

# Robust bootstrap procedures for the chain-ladder method

Kris Peremans<sup>1</sup>, Pieter Segaert<sup>1</sup>, Stefan Van Aelst<sup>1</sup>, and Tim Verdonck<sup>\*1</sup>

<sup>1</sup>Department of Mathematics, KU Leuven, Leuven, Belgium<sup>†</sup>

## Abstract

Insurers are faced with the challenge of estimating the future reserves needed to handle historic and outstanding claims that are not fully settled. A well-known and widely used technique is the chain-ladder method, which is a deterministic algorithm. To include a stochastic component one may apply generalized linear models to the run-off triangles based on past claims data. Analytical expressions for the standard deviation of the resulting reserve estimates are typically difficult to derive. A popular alternative approach to obtain inference is to use the bootstrap technique. However, the standard procedures are very sensitive to the possible presence of outliers. These atypical observations, deviating from the pattern of the majority of the data, may both inflate or deflate traditional reserve estimates and corresponding inference such as their standard errors. Even when paired with a robust chain-ladder method, classical bootstrap inference may break down. Therefore, we discuss and implement several robust bootstrap procedures in the claims reserving framework and we investigate and compare their performance on both simulated and real data. We also illustrate their use for obtaining the distribution of one year risk measures.

**Keywords:** Claims reserving, Generalized linear models, Outliers, Inference, Solvency II

---

<sup>\*</sup>corresponding author. E-mails : kris.peremans@kuleuven.be, pieter.segaert@kuleuven.be, stefan.vanaelst@kuleuven.be, tim.verdonck@kuleuven.be

<sup>†</sup>KU Leuven, Department of Mathematics, Celestijnenlaan 200B box 2400, BE-3001 Leuven

# 1 Introduction

One of the main challenges for insurers is the task of estimating the future reserves needed to handle the liabilities related to current insurance contracts. The recent Solvency II directive provides new guidelines for the calculation of these provisions. A comprehensive discussion on the Solvency II directive and its implications may be found in Dreksler et al. (2015). They report that technical provisions under Solvency II consist of the following three components: claims provisions, premium provisions and a risk margin. Whereas claims provisions relate to claim events that have already occurred, the premium provisions relate to all future claim events covered by current insurance and reinsurance obligations. The risk margin is described as the amount needed to make the value of the technical provisions equivalent to the amount a (re)insurance undertaking would be expected to require in order to take over and meet (re)insurance obligations.

In this article we focus on the claims provision component. Although micro-level claims reserving models (see for example Norberg (1993); Antonio and Plat (2014)) seem to be very promising, Dreksler et al. (2015) state that the current norm is to use deterministic actuarial techniques to assess the undiscounted element of the earned claim reserves and that this is likely to continue in the immediate future. The most popular and widely used technique is the chain-ladder method. A substantial amount of research has already been devoted to exploring the relationship between the chain-ladder method and various stochastic models (see for example Mack (1993, 1994), Renshaw and Verrall (1998), Verrall (1991a, 2000) and Schiegl (2015)).

To approach claims reserving in a stochastic manner, the chain-ladder method is typically reformulated within the context of generalized linear models (GLM). However, assessing the variability of claims reserving predictions and the construction of upper limits with a certain confidence level still remains a difficult task. Analytical expressions for standard errors in GLM depend on strong model assumptions (Mack et al. (2006); Venter (2006) and Wüthrich et al. (2008)). An alternative approach to statistical inference that requires less assumptions is the bootstrap technique (see for example Efron (1981), Hall et al. (1995) and Chatterjee and Lahiri (2011)). Also in the claims reserving framework, the bootstrap technique is very popular and has been studied by various authors such as Ashe (1986), Lowe (1994), Pinheiro et al. (2003), England and Verrall (2006) and Barnett and Zehnwirth (2008).

When applying a claims reserving method in practice it may occur that one or a couple of observations deviate from the main pattern. One might naively expect that if the model assumptions hold for almost all available data, then the results of the classical method also hold approximately. This is unfortunately not the case and it is well known in the robust statistics literature that outlier(s) may heavily influence classical statistical techniques. Verdonck and Debruyne (2011) have shown theoretically that the classical chain-ladder method is very sensitive to outliers and several robust methods have already been developed in the claims reserving framework (see e.g. Brazauskas et al. (2009), Brazauskas (2009), Verdonck et al. (2009), Verdonck and Debruyne (2011), Verdonck and Van Wouwe (2011) and Pitselis et al. (2015)). Robust methods provide estimates for the claim provisions which resemble the classical estimates that would have been obtained if there were no outliers in the data, while they do

not model the outlier generating process. As a consequence of fitting the majority of the data well, robust methods also provide a reliable method to detect outliers. Observations which are flagged as outliers can then be examined in detail by experts to understand their origin. If these flagged observations are errors or one-time events, then the robust method indeed yields reliable estimates of the claim provisions as desired. If the flagged observations are not errors and such deviations may reoccur in the future, then the robust reserve may need to be adjusted to cover such potentially large deviations in the future. We believe that expertise from practitioners in the field of the data is needed to achieve this, possibly combined with techniques from extreme value theory. The idea of combining robust statistics (which typically downweights atypical points) and extreme value statistics (which models the extremes) has also been advocated in Dell'Aquila and Embrechts (2006) and Hubert et al. (2013). It is however outside the scope of this paper to investigate estimators for the extra amount that needs to be added to the robust reserve estimate in such case.

To our knowledge, robust methods in claims reserving have only focused on reliable point estimates and/or outlier detection. However, confidence intervals may also become very unreliable (incorrect coverage) and/or uninformative (very large interval length) when the data are contaminated. In this article we investigate the influence of outliers on the variability of point estimators and we present some state-of-the-art robust bootstrap procedures in the claims reserving framework. These bootstrap techniques provide a complete approximate predictive distribution, from which characteristics such as the mean, variance and quantiles are easily derived. Note that we do not advocate replacing the classical methodology by the robust counterpart, but we advise to always apply both techniques. If the classical and robust method give approximately the same results (and no outliers are detected by the robust method), then it seems safe to continue the analysis with the traditional reserving techniques that will give the most efficient estimates. However, if both methods yield different results, then the robust procedure is helpful to gain insight into the data and will yield more reliable reserve estimates if the outliers are indeed isolated events.

