 1.1: General notation for a 5 by 5 claims triangle

was originally conceived as a purely computational algorithm, but has since been framed as a stochastic model in a variety of ways. Its central assumption 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, however, as it limits the efficacy of classical statistical techniques.

To illustrate this point, consider Table 1.2, which contains the dataset of cumulative payments for a motor insurance account from the UK given in [1] (this will serve as the running example throughout this text). It consists of a 7 by 7 claims triangle with a total of 28 observations. Fig. 1.1 shows the diagnostic plot of standardised residuals against fitted value for the Mack chain ladder, which will be discussed in Chapter 3. The details are not important at this point; all that matters at present is that it should be symmetric around the x-axis and exhibit no structural patterns if the model gives a good fit.

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 1.2: UK Motor cumulative claims triangle from Christofides [1]
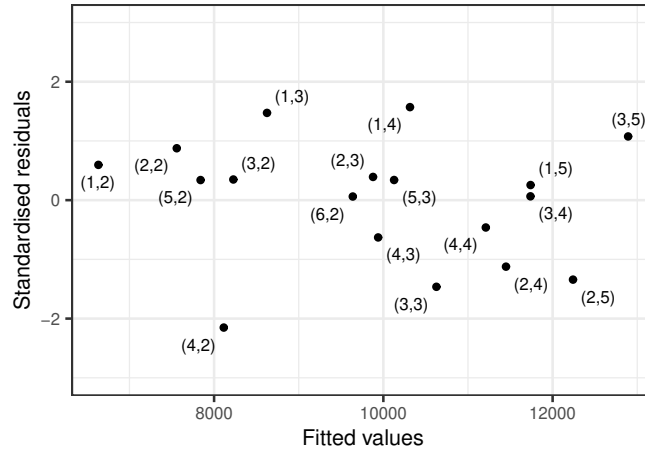
Figure 1.1: Diagnostic plot for original triangle

The same diagnostic plots are shown in Fig. 1.2 for the case where the points $(2,5)$ and $(4,4)$ have been perturbed by a factor 1.5, either by direct multiplication or through simulation from the underlying model. The residual corresponding to the pathological observation has been

highlighted in red. As these examples demonstrate, it is not always feasible to identify deviations from the model assumptions by examining such plots, even for the trained eye.
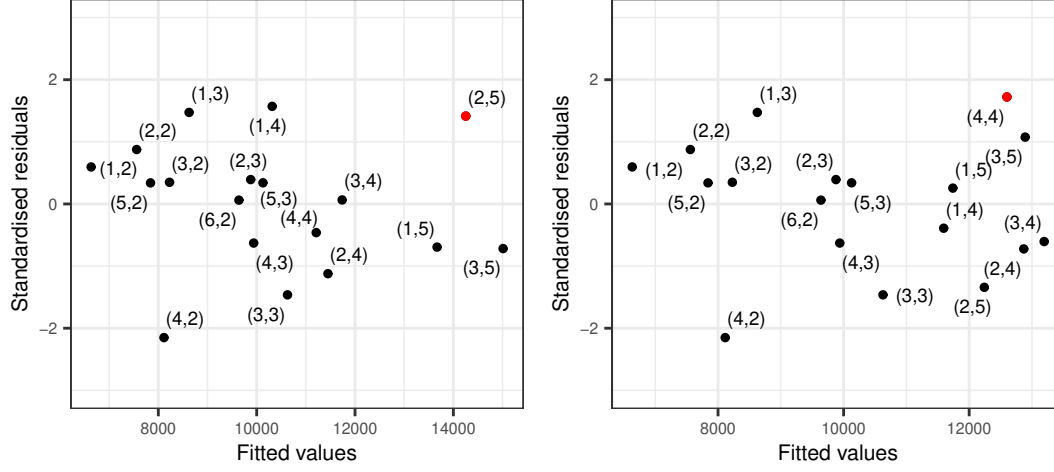
(a) Perturbed directly
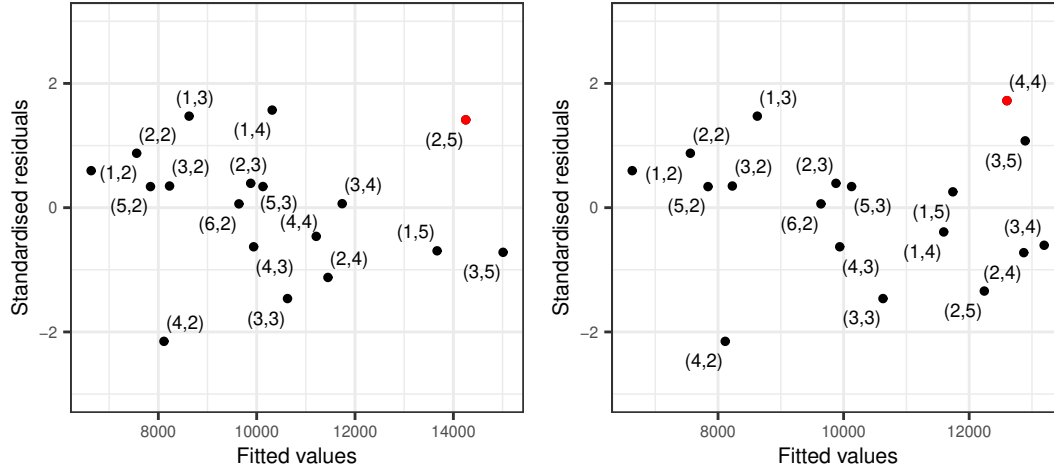
(b) Perturbed according to model

Figure 1.2: Diagnostic plots for perturbed triangles

Our aim in the present study is to investigate whether it is possible to use bootstrap simulations to remedy this problem. The next chapter introduces the bootstrap and explains how it can be used for inference on regression models, which is how we will frame the reserving methods under consideration. In the following two chapters, we will then apply this theory to the most widespread stochastic claims reserving models, the Mack chain ladder and the overdispersed Poisson GLM, in order to study whether it is possible to identify pattern breaks using bootstrap simulation. Specifically, we simulate claims triangles which perfectly follow the model assumptions, perturb these, and generate a bootstrap reserve from the resulting dataset. This

will allow us to investigate how the simulated reserve is impacted by deviations from the model assumptions.

Chapter 2

The bootstrap method

When using a statistical model to describe some dataset in terms of a reduced number of parameters, we are not only interested in obtaining point estimates of these, but also in quantifying their *uncertainty*. In classical statistics, this is usually achieved by starting from the model assumptions and deriving from them the sampling distribution of the estimators analytically. With the exception of the normal distribution, however, the resulting calculations invariably turn out to be intractible, so that one is forced to rely on approximations or asymptotic results. Such estimates tend to be highly sensitive to the simplifying assumptions made in order to obtain them, leading to gross errors when these are violated.

The bootstrap method aims to remedy this problem through the use of numerical simulation. At its core, it is premised on the idea that the empirical distribution of the sample forms a good proxy for that of the population. Consequently, we can approximate sampling from the population by *resampling our data*; to the casual observer, this can give the impression that we're creating new data *ex nihilo*, using our single sample to 'pull ourselves up by our own bootstraps', like Baron Munchausen in the fairy tale.

2.1   Bootstrapping an estimator

To ease the exposition, we begin by describing the bootstrap in the simple case of a classical one-sample estimation problem. Let $X_1, \ldots, X_n$ be an i.i.d. sample drawn from a distribution F, and consider an estimator $\widehat{h(F)} = g(X_1, \ldots, X_n)$ of some quantity $h(F)$ whose uncertainty we wish to estimate using e.g. the variance of the sampling distribution. Depending on which assumptions we are willing to make, we can broadly distinguish between *parameteric* and *nonparametric* approaches to bootstrapping.

The nonparametric bootstrap makes direct use of the sample data, drawing with replacement to produce replicates $X_1^{(b)}, \ldots, X_n^{(b)}$. In other words, the population distribution F is approximated by the *empirical cumulative distribution function*

$$\widehat{F}_n(x) := \sum_{k=1}^{n} I_{\{X_k \leq x\}} \, , \tag{2.1}$$

which is used to generate pseudo-observations. We can then compute the statistic of interest on these, yielding pseudo-realisations $g^{(b)} = g(X_1^{(b)}, \ldots, X_n^{(b)})$ which approximate the sampling distribution of $\widehat{h(F)}$, and can be used for inference about the uncertainty of the estimate. For

instance, if B denotes the total number of bootstrap samples, we can esimate the variance of $\widehat{h(F)}$ by

$$\frac{1}{B-1} \sum_{b=1}^{B} (g^{(b)} - \bar{g})^2 \,, \tag{2.2}$$

with $\bar{g} = \frac{1}{B} \sum_{b=1}^{B} g^{(b)}$. The accuracy of such estimates will ultimately depend on the accuracy of the approximation $F \approx \widehat{F}_n$.

By contrast, the parametric bootstrap starts by fitting a distributional model to the sample data, and then simulates replicates from this with the help of a random number generator. If we assume that F belongs to some family $\{F_{\boldsymbol{\theta}} \mid \boldsymbol{\theta} \in \Theta\}$, then we can use the sample $X_1, \ldots, X_n$ to produce an estimate $\widehat{\boldsymbol{\theta}}$ of the parameter. Plugging this in then gives us $F_{\widehat{\boldsymbol{\theta}}}$, from which we can simulate $X_1^{(b)}, \ldots, X_n^{(b)}$ and $g^{(b)}$ as before. An estimate of the sampling variance can be obtained in the same manner. When the assumptions of its underlying model are met, this leads to estimates which are more efficient; the downside, of course, is an increased error when they are not.

Although we have so far only used it for calculating a single statistic, it is clear that the bootstrap produces a complete *simulated distribution* of the estimator, which can be used for many different forms of inference. This highlights its tremendous potential as a tool for statistical analysis, which explains the rise in popularity of bootstrap methods once personal computers powerful enough to carry out the requisite calculations became widely available. We illustrate this with an example. Figure 2.1 compares the analytic and bootstrap distributions of the maximum likelihood estimators for simulated sample of size $n = 1000$ drawn from a $\Gamma(\alpha, \beta)$-distribution with $\alpha = 2$ and $\beta = 0.5$. The analytic distributions are based on the well-known fact from likelihood theory that the asymptotic distribution of the MLE is given by the multivariate normal distribution $\mathcal{N}(\boldsymbol{\theta}, I(\boldsymbol{\theta})^{-1})$, where $\boldsymbol{\theta}$ is the parameter vector and $I(\boldsymbol{\theta})$ the Fisher information matrix. For the gamma distribution, the latter is given by

$$I(\alpha, \beta) = n \begin{pmatrix} \psi'(\alpha) & -1/\beta \\ -1/\beta & \alpha/\beta^2 \end{pmatrix} \tag{2.3}$$

where $\psi(x) := \frac{\mathrm{d}}{\mathrm{d}x} \log \Gamma(x)$ is the so-called digamma function. We can clearly see that the shape of the bootstrap distribution approaches the analytic one, although it still exhibits an unmistakable bias even at this sample size.
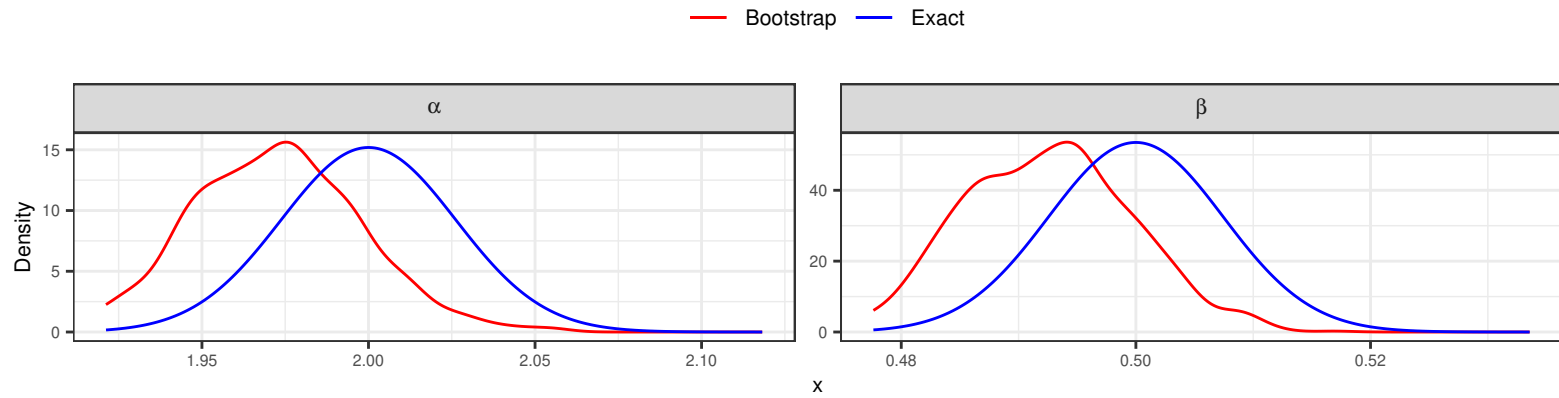
(a) Parametric
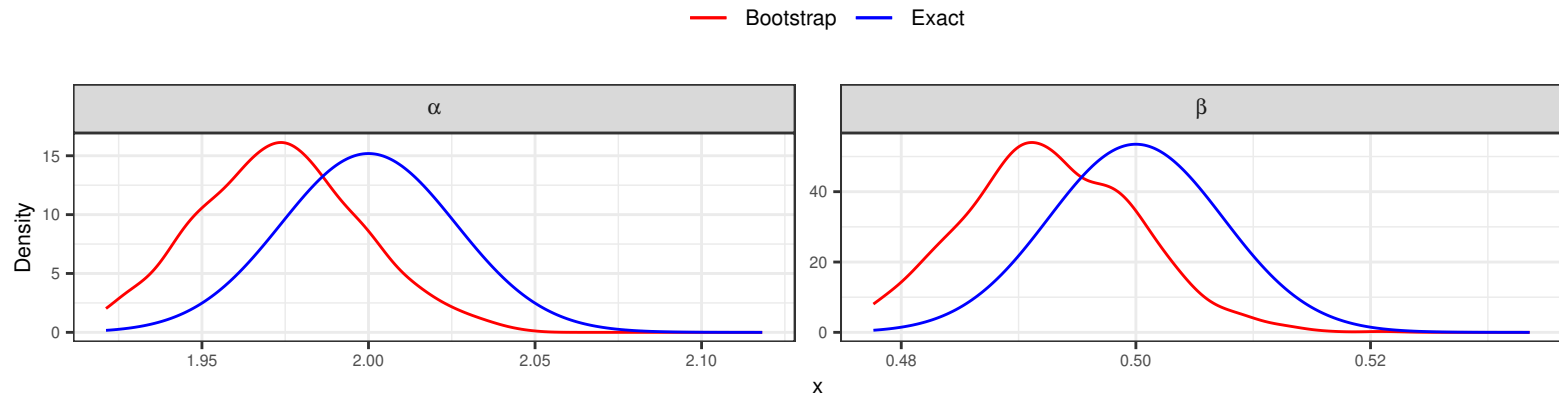
(b) Nonparametric

Figure 2.1: Analytic and bootstrap distribution of MLE for a gamma distribution with $\alpha = 2$ and $\beta = 0.5$

2.2   Bootstrapping a regression model

The principles of the previous section can be applied to bootstrap any kind of data structure, as long as we have a model for the underlying data-generating process (see [2, Chapter 8]). In particular, it is perfectly possible to use the boostrap with regressions; indeed, such methods are well-established in the literature, and will provide the foundation for bootstrapping claims triangles in Chapters 3 and 4.

Consider a set of covariates $X_1, \ldots, X_p$ and a response variable Y whose relationship is modelled by a parametrised mapping $f(X_1, \ldots, X_p; \boldsymbol{\beta})$. Given a sample of pairs $(\mathbf{x_1}, Y_1), \ldots, (\mathbf{x_n}, Y_N)$ and a choice of loss function, we can fit this model to the data, obtaining an estimate $\widehat{\boldsymbol{\beta}}$. For a new value $\mathbf{x_+}$ of the regressors, we then predict the response Y_+ as $f(\mathbf{x_+}; \widehat{\boldsymbol{\beta}})$. It's worth emphasising that there are two different processes at play here, which will correspond to different kinds of bootstrap (cfr. [3, Sections 6.3.3 and 7.2.4]). *Estimation* seeks to *identify* the value of a quantity which is *fixed but unknown*; *prediction* aims to *forecast* the value of a *random variable*.

Under the least squares criterion, for example, we know that the optimal predictor for Y is the conditional expectation $\mathbb{E}[Y \mid X = x]$. This is an ordinary function which returns a real number for any $x \in \mathbb{R}^p$ and can therefore be estimated from a sample. Such an estimate will contain some error, and we need to take this into account when doing inference. If we additionally want to account for the error made when predicting Y using $\mathbb{E}[Y \mid X = x]$, we also have to incorporate the intrinsic stochasticity or *process variance* of the response variable. Prediction is therefore a two-stage procedure involving an intermediate estimation step.

We will illustrate this using the all-familiar linear regression model given by

$$Y_i = \mathbf{x_i}^T \boldsymbol{\beta} + \varepsilon_i, \qquad i \in \{1, \ldots, n\}, \tag{2.4}$$

with $\mathbb{E}[\varepsilon_i] = 0$, $\mathrm{Var}(\varepsilon_i) = \sigma$ and $\mathbb{E}[\varepsilon_i \varepsilon_j] = 0$ for $i \neq j$. Considering the nonparametric bootstrap first, we need to identify a fundamental unit of resampling such that the corresponding variables are interchangeable. One option would be to use the residuals obtained by fitting the model (cfr. [3, Algorithm 6.1]). This approach is sometimes referred to as *semiparametric*, because it only uses the specification of certain aspects of the data distribution in terms of some parameters, but does not assume a specific form for it. Choosing for example

$$r_i := Y_i - \mathbf{x}_i^T \widehat{\boldsymbol{\beta}}, \tag{2.5}$$

we resample these to obtain bootstrap replicates $r_1^{(b)}, \ldots, r_n^{(b)}$, which in turn yield pseudo-responses

$$Y_i^{(b)} := \mathbf{x}_i^T \widehat{\boldsymbol{\beta}} + r_i^{(b)}. \tag{2.6}$$

Refitting the model to the simulated data then gives us bootstrap replicates of the $\widehat{\boldsymbol{\beta}}^{(b)}$ of the regression parameter estimator.

An alternative approach, which is fully nonparametric, is to resample the pairs $(\mathbf{x_i}, Y_i)$ themselves (cfr. [2, Section 9.5], [3, Algorithm 6.2]), thereby approximating the multivariate distribution of $(X_1, \ldots, X_n, Y)$ with the empirical distribution of the data. This has the significant benefit of *parsimony*, making no other assumption beside the i.i.d.-ness of the sample. The model is then refitted to the bootstrap samples $(\mathbf{x_1}^{(b)}, Y_1^{(b)}), \ldots, (\mathbf{x_n}^{(b)}, Y_n^{(b)})$ to produce pseudo-realisations of the regression parameter estimator.

For the parametric case, we have to make an additional assumption regarding the distribution of ϵ in Eq. (2.4), the classic choice being the Gaussian. We then fit the model and use a random number generator to produce bootstrap responses $Y_i^{(b)}$ by drawing from the estimated distribution $\mathcal{N}(\mathbf{x_i}^T \widehat{\boldsymbol{\beta}}, \widehat{\sigma}^2)$. Performing the regression a second time on this pseudo-data then gives us pseudo-replicates of the regression parameter estimator.

2.3 Process variance and predictive distributions

If we want to use the bootstrap for predictive inference on regression models, we must additionally take into account the inherent variability of the dependent variable. Returning once again to the example of normal linear regression, suppose we are interested in predicting the response Y_+ at new value $\mathbf{x}_+$ of the regressors. One way to quantify the accuracy of our forecast is to consider the *prediction error*

$$\delta := Y_+ - \widehat{Y}_+ \,. \tag{2.7}$$

In this expression, the second term can be easily bootstrapped using the methods described in the previous section, which yield replicates $\widehat{\boldsymbol{\beta}}^{(b)}$ for the parameter estimator and hence the predictor via $\widehat{Y}_+^{(b)} = \mathbf{x}_+^T \widehat{\boldsymbol{\beta}}^{(b)}$. It then remains for us to produce bootstrap realisations of the response itself. In the semiparametric approach, this can be achieved by resampling the residuals a second time to obtain pseudo-realisations $r^{(s)}$ for $s = 1, \dots, S$ and adding these (after correct scaling) to the height of the regression hyperplane at $\mathbf{x}_+$. The resulting pseudo-responses $Y_+^{(s)}$ mimic the random fluctuations of Y_+, allowing us to approximate δ with the bootstrap replicates

$$\widehat{\delta}^{(b,s)} := Y_+^{(s)} - \widehat{Y}_+^{(b)} = (\mathbf{x}_+^T \widehat{\boldsymbol{\beta}} + r^{(s)}) - \mathbf{x}_+^T \widehat{\boldsymbol{\beta}}^{(b)} \tag{2.8}$$

(see [3, Algorithm 6.4]). A parametric variant can be obtained by generating pseudo-responses from the fitted model, i.e.

$$Y_+^{(s)} \sim \mathcal{N}(\mathbf{x}_+^T \widehat{\boldsymbol{\beta}}, \widehat{\sigma}^2) \,. \tag{2.9}$$

It is difficult to see how this method could be adapted directly to the pairs bootstrap, however, as it lacks a mechanism for simulating new response values on their own; we will therefore borrow this part from one of the other variants, losing part of its parsimony and robustness in the process.

The bootstrap replicates of the prediction error in Eq. (2.8) can be used for many kinds of predictive inference, from computing prediction intervals to estimating the mean squared error of prediction. While this offers an avenue for incorporating the process variance, it is not the most suitable one for our purposes. Recall from Chapter 1 that we want to investigate how the reserve is impacted by violations of the assumptions of certain actuarial models which can be framed in terms of regression, as we shall see in Chapters 3 and 4; in other words, we need to simulate the distribution of the response itself to study how it is affected. However, we cannot do this just by generating fluctuations around $\mathbf{x}_+^T \widehat{\boldsymbol{\beta}}$, as in Eq. (2.8), as we would then fail to account for the error in our parameter estimates, leading to an underestimation of the prediction uncertainty.

To address this difficulty, we have to borrow a concept from Bayesian school of statistics[1]. Recall that the Bayesian point of view is premised on the idea that the parameters $\boldsymbol{\theta}$ governing a statistical model $p(y \mid \boldsymbol{\theta}, x_1, \dots, x_p)$ are themselves random variables, which are assumed to follow a so-called *prior distribution* $p(\theta)$ expressing the beliefs we have about them before observing any data[2]. When presented with a sample $D = \{(\mathbf{x_1}, Y_1), \dots, (\mathbf{x_n}, Y_n)\}$, we are then led to *update our beliefs* using the formula

$$p(\boldsymbol{\theta} \mid D) \propto p(D \mid \boldsymbol{\theta}) p(\boldsymbol{\theta}) \,, \tag{2.10}$$

which is known as *Bayes' rule*, thereby obtaining a *posterior distribution* $p(\boldsymbol{\theta} \mid D)$ expressing the new likelihood of different values of $\boldsymbol{\theta}$ in light of the observed evidence. For any value of

---

[1] Although we will endeavour in this text to remain agnostic on philosophical questions of probability interpretation, it will become clear shortly why it is necessary to adopt a Bayesian perspective here.

[2] Strictly speaking, this is only one possible paradigm within Bayesian probability, known as subjective Bayes. Objective Bayes dispenses with the subjective aspect of the method through the use of so-called *uninformative priors*. Since this distinction is not important for our exposition, however, we will not pursue it further.

the parameters, the likelihood of the response at a new input, conditional on this value and the sample D, is now given by $p(y_+ \mid \boldsymbol{\theta}, D)$. By marginalising over the posterior distribution, we then incorporate all possible values of $\boldsymbol{\theta}$ in proportion to their likelihood under the data, resulting in the *posterior predictive distribution*.

$$p(y_+ \mid D) = \int p(y_+ \mid \boldsymbol{\theta}) p(\boldsymbol{\theta} \mid D) \, d\boldsymbol{\theta} \,. \tag{2.11}$$

of Y_+ given D. This incorporates both the intrinsic variability of the response as well as our uncertainty regarding the parameters (or, in the classical view, their estimates). By contrast, the *fitted distribution* $p(y_+ \mid \widehat{\boldsymbol{\theta}})$, obtained by plugging in an estimate of the parameters, does not possess this virtue. Moreover, the predictive distribution can be easily integrated into the bootstrap framework; indeed, we can follow the same steps for simulating the prediction errror, but instead of Eq. (2.8) compute pseudo-responses

$$Y_+^{(b,s)} := \mathbf{x}_+^T \widehat{\boldsymbol{\beta}}^{(b)} + r^{(s)} \,. \tag{2.12}$$

This explains why it is difficult to fit this approach within a classical frequentist paradigm, as it is unclear which theoretical quantity these bootstrap replicates would be approximating in that case. We hasten to add that there have been many suggestions to remedy this discrepancy, leading among other things to the idea of a *confidence distribution* which appears to be the solution most widely adopted (see, for example, [4], [5]). As noted by Dickson, Tedesco and Zehnwirth [6], however, the fact remains that there exists a veritable cornucopia of 'frequentist predictive distributions', lending credence to the notion that this concept fits in more naturally with the Bayesian framework.

Similar to Eq. (2.12), we can obtain a bootstrap version of the fitted distribution using

$$Y_+^{(s)} := \mathbf{x}_+^T \widehat{\boldsymbol{\beta}} + r^{(s)} \,, \tag{2.13}$$

which will allow us to compare the two. For the following example, we consider a simulated dataset generated from a normal linear model with $p = 10$, $n = 1000$, $\sigma = 5.7$ and

$$\boldsymbol{\beta} = \begin{bmatrix} 4.1 & 4.4 & -2.1 & 3.3 & 1.4 & 0.2 & 2.4 & -3.7 & 1.6 & 2.1 \end{bmatrix} \,.$$

The predictive and fitted distributions for both the parametric as well as non-parametric bootstrap are shown in Fig. 2.2. We can see in this case that the two are nigh indistinguishable, reflecting the high precision of the parameter estimates on a sample constructed in perfect accordance with the model assumptions. As we shall see in Section 3.4, the results for real data will be quite different.
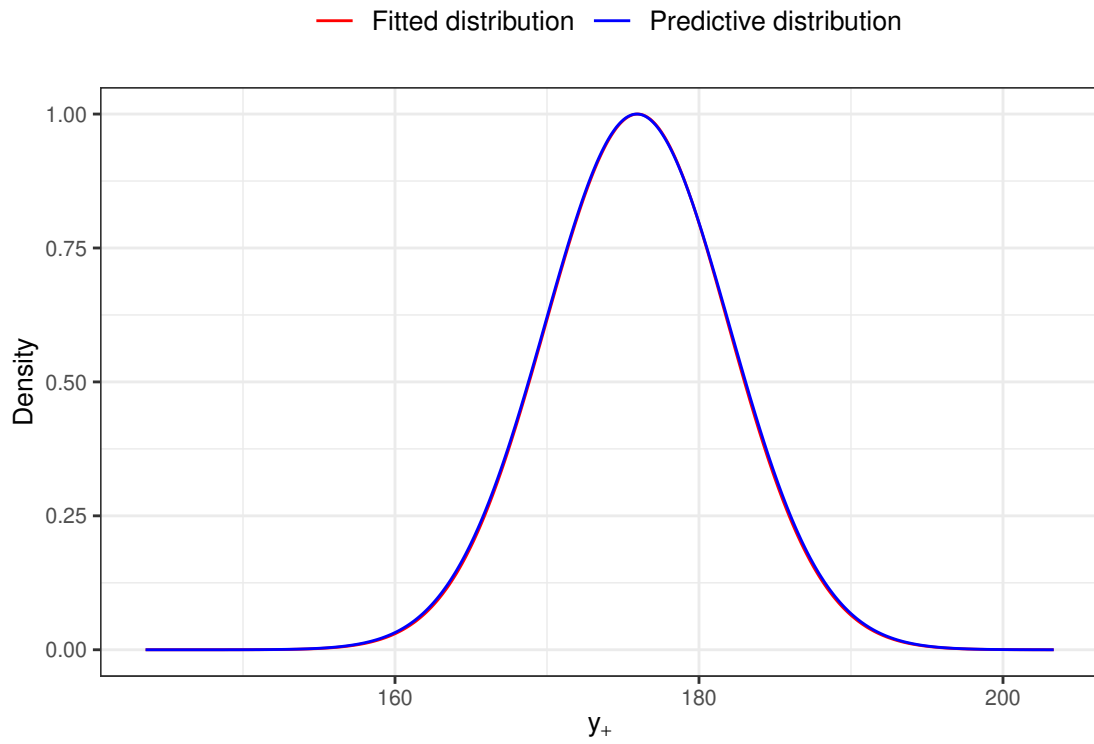
(a) Parametric
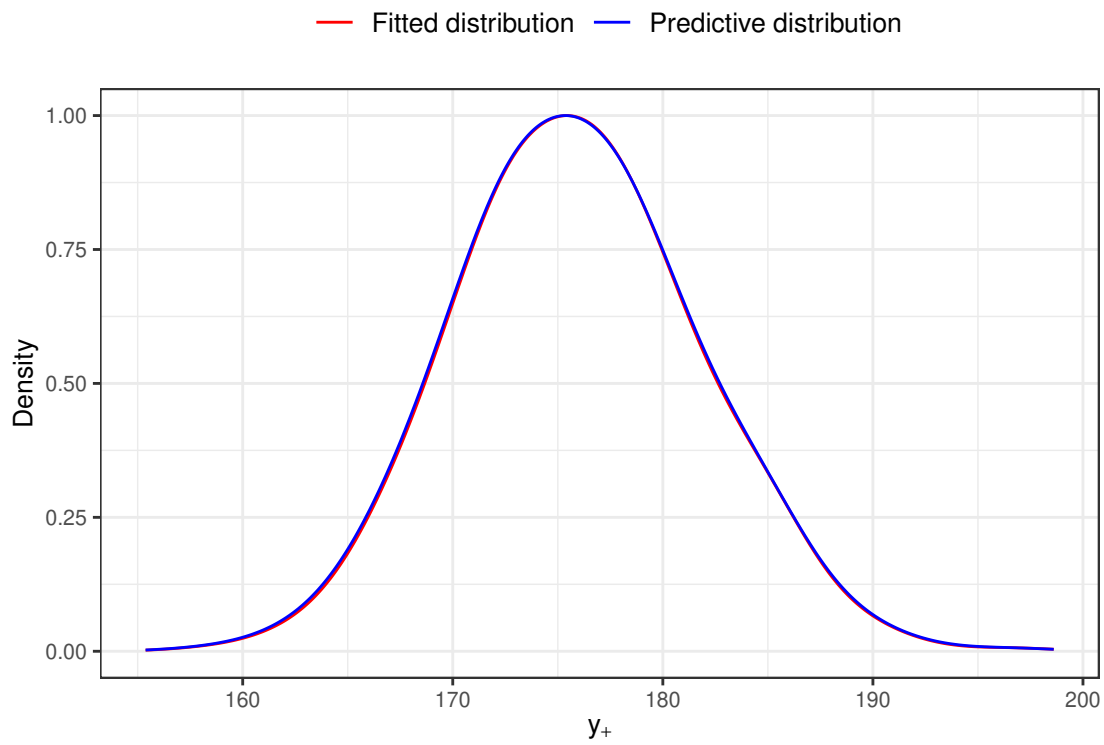
(b) Semiparametric

Figure 2.2: Comparison of the fitted and predictive bootstrap distributions for a normal linear model

Chapter 3

Mack's model

3.1 Introduction

In his seminal paper [7], Mack proposed the following model for cumulative claims triangles, which remains among the most influential in actuarial reserving.