This paper is structured as follows. Section 2 reviews the chain-ladder method and its formulation as a Poisson GLM (which is the default approach to obtain prediction under Solvency II). In this section, we also present the bootstrap procedure in the context of claims reserving. The robust chain-ladder method is briefly described in Section 3 and it is shown that the classical bootstrap method may produce unreliable results when the data contains outlier(s), even when the classical chain-ladder method is replaced by its robust alternative. Therefore, we present several state-of-the-art robust bootstrap methods for GLM within the context of claims reserving in section 4. Their performance is compared by means of a simulation study in section 5 and on a real dataset from a non-life business line in section 6. We also illustrate how the methodology can be applied to investigate one year risk measures and to detect outliers automatically. Some concluding remarks and potential directions for further research are given in section 7.

## 2 Classical chain-ladder method and bootstrapping

In this section we introduce the classical chain-ladder method and briefly explain how this method can be derived from a basic GLM. Then, we describe the classical bootstrap procedure to obtain inference corresponding to the estimates in this GLM. Finally, we briefly describe the robust GLM estimator.

### 2.1 Classical chain-ladder method

Insurers have to build up reserves enabling them to pay outstanding claims and to meet claims arising in the future on the written contracts. The chain-ladder method is a very popular algorithm to compute these reserve estimates. Let  $Y_{ij}$  and  $C_{ij}$  denote respectively the incremental and cumulative claim amount corresponding to accident year  $i$  and development year  $j$  for  $1 \leq i \leq I$  and  $1 \leq j \leq J$ . We suppose that  $I = J = n$  for ease of notation, but all methodology remains valid when  $I \neq J$ . Variables for which  $i + j \leq n + 1$  correspond to past claims data and are used to predict the future claims which correspond to the variables with  $i + j > n + 1$ . We define the total or overall reserve  $R$  as the sum over all future (incremental) claims<sup>1</sup>:

$$R = \sum_{i=2}^n \sum_{j=n-i+2}^n Y_{ij}. \quad (1)$$

$\frac{\text{development year}}{\text{accident year}}$	1	2	...	$j$	...	$J-1$	$J$
1	$Y_{11}$	$Y_{12}$	...	$Y_{1j}$	...	$Y_{1,J-1}$	$Y_{1J}$
2	$Y_{21}$	$Y_{22}$	...	$Y_{2j}$	...	$Y_{2,J-1}$	
$\vdots$							
$i$	$Y_{i1}$	$Y_{i2}$	...	$Y_{ij}$			
$\vdots$							
$I$	$Y_{I1}$						

Table 1: Run-off triangle (incremental claim amounts).

The chain-ladder method uses cumulative data and assumes the existence of development factors  $f_j$  for  $j = 2, \dots, n$  that are estimated by

$$\hat{f}_j = \frac{\sum_{i=1}^{n-j+1} C_{ij}}{\sum_{i=1}^{n-j+1} C_{i,j-1}}.$$

<sup>1</sup>Note that we define the total reserve  $R$  as the total amount of the incremental claims in the lower right triangle (as in key references England and Verrall (2002) and Wüthrich and Merz (2008)), but in practice the final reserve is an amount set by the insurer based on this value.

This leads to the following estimates for the future claim amounts:

$$\begin{aligned}\hat{C}_{i,n-i+2} &= C_{i,n-i+1} \cdot \hat{f}_{n-i+2} & 1 < i \leq n \\ \hat{C}_{ik} &= \hat{C}_{i,k-1} \cdot \hat{f}_k & 3 \leq i \leq n \text{ and } n-i+3 \leq k \leq n.\end{aligned}$$

The chain-ladder method is typically specified in the GLM framework using the following multiplicative model for the incremental claim amounts:

$$Y_{ij} \stackrel{i.i.d.}{\sim} \text{Poisson}(\zeta_i \xi_j), \quad (2)$$

with  $\zeta_i$  the exposure (i.e. expected claims up to the latest observed development year) of accident year  $i$  and  $\xi_j$  the expected claims pattern over the different development periods  $j$ . Since  $\zeta_i$  and  $\xi_j$  can only be determined up to a constant factor, the condition  $\sum_{j=1}^J \xi_j = 1$  is often imposed. Formulation (2) has a multiplicative structure for the mean, i.e.  $\mu_{ij} = E[Y_{ij}] = \zeta_i \xi_j$  and hence it is straightforward to use a log link function such that

$$\log(\mu_{ij}) = \tau + \alpha_i + \beta_j = x_{ij}^t \boldsymbol{\theta}. \quad (3)$$

This structure now has a parameter for each row  $i$  and each column  $j$  and typically the constraint  $\alpha_1 = \beta_1 = 0$  is used (to obtain a more convenient interpretation of the parameters). Formulations (2) and (3) are equivalent and are reparameterisations of the same structure. The parameters of the first have physical interpretations, whereas the statistical analysis of the latter is more straightforward. The relationship between the parameters and the proof of the equivalence between both models is given in Kremer (1982) and Verrall (1991b). The full parameter vector is thus  $\boldsymbol{\theta} = (\tau, \alpha_2, \dots, \alpha_I, \beta_2, \dots, \beta_I)$  and the number of parameters to be estimated equals  $2n - 1$ . The predictor variable  $x_{ij}$  consists of the constant 1 (to include an intercept) and indicators corresponding to each of the rows and columns of the claims triangle except for the first. By inserting parameter estimates  $\hat{\boldsymbol{\theta}}$  into equation (3) and exponentiating, estimates for the future claim amounts and the total reserve are then obtained. When estimating the parameters  $\boldsymbol{\theta}$  by maximum likelihood, the results are identical to the estimates obtained by the deterministic chain-ladder method using development factors as shown in Kremer (1982). For more details we refer to England and Verrall (2002), Hoedemakers et al. (2005) and Wüthrich and Merz (2008).

## 2.2 Classical bootstrap procedure

Point estimates are very useful of course, but the natural next step is to also estimate their precision. We consider the bootstrap methodology to derive inference corresponding to the point estimates. To create bootstrap samples we follow the procedure described in Kaas et al. (2009), which is based on England and Verrall (1999, 2002).