Model 1 (Mack chain ladder).

(**Mack1**) *There exist development factors $f_1, \ldots, f_{I-1}$ such that*

$$\mathbb{E}[C_{ij} \parallel C_{i,j-1}, \ldots, C_{i1}] = \mathbb{E}[C_{ij} \parallel C_{i,j-1}] = f_{j-1} C_{i,j-1} \tag{3.1}$$

for $1 \leq i \leq I$.

(**Mack2**) *There exist variance parameters $\sigma_1, \ldots, \sigma_{I-1}$ such that*

$$\mathrm{Var}[C_{ij} \parallel C_{i,j-1}, \ldots, C_{i1}] = \mathrm{Var}[C_{ij} \parallel C_{i,j-1}] = \sigma_{j-1}^2 C_{i,j-1}, \tag{3.2}$$

for $1 \leq i \leq I$.

(**Mack3**) *The cumulative claims processes $(C_{ij})_j, (C_{i'j})_j$ are independent for $i \neq i'$.*

The development factors are estimated by

$$\widehat{f}_j(\mathcal{D}_I) = \widehat{f}_j(C_{1j}, \ldots, C_{I-j,j}, \ldots, C_{1,j+1}, \ldots, C_{I-j,j+1}) := \frac{\sum_{i=1}^{I-j} C_{i,j+1}}{\sum_{i=1}^{I-j} C_{i,j}}. \tag{3.3}$$

If we define the *single* or *individual* development factors as

$$F_{i,j+1} := \frac{C_{i,j+1}}{C_{ij}}, \tag{3.4}$$

then $\widehat{f}_j$ can be obtained as the weighted average

$$\widehat{f}_j = \frac{\sum_{i=1}^{I-j} C_{ij} F_{i,j+1}}{\sum_{i=1}^{I-j} C_{ij}}. \tag{3.5}$$

The variance parameters are estimated by

$$\widehat{\sigma}_j := \sqrt{\frac{1}{I-j-1} \sum_{i=1}^{I-j} C_{ij} \left(F_{i,j+1} - \widehat{f}_j \right)^2} \tag{3.6}$$

for $j < I - 1$. This formula does not work for $j = I - 1$, as we only have a single pair of observations in the last two columns of the triangle. To remedy this, Mack proposed a simple extrapolation from the previous development years, leading to the estimator

$$\widehat{\sigma}_{I-1} = \sqrt{\min \left\{ \frac{\widehat{\sigma}_{I-2}^4}{\widehat{\sigma}_{I-3}^2}, \widehat{\sigma}_{I-2}^2, \widehat{\sigma}_{I-3}^2 \right\}}, \tag{3.7}$$

which appears to be the most widely adopted one in the literature.

Under the assumptions of the Mack chain ladder, it can be shown (see [8, Lemma 2.5 and Lemma 3.5]) that $\widehat{f}_j$ and $\widehat{\sigma}_j$ are (conditionally) unbiased, and that the $\widehat{f}_j$ are uncorrelated. Predictions for the ultimate claim amounts C_{iI} are obtained by substituting these estimates for the unknown development factors f_j in the conditional expectation. In other words, we predict the ultimate loss using the conditional mean $\mathbb{E}[C_{iI} \parallel C_{i,I+1-i}]$, and estimate the latter by plugging in $\widehat{f}_j$, yielding

$$\widehat{C}_{iI} := \widehat{\mathbb{E}}[C_{i,I} \parallel C_{i,I+1-i}] = C_{i,I+1-i} \prod_{j=I+1-i}^{I-1} \widehat{f}_j. \tag{3.8}$$

From this, we then finally obtain the reserve predictor

$$\widehat{R} = g(\mathcal{D}_I) := \sum_{i=2}^{I} (\widehat{C}_{i,I} - C_{i,I+1-i}). \tag{3.9}$$

The Mack chain ladder is often referred to as "distribution-free" because it only makes assumptions about the first two moments of the random variables in the claims triangle. Indeed, we will now show that this model can be viewed as a series of linear regressions through the origin (i.e. without intercept term), in which case (Mack1) to (Mack3) correspond to the assumptions of the Gauss-Markov theorem[1] guaranteeing optimality of the estimator Eq. (3.3). If we introduce, for any development year $j \in \{1, \ldots, I-1\}$, the notation

$$\mathbf{c_j} := \begin{bmatrix} C_{1,j} \\ \vdots \\ C_{I-j,j} \end{bmatrix}, \tag{3.10}$$

then the first two assumptions of the Mack chain ladder can be equivalently stated as

$$\mathbf{c_{j+1}} = f_j \mathbf{c_j} + \boldsymbol{\varepsilon}, \tag{3.11}$$

with $\boldsymbol{\varepsilon}$ a random vector satisfying

$$\mathbb{E}[\boldsymbol{\varepsilon} \parallel C_{1,j}, \ldots, C_{I-j,j}] = \mathbf{0} \qquad \mathrm{Var}(\boldsymbol{\varepsilon} \parallel C_{1,j}, \ldots, C_{I-j,j}) = \sigma_j^2 \begin{bmatrix} C_{1j} & & \\ & \ddots & \\ & & C_{I-j,j} \end{bmatrix}. \tag{3.12}$$

---

[1]The third assumption is actually slightly stronger than necessary: Gauss-Markov only requires the errors to be uncorrelated.

Consequently, it follows (see [9, Proposition 1.7]) that the weighted least squares method with weights matrix

$$\mathbf{W} = \begin{bmatrix} 1/C_{1j} & & \\ & \ddots & \\ & & 1/C_{I-j,j} \end{bmatrix}, \tag{3.13}$$

leads to an estimator for f_j which has minimal variance in the class of linear unbiased estimators. This estimator is given by

$$\widehat{f}_j^{\mathrm{WLS}} = (\mathbf{c}_{\boldsymbol{j}}^T \mathbf{W} \mathbf{c}_{\boldsymbol{j}})^{-1} \mathbf{c}_{\boldsymbol{j}}^T \mathbf{W} = \frac{\sum_{i=1}^{I-j} C_{i,j+1}}{\sum_{i=1}^{I-j} C_{i,j}}, \tag{3.14}$$

which is the same expression as Eq. (3.3).

3.2   A challenging simulation

Owing to its recursive nature, Mack's model does not readily lend itself to application of the theory from Chapter 2. Unfortunately, previous work in this area is rather scarce, as studies of the Mack chain ladder have tended to focus on deriving closed-form expressions for the uncertainty of the reserve; in particular, much attention has been spent on finding approximations of the conditional *mean square error of prediction* (MSEP). The actuarial literature on bootstrap methods is not very helpful either, as it has mostly been concerned with generalised linear models—even papers such as [10] which address the Mack chain ladder frame it as a GLM. As will become clear shortly, this passes over some subtleties related to the particular structure of Mack's model, and we therefore choose a different approach here. In particular, we will take as our starting point the problem of analytically estimating the MSEP. While this might appear at first glance to be unrelated (even antithetical) to the bootstrap, we will see that theoretical framework associated with it can be easily subverted for our purposes.

The MSEP is a measure for the uncertainty of a given predictive model. It is defined as the Euclidean distance between the predictor and the response in the underlying filtered probability space, i.e.

$$\operatorname*{MSEP}_{R \mid \mathcal{D}_I}(\widehat{R}) := \mathbb{E}\Big[(\widehat{R} - R)^2 \; \big\| \; \mathcal{D}_I\Big] \tag{3.15}$$

for our special case of predicting the reserve. The MSEP admits a decomposition, similar to the familiar bias-variance decomposition from classical statistics, into so-called *parameter* or *estimation error* and *process error*:

$$
\begin{aligned}
\mathbb{E}\Big[(\widehat{R} - R)^2 \; \big\| \; \mathcal{D}_I\Big] &= \mathbb{E}\big[(R - \mathbb{E}[R \| \mathcal{D}_I])^2 \; \big\| \; \mathcal{D}_I\big] + \mathbb{E}\Big[(\mathbb{E}[R \| \mathcal{D}_I] - \widehat{R})^2 \; \big\| \; \mathcal{D}_I\Big] \\
&\quad - 2\mathbb{E}\Big[(R - \mathbb{E}[R \| \mathcal{D}_I])(\mathbb{E}[R \| \mathcal{D}_I] - \widehat{R}) \; \big\| \; \mathcal{D}_I\Big]
\end{aligned}
\tag{3.16}
$$

$$
\begin{aligned}
&= \operatorname{Var}(R \| \mathcal{D}_I) + (\mathbb{E}[R \| \mathcal{D}_I] - \widehat{R})^2 \\
&\quad - 2(\mathbb{E}[R \| \mathcal{D}_I] - \widehat{R})(\mathbb{E}[R - \mathbb{E}[R \| \mathcal{D}_I] \| \mathcal{D}_I])
\end{aligned}
\tag{3.17}
$$

$$
= \underbrace{\operatorname{Var}(R \| \mathcal{D}_I)}_{\text{process error}} + \underbrace{(\mathbb{E}[R \| \mathcal{D}_I] - \widehat{R})^2}_{\text{estimation error}}, \tag{3.18}
$$

corresponding to the two stages of bootstrapping a predictor which we discussed in Section 2.2. Consider now, for any accident year $i \in \{1, \ldots, I\}$, the MSEP for the associated ultimate

$$\operatorname*{MSEP}_{C_{iI} \mid \mathcal{D}_I}(\widehat{C}_{iI}) = (\mathbb{E}[C_{iI} \| \mathcal{D}_I] - \widehat{C}_{iI})^2 + \operatorname{Var}(C_{iI} \| \mathcal{D}_I), \tag{3.19}$$

and suppose we are interested in obtaining a closed-form estimator for it. Such an expression can be derived relatively straightforwardly for the process error from the assumptions of the Mack chain ladder in the following way. We begin by applying the law of total variance in conjunction with (Mack1) and (Mack2) to obtain

$$\text{Var}(C_{iI} \parallel \mathcal{D}_I) = \text{Var}(C_{iI} \parallel C_{i,I+1-i}) \tag{3.20}$$

$$= \mathbb{E}[\text{Var}(C_{iI} \parallel C_{i,I-1}) \parallel C_{i,I+1-i}] + \text{Var}(\mathbb{E}[C_{iI} \parallel C_{i,I-1}] \parallel C_{i,I+1-i}) \tag{3.21}$$

$$= \sigma_{I-1}^2 \mathbb{E}[C_{i,I-1} \parallel C_{i,I+1-i}] + f_{I-1}^2 \text{Var}(C_{i,I-1} \parallel C_{i,I+1-i}) \tag{3.22}$$

$$= \sigma_{I-1}^2 C_{i,I+1-i} \prod_{j=I+1-i}^{I-2} f_j + f_{I-1}^2 \text{Var}(C_{i,I-1} \parallel C_{i,I+1-i}), \tag{3.23}$$

which is a linear recurrence equation of the form

$$x_n = a_{n-1} x_{n-1} + g_{n-1} \tag{3.24}$$

with $x_n = \text{Var}(C_{in} \parallel C_{i,I+1-i})$ and

$$g_{n-1} = \sigma_{n-1}^2 C_{i,I+1-i} \prod_{j=I+1-i}^{n-1} f_j, \qquad a_{n-1} = f_{n-1}^2. \tag{3.25}$$

The general solution is given by

$$x_n = \left(\prod_{j=n_0}^{n-1} a_j \right) \left(x_{n_0} + \sum_{k=n_0}^{n-1} \frac{g_k}{\prod_{l=n_0}^{k} a_l} \right) \tag{3.26}$$

where n_0 denotes the first index of the sequence x_n, in our case $I + 1 - i$. Using the initial condition $x_{I+1-i} = \text{Var}(C_{i,I+1-i} \parallel C_{i,I+1-i}) = 0$, we finally obtain

$$\text{Var}(C_{iI} \parallel \mathcal{D}_I) = \left(\prod_{j=I+1-i}^{I-1} f_j^2 \right) \left(\sum_{k=I+1-i}^{I-1} \frac{\sigma_k^2 C_{i,I+1-i} \prod_{j=I+1-i}^{k-1} f_j}{\prod_{j=I+1-i}^{k} f_j^2} \right) \tag{3.27}$$

$$= \left(\prod_{j=I+1-i}^{I-1} f_j^2 \right) C_{i,I+1-i}^2 \left(\sum_{k=I+1-i}^{I-1} \frac{\sigma_k^2 / f_k^2}{\prod_{j=I+1-i}^{k-1} f_j C_{i,I+1-i}} \right) \tag{3.28}$$

$$= \mathbb{E}[C_{iI} \parallel C_{I+1-i}]^2 \sum_{k=I+1-i}^{I-1} \frac{\sigma_k^2 / f_k^2}{\mathbb{E}[C_{ik} \parallel C_{i,I+1-i}]}, \tag{3.29}$$

which we can estimate by plugging in $\widehat{f}_j$ and $\widehat{\sigma}_j$ for f_j and σ_j, respectively.

For the parameter error, if we use the definitions from the previous section to rewrite it as

$$(\mathbb{E}[C_{iI} \parallel \mathcal{D}_I] - \widehat{C}_{iI})^2 = C_{i,I+1-i}^2 \left(\prod_{j=I+1-i}^{I-1} f_j - \prod_{j=I+1-i}^{I-1} \widehat{f}_j \right)^2 \tag{3.30}$$

$$= C_{i,I+1-i}^2 \left(\prod_{j=I+1-i}^{I-1} f_j^2 + \prod_{j=I+1-i}^{I-1} \widehat{f}_j^2 - 2 \prod_{j=I+1-i}^{I-1} f_j \widehat{f}_j \right), \tag{3.31}$$

it becomes clear that things are more complicated than with process error. Indeed, we cannot simply substitute the $\widehat{f}_j$ for the unknown parameters in this expression as that would cause it to vanish, yielding an estimate which will generally not be accurate. This problem was recognised by Mack himself in [11], and is caused by the fact that the claims triangle observations are used for both estimation and forecasting (see [12, Section 2] for a more general discussion). His suggested solution was to apply some kind of conditional averaging to the $\widehat{f}_j$. Ideally, one would like to condition on all available observations in $\mathcal{D}_I$, but the $\mathcal{D}_I$-measurability of the $\widehat{f}_j$ would then bring us right back where we started. We must therefore use a smaller set in order to allow $\widehat{f}_{I+1-i}, \ldots, \widehat{f}_{I-1}$ to fluctuate around $f_{I+1-i}, \ldots, f_{I-1}$. This corresponds to asking which other values $\widehat{f}_j$ could have taken, given that we fix a certain subset of the data—in other words, it's a resampling scheme on the parameter estimates. Thus, one can obtain an estimate of the parameter error by specifying a mechanism for generating new realisations of $\widehat{f}_j$ (see [13], [8, pp. 44 sqq.]), with different mechanisms yielding different estimates. The literature uses this mostly as a theoretical device to facilitate analytical calculations; for the specific approach developed by Mack, it leads to the estimator

$$\widehat{\mathrm{MSEP}}(\widehat{R_i}) \coloneqq \widehat{C}_{iI} \sum_{j=I+1-i}^{I-1} \frac{\widehat{\sigma}_j^2}{\widehat{f}_j} \left(\frac{1}{\widehat{C}_{ij}} + \frac{1}{\sum_{i=1}^{I-j} C_{ij}} \right) \tag{3.32}$$

(see [11, p. 11]). In this case, however, the theory happens to fit in perfectly with the resampling framework, and we can therefore employ it as a basis for bootstrap procedures. In the remainder of this section, we outline two approaches for estimating Eq. (3.30) and indicate the corresponding resampling methods.

Denote the subset of observations in $\mathcal{D}_I$ up to and including development year k by

$$\mathcal{B}_k \coloneqq \{C_{ij} \in \mathcal{D}_I \mid j \leq k\}, \tag{3.33}$$

and write

$$\mathcal{D}_{I,k}^O \coloneqq \{C_{ij} \in \mathcal{D}_I \mid j > I + 1 - k\} \tag{3.34}$$

for its complement. One option would then be to take the conditional expectation of $\widehat{f}_j$ with respect to $\mathcal{B}_{I+1-i}$, leading to the estimate

$$\mathbb{E}[(\mathbb{E}[C_{iI} \parallel \mathcal{D}_I] - \widehat{C}_{iI})^2 \parallel \mathcal{B}_{I+1-i}] = C_{i,I+1-i}^2 \mathbb{E}\left[\left(\prod_{j=I+1-i}^{I-1} f_j - \prod_{j=I+1-i}^{I-1} \widehat{f}_j \right)^2 \parallel \mathcal{B}_{I+1-i} \right] \tag{3.35}$$

$$= C_{i,I+1-i}^2 \left(\mathbb{E}\left[\prod_{j=I+1-i}^{I-1} \widehat{f}_j^2 \parallel \mathcal{B}_{I+1-i} \right] - \prod_{j=I+1-i}^{I-1} f_j^2 \right), \tag{3.36}$$

where we used the fact that the $\widehat{f}_j$ are uncorrelated. This corresponds to averaging over the distribution of $\mathcal{D}_{I,k}^O$, or, expressed in terms of resampling, to generating new pseudo-observations in the upper right triangle. Borrowing the nomenclature from [13], we call this the *unconditional approach*. Alternatively, we could average each $\widehat{f}_j$ only over the observations after j. This is equivalent to fixing the denominator $\sum_{i=1}^{I-j} C_{ij}$ in the development factor estimator Eq. (3.3) and allowing the numerator $\sum_{i=1}^{I-j} C_{i,j+1}$ to vary. Formally, it corresponds to taking the expectation

with respect to the probability measure defined on $\mathcal{D}^O_{I,i}$ by

$$\mathbb{P}^*_{\mathcal{D}_I}(\{dz_{ij}\}_{i+j\leq I+1}) := \prod_{j=1}^{I-1}\prod_{i=1}^{I-j}\mathbb{P}_{C_{i,j+1}}(dz_{i,j+1} \mid C_{ij} = c_{ij}), \qquad (3.37)$$

yielding the estimate

$$\mathbb{E}_{\mathbb{P}^*_{\mathcal{D}_I}}\left[(\mathbb{E}[C_{iI}\parallel\mathcal{D}_I]-\widehat{C}_{iI})^2\right] = C_{iI}^2\,\mathbb{E}_{\mathbb{P}^*_{\mathcal{D}_I}}\left[\left(\prod_{j=I+1-i}^{I-1}f_j-\prod_{j=I+1-i}^{I-1}\widehat{f}_j\right)^2\right] \qquad (3.38)$$

$$= C_{i,I+1-i}^2\left(\mathbb{E}_{\mathbb{P}^*_{\mathcal{D}_I}}\left[\prod_{j=I+1-i}^{I-1}\widehat{f}_j^2\right]-\prod_{j=I+1-i}^{I-1}f_j^2\right) \qquad (3.39)$$

$$= C_{i,I+1-i}^2\left(\prod_{j=I+1-i}^{I-1}\mathbb{E}\left[\widehat{f}_j^2\parallel\mathcal{B}_j\right]-\prod_{j=I+1-i}^{I-1}f_j^2\right) \qquad (3.40)$$

(see [8, p. 46]). This is referred to as the *conditional approach*, and it corresponds to a scheme in which only the observations from the next period are resampled to produce a new pseudo-realisation of the parameter estimate for the current period.

There has been some controversy about which of these approaches should be preferred, leading to a vigorous discussion in [13]–[16]. As we shall see in Section 3.3, the difference between the results which they produce is negligeable, and so the question is mainly of theoretical interest. Nevertheless, based on the previous exposition, it seems reasonable to prefer whichever method produces resampled parameter estimates approximating the original $\widehat{f}_j$ most closely. In particular, we note that these posses the following property, the proof of which can be found in [14].

Theorem 1. *The squares of two successive development factor estimates in the Mack chain ladder are negatively correlated:*

$$\operatorname{Cov}(\widehat{f}_j, \widehat{f}_{j-1}) < 0. \qquad (3.41)$$

In the conditional approach, the resampled parameter estimates are independent by construction, and so they cannot incorporate this covariance structure. In light of this, it would appear that the unconditional scheme has slightly better theoretical properties. As the empirical difference between the two is minimal, however, the conditional version is a reasonable approximation to fall back on when needed. In the next section, we will see how both approaches give rise to a variety of different bootstrap methods.

3.3   Bootstrap methodology

In Section 2.2, we introduced a taxonomy for the different types of bootstrap, distinguishing between semiparameteric, nonparametric and parametric variants; we now consider how each of these can be applied to the Mack chain ladder. For comparison, the analytical results for the example dataset from Table 1.2 (using the estimator in Eq. (3.32)) are provided in Table 3.1.

For the semiparametric bootstrap, the crucial step is to find a suitable definition for the residuals which ensures that they are interchangeable. The distribution-free nature of the model makes this difficult, however, as it limits the statements we can make about the errors to the

$i \, / \, j$	$\widehat{f}_j^{\,\mathrm{CL}}$	$\widehat{\sigma}_i^{\,\mathrm{CL}}$	$\widehat{R}_i^{\,\mathrm{CL}}$	$\widehat{\mathrm{MSEP}}(\widehat{R}_j)$
2	1.89	2.83	350.9	3.62
3	1.28	3.34	1037.54	22.9
4	1.15	2.98	2044.86	141.98
5	1.1	1.07	3663.4	426.7
6	1.05	0.16	7162.15	692.39
7	1.03	0.02	14396.92	900.58

Table 3.1: Mack chain ladder results for the UK Motor triangle from Table 1.2

first two moments. This can be resolved in one of two ways. On the one hand, we could decide to extrapolate from homogeneity of the mean and variance to homogeneity of the distributions. In that case, the *raw residuals*

$$r_{i,j+1} := C_{i,j+1} - \widehat{C}_{i,j+1} = C_{i,j+1} - \widehat{f}_j C_{ij} \tag{3.42}$$

are not an option, as they suffer from heteroscedasticity:

$$\mathrm{Var}(r_{i,j+1} \parallel C_{ij}) = \sigma_j^2 \left(C_{ij} - \frac{C_{ij}^2}{\sum_{i=1}^{I-j} C_{ij}} \right) . \tag{3.43}$$

We can address this by dividing out the variance and considering the errors

$$\varepsilon_{i,j+1} := \frac{C_{i,j+1} - f_j C_{ij}}{\sigma_j \sqrt{C_{ij}} \sqrt{1 - \frac{C_{ij}}{\sum_{i=1}^{I-j} C_{ij}}}} , \tag{3.44}$$

which satisfy $\mathbb{E}[\varepsilon_{i,j+1} \parallel C_{ij}] = 0$ and $\mathrm{Var}(\varepsilon_{i,j+1} \parallel C_{ij}) = 1$. Provided the sampling variability of the $\widehat{f}_j$ and $\widehat{\sigma}_j$ is not too bad (a non-trivial assumption given the small sample sizes we're usually dealing with), the same should approximately hold for the residuals

$$r_{i,j+1} := \frac{C_{i,j+1} - \widehat{f}_j C_{ij}}{\widehat{\sigma}_j \sqrt{C_{ij}} \sqrt{1 - \frac{C_{ij}}{\sum_{i=1}^{I-j} C_{ij}}}} \tag{3.45}$$

obtained by plugging in these estimators. Note that the factor $\sqrt{1 - \frac{C_{ij}}{\sum_{i=1}^{I-j} C_{ij}}}$ in the denominator corresponds to the leverage adjustment, as can be seen by computing the hat matrix:

$$\mathbf{H} = \mathbf{c}_j (\mathbf{c}_j^T \mathbf{W} \mathbf{c}_j)^{-1} \mathbf{c}_j^T \mathbf{W} \tag{3.46}$$

$$= \frac{1}{\sum_{i=1}^{I-j} C_{ij}} \begin{bmatrix} C_{1j} & \dots & C_{1j} \\ \vdots & \ddots & \vdots \\ C_{I-j,j} & \dots & C_{I-j,j} \end{bmatrix} . \tag{3.47}$$

It's worth emphasising, however, that such an extrapolation should not be made lightly: it is perfectly possible for the error distribution to exhibit heterogeneity in ways other than through its mean and variance (see [2, p. 114] for an example where the *percentiles* vary with the value of the regressor). In light of this, an alternative approach would be to augment our model with some explicit distributional assumptions, which is more transparent and allows us to make

precise statements about errors and residuals. One such augmentation that has been studied in the literature (see [8, p. 49]) is the autoregressive Gaussian time series model

$$C_{i,j+1} = f_j C_{ij} + \sigma_j \sqrt{C_{ij}}\, \varepsilon_{i,j+1}, \qquad \varepsilon_{i,j+1} \sim \mathcal{N}(0,1)\,, \tag{3.48}$$

which can easily be seen to be compatible with (Mack1) to (Mack3). In light of the fact that the Mack chain ladder can be viewed as a series of weighted linear regressions, as we say in Section 3.1, this has the benefit of making available to us the results of the classical theory. We know, for example, that the *externally studentised residuals*

$$r_{i,j+1} := \frac{e_{i,j+1}}{\widehat{\sigma}_{j(i)} \sqrt{1 - \mathbf{H}_{ii}}} \sqrt{\mathbf{W}_{ii}} = \frac{C_{i,j+1} - \widehat{f}_j C_{ij}}{\widehat{\sigma}_{j(i)} \sqrt{C_{ij}} \sqrt{1 - \frac{C_{ij}}{\sum_{i=1}^{I-j} C_{ij}}}}\,, \tag{3.49}$$

where $\widehat{\sigma}_{j(i)}$ denotes the leave-i-out estimator of σ_j, follow a t_{I-j-1} distribution. Similarly, the *standardised* or *internally studentised* residuals

$$r_{i,j+1} := \frac{C_{i,j+1} - \widehat{f}_j C_{ij}}{\widehat{\sigma}_j \sqrt{C_{ij}} \sqrt{1 - \frac{C_{ij}}{\sum_{i=1}^{I-j} C_{ij}}}} \tag{3.50}$$

also share a common (although more complicated) distribution (see [17, pp. 267 sqq.]).

A major shortcoming of Eq. (3.48), however, is that it makes it possible to have a negative realisation in the next step of the time series, in which case all future observations from that point on are undefined because of the square root factor in the second term. This is not merely a theoretical problem either: we have observed this phenomenon in our numerical implementation, where the resampling produces negative pseudo-realisations of certain claim amounts, particularly when the model has been severely perturbed. An obvious way of avoiding this difficulty would be to simply discard the current bootstrap iteration as soon as a negative value is produced. This has the drawback of requiring more computational power, sometimes beyond the realm of what is reasonable. Moreover, we have seen cases in which the probability of having no negative replicates was so vanishingly small as to cause the program to get stuck indefinitely. It is therefore imperative that we develop a different approach, if only as a fall-back in case of problems with Eq. (3.48).

England and Verrall [10, p. 238] discuss an similar difficulty in the context of generating bootstrap realisations of future claim amounts, which we discuss in Section 3.4. To avoid negative draws, the authors recommend using a gamma distribution in place of the Gaussian, choosing parameters that lead to the same mean and variance. This suggests that our woes may be remedied by choosing a different distribution in Eq. (3.48) which is capable of guaranteeing

$$\varepsilon_{i,j+1} > -\frac{f_j \sqrt{C_{ij}}}{\sigma_j}\,. \tag{3.51}$$

Clearly, the support of such a distribution has to be bounded from below. Moreover, it must allow for normalisation, so that the resulting residuals can be made interchangeable by an algebraic transformation. A good candidate satisfying these conditions which readily comes to mind is the *shifted log-normal distribution*

$$\log(\varepsilon_{i,j+1} - \xi_{ij}) \sim \mathcal{N}(m_{ij}, s_{ij}^2) \tag{3.52}$$

for certain parameters ξ_{ij}, m_{ij} and s_{ij}. his has support $(\xi_{ij}, +\infty)$, leading us to choose $\xi_{ij} =$

$-\frac{f_j\sqrt{C_{ij}}}{\sigma_j}$, and we must determine m_{ij} and s_{ij} to satisfy

$$\begin{cases} \mathbb{E}[\varepsilon_{i,j+1} \parallel C_{ij}] = \exp\left(m_{ij} + \frac{s_{ij}^2}{2}\right) + \xi_{ij} = 0 \\ \mathrm{Var}(\varepsilon_{i,j+1} \parallel C_{ij}) = (\exp s_{ij}^2 - 1)\exp(s_{ij}^2 + 2m_{ij}) = 1 \,. \end{cases} \tag{3.53}$$

Solving these equations, we obtain

$$s_{ij} = \sqrt{\log\left(1 + \frac{1}{\xi_{ij}^2}\right)}, \qquad m_{ij} = \log(-\xi_{ij}) - \frac{s_{ij}^2}{2}\,. \tag{3.54}$$

Provided the sampling variability of $\widehat{f}_j$ and $\widehat{\sigma}_j$ is not too severe, this means that the residuals

$$r_{i,j+1} := \frac{\log\left(C_{i,j+1} - \widehat{f}_j C_{ij} + \widehat{f}_j \sqrt{C_{ij}}/\widehat{\sigma}_j\right) - \widehat{m}_{ij}}{\widehat{s}_{ij}} \tag{3.55}$$

$$= \frac{\log\left(C_{i,j+1} - \widehat{f}_j C_{ij} + \widehat{f}_j \sqrt{C_{ij}}/\widehat{\sigma}_j\right) - \log(\widehat{f}_j \sqrt{C_{ij}}/\widehat{\sigma}_j) + \frac{1}{2}\log\left(1 + \widehat{\sigma}_j^2/\widehat{f}_j^2 C_{ij}\right)}{\sqrt{\log\left(1 + \widehat{\sigma}_j^2/\widehat{f}_j^2 C_{ij}\right)}} \tag{3.56}$$

are approximately $\mathcal{N}(0,1)$-distributed.