First, the Pearson residuals, defined as

$$r_{ij} = \frac{y_{ij} - \mu_{ij}}{\sqrt{\mu_{ij}}} \quad \text{for } 1 \leq i, j \leq n \text{ and } i + j \leq n + 1,$$

are calculated by replacing the  $\mu_{ij}$  by their estimates  $\hat{\mu}_{ij} = e^{x_{ij}^t \hat{\theta}}$ . When the number of observations is not very large, the residuals suffer from a small sample bias. To remedy this, England and Verrall (1999) and England (2002) propose to adjust the residuals  $\mathbf{r} = \{r_{ij} | 1 \leq i, j \leq n; i + j \leq n + 1\}$  by multiplying them by a correction factor:

$$\mathbf{r}^E = \sqrt{\frac{N}{N-p}} \mathbf{r},$$

where  $N = n(n+1)/2$  is the sample size and  $p = 2n - 1$  is the number of fitted parameters. Pinheiro et al. (2003) argue that it is better to use an individual adjustment factor for each residual  $\{r_k | k = 1, \dots, N\} = \{r_{ij} | 1 \leq i, j \leq n; i + j \leq n + 1\}$  as opposed to a global correction factor. In correspondence with the classical linear regression model they propose to use the hat matrix  $H$  of the model to standardize the Pearson residuals as follows

$$r_{ij}^P = \frac{r_{ij}}{\sqrt{1 - h_{kk}}}$$

where  $h_{kk}$  is the corresponding element on the diagonal of the hat matrix. For a Poisson GLM, the hat matrix is given by  $H = W^{1/2} X (X^T W X)^{-1} X^T W^{1/2}$ , where  $W$  is a diagonal matrix with elements  $\{\mu_{ij} | 1 \leq i, j \leq n; i + j \leq n + 1\}$  on the diagonal (see e.g. McCullagh and Nelder (1983) for more details). Cordeiro (2004) recently provided general expressions for the expectation and variance of Pearson residuals in the GLM framework up to the first order  $O(N^{-1})$ . Using equations (13) and (14), calculated in Appendix 8.2, we obtain that

$$\begin{aligned} E[\mathbf{r}] &= \text{diag} \left( -\frac{1}{2} (I - H) H W^{-1/2} \right) \\ \text{Var}[\mathbf{r}] &= \text{diag} (I - H) \end{aligned}$$

where ‘diag’ indicates that we extract the diagonal elements of a matrix. These correction terms lead to the following adjusted residuals

$$\mathbf{r}^C = \frac{\mathbf{r} - E[\mathbf{r}]}{\sqrt{\text{Var}[\mathbf{r}]}}.$$

Replacing the parameters  $\mu_{ij}$  in the expressions of  $H$  and  $W$  by their estimates  $\hat{\mu}_{ij}$  yields the expectation and variance correction term of the sample Pearson residuals.

The bootstrap methodology can be applied by using the original Pearson residuals or any of the above explained adjustments. From now on, we use the general notation  $\mathbf{r}^*$  for the residuals of choice, i.e.  $\mathbf{r}^*$  is any of  $\mathbf{r}, \mathbf{r}^E, \mathbf{r}^P$  or  $\mathbf{r}^C$ . In the next sections, all results shown are obtained for  $\mathbf{r}^* = \mathbf{r}^C$ , but the conclusions also hold for the other adjustment options.

After obtaining the residuals, the following bootstrap steps are performed many times, e.g. for  $b = 1, \dots, 10,000$ :

1. Resample with replacement from the residuals  $\mathbf{r}^*$  leading to a new vector of residuals  $\mathbf{r}^{(b)}$ . Since there is only one observation for the last accident year and the last development year, the corresponding residuals (on the cornerpoints of the run-off triangle) will be exactly zero when fitting a GLM. As suggested by England (2002) and Pinheiro et al. (2003), these two residuals are not resampled (whereas all other residuals have equal resampling probabilities).

2. Backtransform the new residuals  $\mathbf{r}^{(b)}$  in order to obtain a new pseudo-history:

$$\mathbf{y}^{(b)} = \mathbf{r}^{(b)}\sqrt{\boldsymbol{\mu}} + \boldsymbol{\mu}.$$

3. Refit the GLM on the new pseudo-history  $\mathbf{y}^{(b)}$  in order to obtain a new total future reserve estimate  $R^{(b)}$ .

### 3 Robust chain-ladder method and classical bootstrapping

The robust chain-ladder method of Verdonck and Debruyne (2011) is briefly described and the effect of outliers on the total reserve estimates obtained by the classical and robust chain-ladder method is illustrated on the well studied claims data set of Taylor and Ashe (1983) as well as on a simulated data set. For both estimators we also evaluate the effect of outliers on an estimated upper limit for the total reserve obtained by using the classical bootstrap procedure.

#### 3.1 Robust chain-ladder method

Verdonck and Debruyne (2011) already showed the sensitivity of the classical chain-ladder method to outliers and therefore proposed a robust alternative, which is based on a general class of  $M$ -estimators of Mallow's type proposed by Cantoni and Ronchetti (2001), with the following set of estimating equations:

$$\sum_{i=1}^n \sum_{j=1}^{n-i+1} \psi(y_{ij}, \mu_{ij}) = \sum_{i=1}^n \sum_{j=1}^{n-i+1} \left[ \psi_c(r_{ij}) w(x_{ij}) \frac{1}{V^{1/2}(\mu_{ij})} \mu'_{ij} - a(\boldsymbol{\theta}) \right] = 0. \quad (4)$$