After selecting one of these variants, the next step is to fit the model and compute the residuals from it. These are then resampled to generate pseudo-residuals $r_{ij}^{(b)}$, from which a bootstrap triangle is obtained by inverting the appropriate formula. The conditional approach employs the original triangle here, whereas the unconditional one uses the pseudo-realisations generated in the previous step. Finally, the model is refitted to the simulated triangle to obtain bootstrap replicates of the development factor and variance parameter estimators $\widehat{\boldsymbol{f}}$ and $\widehat{\boldsymbol{\sigma}}$. The entire procedure is outlined in Algorithms 1 and 2 for the special case of standardised residuals; the others are completely analogous. Results for the example data from Table 1.2 are given in Tables 3.2 to 3.4 for all three types. Note that these feature a bootstrap estimate for the MSEP; this is obtained using the method from Section 2.3 (see Eq. (2.8)), which we will adapt to the Mack chain ladder in Section 3.4. To preserve the continuity of the exposition, however, we have chosen to include them here already. All bootstrap results in chapter were computed with $B = 1000$ and $S = 100$.

Next, we move on to the fully nonparametric bootstrap, which is based on direct resampling of the observed predictor-response pairs. Once again viewing the Mack chain ladder as a series of weighted linear regressions, the most obvious way of adapting this would be to resample the pairs $\{(C_{ij}, C_{i,j+1}) \mid 1 \leq i \leq I - j\}$ at every development index $j \in \{1, \ldots, I-1\}$. In this case, however, the only resampling scheme which can be used is the conditional one. To see why, consider what an implementation of the unconditional approach would look like. Starting at $j = 1$, if the result of resampling the first two columns is given by

$$\{(C_{11}^*, C_{12}^*), \ldots, (C_{I-1,1}^*, C_{I-1,2}^*)\}\,,$$

then this would mean that we have to use $(C_{i2}^*)_{1 \leq i \leq I-1}$ as the regressors in the next step. However, because the pairs are the fundamental unit of resampling for this method, we have to ensure that every bootstrap replicate remain tied to the same response, effectively forcing us to permute entire rows $\{(C_{ik})_{j \leq k \leq I} \mid 1 \leq i \leq I - j\}$ at the jth step of the process. But this creates

Input: Cumulative claims triangle $\mathcal{D}_I$, number of bootstrap samples B

$(\{r_{ij} \mid i+j \leq I+1\}, \widehat{\boldsymbol{f}}, \widehat{\boldsymbol{\sigma}}) \leftarrow \text{FIT}(\mathcal{D}_I)$

for $b \leftarrow 1, B$ **do**

    $\{r_{ij}^{(b)} \mid i+j \leq I+1\} \leftarrow \text{RESAMPLE}(\{r_{ij} \mid i+j \leq I+1\})$

    for $j \leftarrow 1, I-1$ **do**

        for $i \leftarrow 1, I-j$ **do**

            $C_{i,j+1}^{(b)} \leftarrow \widehat{f}_j C_{ij} + \widehat{\sigma}_j \sqrt{C_{ij}} r_{i,j+1}^{(b)}$

            $F_{i,j+1}^{(b)} \leftarrow C_{i,j+1}^{(b)}/C_{ij}$

        end for

        $\widehat{f}_j^{(b)} \leftarrow \sum_{i=1}^{I-j} C_{i,j+1}^{(b)} / \sum_{i=1}^{I-j} C_{ij}$

        if $j < I-1$ **then**

            $\widehat{\sigma}_j^{(b)} \leftarrow \sqrt{\frac{1}{I-j-1} \sum_{i=1}^{I-j} C_{ij} \left(F_{i,j+1}^{(b)} - \widehat{f}_j^{(b)}\right)^2}$

        else

            $\widehat{\sigma}_{I-1}^{(b)} \leftarrow \sqrt{\min\left\{\frac{(\widehat{\sigma}_{I-2}^{(b)})^4}{(\widehat{\sigma}_{I-3}^{(b)})^2}, (\widehat{\sigma}_{I-2}^{(b)})^2, (\widehat{\sigma}_{I-3}^{(b)})^2\right\}}$

        end if

    end for

end for

return $\{(\widehat{\boldsymbol{f}}^{(b)}, \widehat{\boldsymbol{\sigma}}^{(b)}) \mid b = 1, \ldots, B\}$

Algorithm 1: Conditional semiparametric bootstrap for the Mack chain ladder

---

Input: Cumulative claims triangle $\mathcal{D}_I$, number of bootstrap samples B

$(\{r_{ij} \mid i+j \leq I+1\}, \widehat{\boldsymbol{f}}, \widehat{\boldsymbol{\sigma}}) \leftarrow \text{FIT}(\mathcal{D}_I)$

for $b \leftarrow 1, B$ **do**

    $\{r_{ij}^{(b)} \mid i+j \leq I+1\} \leftarrow \text{RESAMPLE}(\{r_{ij} \mid i+j \leq I+1\})$

    for $i \leftarrow 1, I$ **do**

        $C_{i1}^{(b)} \leftarrow C_{i1}$

    end for

    for $j \leftarrow 1, I-1$ **do**

        for $i \leftarrow 1, I-j$ **do**

            $C_{i,j+1}^{(b)} \leftarrow \widehat{f}_j C_{ij}^{(b)} + \widehat{\sigma}_j \sqrt{C_{ij}^{(b)}} r_{i,j+1}^{(b)}$

            $F_{i,j+1}^{(b)} \leftarrow C_{i,j+1}^{(b)}/C_{ij}^{(b)}$

        end for

        $\widehat{f}_j^{(b)} \leftarrow \sum_{i=1}^{I-j} C_{i,j+1}^{(b)} / \sum_{i=1}^{I-j} C_{ij}^{(b)}$

        if $j < I-1$ **then**

            $\widehat{\sigma}_j^{(b)} \leftarrow \sqrt{\frac{1}{I-j-1} \sum_{i=1}^{I-j} C_{ij}^{(b)} \left(F_{i,j+1}^{(b)} - \widehat{f}_j^{(b)}\right)^2}$

        else

            $\widehat{\sigma}_{I-1}^{(b)} \leftarrow \sqrt{\min\left\{\frac{(\widehat{\sigma}_{I-2}^{(b)})^4}{(\widehat{\sigma}_{I-3}^{(b)})^2}, (\widehat{\sigma}_{I-2}^{(b)})^2, (\widehat{\sigma}_{I-3}^{(b)})^2\right\}}$

        end if

    end for

end for

return $\{(\widehat{\boldsymbol{f}}^{(b)}, \widehat{\boldsymbol{\sigma}}^{(b)}) \mid b = 1, \ldots, B\}$

Algorithm 2: Unconditional semiparametric bootstrap for the Mack chain ladder

j	$\widehat{f}_j^B$	$\widehat{\sigma}_j^B$	$\widehat{R}_j^B$	$\widehat{\mathrm{MSEP}}(\widehat{R}_j^B)$
2	1.89	2.69	338.57	22.28
3	1.28	4.86	967.69	118.79
4	1.15	5.58	2009.03	240.38
5	1.10	4.48	3601.84	450.97
6	1.05	1.20	7091.87	661.19
7	1.03	0.54	14305.17	822.05

(a) Conditional

j	$\widehat{f}_j^B$	$\widehat{\sigma}_j^B$	$\widehat{R}_j^B$	$\widehat{\mathrm{MSEP}}(\widehat{R}_j^B)$
2	1.89	2.70	350.79	3.72
3	1.28	3.19	1036.88	22.91
4	1.15	2.76	2043.54	141.99
5	1.10	0.96	3660.72	425.09
6	1.05	0.12	7155.61	694.83
7	1.03	0.03	14385.37	905.82

(b) Unconditional

Table 3.2: Results from the semiparameteric bootstrap with standardised residuals for the data in Table 1.2

j	$\widehat{f}_j^B$	$\widehat{\sigma}_j^B$	$\widehat{R}_j^B$	$\widehat{\mathrm{MSEP}}(\widehat{R}_j^B)$
2	1.89	5.11	338.79	41.04
3	1.28	8.88	961.64	209.89
4	1.15	9.43	2009.08	497.71
5	1.10	6.70	3620.87	950.44
6	1.05	1.87	7112.64	1378.98
7	1.03	0.81	14318.64	1715.18

(a) Conditional

j	$\widehat{f}_j^B$	$\widehat{\sigma}_j^B$	$\widehat{R}_j^B$	$\widehat{\mathrm{MSEP}}(\widehat{R}_j^B)$
2	1.89	5.11	350.64	7.91
3	1.28	6.01	1037.24	49.44
4	1.15	5.27	2045.67	297.21
5	1.10	1.66	3651.50	895.15
6	1.05	0.23	7155.71	1446.47
7	1.03	0.08	14395.86	1887.46

(b) Unconditional

Table 3.3: Results from the semiparameteric bootstrap with studentised residuals for the data in Table 1.2

j	$\widehat{f}_j^B$	$\widehat{\sigma}_j^B$	$\widehat{R}_j^B$	$\widehat{\mathrm{MSEP}}(\widehat{R}_j^B)$
2	1.89	2.38	338.74	20.35
3	1.28	4.47	969.44	110.29
4	1.14	5.15	2023.23	217.67
5	1.10	4.16	3607.15	397.80
6	1.05	1.16	7100.02	586.11
7	1.03	0.52	14313.18	725.15

(a) Conditional

j	$\widehat{f}_j^B$	$\widehat{\sigma}_j^B$	$\widehat{R}_j^B$	$\widehat{\mathrm{MSEP}}(\widehat{R}_j^B)$
2	1.89	2.38	350.95	3.26
3	1.28	2.74	1037.37	20.27
4	1.15	2.45	2044.65	125.25
5	1.10	0.84	3660.69	378.25
6	1.05	0.11	7159.14	615.27
7	1.03	0.03	14393.80	799.03

(b) Unconditional

Table 3.4: Results from the semiparameteric bootstrap with log-normal residuals for the data in Table 1.2

a problem: the last point in the second column does not have a successor in the triangle, and we therefore become stuck in the next step if we had drawn it previously, as illustrated by Fig. 3.1.

Hence, our only option is to resample each pair of successive columns from the original data and use it to compute bootstrap replicates for the corresponding development factor and variance parameter estimators. This procedure is outlined in Algorithm 3, and the results for the data in Table 1.2 are given in Table 3.5. Note that the 'different methods in the simulation step' refer
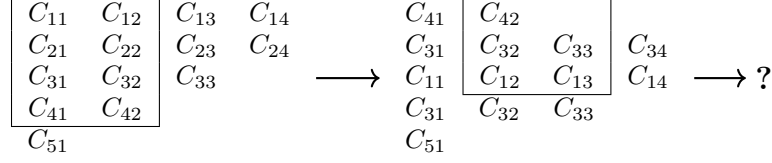
$$
\begin{array}{|cc|cc} \hline C_{11} & C_{12} & C_{13} & C_{14} \\ C_{21} & C_{22} & C_{23} & C_{24} \\ C_{31} & C_{32} & C_{33} \\ C_{41} & C_{42} \\ \hline C_{51} \end{array}
\qquad \longrightarrow \qquad
\begin{array}{c|ccc|c} C_{41} & C_{42} \\ C_{31} & C_{32} & C_{33} & C_{34} \\ C_{11} & C_{12} & C_{13} & C_{14} \\ C_{31} & C_{32} & C_{33} \\ C_{51} \end{array}
\longrightarrow \ ?
$$

Figure 3.1: Example of the failure of unconditional resampling for the pairs bootstrap

to the peculiarity of the pairs bootstrap discussed in Section 2.3, which will be clarified further in Section 3.4.

Input: Cumulative claims triangle $\mathcal{D}_I$, number of bootstrap samples B

for $b \leftarrow 1, B$ **do**

    for $j \leftarrow 1, I-1$ **do**

        $\{(C_{i,j}^{(b)}, C_{i,j+1}^{(b)}) \mid i = 1, \ldots, I-j\} \leftarrow \textsc{resample}(\{(C_{i,j}, C_{i,j+1}) \mid i = 1, \ldots, I-j\})$

        $\widehat{f}_j^{(b)} \leftarrow \sum_{i=1}^{I-j} C_{i,j+1}^{(b)} / \sum_{i=1}^{I-j} C_{ij}^{(b)}$

        for $i \leftarrow 1, I-j$ **do**

            $F_{i,j+1}^{(b)} \leftarrow C_{i,j+1}^{(b)} / C_{ij}^{(b)}$

        end for

        if $j < I-1$ **then**

            $\widehat{\sigma}_j^{(b)} \leftarrow \sqrt{\frac{1}{I-j-1} \sum_{i=1}^{I-j} C_{ij}^{(b)} \left(F_{i,j+1}^{(b)} - \widehat{f}_j^{(b)} \right)^2}$

        else

            $\widehat{\sigma}_{I-1}^{(b)} \leftarrow \sqrt{\min\left\{ \frac{(\widehat{\sigma}_{I-2}^{(b)})^4}{(\widehat{\sigma}_{I-3}^{(b)})^2}, (\widehat{\sigma}_{I-2}^{(b)})^2, (\widehat{\sigma}_{I-3}^{(b)})^2 \right\}}$

        end if

    end for

end for

return $\{(\widehat{\boldsymbol{f}}^{(b)}, \widehat{\boldsymbol{\sigma}}^{(b)}) \mid b = 1, \ldots, B\}$

Algorithm 3: Pairs bootstrap for the Mack chain ladder

Finally, we consider the parametric bootstrap, which relies on simulation from the underlying model to produce pseudo-observations. In light of Eq. (3.48), we might be tempted to implement such a procedure by simply adapting Algorithms 1 and 2 to use $\mathcal{N}(0,1)$-distributed draws from a random number generator in place of the residuals. The drawback of this approach is that it does not generalise well to other distributions, because these rarely admit an additive decomposition into mean and error terms. It is therefore preferable to generate bootstrap pseudo-responses directly from the fitted distribution, i.e.

$$
C_{i,j+1}^{(b)} \sim \mathcal{N}(\widehat{f}_j C_{ij}, \widehat{\sigma}_j^2 C_{ij}) \quad \text{or} \quad C_{i,j+1}^{(b)} \sim \mathcal{N}(\widehat{f}_j C_{ij}^{(b)}, \widehat{\sigma}_j^2 C_{ij}^{(b)}), \tag{3.57}
$$

depending on whether the conditional or the unconditional approach is used. The model is then refitted to the bootstrapped triangle in order to obtain $\widehat{\boldsymbol{f}}^{(b)}$ and $\widehat{\boldsymbol{\sigma}}^{(b)}$. The algorithms and results are given in Algorithms 4 and 5 and Tables 3.6 and 3.7.

Input: Cumulative claims triangle $\mathcal{D}_I$, number of bootstrap samples B

   $(\widehat{\boldsymbol{f}}, \widehat{\boldsymbol{\sigma}}) \leftarrow \text{FIT}(\mathcal{D}_I)$

   for $b \leftarrow 1, B$ **do**

      for $j \leftarrow 1, I-1$ **do**

         for $i \leftarrow 1, I-j$ **do**

            $C_{i,j+1}^{(b)} \leftarrow \text{SAMPLE}(\mathcal{N}(\widehat{f}_j C_{ij}, \widehat{\sigma}_j^2 C_{ij}))$

            $F_{i,j+1}^{(b)} \leftarrow C_{i,j+1}^{(b)}/C_{ij}$

         end for

         $\widehat{f}_j^{(b)} \leftarrow \sum_{i=1}^{I-j} C_{i,j+1}^{(b)} / \sum_{i=1}^{I-j} C_{ij}$

         if $j < I-1$ **then**

            $\widehat{\sigma}_j^{(b)} \leftarrow \sqrt{\frac{1}{I-j-1} \sum_{i=1}^{I-j} C_{ij} \left(F_{i,j+1}^{(b)} - \widehat{f}_j^{(b)}\right)^2}$

         else

            $\widehat{\sigma}_{I-1}^{(b)} \leftarrow \sqrt{\min\left\{\frac{(\widehat{\sigma}_{I-2}^{(b)})^4}{(\widehat{\sigma}_{I-3}^{(b)})^2}, (\widehat{\sigma}_{I-2}^{(b)})^2, (\widehat{\sigma}_{I-3}^{(b)})^2\right\}}$

         end if

      end for

   end for

   return $\{(\widehat{\boldsymbol{f}}^{(b)}, \widehat{\boldsymbol{\sigma}}^{(b)}) \mid b = 1, \ldots, B\}$

Algorithm 4: Conditional parametric bootstrap for the Mack chain ladder

---

Input: Cumulative claims triangle $\mathcal{D}_I$, number of bootstrap samples B

   $(\widehat{\boldsymbol{f}}, \widehat{\boldsymbol{\sigma}}) \leftarrow \text{FIT}(\mathcal{D}_I)$

   for $b \leftarrow 1, B$ **do**

      for $i \leftarrow 1, I$ **do**

         $C_{i1}^{(b)} \leftarrow C_{i1}$

      end for

      for $j \leftarrow 1, I-1$ **do**

         for $i \leftarrow 1, I-j$ **do**

            $C_{i,j+1}^{(b)} \leftarrow \text{SAMPLE}(\mathcal{N}(\widehat{f}_j C_{ij}^{(b)}, \widehat{\sigma}_j^2 C_{ij}^{(b)}))$

            $F_{i,j+1}^{(b)} \leftarrow C_{i,j+1}^{(b)}/C_{ij}^{(b)}$

         end for

         $\widehat{f}_j^{(b)} \leftarrow \sum_{i=1}^{I-j} C_{i,j+1}^{(b)} / \sum_{i=1}^{I-j} C_{ij}^{(b)}$

         if $j < I-1$ **then**

            $\widehat{\sigma}_j^{(b)} \leftarrow \sqrt{\frac{1}{I-j-1} \sum_{i=1}^{I-j} C_{ij}^{(b)} \left(F_{i,j+1}^{(b)} - \widehat{f}_j^{(b)}\right)^2}$

         else

            $\widehat{\sigma}_{I-1}^{(b)} \leftarrow \sqrt{\min\left\{\frac{(\widehat{\sigma}_{I-2}^{(b)})^4}{(\widehat{\sigma}_{I-3}^{(b)})^2}, (\widehat{\sigma}_{I-2}^{(b)})^2, (\widehat{\sigma}_{I-3}^{(b)})^2\right\}}$

         end if

      end for

   end for

   return $\{(\widehat{\boldsymbol{f}}^{(b)}, \widehat{\boldsymbol{\sigma}}^{(b)}) \mid b = 1, \ldots, B\}$

Algorithm 5: Unconditional parametric bootstrap for the Mack chain ladder

j	$\widehat{f}_j^B$	$\widehat{\sigma}_j^B$	$\widehat{R}_j^B$	$\widehat{\text{MSEP}}(\widehat{R}_j^B)$
2	1.89	6.75	350.90	2.54
3	1.28	8.80	1037.58	20.72
4	1.15	6.86	2038.59	136.14
5	1.10	0.77	3663.60	415.48
6	1.05	0.01	7158.19	675.49
7	1.03	0.00	14396.28	878.12

(a) Semiparametric

j	$\widehat{f}_j^B$	$\widehat{\sigma}_j^B$	$\widehat{R}_j^B$	$\widehat{\text{MSEP}}(\widehat{R}_j^B)$
2	1.89	6.74	350.90	2.54
3	1.28	8.81	1037.16	20.69
4	1.15	6.59	2045.14	135.51
5	1.10	0.78	3664.09	414.79
6	1.05	0.01	7164.13	682.63
7	1.03	0.00	14399.01	884.44

(b) Parametric

Table 3.5: Results from the pairs bootstrap with different methods in the simulation step for the data in Table 1.2

j	$\widehat{f}_j^B$	$\widehat{\sigma}_j^B$	$\widehat{R}_j^B$	$\widehat{\text{MSEP}}(\widehat{R}_j^B)$
2	1.89	2.70	337.81	97.30
3	1.28	4.97	968.55	202.59
4	1.14	5.43	2026.69	583.82
5	1.10	4.42	3612.92	916.39
6	1.05	1.24	7088.42	1227.17
7	1.03	0.55	14313.72	1416.73

(a) Conditional

j	$\widehat{f}_j^B$	$\widehat{\sigma}_j^B$	$\widehat{R}_j^B$	$\widehat{\text{MSEP}}(\widehat{R}_j^B)$
2	1.89	2.72	350.77	6.90
3	1.28	3.19	1037.25	19.24
4	1.15	2.82	2048.48	121.06
5	1.10	0.94	3667.20	389.40
6	1.05	0.12	7161.25	619.39
7	1.03	0.03	14393.47	789.32

(b) Unconditional

Table 3.6: Results from the parametric bootstrap with normal distribution for the data in Table 1.2

j	$\widehat{f}_j^B$	$\widehat{\sigma}_j^B$	$\widehat{R}_j^B$	$\widehat{\text{MSEP}}(\widehat{R}_j^B)$
2	1.89	2.70	337.75	96.58
3	1.28	4.94	962.76	200.40
4	1.15	5.54	2022.25	593.79
5	1.10	4.53	3615.22	936.06
6	1.05	1.21	7095.24	1237.38
7	1.03	0.54	14313.25	1434.88

(a) Conditional

j	$\widehat{f}_j^B$	$\widehat{\sigma}_j^B$	$\widehat{R}_j^B$	$\widehat{\text{MSEP}}(\widehat{R}_j^B)$
2	1.89	2.70	350.93	7.30
3	1.28	3.18	1037.30	19.84
4	1.15	2.73	2042.37	122.93
5	1.10	0.95	3653.75	376.26
6	1.05	0.13	7156.47	610.36
7	1.03	0.03	14394.57	781.67

(b) Unconditional

Table 3.7: Results from the parametric bootstrap with gamma distribution for the data in Table 1.2

3.4   Incorporating the process error

We end this chapter by discussing how the process error can be incorporated into the bootstrap procedures from the previous section. Recall that our aim is ultimately to obtain a predictive distribution for the reserve incorporating both parameter and process error, which we can then use to study the impact of violations of the Mack chain ladder assumptions. Following the

procedure outlined in Section 2.3, this means that we have to generate pseudo-realisations of the reserve which fluctuate around point estimates obtained using bootstrapped development factors and variance parameters. In light of Eq. (3.9), the latter are given by

$$R^{(b)} := \sum_{i=2}^{I} (C_{iI}^{(b)} - C_{i,I+1-i}) \tag{3.58}$$

with $\widehat{C}_{iI}^{(b)} = C_{i,I+1-i} \prod_{j=I+1-i}^{I-1} \widehat{f}_j^{(b)}$. Bootstrap methods are always derived from the underlying data-generating process; it therefore follows that we must generate pseudo-realisations of the future claim amounts in the lower triangle $\mathcal{D}_I^{\complement} = \{C_{ij} \mid i+j > I+1\}$ in order to simulate the process error of the reserve.

The procedures from the previous section can be easily adapted to achieve this. For the semiparametric bootstrap, we resample the residuals a second time and again use Eq. (3.48), substitute the bootstrap replicates of the parameter estimators this time, i.e.

$$C_{i,j+1}^{(s)} = \widehat{f}_j^{(b)} C_{ij} + \widehat{\sigma}_j^{(b)} \sqrt{C_{ij}} r_{i,j+1}^{(s)} \quad \text{or} \quad C_{i,j+1}^{(s)} = \widehat{f}_j^{(b)} C_{ij}^{(s)} + \widehat{\sigma}_j^{(b)} \sqrt{C_{ij}^{(s)}} r_{i,j+1}^{(s)}, \tag{3.59}$$

depending on whether the conditional or unconditional approach is used. A parametric version is obtained simulating from the distribution obtained by plugging in the $\widehat{f}_j^{(b)}$ and $\widehat{\sigma}_j^{(b)}$, and the pairs bootstrap must borrow from one of them, as described in Section 2.3.

The same problems encountered in Section 3.3 regarding the possibility of negative draws crop up here as well. For the semiparametric bootstrap, they can be solved in the same way using the shifted log-normal residuals from Eq. (3.55), which by construction cannot lead to negative pseudo-responses. For the parametric bootstrap, we follow the previously mentioned suggestion in [10] and simulate future claim amounts from a gamma distribution. In the conditional approach, for example, if $C_{ij}^{(s)} \sim \Gamma(\alpha_{ij}^{(b)}, \beta_{ij}^{(b)})$, then we need to have

$$\mathbb{E}[C_{ij} \parallel C_{i,j-1}] = \frac{\alpha_{ij}^{(b)}}{\beta_{ij}^{(b)}} = f_{j-1}^{(b)} C_{i,j-1} \quad \text{and} \quad \text{Var}(C_{ij} \parallel C_{i,j-1}) = \frac{\alpha_{ij}^{(b)}}{(\beta_{ij}^{(b)})^2} = (\sigma_{j-1}^{(b)})^2 C_{i,j-1}, \tag{3.60}$$

and must therefore choose parameters

$$\alpha_{ij}^{(b)} = \frac{(f_{j-1}^{(b)})^2 C_{i,j-1}}{(\sigma_{j-1}^{(b)})^2} \quad \text{and} \quad \beta_{ij}^{(b)} = \frac{f_{j-1}^{(b)}}{(\sigma_{j-1}^{(b)})^2}. \tag{3.61}$$

The solution adopted for the pairs bootstrap will depend on which of the two other variants is used in the simulation step. We will use the parametric approach for the simulations in Chapter 5.

In Section 2.3, we gave not much difference could be discerned between the fitted and predictive distributions. This stands in stark contrast to Figs. 3.2 to 3.5, which compare these for the reserve under different bootstrap configurations. We can clearly see that fitted distribution severely underestimates the uncertainty of the reserve in all cases.
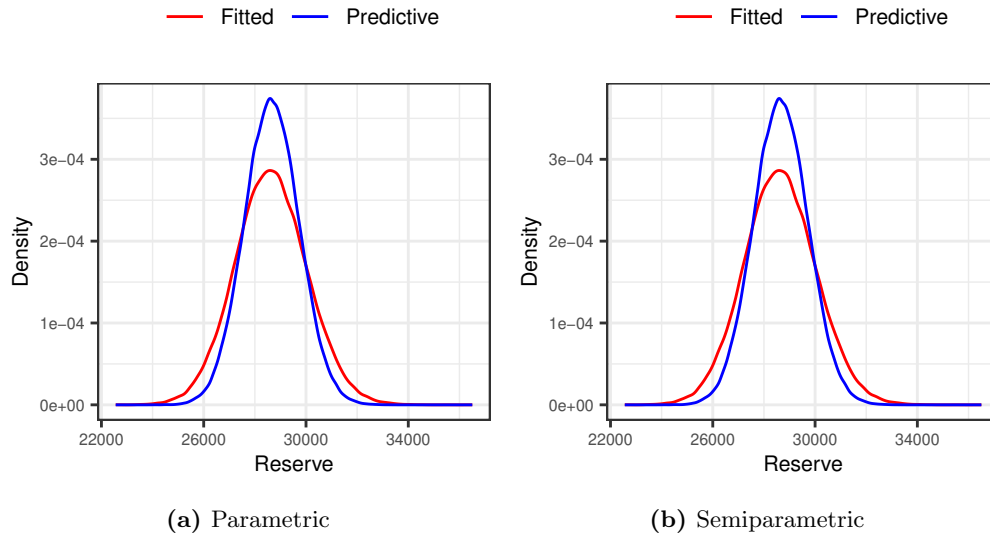
(a) Parametric

(b) Semiparametric

Figure 3.2: Comparison of the fitted and predictive distribution for the pairs bootstrap with different simulation methods
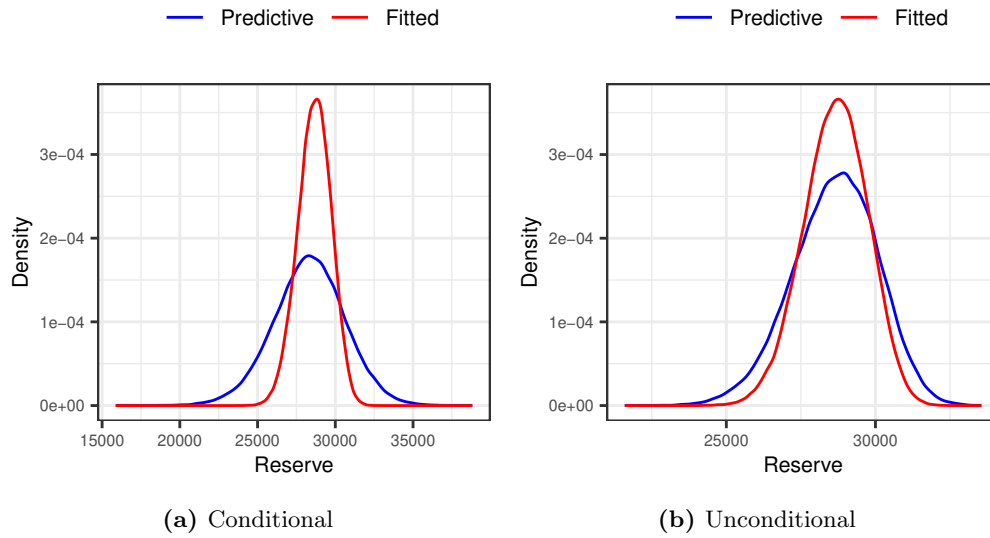