The vector  $\boldsymbol{\theta} = (c, \alpha_2, \dots, \alpha_I, \beta_2, \dots, \beta_I)$  as in equation (3) and  $\mu'_{ij}$  denotes the derivative of  $\mu_{ij}$  with respect to  $\boldsymbol{\theta}$ . The constant  $a(\boldsymbol{\theta}) = \frac{2}{n(n+1)} \sum_{i=1}^n \sum_{j=1}^{n-i+1} E[\psi_c(r_{ij})] \omega(x_{ij}) \frac{1}{V^{1/2}(\mu_{ij})} \mu'_{ij}$  ensures Fisher consistency and

$$\psi_c(r_{ij}) = \begin{cases} r_{ij} & |r_{ij}| \leq c \\ c \cdot \text{sign}(r_{ij}) & |r_{ij}| \geq c, \end{cases}$$

is the Huber function. The weight function  $w(x_{ij})$  gives less weight to observations with outlying predictors, whereas the Huber function limits the effect of observations with an outlying response by truncating the residual. Therefore, the influence function of deviations on the response and on the predictors is bounded separately. Note that for a log-link Poisson GLM one obtains  $V^{1/2}(\mu_{ij}) = \sqrt{\mu_{ij}}$ . Cantoni and Ronchetti (2001) showed that the estimating equations may equivalently be written as

$$\sum_{i=1}^n \sum_{j=1}^{n-i+1} [(\psi_c(r_{ij}) - E[\psi_c(r_{ij})]) w(x_{ij}) \sqrt{\mu_{ij}} x_{ij}] = 0 \quad (5)$$

with

$$\begin{aligned} E[\psi_c(r_{ij})] &= c [\text{P}(Y_{ij} \geq j_2 + 1) - \text{P}(Y_{ij} \leq j_1)] \\ &\quad + \sqrt{\mu_{ij}} [\text{P}(Y_{ij} = j_1) - \text{P}(Y_{ij} = j_2)] \end{aligned} \quad (6)$$

and  $j_1 = \lfloor \mu_{ij} - c\sqrt{\mu_{ij}} \rfloor$  and  $j_2 = \lfloor \mu_{ij} + c\sqrt{\mu_{ij}} \rfloor$ .

### 3.2 Illustration of the effect of outlier(s)

In this section we illustrate the classical and robust techniques on the widely studied data set of Taylor and Ashe (1983), shown in table 2. The classical and robust total reserve estimates

	1	2	3	4	5	6	7	8	9	10
1	357,848	766,940	610,542	482,940	527,326	574,398	146,342	139,950	227,229	67,948
2	352,118	884,021	933,894	1,183,289	445,745	320,996	527,804	266,172	425,046	
3	290,507	1,001,799	926,219	1,016,654	750,816	146,923	495,992	280,405		
4	310,608	1,108,250	776,189	1,562,400	272,482	352,053	206,286			
5	443,160	693,190	991,983	769,488	504,851	470,639				
6	396,132	937,085	847,498	805,037	705,960					
7	440,832	847,631	1,131,398	1,063,269						
8	359,480	1,061,648	1,443,370							
9	376,686	986,608								
10	344,014									

Table 2: claims data from Taylor and Ashe (1983).

equal respectively 18,680,856 and 18,562,327. To study the effect of an outlier on the classical and robust chain-ladder method, one observation  $y_{ij}$  is adjusted to  $\kappa y_{ij}$  for  $\kappa \in [0, 10]$ . We calculate the total reserve estimates before and after this adjustment, respectively denoted by  $\hat{R}$  and  $\hat{R}^\kappa$ . For observations  $y_{41}$ ,  $y_{43}$  and  $y_{64}$ , we plot the ratio  $\frac{\hat{R}^\kappa}{\hat{R}}$  in Figure 1. It is

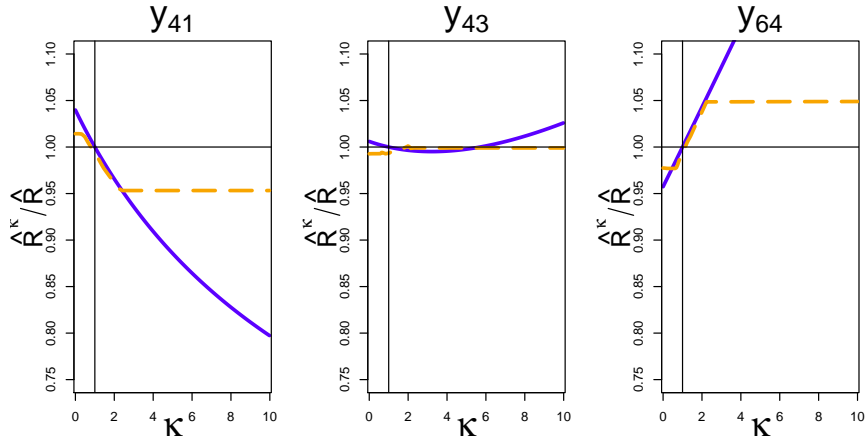


Figure 1: Influence of  $\kappa$  on the ratio between the original total reserve estimate, using  $y_{ij}$ , and the total reserve estimate corresponding to  $\kappa$ , using  $\kappa y_{ij}$  for the classical (blue solid line) and the robust (orange dashed line) chain-ladder method. Left, middle and right plot correspond respectively to  $y_{41}$ ,  $y_{43}$  and  $y_{64}$ .

immediately clear that the influence of an outlier on the robust chain-ladder method (orange dashed line) is very limited in all cases, even if  $\kappa$  becomes very large. On the other hand, the outlier has a large, unbounded effect on the classical chain-ladder method (blue solid line). Depending on the location of the outlier in the run-off triangle, the estimated reserve  $\hat{R}^\kappa$  can



largely increase or decrease. For example, an unusually large claim amount in the past ( $\kappa > 1$ ) cannot only result in a much bigger reserve estimate, but may also result in a reserve estimate that is much smaller even when no anomalies occur in the future anymore. This illustrates the sensitivity of the classical methodology and hence, there is no guarantee that the estimated classical reserve suffices to cover future claim amounts.

We now investigate the effect of an outlier on upper limits for the total reserve estimated by the classical bootstrap procedure. For the classical chain-ladder the distribution of the obtained classical bootstrap total reserve estimates is shown in orange (light color) in Figure 2. The 99.5% quantile of this bootstrap distribution corresponds to a total reserve estimate of approximately 27.8 million (indicated by a vertical line). To illustrate the effect of an outlier, we multiply observation  $y_{27}$  in Table 2 by ten. From the corresponding bootstrap distribution, which is shown in blue (dark color) in Figure 2, it is seen that now a 99.5% quantile of roughly 55 million is obtained. This clearly indicates the sensitivity of the classical chain-ladder method in combination with the classical bootstrap procedure.

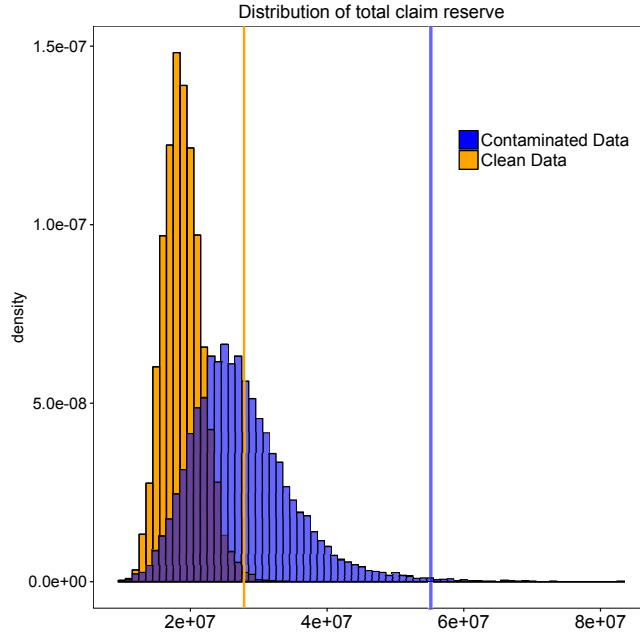


Figure 2: Distribution of bootstrap total reserve estimates based on the classical chain-ladder for the data of Taylor and Ashe (1983). total reserve estimates based on the clean data are shown in orange (light color) while total reserve estimates based on the contaminated data are shown in blue (dark color). Vertical lines indicate the corresponding 99.5% quantiles.

To limit the effect of outlying observations in the obtained bootstrap distribution, we now combine the robust chain-ladder method with the traditional bootstrap procedure and apply it to the contaminated data set of Taylor and Ashe (1983) (i.e.  $y_{27}$  is multiplied by ten). Note that for the adjusted residuals  $\mathbf{r}^P$  and  $\mathbf{r}^C$ , we need the hat matrix corresponding to the robust Poisson GLM estimator of Cantoni and Ronchetti (2001). This expression is derived in Appendix 8.1. To compare the distribution of the obtained total reserves for the

contaminated data with the previously obtained *original* reserve estimates (i.e. the reserves obtained by bootstrapping the classical chain-ladder method on the non-contaminated data), the quantile-quantile plot is given in Figure 3. As expected, a large range of the obtained quantiles are now relatively close to the corresponding original quantiles (since we plugged in the robust chain-ladder method). However, the upper quantiles of the robust chain-ladder method on contaminated data deviate significantly from their original counterparts. This is caused by the resampling with replacement in the bootstrap process. During the resampling process, it might occur that the outliers are selected so many times in a bootstrap sample that the robust method can no longer withstand this huge amount of contamination. This will happen several times if we generate 10,000 resamples, yielding unreliable bootstrap estimates that affect the bootstrap distribution, especially if one is interested in an upper quantile that lies far in the tail of the distribution.

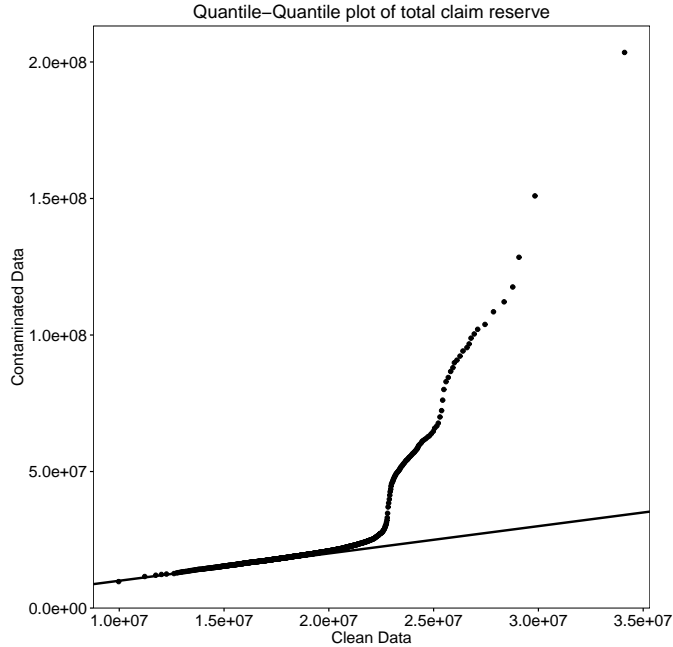


Figure 3: Quantile-quantile plot of total claim reserve distribution comparing the traditional procedure on clean data with the robust chain-ladder method combined with the classical bootstrap technique on contaminated data.

This sensitivity of the bootstrap technique, even in combination with robust point estimators, is well-known in the robust statistics literature and was first discussed by Stromberg (1997). This seminal paper gave rise to the development of several robust alternatives for the classical bootstrap. In the next section we discuss the robust bootstrap methods that have been proposed (up to our knowledge) and we adapt them to the claims reserving framework.

To end this section, we compare the outlier detection capacity of the classical and robust chain ladder on the simulated data triangle in Table 3 which was generated from the benchmark model (see section 5.1.1). In this triangle we replaced observations  $y_{16}$ ,  $y_{36}$ ,  $y_{61}$ ,  $y_{65}$  and  $y_{24}$  respectively by 33,000, 35,000, 120,000 and 65,000 and 7000. The classical and robust total

reserve estimates are 314,240 and 155,086, respectively. Hence, inspection of the data is needed to understand this difference. In practice, one often tries to detect outliers using diagnostics starting from a fit obtained by the classical estimation method. The Pearson residuals obtained from the classical chain-ladder method are given in Table 4. Besides both corner points and  $r_{51}$ , all residuals have an absolute value much larger than 3 (which is typically used as a benchmark in a classical outlier detection rule). Apparently, the classical method is so heavily affected by the outliers that the fitted model does not allow to detect the deviating observations anymore. For example, all observations in row two and six and column six have become inflated, making it impossible to flag observation  $y_{24}$  as an outlier. This is called the *masking* effect. On the other hand, observation  $y_{62}$  which is a regular observation yields the largest residual and hence appears to be the most outlying observation. This effect is known as *swamping*. To avoid these adverse effects, the goal of robust statistics is to obtain a fit which is similar to the fit we would have found without outliers in the data. An important benefit is that it then becomes possible again to detect the outliers by their large deviation from the robust fit. The robust chain-ladder method automatically provides a robustness weight for each observation, which is given by

$$\frac{\psi_c(r_{ij})}{r_{ij}}. \quad (7)$$

If a residual  $r_{ij}$  is large compared to its Huberized value  $\psi_c(r_{ij})$ , then the observation receives a low weight and is identified as a potential outlier. The robustness weights obtained by fitting the robust chain-ladder method to the data are shown in Table 5. The outlying cells are now correctly flagged by the robustness weights and can be studied in more detail.