(a) Conditional
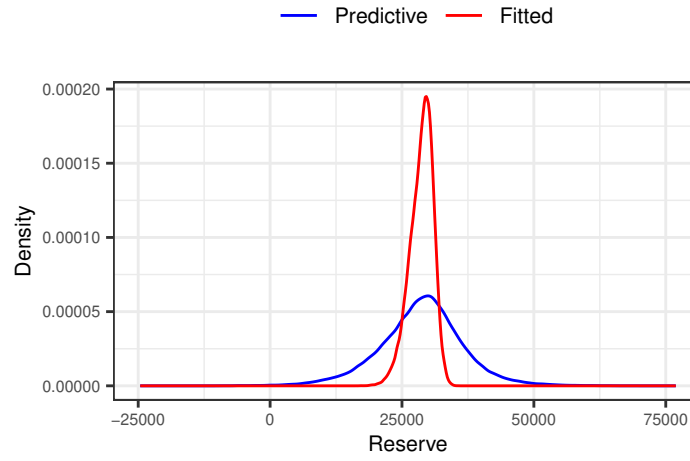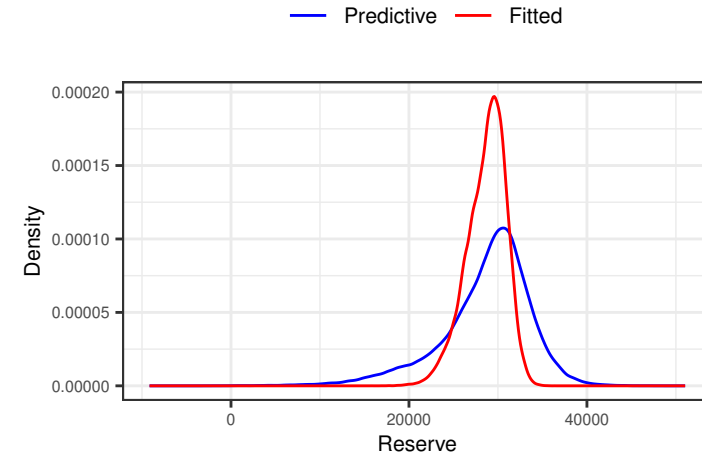

(b) Unconditional

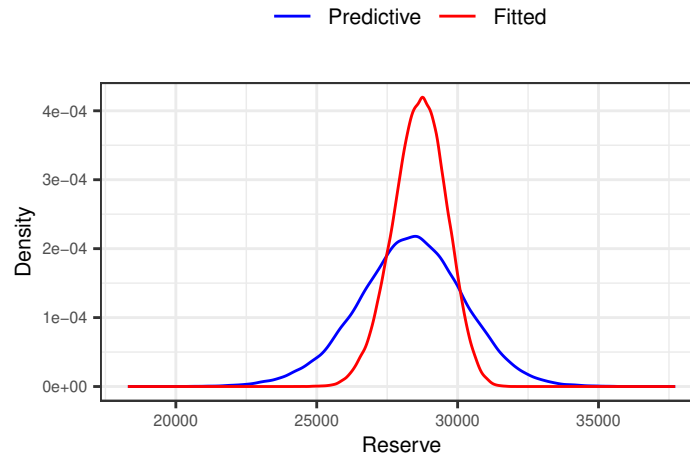**Figure 3.3:** Comparison of the fitted and predictive distribution for the semiparametric bootstrap with standardised residuals

(a) Conditional, studentised residuals

(b) Unconditional, studentised residuals

(c) Conditional, log-normal residuals
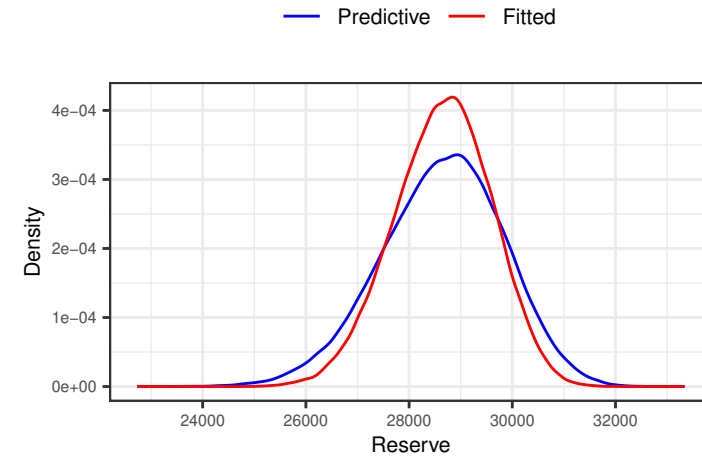
(d) Unconditional, log-normal residuals

Figure 3.4: Comparison of the fitted and predictive distribution for the semiparametric bootstrap
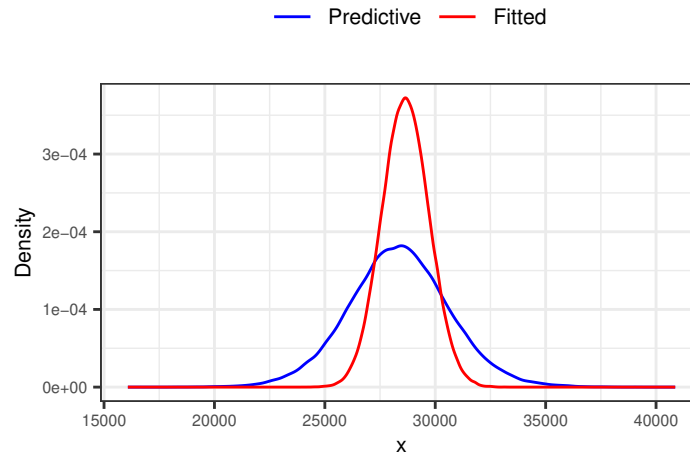
(a) Conditional, normal distribution

(b) Unconditional, normal distribution

(c) Conditional, gamma distribution

(d) Unconditional, gamma distribution

Figure 3.5: Comparison of the fitted and predictive distribution for the parametric bootstrap

Chapter 4

Overdispersed Poisson GLM

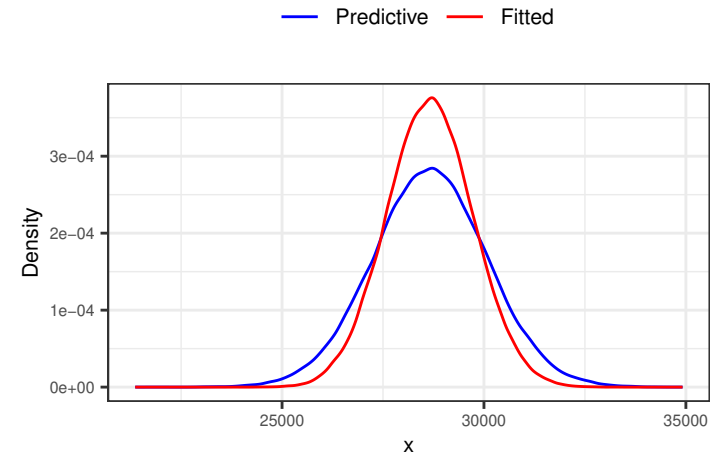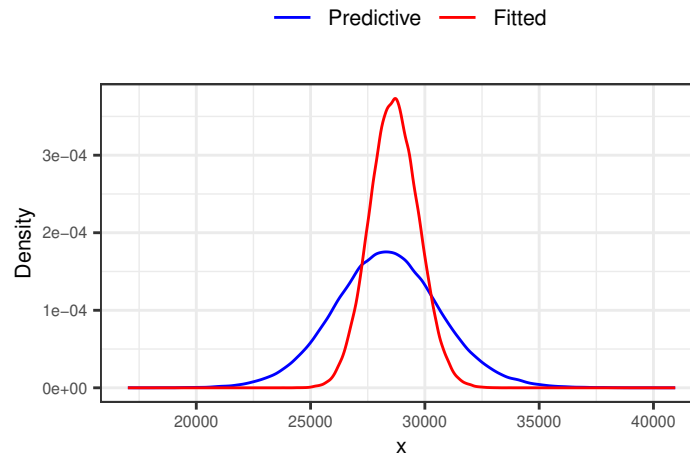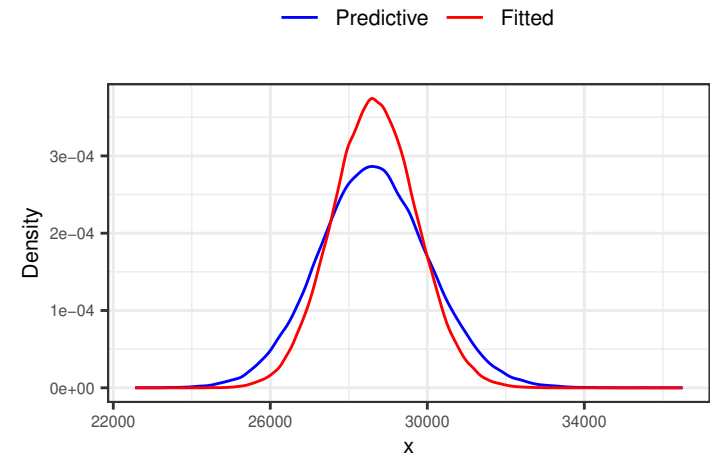## 4.1   Introduction

The overdispersed Poisson model (ODP), proposed by Renshaw and Verrall in [18], belongs to the family of so-called *generalised linear models* (GLM). In contrast to the Mack chain ladder, it describes the incremental claims X_{ij}, and is non-recursive: the diffferent observations in the triangle are modelled independently from each other. To ease the exposition, we will only state the assumptions for the ordinary Poisson model (i.e. without overdispersion) at this point; after this concept has been introduced in Section 4.2.2, we will then adjust the model accordingly.

Model 2 (Poisson GLM).

    (Pois1) *The incremental claims are independent from each other.*

    (Pois2) *There exist parameters c, $a_1, \ldots, a_I$ and $b_1, \ldots, b_I$ such that*

$$\log(\mathbb{E}[X_{ij}]) = c + a_i + b_j \, , \tag{4.1}$$

with $a_1 = b_1 = 0$.

    (Pois3) *The incremental claims follow a Poisson distribution:*

$$X_{ij} \sim \mathrm{Pois}(e^{c+a_i+b_j}) \, . \tag{4.2}$$

The condition $a_1 = b_1 = 0$ is necessary to obtain an identifiable model; without it, we could derive, from any set of parameters $c, a_1, \ldots, a_I, b_1, \ldots, b_I$ satisfying the assumptions, an infinite number of alternatives $c + a' + b', a_1 - a', \ldots, a_I - a', b_1 - b', \ldots, b_I - b'$ for $a', b' \in \mathbb{R}$. We therefore have two superfluous degrees of freedom, which can be eliminated by imposing the same number of conditions on the parameters.

By defining $\xi_i := e^{c+a_i}$ and $\gamma_j := e^{b_j}$, we get a different parametrisation of the model which has a multiplicative structure for the mean, i.e.

$$\mathbb{E}[X_{ij}] = \xi_i \gamma_j \, . \tag{4.3}$$

This is sometimes preferred to the additive one for reasons of interpretability. Indeed, it is clear that the multiplicative form has one fewer degree of freedom than the linear one, and if we remove this by imposing the constraint

$$\sum_{j=1}^{I} \gamma_j = 1 \tag{4.4}$$

then we can view the ξ_i as expected ultimate claim amounts, and the γ_j as the expected development pattern.

As mentioned in the introduction, stochastic claims reserving models have to reproduce the chain ladder point predictions in order to be acceptable to practitioners. While less obvious than for the Mack chain ladder, it can be shown that the Poisson GLM also satisfies this requirement (see [8, Lemma 2.16]).

4.2 Generalised linear models

GLMs were first conceived by Nelder and Wedderburn in [19] as a way of unifying the many disparate generalisations of linear regression with Gaussian errors which were then in existence. These sought to extend the classical model by allowing the use of different functional forms for the conditional mean and different distributions for the response, thus making it suited to modelling counts data (Poisson regression) or the probability of binary events (logistic regression), among others. For a set of covariates $X_1, \ldots, X_p$ and a response variable Y, a GLM consists of three parts:

1. The *random component*, a distribution for response Y belonging to the so-called *exponential dispersion model* family (EDM), which consists of all probability distributions whose density (with respect either to the Lebesgue or counting measure) has the form

$$p(y \mid \theta, \phi) = \exp\left\{ \frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi) \right\}, \tag{4.5}$$

   where a, b and c are known functions, and b is at least twice differentiable. We call θ the *canonical parameter* of the distribution and ϕ the *dispersion parameter*.

2. The *systematic component*, a predictor $\eta := \mathbf{x}^T \boldsymbol{\beta}$ which is a linear function of the covariates.

3. A monotonic differentiable link function $g : \mathbb{R} \to \mathbb{R}$ giving the relation between the conditional expectation and the linear predictor,

$$\mu := \mathbb{E}[Y \parallel X_1, \ldots, X_p] = g^{-1}(\eta). \tag{4.6}$$

The Gaussian distribution $\mathcal{N}(\mu, \sigma^2)$ can be seen to belong to the EDM family by rewriting its density as

$$\frac{1}{\sqrt{2\pi}\sigma} \exp\left\{ -\frac{1}{2}\left(\frac{y - \mu}{\sigma} \right)^2 \right\} = \exp\left\{ -\frac{y^2}{2\sigma^2} + \frac{y\mu}{\sigma^2} - \frac{\mu^2}{2\sigma^2} - \log(\sqrt{2\pi}\sigma) \right\} \tag{4.7}$$

$$= \exp\left\{ \frac{y\mu - \mu^2/2}{\sigma^2} - \frac{y^2}{2\sigma^2} - \log(\sqrt{2\pi}\sigma) \right\}, \tag{4.8}$$

which is of the form Eq. (4.5) with $\theta = \mu$, $b(\theta) = \frac{\theta^2}{2}$, $\phi = \sigma^2$, $a(\phi) = \phi$ and $c(y, \sigma) = -\frac{y^2}{2\sigma^2} - \log(\sqrt{2\pi}\sigma)$. Thus, the familiar normal linear model can be obtained from the GLM framework by choosing the response distribution $\mathcal{N}(\mu, \sigma^2)$ and link function $g(\mu) = \mu$.

4.2.1 Estimation

The EDM family has a number of properties which greatly facilitate the computations involved in estimation. Recall from likelihood theory that $l(\theta \mid y, \phi) := \log p(y \mid \theta, \phi)$ satisfies

$$\mathbb{E}\left[\frac{\partial l(\theta \mid Y)}{\partial \theta} \right] = 0, \qquad \mathrm{Var}\left(\frac{\partial l(\theta \mid Y)}{\partial \theta} \right) = -\mathbb{E}\left[\frac{\partial^2 l(\theta \mid Y)}{\theta^2} \right], \tag{4.9}$$

where $\frac{\partial l(\theta|Y)}{\partial \theta}$ is known as the *score function*. Using Eq. (4.5), we then find that

$$\mathbb{E}\left[\frac{Y - b'(\theta)}{a(\phi)}\right] = 0\,, \qquad \mathrm{Var}\left(\frac{Y - b'(\theta)}{a(\phi)}\right) = -\mathbb{E}\left[\frac{-b''(\theta)}{a(\phi)}\right]\,, \tag{4.10}$$

from which we obtain the elegant relations

$$\mu = b'(\theta)\,, \qquad \mathrm{Var}(Y) = a(\phi)b''(\theta)\,. \tag{4.11}$$

Observe that this implies that $\frac{d\mu}{d\theta} = b''(\theta) > 0$ (because the variance is always positive), which means that $\theta \mapsto \mu(\theta)$ is one-to-one and therefore invertible. In particular, we can always write the likelihood as function of the mean. The function $V(\mu) := b''((b')^{-1}(\mu))$ is called the *variance function* and determines how the scale of the response varies as a function of its mean.

Special care has to be taken with the parameter ϕ, as it occupies a rather awkward position in GLM theory.

The problem arises from the desire of the progenitors of the GLM framework to incorporate two-parameter distributions such as the normal and gamma distribution into a paradigm which can fundamentally only handle a single parameter gracefully (the more flexible approach of *vector GLMs* is an attempt to remedy this; see [20, Chapter 2] for a general discussion). The dispersion is therefore regelated to the role of nuisance parameter and subjected to severe (and often unrealistic) constraints. In essence, one would like ϕ to be constant as a function of the covariates, but this precludes certain special cases such as binomial regression with a different number of trials for each observation in the sample which practitioners wanted to accommodate. In the end, the compromise which seems to have gained the widest acceptance in the literature is to allow the function a in the denominator of Eq. (4.5) to vary across different sample responses as $a_i(\phi) = \phi/w_i$, where w_i is a known weight. The parameter ϕ itself is then considered as known, and estimated outside of the GLM framework, most commonly using the Pearson statistic

$$\widehat{\phi} := \frac{1}{n - p} \sum_{i=1}^{N} \frac{(Y_i - \widehat{\mu_i})^2}{V(\widehat{\mu_i})}\,. \tag{4.12}$$

Given a sample $(\mathbf{x_1}, Y_1), \ldots, (\mathbf{x_N}, Y_N)$, the standard way to fit a GLM is by means of maximum likelihood estimation (MLE). The joint log-likelihood of the sample is given by

$$l(\boldsymbol{\beta} \mid \mathbf{y}, \phi) = \sum_{i=1}^{N} \frac{y_i \theta_i - b(\theta_i)}{a_i(\phi)} + c(y_i, \phi)\,, \tag{4.13}$$

which we must differentiate with respect to β_j to obtain the likelihood equations. An application of the chain rule gives us

$$\frac{\partial l(\boldsymbol{\beta} \mid \mathbf{y}, \phi)}{\partial \beta_j} = \sum_{i=1}^{N} \frac{\partial l(\boldsymbol{\beta} \mid y_i, \phi)}{\partial \theta_i} \frac{\partial \theta_i}{\partial \mu_i} \frac{\partial \mu_i}{\partial \eta_i} \frac{\partial \eta_i}{\partial \beta_j} \tag{4.14}$$

$$= \sum_{i=1}^{N} \frac{y_i - b_i'(\theta)}{a_i(\phi)} \frac{1}{b_i''(\theta_i)} \frac{\partial \mu_i}{\partial \eta_i} x_{ij} \tag{4.15}$$

$$= \sum_{i=1}^{N} \frac{y_i - \mu_i}{\mathrm{Var}(Y_i)} \frac{\partial \mu_i}{\partial \eta_i} x_{ij}\,, \tag{4.16}$$

and setting this equal to 0 yields a system of p (usually nonlinear) equations. It is generally impossible to solve these analytically, and so we must resort to numerical methods. In particular, we use a modified version of the Newton-Raphson algorithm known as *Fisher scoring*,

which replaces the negative Hessian of the log-likelihood, called the *observed information*, by its expectation

$$\mathcal{I}_{jk} := \mathbb{E}\left[-\frac{\partial^2 l(\boldsymbol{\beta} \mid \mathbf{y}, \phi)}{\partial \beta_j \partial \beta_k}\right],\tag{4.17}$$

which is known as the *Fisher information matrix*. Thus, starting from an initial guess $\widehat{\boldsymbol{\beta}}^{(0)}$ for the parameters, we compute a successive approximations via

$$\widehat{\boldsymbol{\beta}}^{(k+1)} = \widehat{\boldsymbol{\beta}}^{(k)} + \mathcal{I}(\widehat{\boldsymbol{\beta}}^{(k)})^{-1} \nabla l(\widehat{\boldsymbol{\beta}}^{(k)} \mid \mathbf{y}, \phi).\tag{4.18}$$

Similarly to Eq. (4.9), it can be shown that

$$\mathbb{E}\left[\frac{\partial^2 l(\boldsymbol{\beta} \mid \mathbf{y}, \phi)}{\partial \boldsymbol{\beta} \, \partial \boldsymbol{\beta}^T}\right] = -\mathrm{Var}\big(\nabla l(\boldsymbol{\beta} \mid \mathbf{y}, \phi) \nabla l(\boldsymbol{\beta} \mid \mathbf{y}, \phi)^T\big) < 0\,,\tag{4.19}$$

from which we also see that the log-likelihood is concave, and will therefore have a global maximum. Using the fact that the Y_i are independent, so that $\mathbb{E}[(Y_i - \mu_i)(Y_l - \mu_l)] = 0$ for $i \neq l$, we then obtain

$$I_{jk} = \mathbb{E}\left[\left(\sum_{i=1}^{N} \frac{Y_i - \mu_i}{\mathrm{Var}(Y_i)} \frac{\partial \mu_i}{\partial \eta_i} x_{ij}\right)\left(\sum_{l=1}^{N} \frac{Y_l - \mu_l}{\mathrm{Var}(Y_l)} \frac{\partial \mu_l}{\partial \eta_l} x_{lk}\right)\right]\tag{4.20}$$

$$= \sum_{i=1}^{N}\left(\frac{\partial \mu_i}{\partial \eta_i}\right)^2 \frac{\mathbb{E}\big[(Y_i - \mu_i)^2\big]}{\mathrm{Var}(Y_i)^2} x_{ij} x_{ik}\tag{4.21}$$

$$= \sum_{i=1}^{N}\left(\frac{\partial \mu_i}{\partial \eta_i}\right)^2 \frac{x_{ij} x_{ik}}{\mathrm{Var}(Y_i)}\tag{4.22}$$

$$= \mathbf{x}_j^T \mathbf{W} \mathbf{x}_k\tag{4.23}$$

where $\mathbf{W}^{(k)}$ is a diagonal matrix with

$$\mathbf{W}_{ii}^{(k)} = \frac{1}{\mathrm{Var}(Y_i)}\left(\frac{\partial \mu_i}{\partial \eta_i}\right)^2_{\widehat{\boldsymbol{\beta}}^{(k)}}.\tag{4.24}$$

Hence we have $\mathcal{I}(\widehat{\boldsymbol{\beta}}^{(k)}) = \mathbf{X}^T \mathbf{W}^{(k)} \mathbf{X}$ and we see from Eq. (4.16) that

$$\nabla l(\boldsymbol{\beta} \mid \mathbf{y}, \phi) = \mathbf{X}^T \mathbf{W} \tilde{\mathbf{z}}\tag{4.25}$$

with $\tilde{\mathbf{z}}_i = (y_i - \mu_i)\left(\frac{\partial \eta_i}{\partial \mu_i}\right)$. Multiplying both sides of 4.18 by $\mathcal{I}(\boldsymbol{\beta}^{(k)})$ and using Eqs. (4.23) to (4.25), we finally obtain

$$\mathbf{X}^T \mathbf{W}^{(k)} \mathbf{X} \widehat{\boldsymbol{\beta}}^{(k+1)} = \mathbf{X}^T \mathbf{W}^{(k)} \mathbf{z}\tag{4.26}$$

with $\mathbf{z} = \mathbf{X}\widehat{\boldsymbol{\beta}}^{(k)} + \tilde{\mathbf{z}}$ and all quantities evaluated at the current estimate $\boldsymbol{\beta}^{(k)}$ of the parameter vector. In other words, the Fisher scoring is equivalent to a series of weighted least squares problems, where the new parameter estimates are obtained by regressing the vector $\mathbf{z}$ on the original covariates $\mathbf{x}_1, \ldots, \mathbf{x}_N$ using weight matrix $\mathbf{W}$, and $\mathbf{z}$ and $\mathbf{W}$ are determined by the current estimate $\boldsymbol{\beta}^{(k)}$—hence why the algorithm is called *iteratively reweighted least squares* (IRWLS).

This procedure can be specialised to the particular case of the Poisson GLM in the following way. First, in order to obtain the matrix-vector form used above, we must flatten the tabular

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
2009	3	1946
2010	3	2023
2011	3	2320
2007	4	1712
2008	4	1180
2009	4	1522
2010	4	1320
2007	5	1059
2008	5	956
2009	5	1238
2007	6	587
2008	6	629
2007	7	340

Table 4.1: UK Motor incremental claims triangle in long format

response (using, for example, the colexicographical ordering $(i,j) \mapsto jI + i$, i.e. column-major order). This gives us a triangle in long format, as illustrated in Table 4.1 for the incremental form of the UK Motor dataset from Christofides [1]. If we define the parameter vector

$$\boldsymbol{\beta} := \begin{bmatrix} c & a_2 & \cdots & a_I & b_2 & \cdots & b_I \end{bmatrix}^T , \tag{4.27}$$

then Eq. (4.1) can be rewritten as

$$\log(\mu_{ij}) = c + a_i + b_j = (\mathbf{e_1} + \mathbf{e_i} + \mathbf{e_{I+j-1}})^T \boldsymbol{\beta} \tag{4.28}$$

where $\mathbf{e_k}$ denotes the kth standard basis vector in $\mathbb{R}^{(2I-1)}$. Hence we see that the covariates are binary vectors of length $2I - 1$, with the position of the nonzero entries determined by the indices of the observation in the triangle, forming the rows of a very sparse design matrix. As the Poisson GLM uses the log link, we have $\mu_{ij} = e^{\eta_{ij}}$ and

$$\frac{\partial \mu_{ij}}{\partial \eta_{ij}} = e^{\eta_{ij}} , \tag{4.29}$$

$$\underbrace{\begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ 1 & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & 0 & 0 & \dots & 1 \end{bmatrix}}_{\mathbf{X}^T} \underbrace{\begin{bmatrix} e^{\widehat{c}^{(k)}} & & & \\ & e^{\widehat{c}^{(k)}+\widehat{a}_1^{(k)}} & & \\ & & \ddots & \\ & & & e^{\widehat{c}^{(k)}+\widehat{a}_I^{(k)}+\widehat{b}_I^{(k)}} \end{bmatrix}}_{\mathbf{W}^{(k)}} \underbrace{\begin{bmatrix} 1 & 1 & 1 & \dots & 1 \\ 0 & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \dots & 1 \end{bmatrix}}_{\mathbf{X}} \underbrace{\begin{bmatrix} \widehat{c}^{(k)} \\ \widehat{a}_2^{(k)} \\ \vdots \\ \widehat{a}_I^{(k)} \\ \widehat{b}_2^{(k)} \\ \vdots \\ \widehat{b}_I^{(k)} \end{bmatrix}}_{\widehat{\boldsymbol{\beta}}^{(k)}}$$

$$= \underbrace{\begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ 1 & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & 0 & 0 & \dots & 1 \end{bmatrix}}_{\mathbf{X}^T} \underbrace{\begin{bmatrix} e^{\widehat{c}^{(k)}} & & & \\ & e^{\widehat{c}^{(k)}+\widehat{a}_1^{(k)}} & & \\ & & \ddots & \\ & & & e^{\widehat{c}^{(k)}+\widehat{a}_I^{(k)}+\widehat{b}_I^{(k)}} \end{bmatrix}}_{\mathbf{W}^{(k)}} \mathbf{z}$$

Figure 4.1: IRWLS equation for Poisson GLM in matrix form

from which, using Eq. (4.24), we finally obtain

$$\mathbf{W}_{ii} = \frac{1}{e^{\eta_{ij}}}(e^{\eta_{ij}})^2 = e^{\eta_{ij}} \,, \tag{4.30}$$

giving us all the components of the IRWLS algorithm. Fig. 4.1 shows in matrix form.

4.2.2 Quasi-likelihood methods

We have assumed up to this point that a GLM requires us to specify an exact distribution for the response variable. In many practical situations, however, this is either infeasible or leads to unrealistic models. An example which is particularly common with count data is a phenomenon known as *overdispersion*, where the variability of the data is greater than would be suggested by e.g. the Poisson or binomial distribution. Recall that the variance of a Pois(λ) distribution is λ, and that of a $B(n,p)$ distribution is $np(1-p)$; in both cases, it is fully determined by the mean, and we have no degree of freedom with which to adjust it in order to obtain a better fit to the data, as would be the case with the normal distribution, for example.

To remedy this, an extension can be made to the GLM framework, which only relies on the specification of a relation between mean and variance. Recall from above that the MLE works by setting the score equal to 0. If we write the likelihood in terms of $\boldsymbol{\mu}$, this will have components

$$\frac{\partial l(\boldsymbol{\mu} \mid \mathbf{y}, \phi)}{\partial \mu_j} = \frac{y_j - \mu_j}{\phi V(\mu_j)} \,, \tag{4.31}$$

and is therefore completely determined by $V(\cdot)$. Suppose now, conversely, that we start from

$V(\cdot)$. We could then define functions

$$Q_i(\mu \mid y_i, \phi) := \int_{y_i}^{\mu} \frac{y_i - u}{\phi V(u)} \, du \,, \tag{4.32}$$

and estimate $\boldsymbol{\mu}$ (and therefore $\boldsymbol{\beta}$) by minimising

$$Q(\boldsymbol{\mu} \mid \mathbf{y}, \phi) := \sum_{i=1}^{N} Q_i(\mu_i \mid y_i, \phi) \,. \tag{4.33}$$

It must be stressed that Q has no probabilistic significance: it does not, in general, correspond to the log-likelihood of any distribution. Rather, it functions a device to obtain estimates of the desired parameters, fulfilling in this a similar role to that of the log-likelihood, which is why we refer to it as a *quasi-likelihood function*[1]. This extension of the classical GLM framework allows for more flexibility in our modelling, since we only have to assume a particular mean-variance relation for the response, not a distribution. This finally allows us to state the main model of this chapter.

Model 3 (overdispersed Poisson GLM).

   (ODP1) *The incremental claims are independent from each other.*

   (ODP2) *There exist parameters c, $a_1, \ldots, a_I$ and $b_1, \ldots, b_I$ such that*

$$\log(\mu_{ij}) = c + a_i + b_j \,, \tag{4.34}$$

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

   (ODP3) *There exists a parameter ϕ such that*

$$\mathrm{Var}(X_{ij}) = \phi \mu_{ij} \,. \tag{4.35}$$

It is usual to identify quasi-likelihood models derived from specific distributions (i.e. using the same variance function) by prefixing 'quasi' to their name, e.g. quasi-Poisson or quasi-binomial. The quasi-model will yield the same point estimates for the parameters as the ordinary variant, provided the observations belong to the support of the latter. For data consisting entirely of positive integers (like the triangle from Table 1.2), it thus follows from the remark at the end of Section 4.1 that the overdispersed Poisson GLM yields the same predictions as the chain ladder, although it differs significantly when it comes to the error/variance, as can be seen from Table 4.2

More generally, it can be shown (see [18, Section 2]) that the chain ladder results will be reproduced as long as the additional condition

$$\sum_{i=1}^{I} X_{ij} \geq 0 \tag{4.36}$$

is satisfied for $j \in \{1, \ldots, I\}$. The overdispersed Poisson GLM is therefore robust to the presence of a limited number of negative claim amounts, which is sometimes observed in practice. Moreover, it lifts the unrealistic restriction that the response values must be integers, and gives us a way of accounting for overdispersion, which is a feature of many run-off triangles.