	1	2	3	4	5	6	7	8	9	10
1	9891	7914	6092	4842	3812	3012	2370	1842	1522	1139
2	10373	8210	6624	5304	4146	3218	2486	1939	1525	
3	10890	8611	6702	5345	4119	3328	2502	2054		
4	11463	9036	7091	5596	4316	3429	2783			
5	12050	9385	7424	5828	4425	3533				
6	12675	9793	7594	6026	4623					
7	13116	10185	7928	6315						
8	13473	10652	8234							
9	14205	11101								
10	14598									

Table 3: Simulated run-off triangle.

## 4 Robust bootstrap methods

To obtain robust inference, Stromberg (1997) proposed to use robust measures of location and scale of the classical bootstrap distribution corresponding to the robust estimators. Typically, the median and the median absolute deviation (see Hampel (1974)) are applied instead of the classical mean and standard deviation. Obviously, we cannot use this method to calculate quantiles of the bootstrap distribution. Stromberg (1997) also proposed the deleted- $d$  jackknife procedure, in which a bootstrap sample consists of  $N - d$  data points instead of  $N$  data points

	1	2	3	4	5	6	7	8	9	10
1	-62.0	8.4	10.3	9.9	-61.1	97.9	-11.6	-6.5	-7.8	-0.0
2	-12.0	53.3	55.0	83.4	-27.3	-82.3	12.9	15.4	9.8	
3	-63.3	9.2	12.0	11.8	-63.7	96.7	-13.3	-5.6		
4	-3.7	63.8	61.5	54.9	-26.3	-81.6	18.8			
5	-0.6	66.7	65.2	57.6	-26.4	-82.2				
6	120.9	-129.2	-109.2	-96.5	89.1					
7	-46.3	28.1	29.1	27.0						
8	-41.3	35.6	35.0							
9	-29.3	46.7								
10	0.0									

Table 4: Pearson residuals obtained after fitting a Poisson GLM to the data in Table 3.

	1	2	3	4	5	6	7	8	9	10
1	1.00	1.00	1.00	1.00	1.00	0.00	1.00	1.00	1.00	1.00
2	0.70	1.00	0.81	0.05	0.73	1.00	1.00	1.00	1.00	
3	1.00	1.00	1.00	1.00	1.00	0.00	0.90	1.00		
4	1.00	1.00	1.00	1.00	1.00	1.00	1.00			
5	1.00	1.00	1.00	1.00	1.00	0.75	0.76			
6	0.00	1.00	1.00	1.00	0.00					
7	1.00	1.00	1.00	1.00						
8	1.00	1.00	1.00							
9	1.00	1.00								
10	1.00									

Table 5: Robust weights obtained after fitting a robust Poisson GLM to the data in Table 3.

to increase the probability of sampling a clean subset. However, in our framework this poses problems as the resampled triangle would then be incomplete.

On the other hand, Singh (1998) proposed to robustify the bootstrap procedure itself by resampling from a winsorized version of the original sample. In a regression context, one may resample either the residuals or the original observations themselves. Due to the specific structure of a claims triangle, we resample the residuals. To winsorize the residuals the quantiles  $Q_c$  and  $Q_{1-c}$  of the residuals are determined for  $c \in ]0, 0.5[$ . The  $c$  percent smallest residuals are then replaced by  $Q_c$  whereas the  $c$  percent largest ones are replaced by  $Q_{1-c}$ .

More recently, Salibian-Barrera and Zamar (2002) and Salibian-Barrera et al. (2008) proposed a fast and robust bootstrap method that is asymptotically consistent under reasonable regularity conditions. The corresponding R-package (Van Aelst and Willems (2013)) considers three multivariate settings: principal component analysis, Hotelling tests and multivariate regression. Another recent alternative is proposed by Amado and Pires (2004) and Amado et al. (2014). They suggest to resample observations with different probabilities such that potentially harmful observations have low sampling probability. They propose to determine these probabilities based on the influence function of the estimator. We now discuss both state-of-the-art robust bootstrap methods in more detail and describe their implementation in the claims reserving framework.

## 4.1 Fast and Robust Bootstrap

For each pseudo-history, most robust bootstrap procedures need to recalculate the robust estimator, which in itself is already computationally intensive compared to the classical estimator. An important aim of the fast and robust bootstrap (FRB) is to reduce this computational load (Salibian-Barrera and Zamar, 2002; Salibian-Barrera et al., 2008).

Within the context of claims reserving the FRB procedure can be described as follows. Let  $\boldsymbol{\theta} \in \mathbb{R}^{2n-1}$  be the parameter of interest. Note that  $\boldsymbol{\theta}$  consists of the intercept  $c$  and the parameters  $\alpha_i$  and  $\beta_j$  for  $2 \leq i, j \leq n$ , as before. Now, suppose that  $\hat{\boldsymbol{\theta}}$  is a solution of a fixed-point equation  $\hat{\boldsymbol{\theta}} = g_N(\hat{\boldsymbol{\theta}})$  where  $g_N$  depends on the sample of size  $N = n(n+1)/2$  composed of the historical claims data. Similarly, for a resampled pseudo-history the corresponding estimator  $\hat{\boldsymbol{\theta}}^{(b)}$  solves the equation  $\hat{\boldsymbol{\theta}}^{(b)} = g_N^{(b)}(\hat{\boldsymbol{\theta}}^{(b)})$  where the function  $g_N^{(b)}$  now depends on the bootstrap sample. As a first step one may consider  $\tilde{\boldsymbol{\theta}}^{(b)} = g_N^{(b)}(\hat{\boldsymbol{\theta}})$  as an approximation for the bootstrap replicate  $\hat{\boldsymbol{\theta}}^{(b)}$  of the estimator  $\hat{\boldsymbol{\theta}}$ . This simple procedure avoids having to recompute the robust estimator for each bootstrap sample, which may be computationally hard. However, these approximate bootstrap replicates typically underestimate the true variability since all bootstrap approximations  $\tilde{\boldsymbol{\theta}}^{(b)}$  are based on the same initial estimate  $\hat{\boldsymbol{\theta}}$ . Therefore, Salibian-Barrera et al. (2008) proposed a linear correction resulting in the FRB bootstrap replicates

$$\hat{\boldsymbol{\theta}}_{\text{FRB}}^{(b)} = \hat{\boldsymbol{\theta}} + \left[ I_p - \nabla g_N(\hat{\boldsymbol{\theta}}) \right]^{-1} \cdot \left( g_N^{(b)}(\hat{\boldsymbol{\theta}}) - \hat{\boldsymbol{\theta}} \right),$$

where  $\nabla g_N(\hat{\boldsymbol{\theta}})$  corresponds to the gradient of  $g_N$  evaluated at  $\hat{\boldsymbol{\theta}}$  and  $I_p$  denotes the identity matrix of dimension  $p = 2n - 1$ .

This procedure was later extended by Camponovo et al. (2012) to accommodate  $M$ -estimators defined by the solution of equations of the following type

$$\psi_N(\hat{\boldsymbol{\theta}}) = \sum_{k=1}^N \psi(\hat{\boldsymbol{\theta}}) = 0.$$

Fast and robust bootstrap estimates are then obtained as follows:

$$\hat{\boldsymbol{\theta}}_{\text{FRB}}^{(b)} = \hat{\boldsymbol{\theta}} - \left[ \nabla \psi_N(\hat{\boldsymbol{\theta}}) \right]^{-1} \cdot \psi_N^{(b)}(\hat{\boldsymbol{\theta}}).$$

The FRB paradigm can be used for the Poisson GLM estimator of Cantoni and Ronchetti (2001). Based on the estimating equations (5) we define

$$\psi_N(\hat{\boldsymbol{\theta}}) = \sum_{i=1}^n \sum_{j=1}^{n-i+1} \left[ (\psi_c(r_{ij}) - E[\psi_c(r_{ij})]) w(x_{ij}) \sqrt{\mu_{ij}} x_{ij} \right]. \quad (8)$$

To calculate  $\nabla \psi_N(\hat{\boldsymbol{\theta}})$  we readily obtain

$$\frac{\partial}{\partial \boldsymbol{\theta}} (\sqrt{\mu_{ij}}) = \frac{1}{2} \mu_{ij}^{1/2} x_{ij} \quad \text{and} \quad \frac{\partial}{\partial \boldsymbol{\theta}} (\psi_c(r_{ij})) = -\frac{1}{2} I_{|r_{ij}| \leq c} \left( \frac{y_{ij} + \mu_{ij}}{\mu_{ij}^{1/2}} \right) x_{ij},$$

with  $I$  the indicator function. It then follows that

$$\begin{aligned} \nabla \psi_N(\hat{\theta}) = \sum_{i=1}^n \sum_{j=1}^{n-i+1} w(x_{ij}) x_{ij} & \left[ \left( -\frac{1}{2} I_{|r_{ij}| \leq c} \left( \frac{y_{ij} + \mu_{ij}}{\mu_{ij}^{1/2}} \right) x_{ij} - \frac{\partial}{\partial \theta} (E[\psi_c(r_{ij})]) \right) \sqrt{\mu_{ij}} \right. \\ & \left. + (\psi_c(r_{ij}) - E[\psi_c(r_{ij})]) \frac{1}{2} \mu_{ij}^{1/2} x_{ij} \right]^t. \end{aligned}$$

The term  $E[\psi_c(r_{ij})]$  however is not differentiable when using its exact expression. Therefore, we use the  $\chi^2$  distribution as an approximation for the Poisson distribution and afterwards apply the Wilson-Hilferty approximation. The detailed derivation can be found in Appendix 8.3.

## 4.2 Influence Function Bootstrap

Amado and Pires (2004) propose to determine resampling probabilities for the observations by using the influence function corresponding to the classical estimator. The definition of the influence function is based on the concept of a statistical functional, which we recall first. Let  $T_N$  denote a  $p$ -dimensional estimator defined for any sample size  $N$ , with  $N = (n+1)n/2$  and  $p = 2n - 1$  in our setting. A statistical functional  $T$  corresponding to the estimator  $T_N$  is a map defined on  $p$ -variate distributions  $G$  on  $\mathbb{R}^p$  such that  $T(G_N(\{y_{ij} | 1 \leq i, j \leq n; i+j \leq n+1\})) = T_N(\{y_{ij} | 1 \leq i, j \leq n; i+j \leq n+1\})$  where  $G_N$  is the empirical distribution of the sample  $\{y_{ij} | 1 \leq i, j \leq n; i+j \leq n+1\}$ .

Let  $\{Y_{ij} | 1 \leq i, j \leq n; i+j \leq n+1\} \sim F$  and assume that the marginal distribution of  $Y_{ij}$ , denoted by  $F_{\mu_{ij}} = F_{i,j}$ , is Poisson distributed with mean  $\mu_{ij}$ , see equation (3). For any  $1 \leq l, m \leq n$  with  $l+m \leq n+1$  and  $z \geq 0$ , we define  $F_{l,m,\varepsilon,z}$  such that

$$\begin{cases} Y_{ij} \sim F_{i,j} & \forall (i,j) \neq (l,m) \\ Y_{l,m} \sim (1-\varepsilon)F_{l,m} + \varepsilon\Delta_z \end{cases}$$

where  $\Delta_z$  is a Dirac measure putting all its mass at value  $z$ .

The influence function of the functional  $T$  at a distribution  $F$  is then defined as

$$\text{IF}([z, l, m], T, F) = \lim_{\varepsilon \downarrow 0} \frac{T(F_{l,m,\varepsilon,z}) - T(F)}{\varepsilon}.$$

Intuitively, the influence function measures the change of the estimator when the model distribution is perturbed by an infinitesimal small amount of contamination at location  $z$ .