---

[1]It would actually be more correct to call Q a quasi-*log*-likelihood, but the current nomenclature has been widely adopted and the literature seems to have resigned itself to it.

i/j	$\widehat{a}_i$	$\widehat{b}_j$	$\widehat{R}_j$	$\widehat{\mathrm{MSEP}}(\widehat{R}_j)$	
				Poisson	ODP
2	0.03	-0.12	350.9	27.07	125.81
3	0.1	-0.63	1037.54	44.12	205.08
4	0.03	-1.03	2044.86	60	278.85
5	0.09	-1.31	3663.4	83.22	386.79
6	0.28	-1.86	7162.15	130.22	605.27
7	0.49	-2.43	14396.92	249.17	1158.12
$\widehat{c}$					8.26
$\widehat{\phi}$					21.6

Table 4.2: ODP and Poisson GLM results for the UK Motor triangle from Table 1.2

4.3 Bootstrap methodology

Developing bootstrap procedures for the overdispersed Poisson GLM is in some respects easier than for the Mack chain ladder. The absence of a recursive structure makes it more straightforward to reason about resampling, and there is a reasonable amount of literature on the subject to draw upon (see for instance [21] and [10]). As in Section 3.3, we shall take Section 2.2 as our starting point, adjusting the methods described there to obtain nonparametric, semiparametric and parametric bootstraps for the overdispersed Poisson GLM.

The nonparametric variant, which involves resampling predictor-response pairs, is the most difficult to adapt to the present model. To understand why, recall that the typical size of a claims triangle is quite limited; consequently, such resampling can easily lead to a situation in which there are no observations for a given row or column, in which case no estimates can be obtained for the corresponding parameters. This is especially likely for the lower left and upper right corner points, which are the only observations in their row and column, respectively.

In order to avoid this difficulty, we therefore need to find a way of guaranteeing that every row and column contain at least one observation. The crudest option would be to discard any resampled triangle which is defective, but this method quite inefficient; the same result can be achieved at far lower computational cost by constructing resampled triangles to satisfy the desired property in the first place. We experimented with three ways of doing this: preserving the corner points, preserving the first row and column, and randomly selecting an observation in every row and column to preserve; the estimates for the MSEP obtained from all three are shown in Table 4.3. None of these turn out to produce satisfactory results, however, as seen from comparison with Table 4.2. We will therefore exclude the pairs bootstrap from consideration for the overdispersed Poisson GLM.

For the semiparameteric bootstrap, the essential step, as in Section 3.3, is to find a satisfactory definition for the residuals such that they are i.i.d. We have to deal with an additional complication here, however, as it generally not possible to write the response as the sum of mean and error terms for distributions other than the Gaussian, meaning that there are no 'natural residuals' for most GLMs[2]. Consequently, a multitude of different types are available, reflecting different strategies for producing suitable diagnostic quantities; we will consider two of them in particular.

---

[2]This problem was recognised early on in the literature on GLM bootstrapping, see [22].

Origin	$\widehat{\mathrm{MSEP}}(\widehat{R}_j^B)$		
	First row and column	Corner points	Randomised
2	88.37	85.99	88.40
3	153.71	155.55	174.22
4	282.34	211.79	270.11
5	436.18	272.94	370.45
6	744.95	389.42	548.55
7	978.89	548.92	907.98

Table 4.3: MSEP estimates from the pairs bootstrap with different implementation strategies for the data in Table 1.2

The *Pearson residuals*

$$r_{ij} := \frac{X_{ij} - \widehat{\mu}_{ij}}{\sqrt{V(\widehat{\mu}_{ij})}}, \tag{4.37}$$

attempt to deal with the inherent heteroscedasticity of the response by dividing out the component of the variance which is specific to each observation, similarly to the standardised residuals in the context of weighted linear regression. Extending this analogy further, we can adjust Eq. (4.37) to account for the leverage of the observation, i.e.

$$\tilde{r}_{ij} := \frac{X_{ij} - \widehat{\mu}_{ij}}{\sqrt{V(\widehat{\mu}_{ij})(1 - h_{ij})}}, \tag{4.38}$$

where h_{ij} is the appropriate diagonal element in the hat matrix

$$\mathbf{H} = \mathbf{X}(\mathbf{X}\mathbf{W}^{(k_f)}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{W}^{(k_f)}\mathbf{z}^{(k_f)} \tag{4.39}$$

corresponding to the final iteration k_f of the IRWLS algorithm.

A second type of residuals is based on a goodness-of-fit measure known as the *deviance*, which can be derived from Eq. (4.31) by observing that the mean parametrisation of the log-likelihood is maximised at $\boldsymbol{\mu} = \mathbf{y}$, so that the quantity

$$D(\mathbf{y}, \boldsymbol{\mu}) := \sum_{i=1}^{N} d(y_i, \mu_i) := 2\sum_{i=1}^{N}(l(y_i \mid y_i) - l(\widehat{\mu}_i \mid y_i)) \tag{4.40}$$

expresses the departure of the model from a perfect fit. The functions $D(\mathbf{y}, \boldsymbol{\mu})$ and $d(y_i, \mu_i)$ are called the *total* and *unit deviance*, respectively, and the *deviance residuals* are then defined as

$$r_{ij} := \mathrm{sign}(x_{ij} - \widehat{\mu}_{ij})\sqrt{d(x_{ij}, \widehat{\mu}_{ij})}, \tag{4.41}$$

which can also be leverage-adjusted to obtain

$$\tilde{r}_{ij} := \frac{\mathrm{sign}(x_{ij} - \widehat{\mu}_{ij})\sqrt{d(x_{ij}, \widehat{\mu}_{ij})}}{\sqrt{1 - h_{ij}}}, \tag{4.42}$$

where H is defined as in Eq. (4.39). Under certain assumptions (see [23, Section 7.5]), it can be shown that both the Pearson and deviance residuals converge asymptotically towards a normal distribution, which forms the basis of their use as diagnostic quantities.

i/j	$\widehat{a}_i^B$	$\widehat{b}_j^B$	$\widehat{R}_j^B$	$\widehat{\mathrm{MSEP}}(\widehat{R}_j^B)$	i/j	$\widehat{a}_i^B$	$\widehat{b}_j^B$	$\widehat{R}_j^B$	$\widehat{\mathrm{MSEP}}(\widehat{R}_j^B)$
2	0.03	−0.12	353.41	121.96	2	0.03	−0.11	417.16	85.25
3	0.10	−0.63	1040.90	196.06	3	0.10	−0.61	1201.75	136.75
4	0.03	−1.03	2040.92	267.20	4	0.03	−1.00	2321.53	186.85
5	0.09	−1.32	3662.56	368.93	5	0.09	−1.26	4084.83	258.45
6	0.28	−1.87	7149.55	577.77	6	0.27	−1.78	7841.82	404.07
7	0.49	−2.45	14420.05	1116.74	7	0.48	−2.32	15452.66	757.83

| (a) Pearson | (b) Deviance |

Table 4.4: Results from the semiparametric bootstrap with different residual types for the data in Table 1.2

Algorithm 6 then outlines the semiparametric bootstrap procedure for the overdispersed Poisson GLM, which follows the familiar pattern of fit-resample-refit. The notation $\delta(x_{ij}, \mu_{ij})$ refers to one of the residuals which we just defined; inverting them to obtain pseudo-responses requires solving the equations

$$\delta(x, \widehat{\mu}_{ij}) = r_{ij}^{(b)}, \tag{4.43}$$

which are nontrivial for the deviance residuals. It should also be noted that the Pearson and deviance residuals corresponding to the lower left and upper right corner will always have a value of zero; we follow the general trend of the literature (see [21, p. 706] or [24, p. 6]) in excluding them from the resampling process.

Input: Incremental claims triangle $\mathcal{D}_I$, number of bootstrap samples B

$\quad \{(r_{ij}, \widehat{\mu}_{ij}) \mid i + j \leq I + 1\} \leftarrow \mathrm{FIT}(\mathcal{D}_I)$

$\quad \textbf{for } b \leftarrow 1, B \textbf{ do}$

$\quad\quad \{r_{ij}^{(b)} \mid i + j \leq I + 1\} \leftarrow \mathrm{RESAMPLE}(\{r_{ij} \mid i + j \leq I + 1\})$

$\quad\quad \textbf{for } j \leftarrow 1, I \textbf{ do}$

$\quad\quad\quad \textbf{for } i \leftarrow 1, I + 1 - j \textbf{ do}$

$\quad\quad\quad\quad X_{ij}^{(b)} \leftarrow \mathrm{SOLVE}(\delta(x, \widehat{\mu}_{ij}) = r_{ij}^{(b)})$

$\quad\quad\quad \textbf{end for}$

$\quad\quad \textbf{end for}$

$\quad\quad \widehat{\boldsymbol{\beta}}^{(b)} := (\widehat{c}^{(b)}, \widehat{a}_2^{(b)}, \dots, \widehat{a}_I^{(b)}, \widehat{b}_2^{(b)}, \dots, \widehat{b}_I^{(b)}) \leftarrow \mathrm{FIT}(\mathcal{D}_I^{(b)})$

$\quad \textbf{end for}$

$\quad \textbf{return } \{\widehat{\boldsymbol{\beta}}^{(b)} \mid b = 1, \dots, B\}$

Algorithm 6: Semiparametric bootstrap for the overdispersed Poisson GLM

The results of applying the semiparameteric bootstrap to the data in Table 1.2 are shown in Table 4.4. As in Section 3.3, we note that the MSEP is obtained using Eq. (2.8), which we discuss below in Section 4.4; it is included here in the interest of continuity. All bootstrap results in chapter were computed with $B = 1000$ and $S = 100$.

Comparing Table 4.4 to Table 4.2, we clearly see that the variant with deviance residuals underestimates the uncertainty of the reserve. The reason for this is a technical limitation which

was pointed out in [25, pp. 13–14]: for positive response data, we can only resample residuals which satisfy the condition

$$r_{ij} > -\sqrt{2\min(\mathcal{D}_I)}\,. \tag{4.44}$$

Larger residuals are more likely to violate this, but discarding them introduces bias in the bootstrap, causing the predictive distribution to become narrower than is warranted by the model. Consequently, we will not consider deviance residuals in the remainder of this study.

Finally, we consider the parametric bootstrap, which can be adapted without difficulty from the theory in Section 2.2: fit the model to the data to obtain estimates for c, a_i, b_j and ϕ, and use a random number generator to simulate pseudo-responses from the fitted distribution. Because the overdispersed Poisson GLM is a quasi-model, we can choose any distribution as long as it satisfies

$$\text{Var}(X_{ij}) = \widehat{\phi}\, V(\widehat{\mu}_{ij}) = \widehat{\phi}\, \widehat{\mu}_{ij}\,. \tag{4.45}$$

An obvious candidate would be the normal distribution $\mathcal{N}(\widehat{\mu}_{ij}, \phi\widehat{\mu}_{ij})$; alternatives proposed in the literature (see [26]) include the gamma distribution $\Gamma(\alpha_{ij}, \beta_{ij})$ with

$$\alpha_{ij} = \frac{\widehat{\mu}_{ij}}{\widehat{\phi}} \quad \text{and} \quad \beta_{ij} = \frac{1}{\widehat{\phi}}\,, \tag{4.46}$$

and the scaled Poisson distribution

$$Y = \widehat{\phi} X\,, \quad X \sim \text{Pois}(\widehat{\mu}_{ij}/\widehat{\phi})\,, \tag{4.47}$$

which satisfies $\mathbb{E}[Y] = \widehat{\mu}_{ij}$ and $\text{Var}(Y) = \phi\widehat{\mu}_{ij}$. The model is then refitted to the simulated triangle to obtain $\widehat{c}^{(b)}$, $\widehat{a}_i^{(b)}$, $\widehat{b}_j^{(b)}$ and $\widehat{\phi}^{(b)}$. Algorithm 7 outlines the procedure for the normal distribution, and the results for the three choices of distribution are shown in Tables 4.5 and 4.6

i/j	$\widehat{a}_i^B$	$\widehat{b}_j^B$	$\widehat{R}_j^B$	$\widehat{\text{MSEP}}(\widehat{R}_j^B)$		i/j	$\widehat{a}_i^B$	$\widehat{b}_j^B$	$\widehat{R}_j^B$	$\widehat{\text{MSEP}}(\widehat{R}_j^B)$
2	0.03	−0.12	350.26	125.91		2	0.03	−0.12	345.95	123.28
3	0.10	−0.63	1038.71	207.26		3	0.10	−0.63	1035.70	202.69
4	0.04	−1.03	2053.19	278.32		4	0.03	−1.03	2045.65	280.66
5	0.09	−1.31	3676.36	385.16		5	0.09	−1.32	3662.17	389.09
6	0.28	−1.86	7187.67	607.74		6	0.28	−1.87	7164.46	604.11
7	0.49	−2.47	14428.44	1161.72		7	0.49	−2.47	14395.83	1143.86

(a) Normal	(b) Gamma

Table 4.5: Results from the parametric bootstrap with different distributions for the data in Table 1.2

4.4 Incorporating the process error

The only remaining question now is how to incorporate the intrinsic variability of the response in the bootstrap methods from the previous section. In light of Section 2.3, we want to achieve this using the predictive distribution, just as in Section 3.4. This involves simulating the stochastic fluctuation of the reserve

$$R = \sum_{\substack{1 \le i,j \le I \\ i+j > I+1}} X_{ij} \tag{4.48}$$

i/j	$\widehat{a}_i^B$	$\widehat{b}_j^B$	$\widehat{R}_j^B$	$\widehat{\mathrm{MSEP}}(\widehat{R}_j^B)$
2	0.03	-0.12	347.70	126.24
3	0.10	-0.63	1030.26	204.74
4	0.03	-1.03	2039.72	274.15
5	0.09	-1.31	3658.26	383.81
6	0.28	-1.88	7169.73	604.00
7	0.49	-2.47	14390.55	1173.80

Table 4.6: Results from the parametric bootstrap with scaled Poisson distribution for the data in Table 1.2

Input: Incremental claims triangle $\mathcal{D}_I$, number of bootstrap samples B

$\{\widehat{\mu}_{ij} \mid i + j \leq I + 1\} \leftarrow \mathrm{FIT}(\mathcal{D}_I)$

for $b \leftarrow 1, B$ **do**

    for $j \leftarrow 1, I$ **do**

        for $i \leftarrow 1, I + 1 - j$ **do**

            $X_{ij}^{(b)} \leftarrow \mathrm{SAMPLE}(\mathcal{N}(\widehat{\mu}_{ij}, \phi\widehat{\mu}_{ij}))$

        end for

    end for

    $\widehat{\boldsymbol{\beta}}^{(b)} := (\widehat{c}^{(b)}, \widehat{a}_2^{(b)}, \ldots, \widehat{a}_I^{(b)}, \widehat{b}_2^{(b)}, \ldots, \widehat{b}_I^{(b)}) \leftarrow \mathrm{FIT}(\mathcal{D}_I^{(b)})$

end for

return $\{\widehat{\boldsymbol{\beta}}^{(b)} \mid b = 1, \ldots, B\}$

Algorithm 7: Parametric bootstrap for the overdispersed Poisson GLM

around the point estimates obtained using the bootstrapped parameters $\widehat{c}^{(b)}$, $\widehat{a}_i^{(b)}$, $\widehat{b}_j^{(b)}$ and $\widehat{\phi}^{(b)}$. Consequently, we need to simulate pseudo-realisations of the lower triangle $\mathcal{D}_I^c$, which can be done by modifying Algorithms 6 and 7 in a straightforward way. For the semiparametric bootstrap, we resample the residuals a second time and invert the appropriate formula to obtain pseudo-responses; for the parametric bootstrap, we simulate them from the fitted model obtained by plugging in the bootstrap replicates of the parameters. Figures 4.2 and 4.3 show the resulting predictive distributions for the data in Table 1.2.
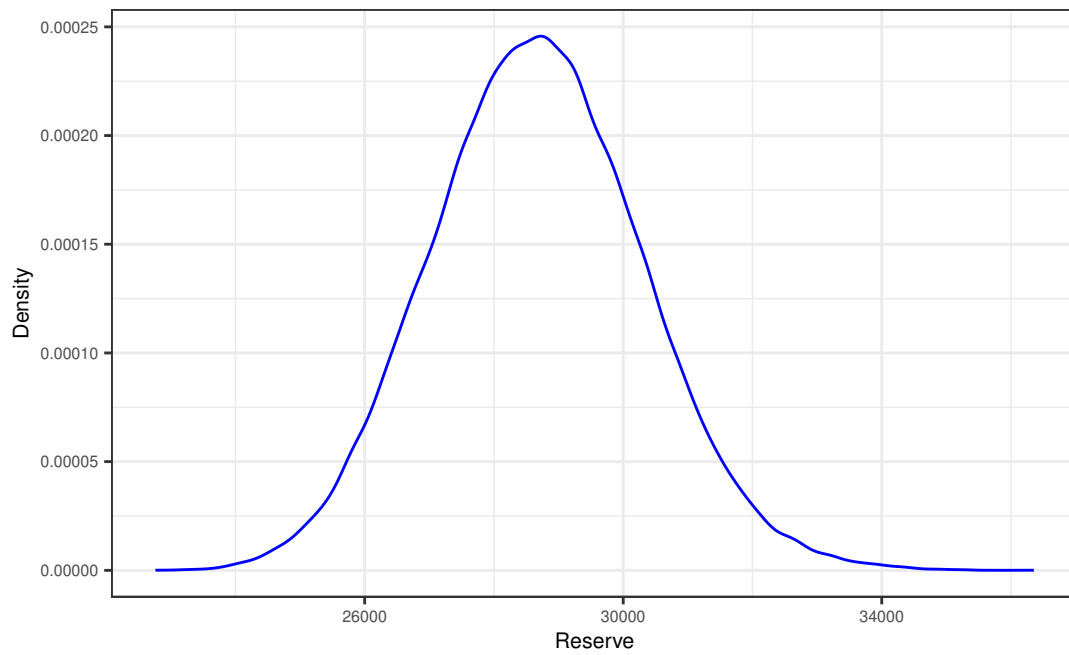
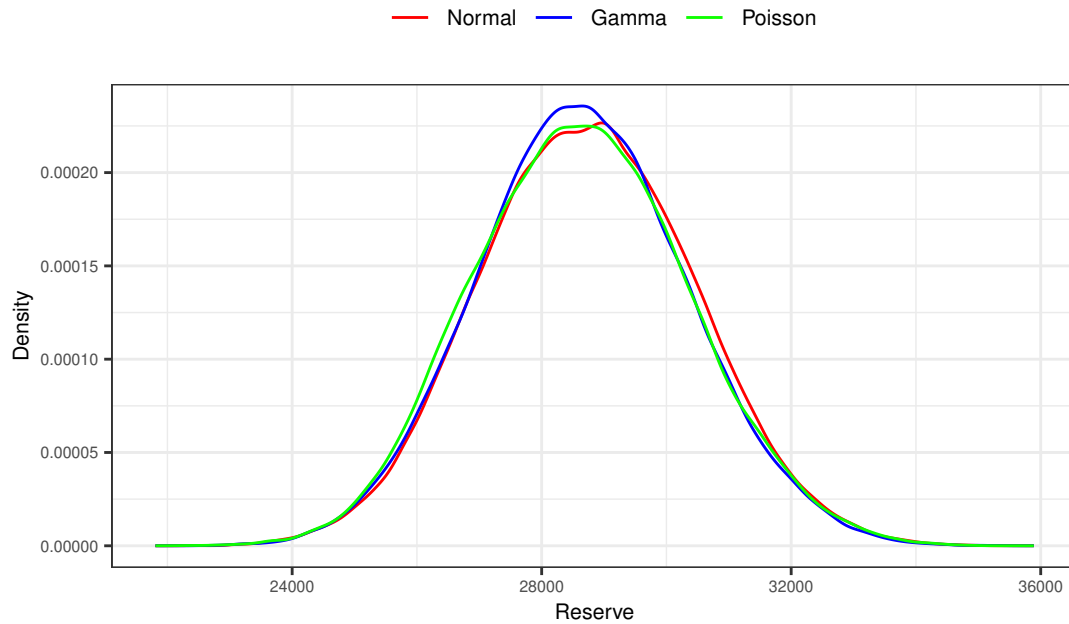**Figure 4.2:** Predictive distribution for the semiparametric bootstrap

Figure 4.3: Predictive distribution for the parametric bootstrap with different distributions

Chapter 5

Simulation study

As explained in the Chapter 1 and Section 2.3, we want to study the impact of deviations from the model assumptions of the Mack chain ladder and the overdispersed Poisson GLM on the predictive distribution of the total outstanding claims. To this end, we will simulate a large number of run-off triangles, following a similar approach to the one described in [27], which satisfy these assumptions by construction, except for a number of deviating observations (we will call such data *contaminated*). We will then use the bootstrap methods described in Chapters 3 and 4 to obtain the predictive distribution while excluding a given subset of points from the resampling process. Our aim is to find out whether predictive distributions computed from data in which the correct points have been discarded (which we will refer to as the *cleaned* data) differ significantly from those obtained after removing arbitrary observations. This would then suggest a way of detecting irregularities in new triangles: compare for every point (resp. calendar, origin period) the predictive bootstrap distribution obtained with and without this point; if a significant difference can be detected, then this subset constitutes a deviation from the model.

5.1 Mack's model

Recall from Section 3.1 that the Mack chain ladder is recursive in nature, modelling the cumulative claim amounts in the next development period as a function of those in the previous one. Hence, it requires that observations for the first development period be given initially. This could be done using a separate model, but we choose here the simpler approach of borrowing the first column from the UK Motor example data in Table 1.2. Similarly, we choose as the development factors $f_1, \ldots, f_{I-1}$ and variance parameters $\sigma_1, \ldots, \sigma_{I-1}$ of the simulated triangles the fitted ones from this dataset.

In order to obtain a synthetic triangle which satisfies (Mack1) to (Mack3) except at a single point (i^*, j^*) (where $i^* + j^* \leq I+1$ and $j^* > 1$), we can then start from the initial column and use any suitable data-generating process (such as Eq. (3.48)) to recursively generate the next anti-diagonal of the triangle. For the non-contaminated rows, we continue this process until reaching the current calendar year. In the contaminated row, we stop at column $j^* - 1$ and use c_μ and c_σ to draw an observation from the perturbed distribution $\mathcal{N}(c_\mu f_{j^*-1} C^*_{i^*, j^*-1}, (c_\sigma \sigma_{j^*-1})^2 C^*_{i^*, j^*-1})$, after which we resume the previous process. The whole procedure is outlined in Algorithm 8. Except for some additional bookkeeping, simulating triangles where an entire origin or calendar period has been perturbed can be done in exactly the same way.

Input: Development factors $f_1, \ldots, f_{I-1}$, variance parameters $\sigma_1, \ldots, \sigma_{I-1}$, perturbed point
  (i^*, j^*) with $i^* + j^* \leq I + 1$ and $j^* > 1$, perturbation factors c_μ, c_σ
  for all $1 \leq i, j \leq I$ **do**
      $C_{ij} \leftarrow$ NA
  end for
  for $i \leftarrow 1, I$ **do**
      $C_{i1}^* \leftarrow C_{i1}$
      if $i \neq i^*$ **then**
          for $j \leftarrow 2, I + 1 - i$ **do**
              $C_{ij}^* \leftarrow \text{SAMPLE}(\mathcal{N}(f_j C_{i,j-1}^*, \sigma_j^2 C_{i,j-1}^*))$
          end for
      else
          for $j \leftarrow 2, j^* - 1$ **do**
              $C_{i^*j}^* \leftarrow \text{SAMPLE}(\mathcal{N}(f_j C_{i^*,j-1}^*, \sigma_j^2 C_{i^*,j-1}^*))$
          end for
          $C_{i^*j^*}^* \leftarrow \text{SAMPLE}(\mathcal{N}(c_\mu f_j C_{i^*,j^*-1}^*, (c_\sigma \sigma_j)^2 C_{i^*,j^*-1}^*))$
          for $j \leftarrow j^* + 1, I + 1 - i$ **do**
              $C_{i^*j}^* \leftarrow \text{SAMPLE}(\mathcal{N}(f_j C_{i^*,j-1}^*, \sigma_j^2 C_{i^*,j-1}^*))$
          end for
      end if
  end for
  return $(C_{ij}^*)_{1 \leq i,j \leq I}$

Algorithm 8: Simulating a cumulative claims triangle with a single perturbed point according
to the Mack chain ladder

5.2   Overdispersed Poisson GLM

Simulating triangles conforming to the overdispersed Poisson GLM is significantly easier than for
the Mack chain ladder, as we don't have to deal with recursion. Using the fitted parameters from
Table 4.1 for ϕ, c, $a_2, \ldots, a_I$ and $b_2, \ldots, b_I$, we generate an observation for every cell (i, j) in the
upper triangle by drawing from an appropriate distribution. Triangles with a single contaminated
point can then be obtained by applying the perturbation before calling the random number
generator. Again, similar algorithms can be used for perturbed origin or calendar periods.

5.3   Setup

We are now ready to discuss the general setup of the simulation study. Using Algorithms 8
and 9, we generate claims triangles which perfectly satisfy the assumptions of the Mack chain
ladder or the overdispersed Poisson GLM, except for a selected number of points. In particular,
we consider the three cases of (i) a single perturbed observation (ii) a perturbed calendar period
and (iii) a perturbed origin period. In each instance, we carry out one of the bootstrap procedures
outlined in Chapters 3 and 4 while excluding a subset of observations of the same type as the

Input: Intercept c, development period parameters $a_2, \ldots, a_I$, origin period parameters $b_2, \ldots, b_I$, dispersion parameter ϕ, perturbed point (i^*, j^*) with $i^* + j^* \leq I + 1$ and $j^* > 1$, perturbation factor c_λ
$a_1 \leftarrow 0$, $b_1 \leftarrow 0$
for all $1 \leq i, j \leq I$ **do**
    $\mu_{ij} \leftarrow \exp(c + a_i + b_j)$
    $X_{ij} \leftarrow \text{NA}$
end for
for $i \leftarrow 1, I$ **do**
    for $j \leftarrow 1, I + 1 - i$ **do**
        if $i = i^*$ and $j = j^*$ **then**
            $X_{ij}^* \leftarrow \text{SAMPLE}(\mathcal{N}(c_\lambda \mu_{ij}, \phi c_\lambda \mu_{ij}))$
        else
            $X_{ij}^* \leftarrow \text{SAMPLE}(\mathcal{N}(\mu_{ij}, \phi \mu_{ij}))$
        end if
    end for
end for
return $(X_{ij}^*)_{1 \leq i, j \leq I}$

Algorithm 9: Simulating an incremental claims triangle with single perturbed point according to the overdispersed Poisson GLM

contaminated one, and then compare the resulting predictive distributions. Our hope is that we can detect a significant difference between the predictive distribution obtained when excluding the right subset and the ones where a different subset has been removed.

The way in which observations can be excluded when carrying out the bootstrap is specific to each method, and raises a number of practical issues. For the semiparametric bootstraps, the most obvious way is to simply remove the corresponding residuals from the resampling pool. In the case of the Mack chain ladder, this implies that the first column cannot contain contaminated points; this makes sense, as this model does not make any statements regarding the behaviour of the initial observations. While odd points in the triangle may be considered *outliers* in a general sense, they do not necessarily constitute deviations from the model assumptions. To minimise the impact of this extraneous factor on the results, we have taken the approach of borrowing this initial column from the benchmark dataset in Table 1.2, as explained in Section 5.1.

For the parametric bootstrap, it is not as clear how we should go about excluding certain points. Obviously, our goal is to nullify the impact of these observations on the bootstrap result, which suggests removing them from the initial fitting process when obtaining the parametric model used to simulate new pseudo-data. We have to be careful here, however, as this can sometimes lead to situations where the model cannot be fit at all. In particular, for the Mack chain ladder, we cannot exclude the upper right corner point, because we cannot estimate the final development factor otherwise. The approach outlined here is therefore not suitable to detecting deviations for this point. This is no great loss, however: the final development year requires special attention in any case, as it is very tied up with issues of tail factors and extrapolation. For the Poisson GLM, the restriction is even more stringent; here, we cannot exclude the lower

left corner either, because the resulting model would have no parameter corresponding to the final origin period, and consequently cannot be used for projection.

Finally, for the pairs bootstrap, we simply exclude the affected points from the resampling process.

5.4   Implementation and results

The code for carrying out the simulations and generating the examples from previous chapters has been mostly implemented in the `R` language [28]. The computational power required to execute the large number of possible configurations exceeds the capacity of this high-level scripting language, however: the results dataframe for the Mack chain ladder contains 37,602,000 rows, for example. It was therefore decided to write the most numerically intensive parts of the simulation in `Fortran` and link it to `R` through a `C++` wrapper using the `Rcpp` package [29]. For reproducibility and convenience, all software has been bundled together into a package, which is available at `https://github.com/OthmanElHammouchi/claimsBoot`. Instructions for installing are provided there as well.

The results are mostly presented as faceted density plots comparing, for different configurations, the predictive density of the reserve when the contaminated observations have been excluded (shown in red) with the predictive densities obtained after excluding model-conforming points. All simulations were done with $B = 1000$ and $S = 1$. As it would be impractical to show all of them, we have endeavoured to select a representative sample for inclusion. The reader is encouraged to download the code and experiment with it for him- or herself.

5.4.1   Mack's model

In experiments where a single point has been perturbed, the semiparametric bootstrap performs poorly when used with standardised and log-normal residuals, as shown by Figs. 5.1 to 5.4. The predictive distribution does not differ significantly between the cleaned and contaminated data; this holds across different perturbation factors, and regardless of whether the conditional or unconditional approach is used.

Closer inspection of the simulation reveals the reason for this: standardised and log-normal residuals are not very sensitive to anomalies in a single observation. Consider for instance the triangle in Table 5.1, which is perturbed in point $(1, 2)$ with $c_\mu = 100$ and $c_\sigma = 1$, leading to hugely inflated values in the first row. The corresponding residual in Tables 5.2a and 5.2b does not exhibit a deviation of commensurate severity, however, and so excluding it does.

This stands in stark contrast with the studentised residuals, which are strongly affected by single point deviations in the simulated triangle, as can be seen in Table 5.3. Consequently, the cleaned triangle stands out from the contaminated ones. This is illustrated in Table 5.4, which contains the means of the predictive distribution based on different outlier and excluded points with $c_\mu = 2$ and $c_\sigma = 1$. The minimum in each row is highlighted in red. This lines up almost perfectly with diagonal, which means that the predictive reserve mean of the cleaned triangle tends to stand out when using studentised residuals.

The parametric bootstrap shows good performance across the board. The choice of distribution does not appear to have a noticeable impact, although we do observe a substantial difference between the conditional and unconditional approaches, with the latter generally outperforming the former, often drastically, as shown in Fig. 5.5. Lastly, the pairs bootstrap is better at distinguishing contaminated from cleaned data than the semiparametric bootstrap with log-normal or standardised residuals, but does not outperform the one using studentised residuals or the parametric bootstrap. As Table 5.5 shows, the mean of the predictive distribution simulated on

the basis of the cleaned data stands out for every outlier point, but often only barely. In all cases, differences tend to be noticeable for later development periods.

The same general trends are observed in a more pronounced form when a calendar or origin period is perturbed, owing to the larger number of deviating points. Consequently, even the semiparametric bootstrap with log-normal or standardised residuals shows serviceable results, as can be seen in Figs. 5.7a, 5.7b, 5.8a and 5.8b. Later periods universally show stronger effects, because they contain more observations; for the first period, no bootstrap method can distinguish between clean and contaminated data.

Origin	Dev						
	1	2	3	4	5	6	7
1	3511	663172.28	847468.65	972420.52	1069036.05	1123357.24	1154289.54
2	4001	7455.21	9760.10	11230.45	12303.74	12949.43	
3	4355	8291.40	10770.20	12376.93	13605.41		
4	4295	7850.70	9992.00	10968.64			
5	4150	7842.25	9869.06				
6	5102	9389.75					
7	6283						

Table 5.1: Simulated triangle where observation C_{12} has been perturbed with $c_\mu = 100$ and $c_\sigma = 1$

Origin	Dev					
	2	3	4	5	6	7
2	2.24	−0.63	0.88	1.00	−1	
3	−0.39	1.43	0.13	−1.41	1	
4	−0.41	1.01	0.08	−0.02		
5	−0.40	−0.26	−1.73			
6	−0.40	−0.94				
7	−0.45					

(a) Standardised

Origin	Dev					
	2	3	4	5	6	7
2	2.04	−0.13	0.16	0.15	−0.11	
3	−1.16	1.42	0.14	−1.41	0.99	
4	−1.19	1.01	0.09	−0.02		
5	−1.21	−0.25	−1.74			
6	−1.17	−0.93				
7	−1.31					

(b) Log-normal

Table 5.2: Residuals corresponding to the triangle in Table 5.1

Origin	Dev					
	2	3	4	5	6	7
2	4834.13	−0.58	0.83	0.99		
3	−0.35	1.78	0.11	−36.03		
4	−0.37	1.02	0.07	−0.02		
5	−0.37	−0.23	−19.38			
6	−0.36	−0.92				
7	−0.41					

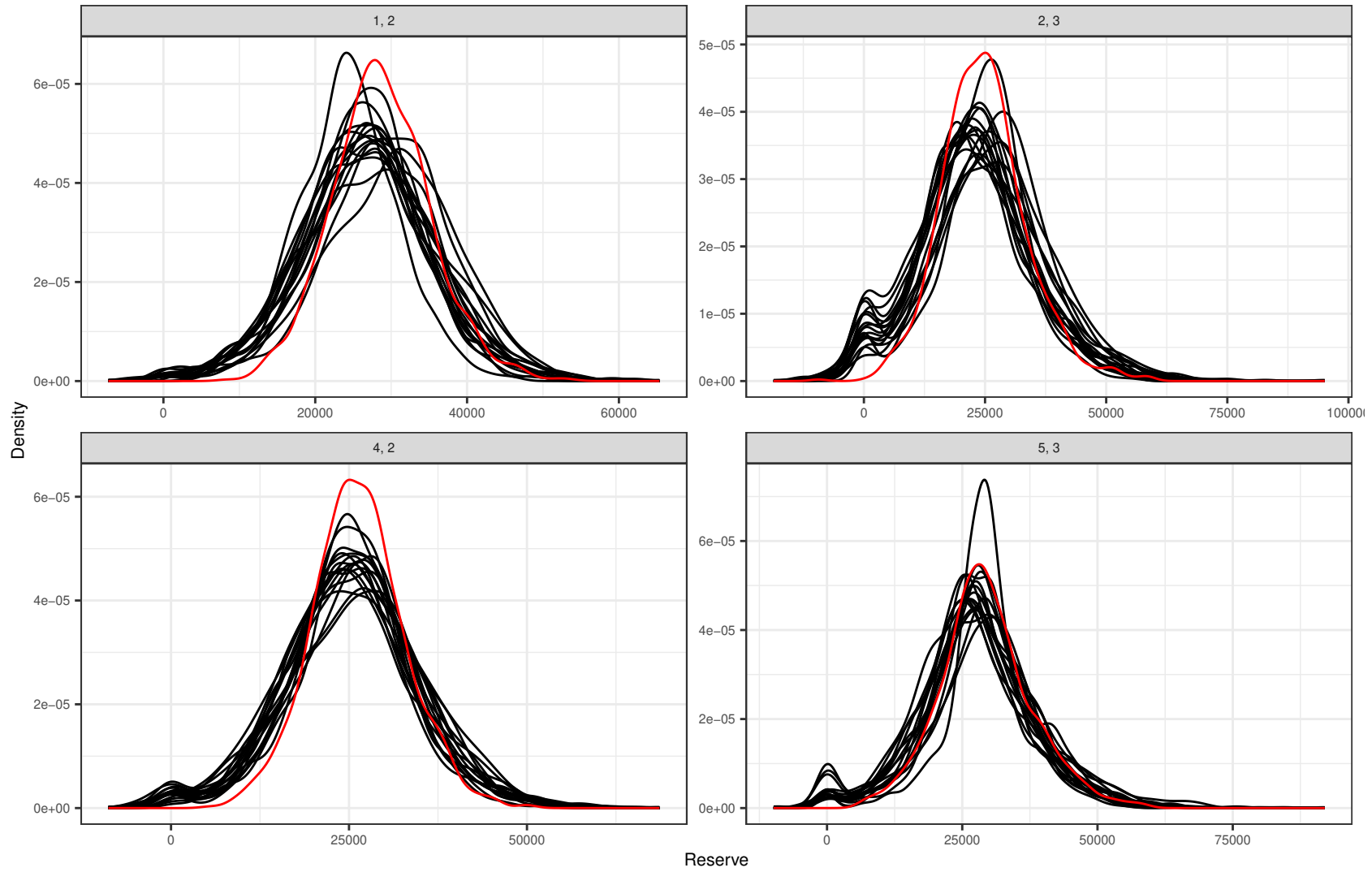**Table 5.3:** Studentised residuals corresponding to the triangle in Table 5.1

Figure 5.1: Density plots of predictive distributions obtained using a conditional semiparametric bootstrap of the Mack chain ladder with standardised residuals for different deviating points. The curves in each facet correspond to densities obtained while excluding a different point from the resampling process, with the result for the cleaned dataset highlighted in red. For all subfigures $c_\mu = c_\sigma = 0.5$.
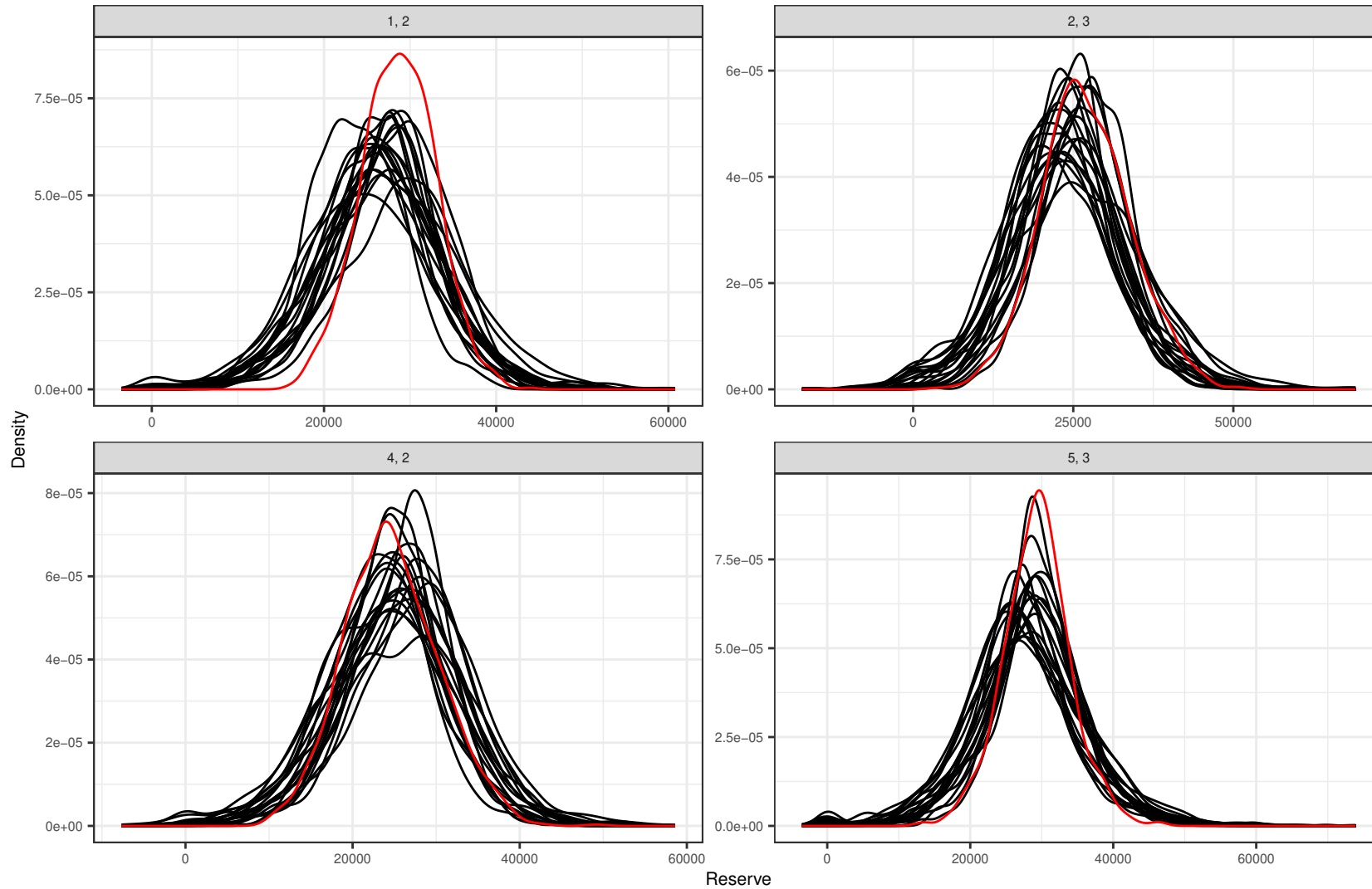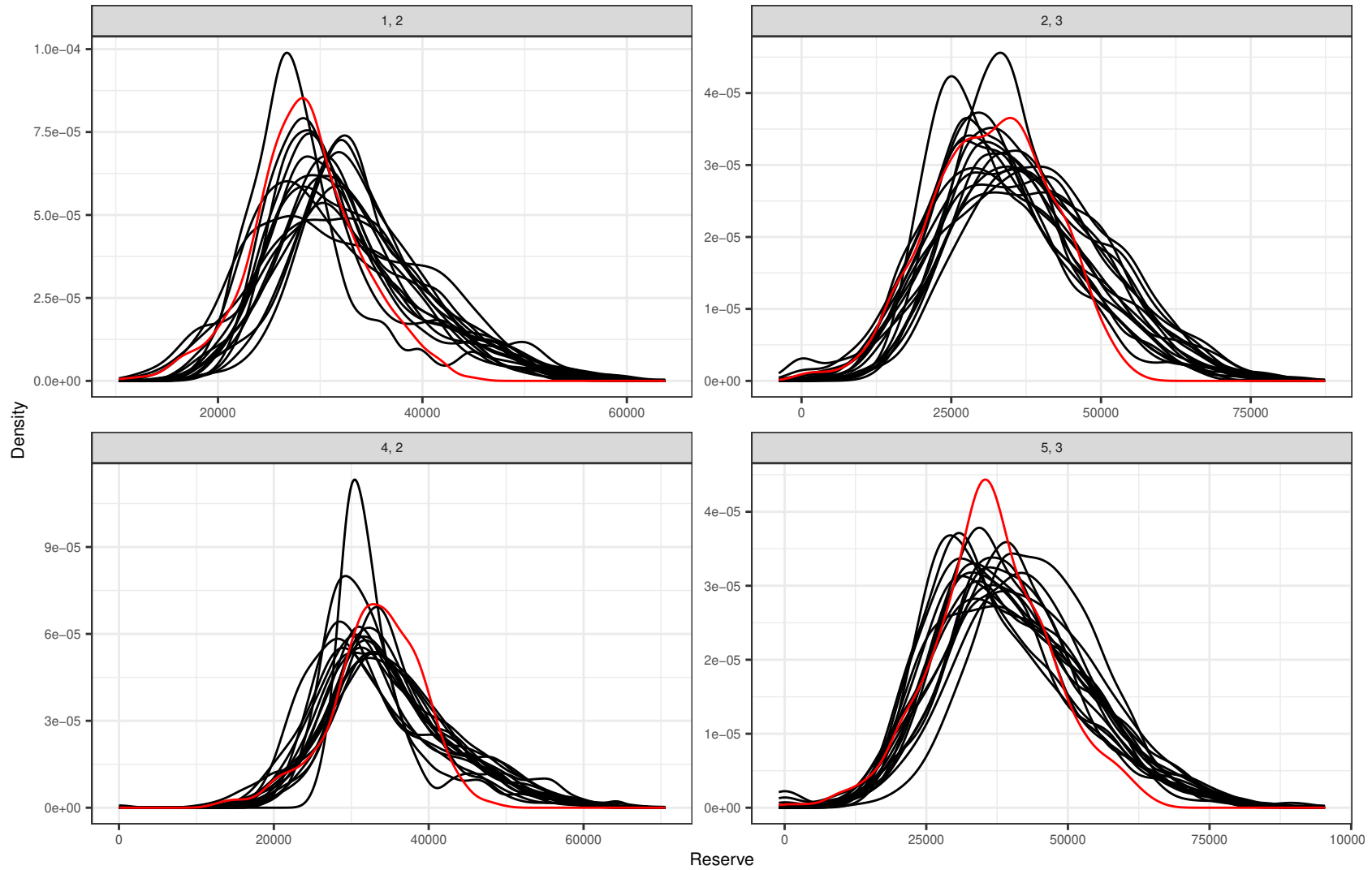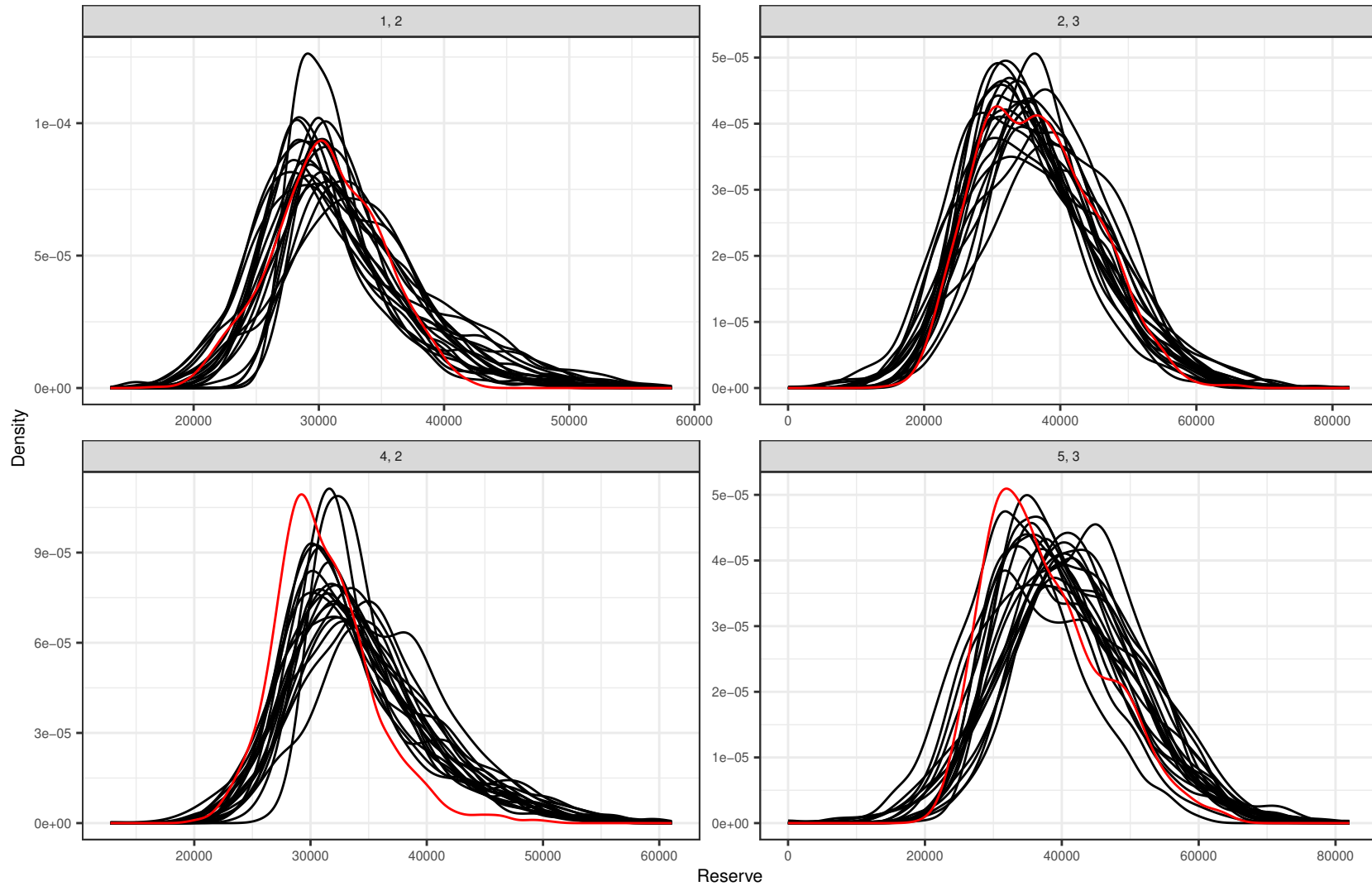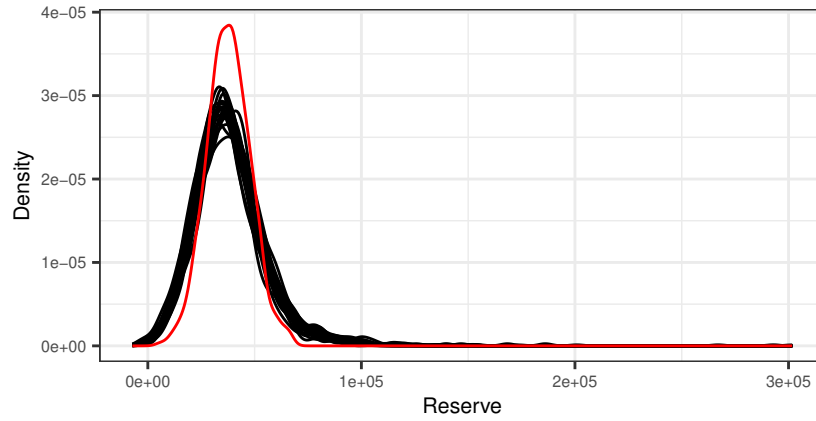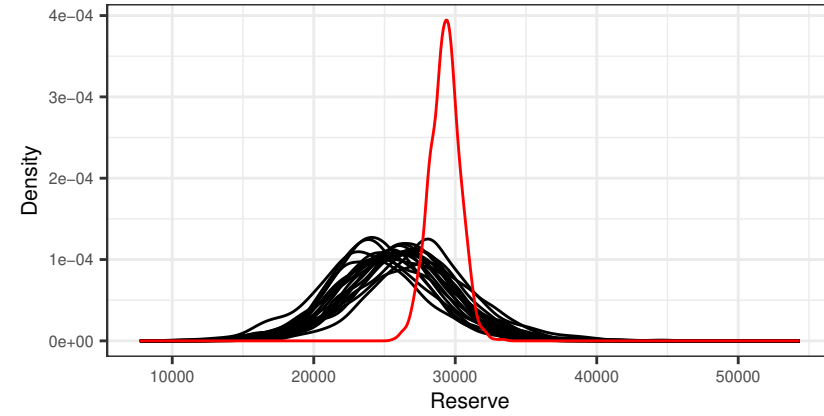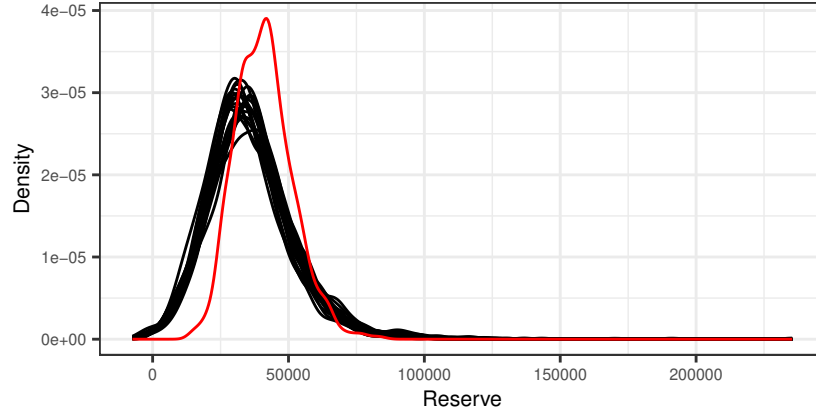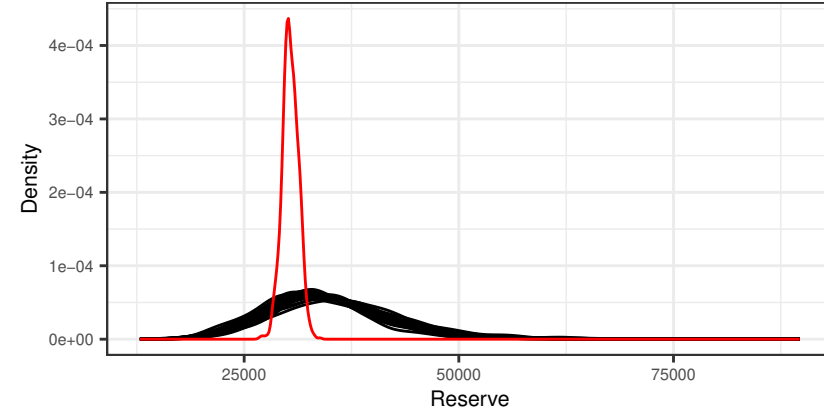
Figure 5.2: Density plots of predictive distributions obtained using a conditional semiparametric bootstrap of the Mack chain ladder with log-normal residuals for different deviating points. The curves in each facet correspond to densities obtained while excluding a different point from the resampling process, with the result for the cleaned dataset highlighted in red. For all subfigures $c_\mu = 0.5$ and $c_\sigma = 2$.

Figure 5.3: Density plots of predictive distributions obtained using an unconditional semiparametric bootstrap of the Mack chain ladder with standardised residuals for different deviating points. The curves in each facet correspond to densities obtained while excluding a different point from the resampling process, with the result for the cleaned dataset highlighted in red. For all subfigures $c_\mu = 2$ and $c_\sigma = 0.5$.

Figure 5.4: Density plots of predictive distributions obtained using an unconditional semiparametric bootstrap of the Mack chain ladder with log-normal residuals for different deviating points. The curves in each facet correspond to densities obtained while excluding a different point from the resampling process, with the result for the cleaned dataset highlighted in red. For all subfigures $c_\mu = c_\sigma = 2$.

(a) Conditional, normal distribution, $c_\mu = 2$

(b) Unconditional, normal distribution, $c_\mu = 0.5$

(c) Conditional, gamma distribution, $c_\mu = 0.5$

(d) Unconditional, gamma distribution, $c_\mu = 2$

Figure 5.5: Density plots of predictive distributions obtained using a parametric bootstrap of the Mack chain ladder with deviating point $(i^*, j^*) = (4, 2)$. The curves in each facet correspond to densities obtained while excluding a different point from the resampling process, with the result for the cleaned dataset highlighted in red. For all subfigures, $c_\sigma = 1$.

Deviating	Excluded																	
	(1, 2)	(2, 2)	(3, 2)	(4, 2)	(5, 2)	(6, 2)	(1, 3)	(2, 3)	(3, 3)	(4, 3)	(5, 3)	(1, 4)	(2, 4)	(3, 4)	(4, 4)	(1, 5)	(2, 5)	(3, 5)
(1, 2)	2e+03	219.29	1e+03	275.55	258.79	125.66	150.80	280.92	243.98	299.74	312.52	719.08	175.85	162.29	212.13	219.21	428.30	193.71
(2, 2)	543.74	34.96	514.50	437.14	2e+03	230.12	396.41	764.36	209.35	282.48	320.86	804.80	131.92	584.82	286.52	6e+03	297.47	2e+03
(3, 2)	182.99	3e+03	37.31	483.43	467.83	241.62	545.59	726.44	272.80	177.99	279.90	359.43	153.36	278.10	141.93	428.18	284.43	165.91
(4, 2)	366.53	243.56	234.10	38.32	146.77	311.62	399.59	339.10	283.71	136.05	330.86	665.72	498.99	175.34	148.15	999.43	384.48	349.31
(5, 2)	402.16	147.21	145.60	341.70	37.76	597.28	141.24	227.13	186.35	476.86	724.56	776.12	1e+03	190.78	176.31	360.09	315.54	378.15
(6, 2)	384.64	498.07	177.62	377.37	132.25	68.30	182.33	376.80	335.74	215.74	497.38	103.79	335.67	243.03	192.55	277.35	260.58	229.31
(1, 3)	209.29	255.28	127.41	422.40	365.59	251.60	38.68	504.08	266.62	328.41	575.94	2e+03	238.50	331.84	168.43	140.68	235.28	288.63
(2, 3)	263.77	166.83	416.34	283.49	134.10	170.86	333.84	44.77	297.97	183.68	231.22	230.65	145.56	321.50	329.59	94.05	183.79	466.67
(3, 3)	171.25	288.31	335.53	292.72	238.50	326.28	282.13	287.98	63.27	313.81	538.01	345.52	1e+03	563.04	284.47	1e+03	2e+03	214.76
(4, 3)	303.37	138.11	170.32	208.45	234.14	424.47	593.07	434.98	540.24	41.19	870.30	1e+03	255.15	806.90	196.63	263.68	280.08	425.28
(5, 3)	223.84	549.37	484.85	140.58	159.31	124.67	326.74	170.80	310.05	385.67	34.99	811.91	122.26	247.51	197.32	172.24	127.31	214.01
(1, 4)	326.15	513.85	895.42	256.74	8e+03	278.81	231.64	3e+03	199.07	228.24	579.38	56.55	394.18	367.45	1e+03	403.11	520.22	229.81
(2, 4)	850.73	756.43	848.70	894.14	2e+03	384.94	584.89	500.55	280.97	286.67	400.13	223.68	67.76	772.88	1e+03	316.27	468.90	2e+03
(3, 4)	304.79	213.68	419.24	376.67	864.85	2e+03	180.49	764.56	2e+03	428.15	364.83	537.76	2e+04	56.89	912.01	389.27	228.16	266.18
(4, 4)	139.09	649.76	208.80	244.51	140.37	893.08	1e+03	204.89	385.52	200.13	427.89	355.60	194.15	362.30	28.66	348.71	317.82	217.54
(1, 5)	1e+03	3e+03	4e+04	2e+03	4e+03	830.36	761.26	1e+04	2e+04	4e+04	2e+03	8e+03	4e+03	1e+03	2e+04	71.46	424.43	6e+04
(2, 5)	2e+03	2e+03	2e+03	4e+03	1e+03	1e+03	4e+03	1e+03	1e+04	1e+05	824.15	1e+03	2e+05	6e+04	2e+03	2e+03	109.64	925.88
(3, 5)	2e+03	4e+03	6e+06	2e+04	7e+03	3e+05	650.20	5e+04	2e+03	425.55	2e+05	1e+03	3e+03	3e+03	1e+05	1e+03	3e+03	35.83

Table 5.4: Means of the predictive distribution of the reserve obtained using a conditional semiparametric bootstrap of the Mack chain ladder with studentised residuals for different deviating and excluded points (thousands). The result for the cleaned dataset is highlighted in red.
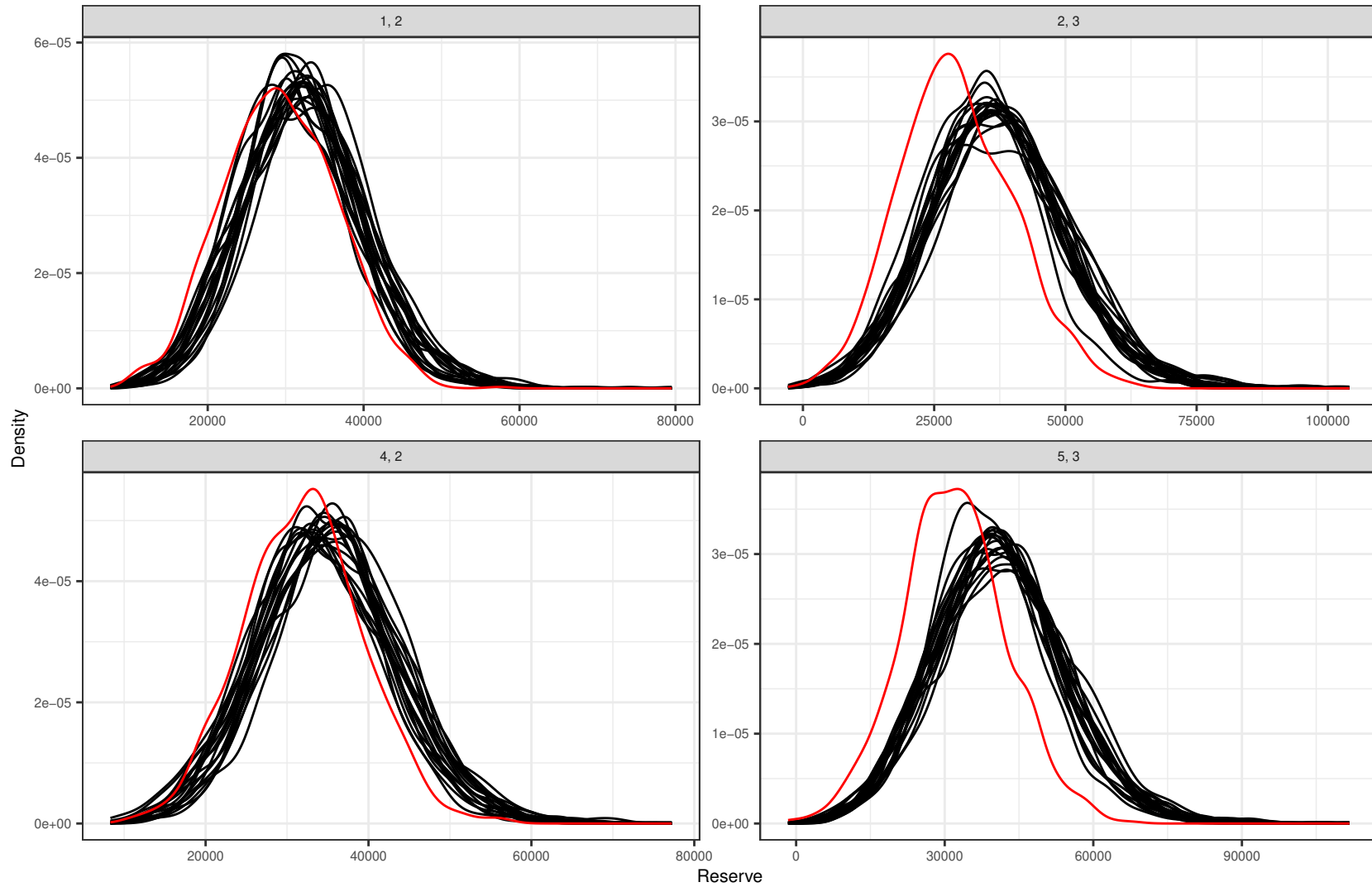
Figure 5.6: Density plots of predictive distributions obtained using a pairs bootstrap of the Mack chain ladder for different deviating points. The curves in each facet correspond to densities obtained while excluding a different point from the resampling process, with the result for the cleaned dataset highlighted in red.

Outlier	Excluded																			
	(1, 2)	(2, 2)	(3, 2)	(4, 2)	(5, 2)	(6, 2)	(1, 3)	(2, 3)	(3, 3)	(4, 3)	(5, 3)	(1, 4)	(2, 4)	(3, 4)	(4, 4)	(1, 5)	(2, 5)	(3, 5)	(1, 6)	(2, 6)
(1, 2)	29.28	32.66	32.79	33.59	33.30	33.18	32.03	33.14	30.58	32.51	32.30	31.27	31.04	34.36	30.58	31.94	32.24	30.30	30.96	31.48
(2, 2)	35.23	27.12	33.94	37.45	33.72	32.12	31.98	30.82	32.54	31.63	33.53	33.17	33.15	32.61	32.40	32.87	33.08	31.65	31.05	32.62
(3, 2)	32.68	32.99	29.76	34.34	33.59	34.02	34.37	32.70	31.03	33.84	33.51	30.92	32.38	34.67	33.33	33.76	33.38	33.72	33.72	31.74
(4, 2)	35.33	33.88	36.23	31.99	34.38	36.40	34.37	36.97	33.61	33.35	33.71	34.47	36.60	35.57	36.33	34.39	34.16	32.83	35.98	35.46
(5, 2)	35.80	36.12	35.39	36.90	32.70	36.71	34.34	36.35	35.51	33.12	34.61	37.88	36.75	38.60	33.11	37.66	36.08	35.10	38.09	35.41
(6, 2)	40.77	39.97	39.23	42.90	39.25	34.86	42.56	39.52	39.68	42.03	39.83	37.78	39.00	40.28	40.80	41.01	42.64	41.42	41.82	40.66
(1, 3)	34.92	35.58	35.47	34.82	35.50	35.37	27.92	36.36	35.76	35.64	37.11	37.68	34.69	34.16	36.29	33.68	37.10	36.01	35.02	35.36
(2, 3)	36.07	37.02	37.13	36.40	35.78	36.51	37.19	29.27	39.36	38.44	38.94	36.38	38.54	35.73	36.99	36.49	37.78	33.34	36.64	37.29
(3, 3)	36.37	40.24	39.15	37.17	36.94	37.02	38.12	39.90	30.94	40.36	39.07	38.88	37.28	37.38	39.56	37.82	39.70	38.15	37.69	38.84
(4, 3)	38.30	40.83	39.54	38.59	39.68	38.91	41.41	41.44	43.08	31.65	43.06	39.29	37.98	42.38	38.53	39.73	38.57	38.48	39.68	39.38
(5, 3)	41.20	40.82	38.08	40.54	41.33	38.84	40.10	42.86	41.16	43.69	31.98	41.23	41.37	41.57	41.88	41.37	41.23	40.59	38.80	39.09
(1, 4)	41.86	42.98	40.50	42.82	39.97	38.79	40.46	40.26	42.17	41.35	39.50	30.94	45.39	45.33	44.26	40.19	41.18	41.71	39.56	41.27
(2, 4)	42.13	40.03	44.93	40.57	44.02	42.89	43.81	43.58	43.52	44.77	41.24	45.93	30.50	46.56	47.74	43.53	43.43	43.13	45.52	42.03
(3, 4)	45.10	43.38	45.58	43.95	44.12	43.94	41.02	42.93	44.15	43.20	46.13	49.22	48.02	31.21	49.61	43.46	44.91	44.93	44.20	44.30
(4, 4)	45.99	46.90	44.42	45.02	44.67	43.31	44.96	47.11	43.62	44.69	45.81	50.26	50.84	47.89	32.98	47.17	47.52	44.22	44.94	45.56
(1, 5)	51.29	51.65	52.82	51.13	50.95	50.36	50.92	50.48	55.08	50.75	52.39	51.14	52.07	50.89	53.53	30.85	58.80	58.99	53.92	51.67
(2, 5)	54.64	54.70	52.29	54.70	50.49	56.70	55.62	51.87	56.59	53.49	50.90	55.32	52.85	55.61	54.29	65.36	34.06	66.17	53.52	53.91
(3, 5)	57.07	56.63	53.98	58.70	55.85	54.30	54.03	60.96	61.07	55.24	55.24	55.43	58.33	57.23	55.95	62.04	66.74	36.22	58.83	57.54
(1, 6)	80.39	74.36	76.74	80.77	75.70	69.91	74.78	78.02	79.15	76.06	78.59	73.50	76.88	78.62	74.89	74.78	73.72	72.30	44.02	109.13
(2, 6)	80.68	78.98	82.33	77.46	78.31	83.51	80.84	83.42	81.07	80.07	79.70	78.50	75.68	80.95	84.15	78.92	79.79	80.07	108.42	45.57

Table 5.5: Means of the predictive distribution of the reserve obtained using a pairs bootstrap of the Mack chain ladder for different deviating and excluded points (thousands). The result for the cleaned dataset is highlighted in red.
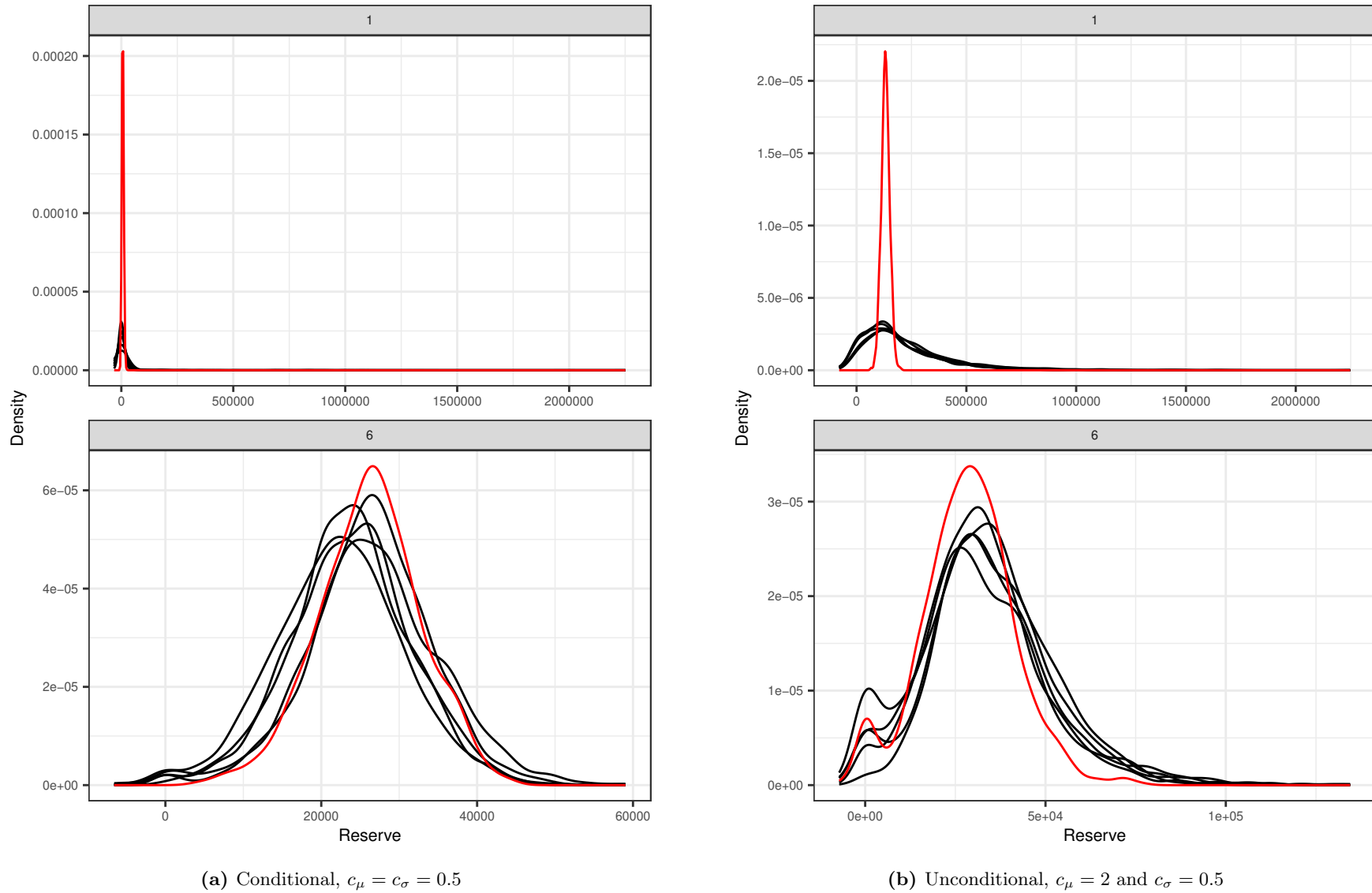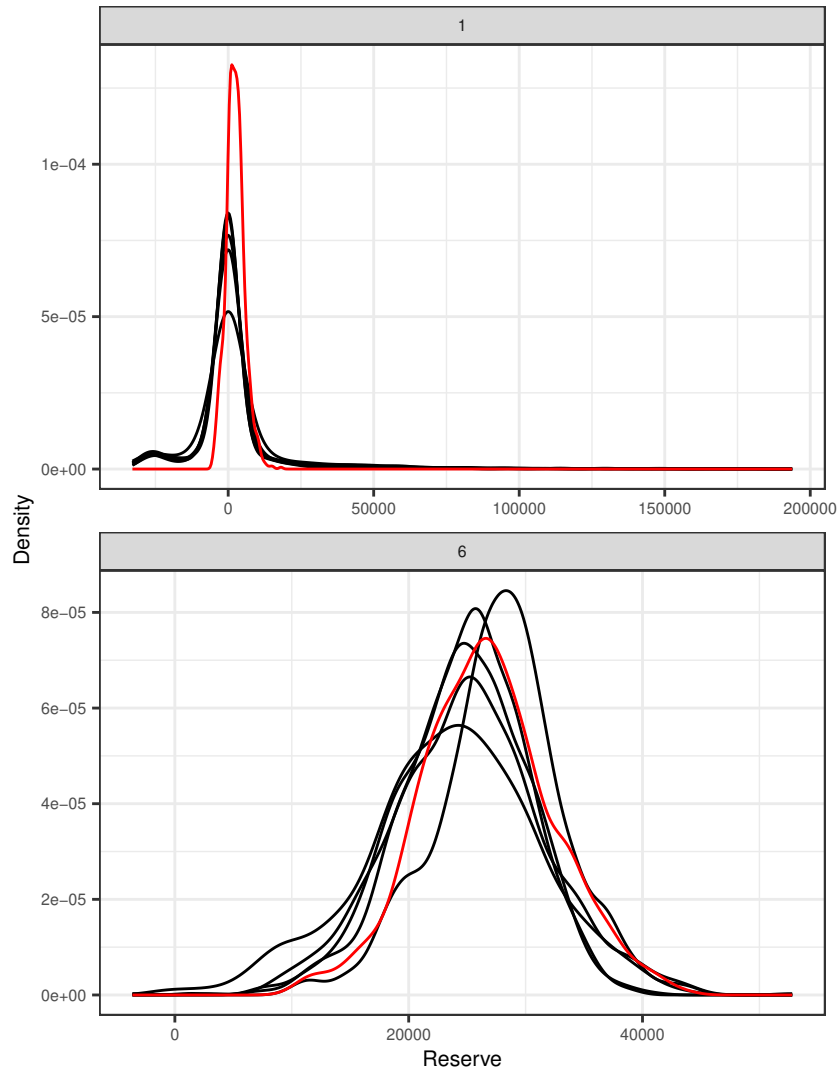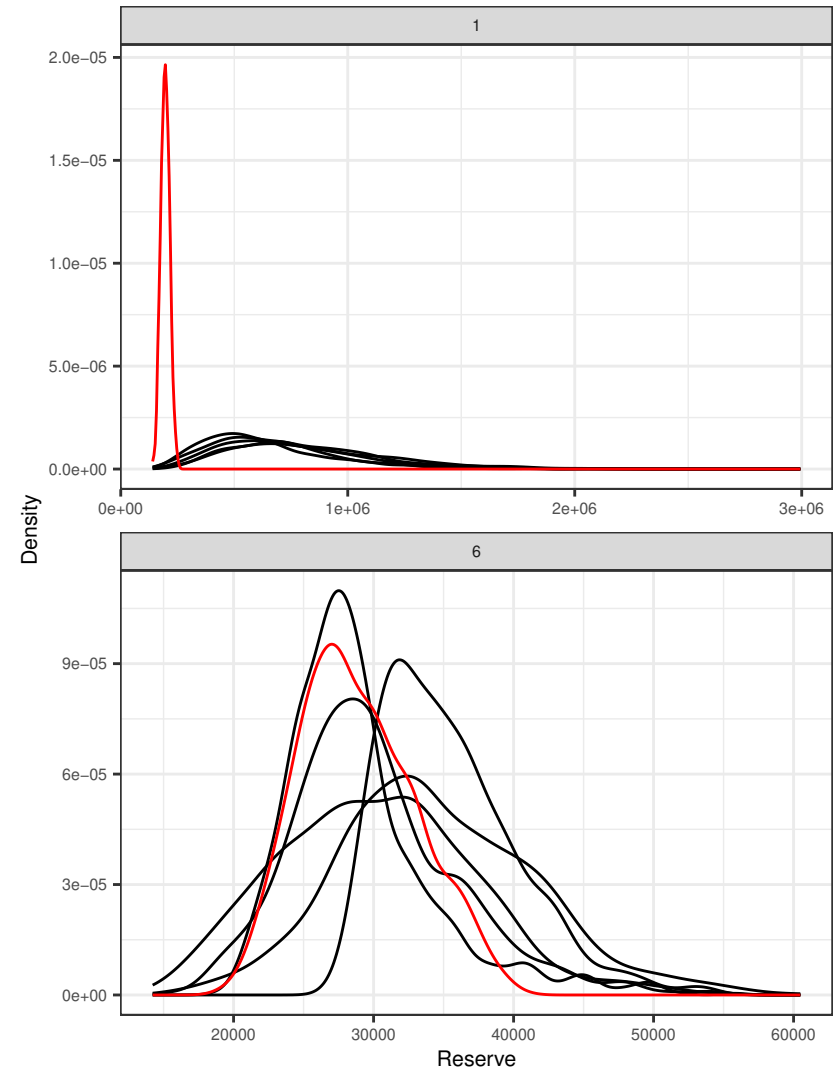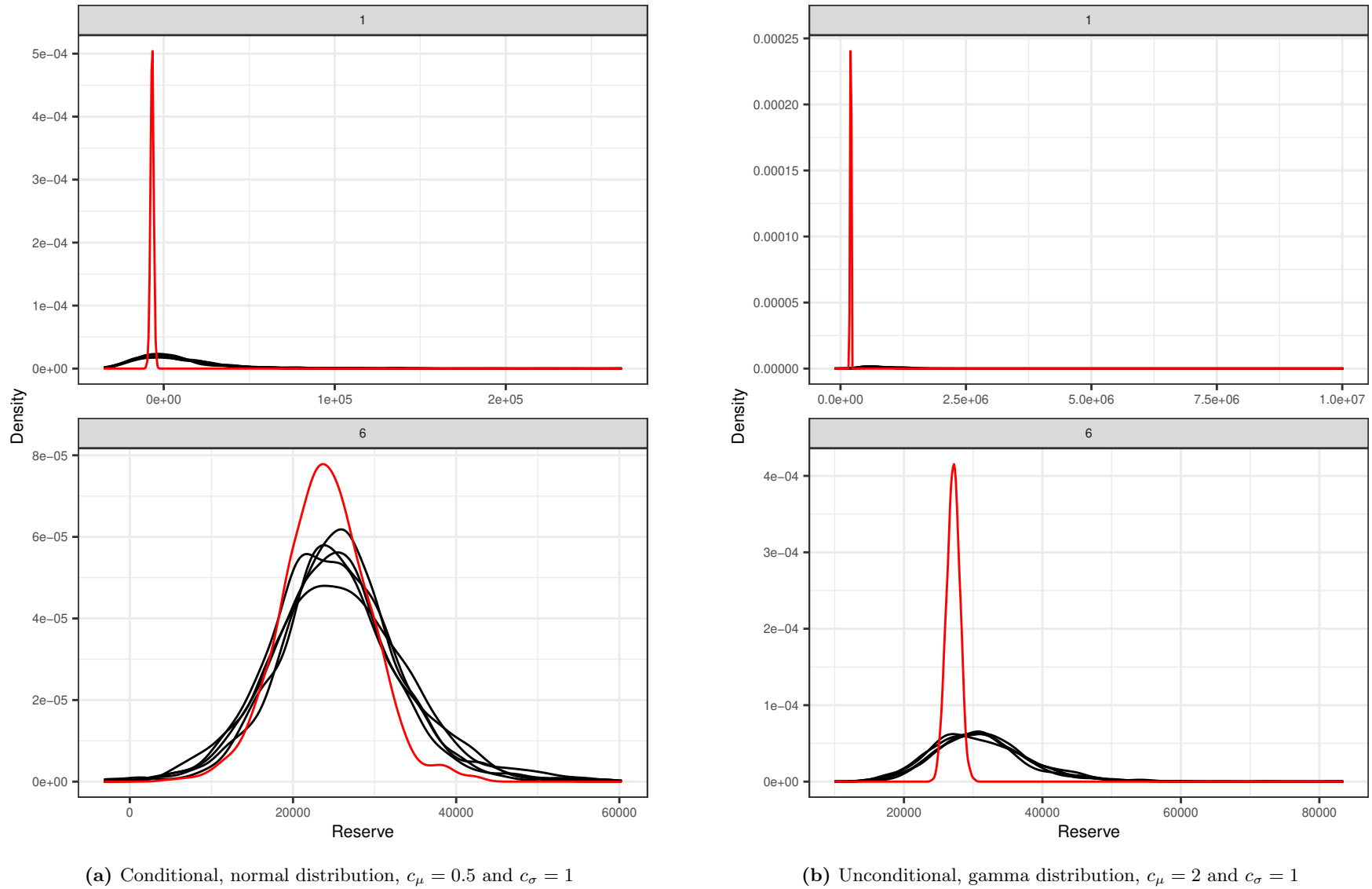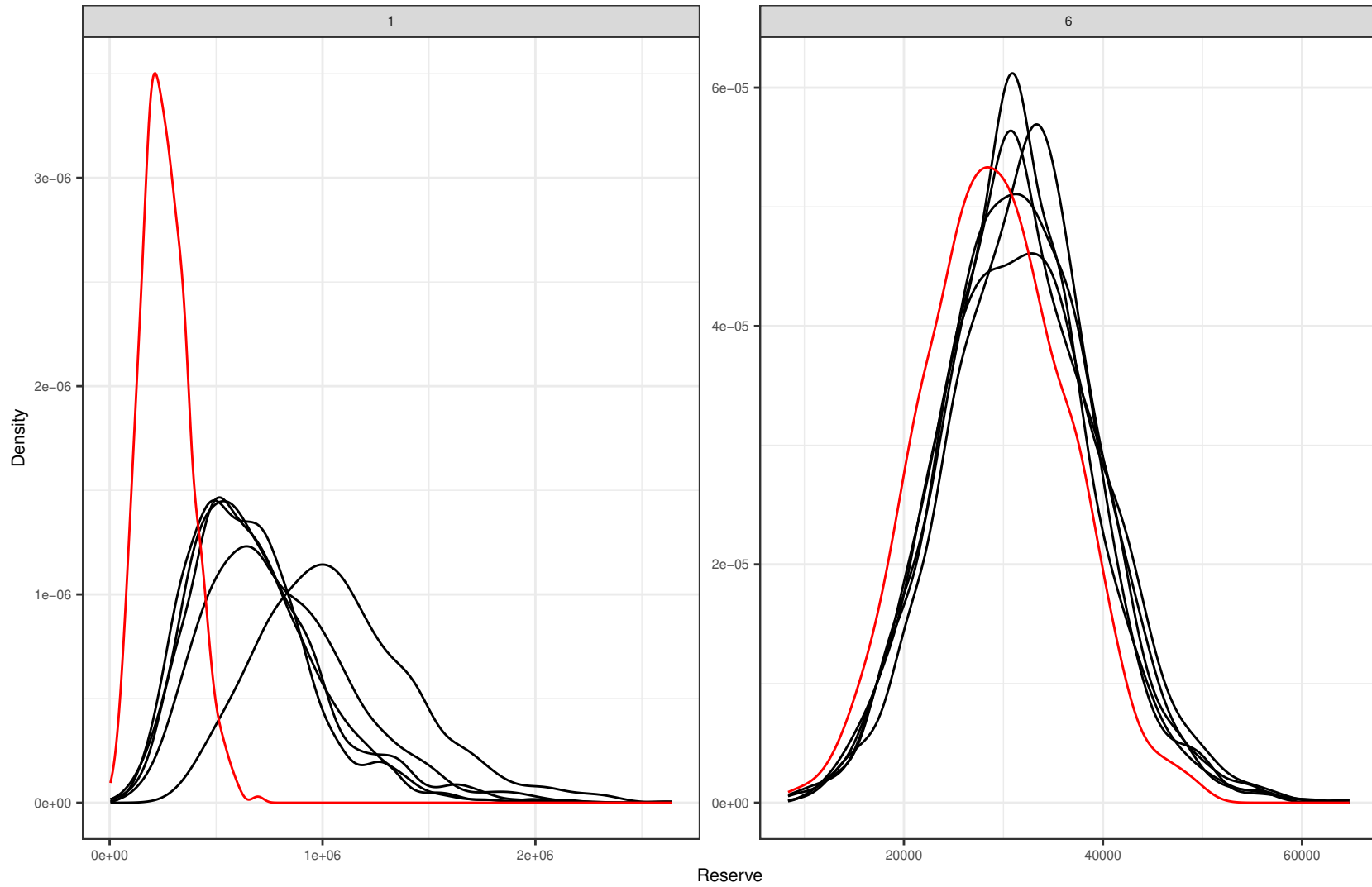
(a) Conditional, $c_\mu = c_\sigma = 0.5$

(b) Unconditional, $c_\mu = 2$ and $c_\sigma = 0.5$

Figure 5.7: Density plots of predictive distributions obtained using a semiparametric bootstrap of the Mack chain ladder with standardised residuals for different deviating calendar periods. The curves in each facet correspond to densities obtained while excluding a different calendar period from the resampling process, with the result for the cleaned dataset highlighted in red.

(a) Conditional, $c_\mu = 0.5$ and $c_\sigma = 2$

(b) Unconditional, $c_\mu = c_\sigma = 2$

Figure 5.8: Density plots of predictive distributions obtained using a semiparametric bootstrap of the Mack chain ladder with log-normal residuals for different deviating calendar periods. The curves in each facet correspond to densities obtained while excluding a different calendar period from the resampling process, with the result for the cleaned dataset highlighted in red.

(a) Conditional, normal distribution, $c_\mu = 0.5$ and $c_\sigma = 1$

(b) Unconditional, gamma distribution, $c_\mu = 2$ and $c_\sigma = 1$

Figure 5.9: Density plots of predictive distributions obtained using a parametric bootstrap of the Mack chain ladder for different deviating calendar periods. The curves in each facet correspond to densities obtained while excluding a different calendar period from the resampling process, with the result for the cleaned dataset highlighted in red.

Figure 5.10: Density plots of predictive distributions obtained using a pairs bootstrap of the Mack chain ladder for different deviating calendar periods. The curves in each facet correspond to densities obtained while excluding a different calendar period from the resampling process, with the result for the cleaned dataset highlighted in red.

5.4.2   Overdispersed Poisson GLM

In the single point simulation, the semiparametric bootstrap performs rather poorly compared to the parametric one, regardless of which distribution is used in the latter. As in Section 5.4.1, we can trace the cause of this by analysing how the residuals are affected by perturbations. Table 5.6 shows a simulated triangle containing a perturbation at the point $(1, 1)$. The Pearson residuals for the original and perturbed triangles are shown in Table 5.7. As we can see, the contamination does not remain localised, but spreads across the whole triangle of residuals, leading to a situation in which the affected point no longer stands out.

Origin	Dev						
	1	2	3	4	5	6	7
1	376.31	6.73	8.99	10.70	11.76	12.35	12.69
2	4.00	7.70	9.98	11.16	12.12	12.75	
3	4.36	8.29	10.23	11.76	12.99		
4	4.30	7.75	9.77	11.09			
5	4.15	7.90	10.22				
6	5.10	9.65					
7	6.28						

Table 5.6: Simulated triangle where observation X_{11} has been perturbed with $c_\lambda = 100$ (thousands)

Origin	Dev						
	1	2	3	4	5	6	7
1	−4.06	−4.47	−1.21	2.68	2.38	1.94	0
2	−0.21	1.36	2.59	0.22	−1.54	−1.87	
3	1.27	2.40	−1.15	−0.90	−0.73		
4	3.53	0.69	−0.87	−1.87			
5	−0.64	−0.24	0.63				
6	−0.05	0.04					
7	0.00						

(a) Original

Origin	Dev						
	1	2	3	4	5	6	7
1	195.62	−144.73	−139.01	−124.02	−102.99	−65.62	0
2	−171.65	48.45	76.31	99.05	129.01	178.47	
3	−155.27	70.60	95.29	124.99	161.95		
4	−127.27	90.19	121.01	151.03			
5	−102.47	123.92	166.30				
6	−70.27	202.37					
7	0.00						

(b) Perturbed

Table 5.7: Example of Pearson residuals obtained in the simulation
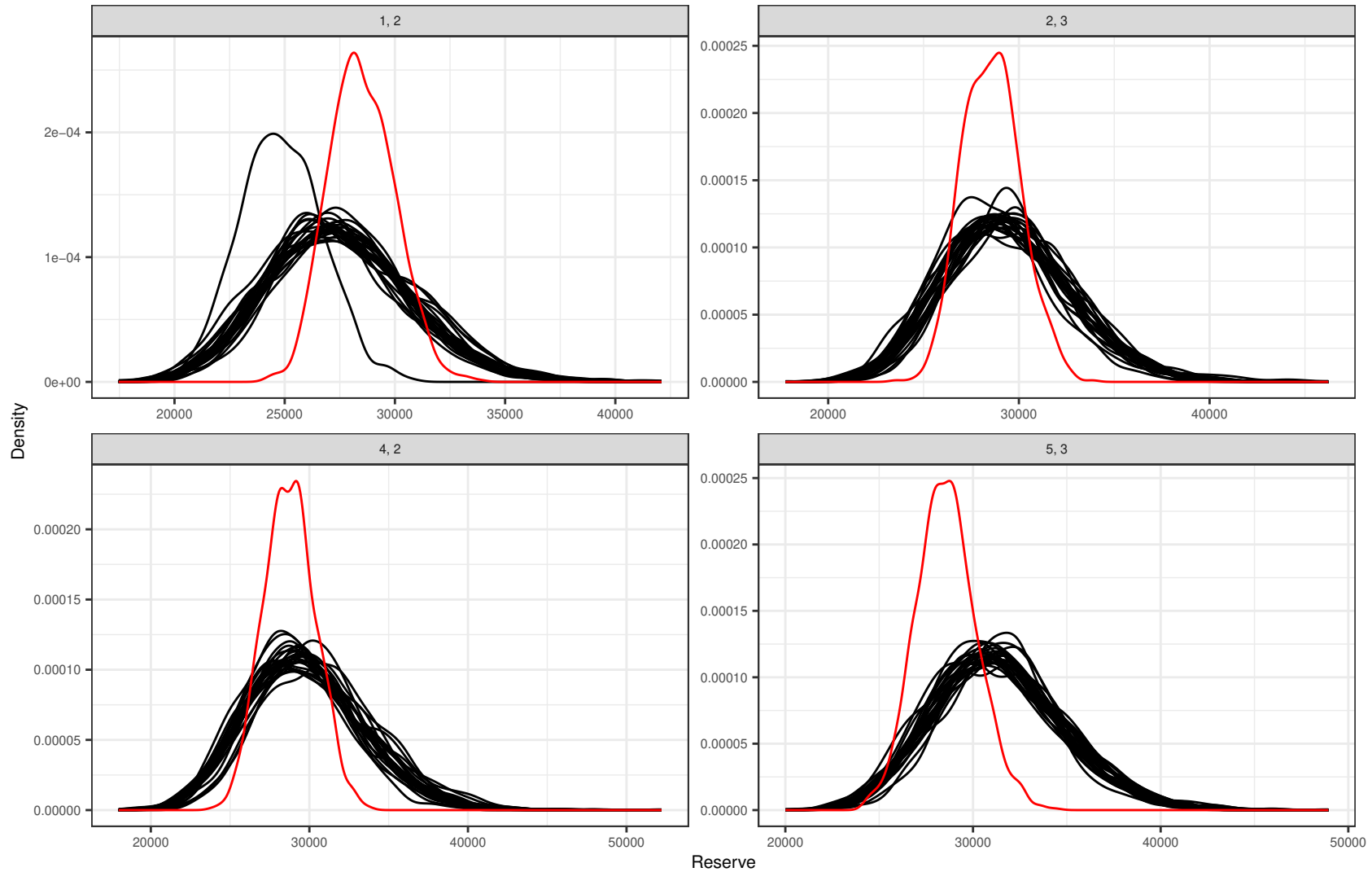
Figure 5.11: Density plots of predictive distributions obtained using a parametric bootstrap of the overdispersed Poisson GLM with gamma distribution for different deviating points. The curves in each facet correspond to densities obtained while excluding a different point from the resampling process, with the result for the cleaned dataset highlighted in red. For all subfigures, $c_\lambda = 0.5$.
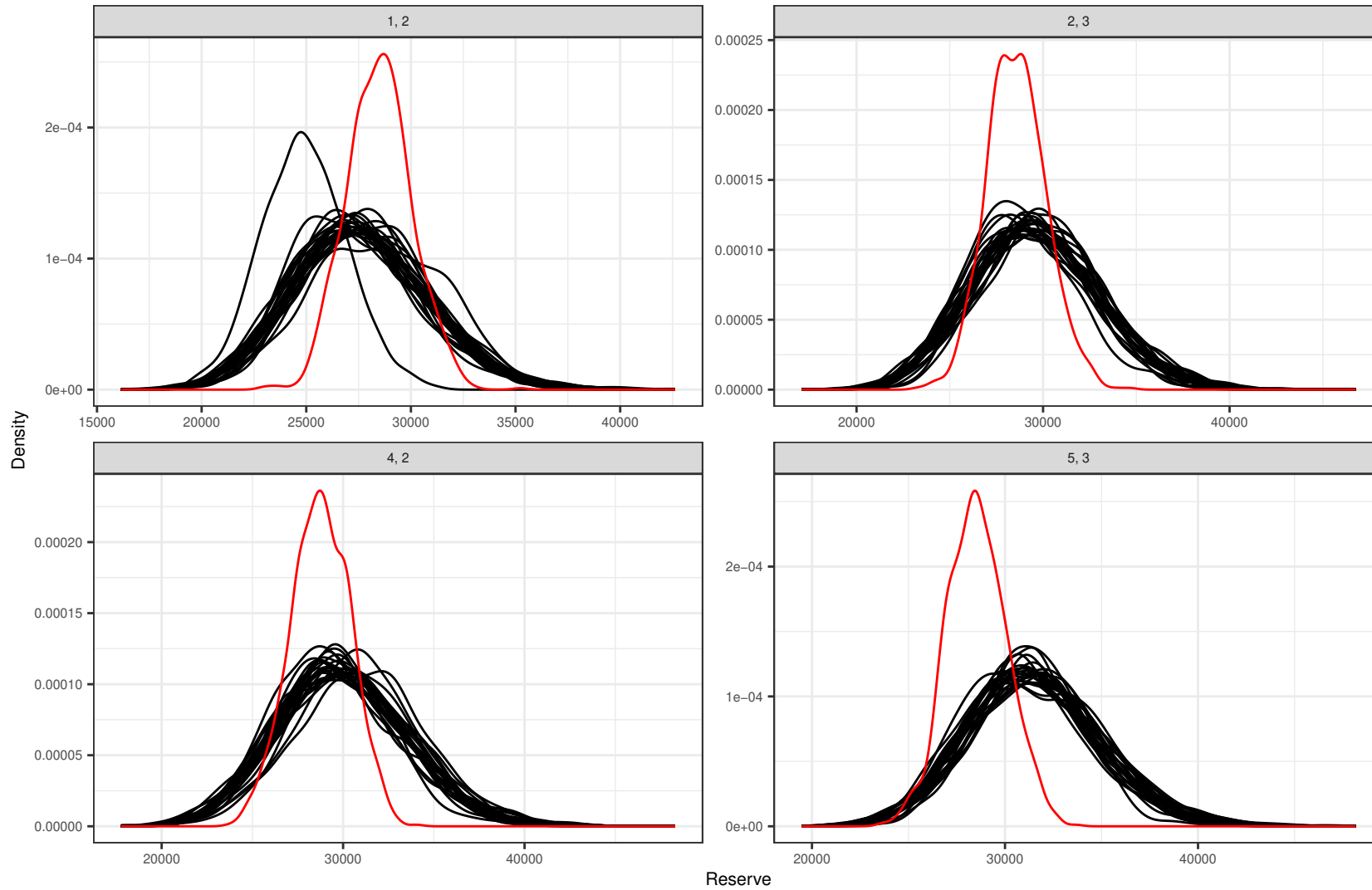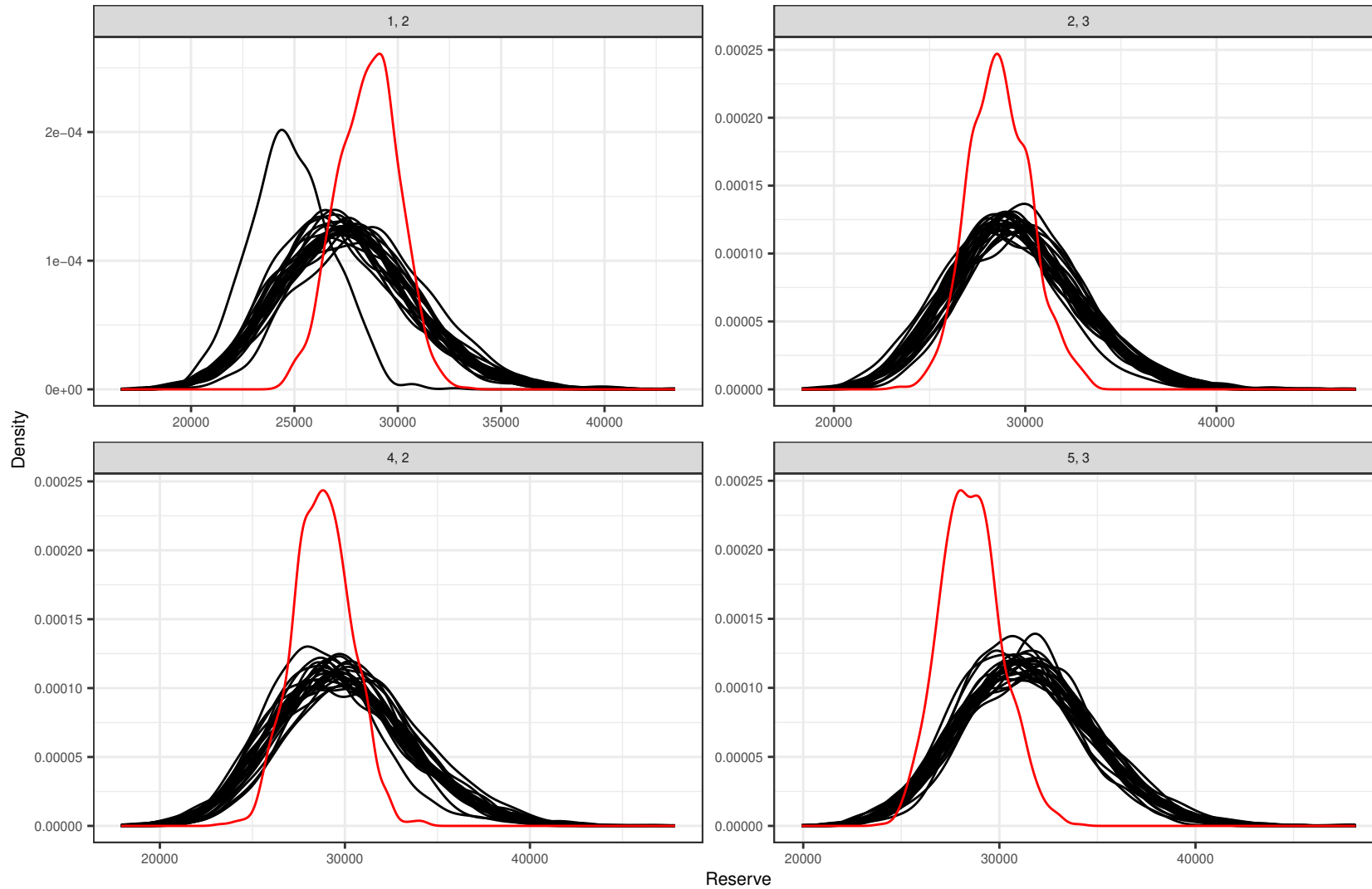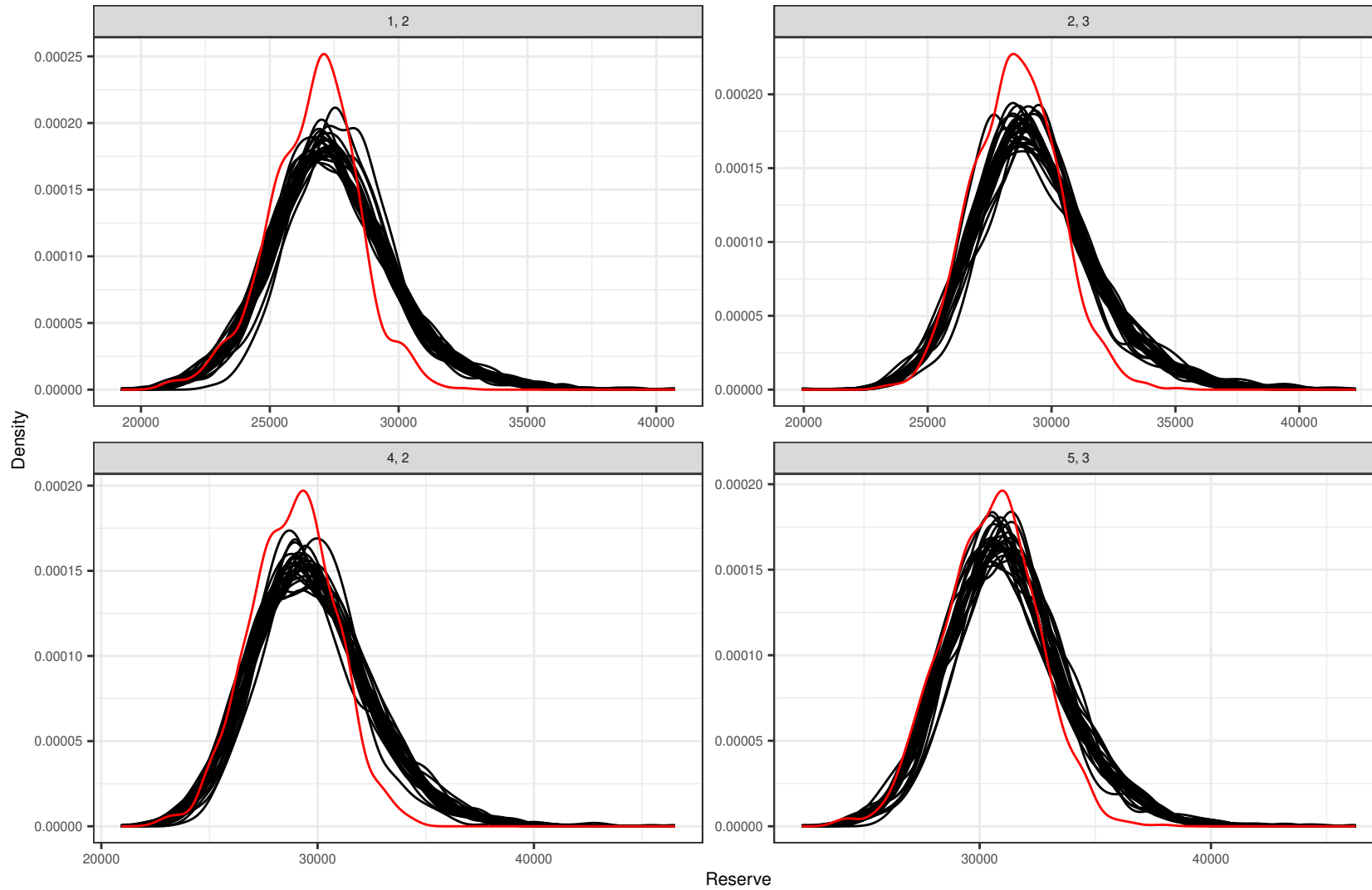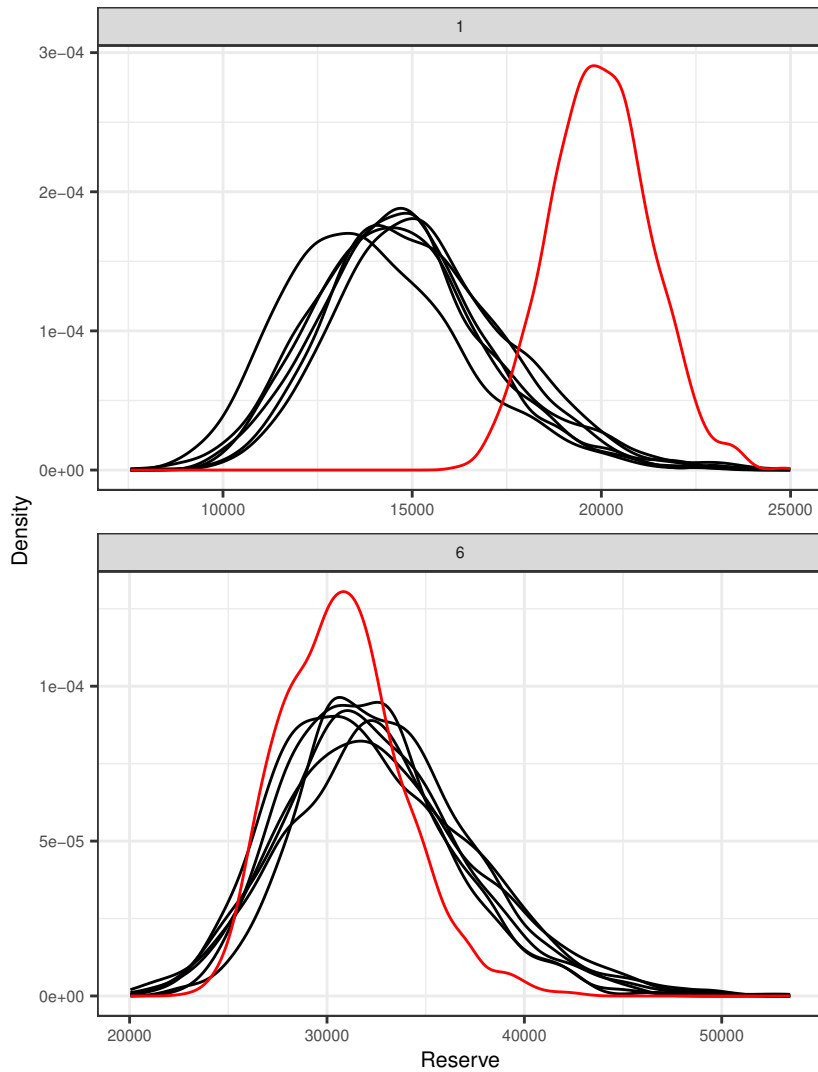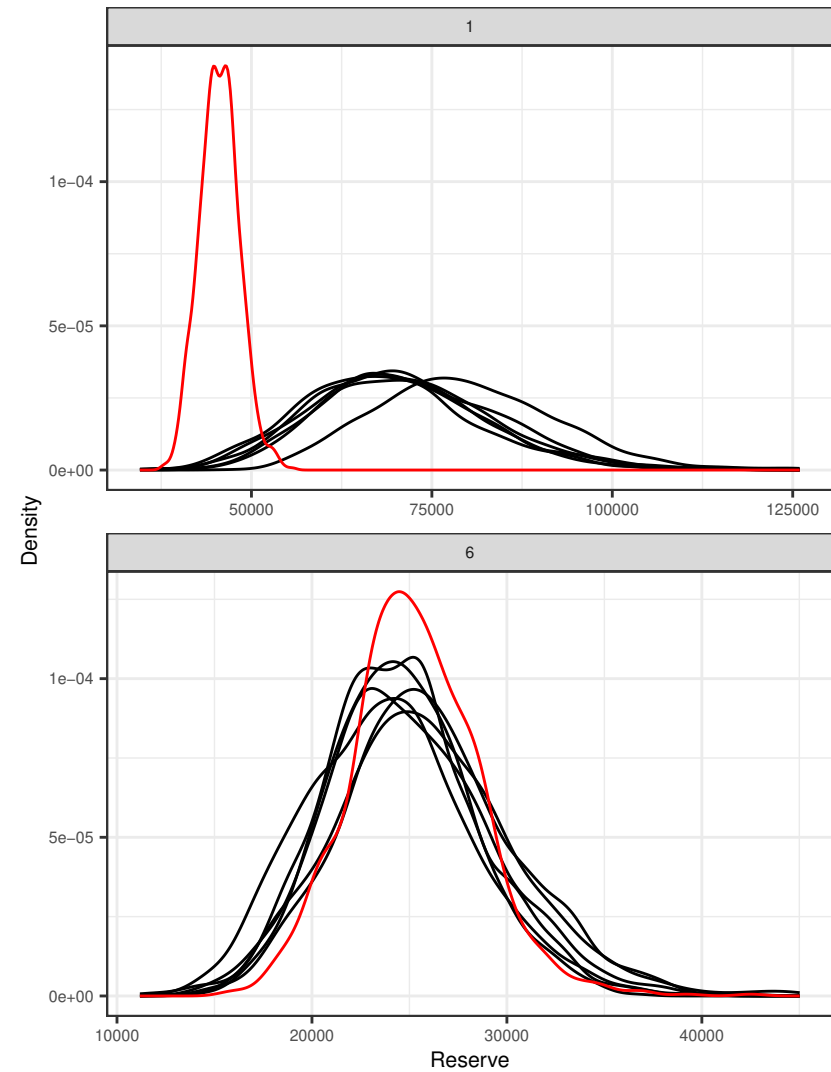
Figure 5.12: Density plots of predictive distributions obtained using a parametric bootstrap of the overdispersed Poisson GLM with normal distribution for different deviating points. The curves in each facet correspond to densities obtained while excluding a different point from the resampling process, with the result for the cleaned dataset highlighted in red. For all subfigures, $c_\lambda = 2$.
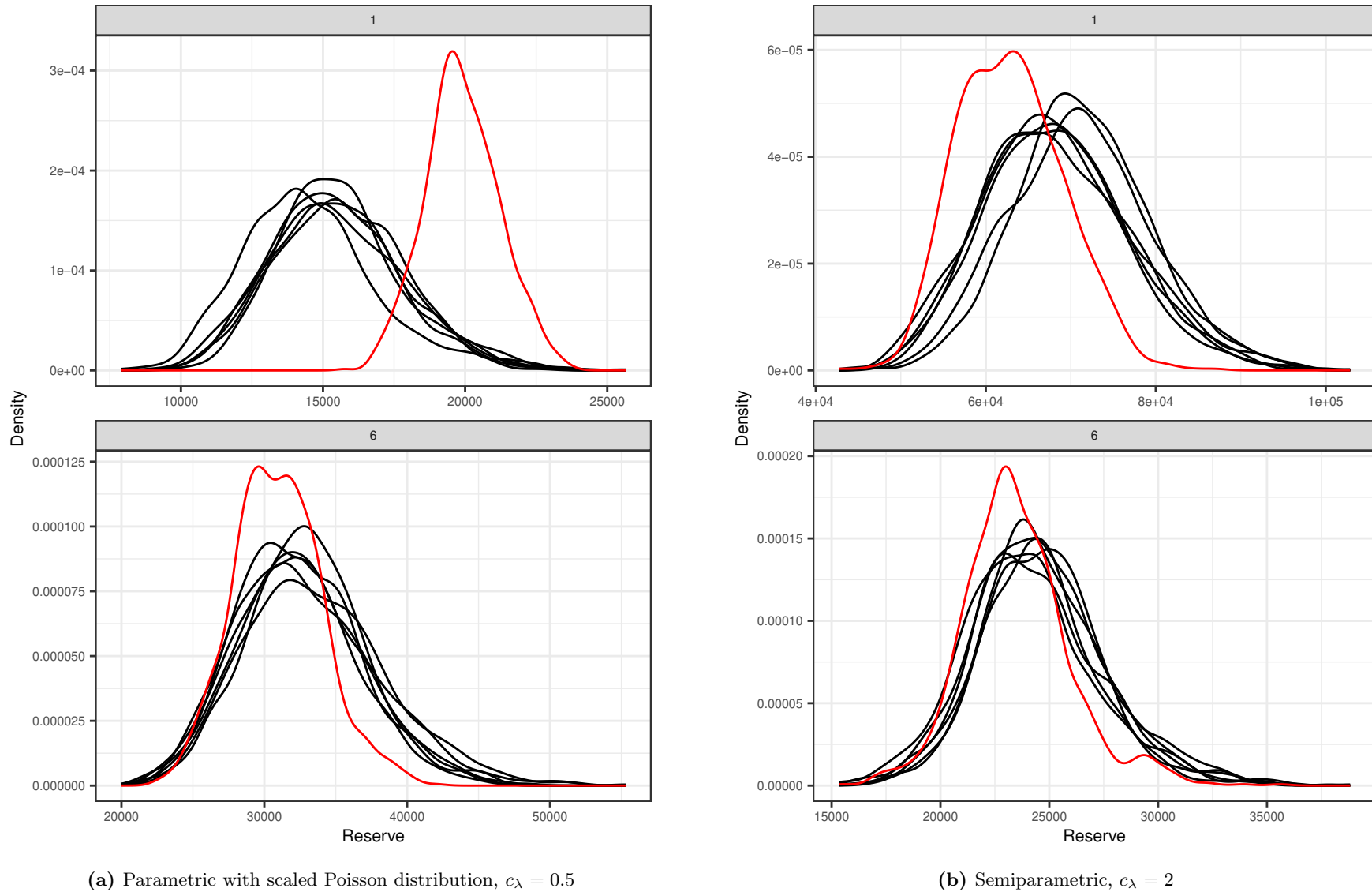
Figure 5.13: Density plots of predictive distributions obtained using a parametric bootstrap of the overdispersed Poisson GLM with scaled Poisson distribution for different deviating points. The curves in each facet correspond to densities obtained while excluding a different point from the resampling process, with the result for the cleaned dataset highlighted in red. For all subfigures, $c_\lambda = 0.5$.

Figure 5.14: Density plots of predictive distributions obtained using a semiparametric bootstrap of the overdispersed Poisson GLM with Pearson residuals for different deviating points. The curves in each facet correspond to densities obtained while excluding a different point from the resampling process, with the result for the cleaned dataset highlighted in red. For all subfigures, $c_\lambda = 2$.

(a) Gamma distribution, $c_\lambda = 0.5$

(b) Normal distribution, $c_\lambda = 2$

Figure 5.15: Density plots of predictive distributions obtained using a parametric bootstrap of the overdispersed Poisson GLM for different deviating calendar periods. The curves in each facet correspond to densities obtained while excluding a different calendar period from the resampling process, with the result for the cleaned dataset highlighted in red.

(a) Parametric with scaled Poisson distribution, $c_\lambda = 0.5$

(b) Semiparametric, $c_\lambda = 2$

Figure 5.16: Density plots of predictive distributions obtained using parametric and semiparametric bootstraps of the overdispersed Poisson GLM for different deviating calendar periods. The curves in each facet correspond to densities obtained while excluding a different calendar period from the resampling process, with the result for the cleaned dataset highlighted in red.

Conclusion

We have investigated the application of bootstrap methods to the problem of detecting deviations from the model assumptions for two major actuarial reserving models, the Mack chain ladder and the overdispersed Poisson GLM. By simulating datasets which exhibit a myriad of different irregularities and studying their effect on the predictive distribution, we managed to identify a number of bootstrap configurations which appear promising for this purpose. The strength of the observed effect is tied to the sensitivity of a bootstrap procedure to the presence of contamination in the data. Parametric bootstraps were found to perform very well: even deviations in a single point caused a significant change in the predictive distribution. For semiparametric bootstraps, the results strongly depend on the type of residuals employed. Finally, the fully nonparametric pairs bootstrap, which can only be used with Mack's model, showed a reasonable performance, though not as good as the parametric variant. These results suggest a way of flagging suspicious datapoints by reverse-engineering the simulation process: instead of starting from synthetic triangles, we use real data and generate a predictive distribution while excluding all observations by turn; the ones for which effect is observed can then be identified as requiring special attention. Although this falls outside the purview of our investigation, it suggests an interesting avenue for future research.

Bibliography

[1] S. Christofides, "Section d5. regression models based on log-incremental payments", in *Claims Reserving Manual*, vol. 2, Institute of Actuaries, 1997.

[2] B. Efron and R. Tibshirani, *An introduction to the bootstrap.* Boca Raton, Fla Chapman & Hall/Crc, 1998, ISBN: 9780412042317.

[3] A. C. Davison and D. V. Hinkley, *Bootstrap methods and their application.* Cambridge University Press, 1997.

[4] O. E. Barndorff-Nielsen and D. R. Cox, "Prediction and asymptotics", *Bernoulli*, vol. 2, no. 4, pp. 319–340, 1996, ISSN: 13507265.

[5] J. F. Lawless and M. Fredette, "Frequentist prediction intervals and predictive distributions", *Biometrika*, vol. 92, no. 3, pp. 529–542, 2005, ISSN: 00063444.

[6] D. C. M. Dickson, L. M. Tedesco and B. Zehnwirth, "Predictive aggregate claims distributions", *The Journal of Risk and Insurance*, vol. 65, no. 4, pp. 689–709, 1998, ISSN: 00224367, 15396975. (visited on 18/04/2023).

[7] T. Mack, "Distribution-free calculation of the standard error of chain ladder reserve estimates", *Astin Bulletin*, vol. 23, no. 2, 1993.

[8] M. V. Wüthrich and M. Merz, *Stochastic Claims Reserving Methods in Insurance.* John Wiley & Sons, 2008.

[9] F. Hayashi, *Econometrics.* Princeton, Nj ; Oxford: Princeton University Press, 2000, ISBN: 9780691010182.

[10] P. D. England and R. J. Verrall, "Predictive distributions of outstanding liabilities in general insurance", *Annals of Actuarial Science*, vol. 1, no. 1, 2006.

[11] T. Mack, "Section d6. measuring the variability of chain ladder reserve estimates", in *Claims Reserving Manual*, vol. 2, Institute of Actuaries, 1997.

[12] M. Lindholm, F. Lindskog and F. Wahl, "Estimation of conditional mean squared error of prediction for claims reserving", *Annals of Actuarial Science*, vol. 14, no. 1, pp. 93–128, 2020. DOI: 10.1017/S174849951900006X.

[13] M. V. Wüthrich, M. Buchwalder, H. Bühlmann and M. Merz, "The mean square error of prediction in the chain ladder reserving method (mack and murphy revisited)", *Astin Bulletin*, vol. 36, no. 2, 2006.

[14] T. Mack, G. Quarg and C. Braun, "The mean square error of prediction in the chain ladder reserving method – a comment", *ASTIN Bulletin: The Journal of the IAA*, vol. 36, no. 2, pp. 543–552, 2006. DOI: 10.1017/S051503610001463X.

[15] A. Gisler, "The estimation error in the chain-ladder reserving method: A bayesian approach", *ASTIN Bulletin: The Journal of the IAA*, vol. 36, no. 2, pp. 554–565, 2006. DOI: 10.1017/S0515036100014653.

[16] G. G. Venter, "Discussion of the mean square error of prediction in the chain ladder reserving method", *ASTIN Bulletin: The Journal of the IAA*, vol. 36, no. 2, pp. 566–571, 2006. DOI: 10.1017/S0515036100014665.

[17] G. A. F. Seber and A. J. Lee, *Linear regression analysis*. Wiley-Interscience, 2003, ISBN: 9780471415404.

[18] A. Renshaw and R. Verrall, "A stochastic model underlying the chain-ladder technique", *British Actuarial Journal*, vol. 4, no. 4, pp. 903–923, 1998. DOI: 10.1017/S1357321700000222.

[19] J. A. Nelder and R. W. M. Wedderburn, "Generalized linear models", *Journal of the Royal Statistical Society. Series A (General)*, vol. 135, no. 3, pp. 370–384, 1972, ISSN: 00359238.

[20] T. W. Yee, *Vector Generalized Linear and Additive Models*. Springer, Sep. 2015, ISBN: 9781493928187.

[21] P. J. R. Pinheiro, J. M. A. e Silva and M. de Lourdes Centeno, "Bootstrap methodology in claim reserving", *The Journal of Risk and Insurance*, vol. 70, no. 4, pp. 701–714, 2003, ISSN: 00224367, 15396975.

[22] L. H. Moulton and S. L. Zeger, "Bootstrapping generalized linear models", *Computational Statistics & Data Analysis*, vol. 11, no. 1, pp. 53–63, 1991, ISSN: 0167-9473. DOI: https://doi.org/10.1016/0167-9473(91)90052-4.

[23] P. Dunn and G. Smyth, *Generalized Linear Models With Examples in R* (Springer Texts in Statistics). Springer New York, 2018, ISBN: 9781441901187. [Online]. Available: https://books.google.be/books?id=tBh5DwAAQBAJ.

[24] K. Peremans, P. Segaert, S. Van Aelst and T. Verdonck, "Robust bootstrap procedures for the chain-ladder method", *Scandinavian Actuarial Journal*, pp. 1–28, Dec. 2016. DOI: 10.1080/03461238.2016.1263236.

[25] T. Hartl, "Bootstrapping generalized linear models for development triangles using deviance residuals", 2010.

[26] P. England, "Addendum to "analytic and bootstrap estimates of prediction errors in claims reserving"", *Insurance: Mathematics and Economics*, vol. 31, no. 3, pp. 461–466, 2002, ISSN: 0167-6687. DOI: https://doi.org/10.1016/S0167-6687(02)00161-0.

[27] M. Schiegl, "Simulation", in *Handbook on Loss Reserving*, M. Radtke, K. D. Schmidt and A. Schnaus, Eds. Cham: Springer International Publishing, 2016, pp. 265–274, ISBN: 978-3-319-30056-6. DOI: 10.1007/978-3-319-30056-6_35.

[28] R Core Team, *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria, 2021. [Online]. Available: https://www.R-project.org/.

[29] D. Eddelbuettel and R. François, "Rcpp: Seamless R and C++ integration", *Journal of Statistical Software*, vol. 40, no. 8, pp. 1–18, 2011. DOI: 10.18637/jss.v040.i08.

Index