Amado and Pires (2004) propose to measure the effect of each data point on the parameter estimates by estimating the value of the population influence function at each observation using estimates of the model parameters based on the sample. However, if a non-robust estimator is employed then an outlier may already have such a high effect on the sample estimates themselves that it results in a low value for the influence function evaluated at the outlier. This phenomenon is known as masking in the robust statistics literature. Conversely, the influence function of a robust estimator will also have small values at the outliers, since outliers should

have a low effect on robust estimators (Pison and Van Aelst, 2004). Therefore, Amado and Pires (2004) consider the influence function of the classical non-robust estimator (denoted by  $T_{\boldsymbol{\theta}}^{nr}$ ) using parameter estimates  $\hat{\boldsymbol{\theta}}^r$  obtained by a robust estimator, i.e. the influence function is evaluated at  $F_{\hat{\boldsymbol{\theta}}^r}$ . By combining robust parameter estimates with the influence function of the non-robust estimator, its values for outliers should be very high compared to regular observations. This idea is formalized by the Robust Empirical Standardized Influence Function (RESIF) that is defined as:

$$\text{RESIF}([z, l, m], T_{\boldsymbol{\theta}}^{nr}, F_{\hat{\boldsymbol{\theta}}^r}) = \left[ \text{IF}([z, l, m], T_{\boldsymbol{\theta}}^{nr}, F_{\hat{\boldsymbol{\theta}}^r})^t \cdot V^{-1}(T_{\boldsymbol{\theta}}^{nr}, F_{\hat{\boldsymbol{\theta}}^r}) \cdot \text{IF}([z, l, m], T_{\boldsymbol{\theta}}^{nr}, F_{\hat{\boldsymbol{\theta}}^r}) \right]^{\frac{1}{2}} \quad (9)$$

where

$$V(T_{\boldsymbol{\theta}}^{nr}, F_{\hat{\boldsymbol{\theta}}^r}) = E_{F_{\hat{\boldsymbol{\theta}}^r}} \left[ \text{IF}([z, l, m], T_{\boldsymbol{\theta}}^{nr}, F_{\hat{\boldsymbol{\theta}}^r}) \cdot \text{IF}([z, l, m], T_{\boldsymbol{\theta}}^{nr}, F_{\hat{\boldsymbol{\theta}}^r})^t \right].$$

Based on the RESIF function the bootstrap procedure is adjusted. We now discuss this procedure in more detail in the context of claims reserving.

1. Calculate the values:

$$\text{RESIF}_{ij} = \text{RESIF}([y_{ij}, i, j], T_{\boldsymbol{\theta}}^{nr}, F_{\hat{\boldsymbol{\theta}}^r})$$

for each of the observations in the sample  $\{y_{ij} | 1 \leq i, j \leq n; i + j \leq n + 1\}$ .

2. Choose constants  $d, \gamma$  and  $c$  and calculate for each observation  $y_{ij}$  a weight  $w_{ij}$  as follows:

$$w_{ij} = I_{[0, c]}(|\text{RESIF}_{ij}|) + I_{]c, \infty[}(|\text{RESIF}_{ij}|) \cdot \eta_{d, \gamma}(c, |\text{RESIF}_{ij}|)$$

where

$$\eta_{d, \gamma}(c, x) = \left[ 1 + \frac{(x - c)^2}{\gamma d^2} \right]^{-\frac{\gamma+1}{2}} \quad (10)$$

and  $I_{[0, c]}$  indicates the indicator function on the interval  $[0, c]$ .

3. Compute the resampling probability

$$p_{ij} = \frac{w_{ij}}{\sum_{k=1}^n \sum_{\ell=1}^{n-i+1} w_{k\ell}} \quad 1 \leq i, j \leq n \text{ and } i + j \leq n + 1$$

for each observation.

The bootstrap procedure is then initiated by resampling the residuals  $\mathbf{r}^*$  using these resampling probabilities. Note that the residuals of the two corner points are still not resampled.

The RESIF for the Poisson GLM estimator of Cantoni and Ronchetti (2001) is derived in Appendix 8.4 and the result is given in Theorem 1.

**Theorem 1.** *The Robust Empirical Standardized Influence Function of the M-estimator defined by equation (8) for the Poisson GLM is given by*

$$\text{RESIF}([z, l, m], T_{\boldsymbol{\theta}}^{nr}, F_{\hat{\boldsymbol{\theta}}^r}) = \left[ \frac{(z - \hat{\mu}_{lm})}{\hat{\mu}_{lm}} \right]^{1/2} \quad 1 \leq l, m \leq n \text{ and } l + m \leq n + 1.$$

Amado and Pires (2004) propose to apply the classical GLM estimator to calculate the parameter estimates corresponding to the bootstrap samples. This drastically reduces the computation time that is needed for the bootstrap. The robust initial estimate combined with the weights based on the influence function protect against outliers in the bootstrap samples. The resulting procedure is both fast and robust. Note that the consistency of this methodology has not been formally proven.

## 5 Simulation Study

To compare the different bootstrap procedures, their performance is evaluated in an extensive simulation study. Three different models to generate complete run-off triangles (lower and upper part) are considered. We start with a simple benchmark model already allowing us to study several properties of the various bootstrap methods. Then, we turn our attention to more realistic simulation models from the claims reserving literature. To simulate run-off triangles in the claims reserving framework we focus on the models proposed by Schiegl (2004) and Cowell (2009). Before discussing the simulation setup and results, we briefly describe the three data generating models and their corresponding parameters. Values for these parameters were chosen to obtain realistically looking data sets (based on the proposed values in Schiegl (2004) and Cowell (2009)). Note that other constants than below were also studied, but they led to similar results and conclusions and hence were omitted.

### 5.1 Data generating models

#### 5.1.1 Benchmark model

The benchmark model is a simplification of the model proposed by Schiegl (2004). The entries in cell  $i, j$  ( $i = 1, \dots, 10$  and  $j = 1, \dots, 10$ ) are simulated from a Poisson distribution with parameter  $\lambda_{ij}$  defined by

$$\lambda_{ij} = \lambda_0(1 + (i - 1)\eta_2) \exp\left(\frac{2(j - 1)}{n} \log(\eta_1)\right), \quad (11)$$

where  $\lambda_0 = 10,000$  gives the expected value of claim amounts at the first accident and development year,  $\eta_1 = 0.3$  indicates the reduction of claim amounts at development period  $\frac{n}{2}$  compared to the claim amount in the first development year and  $\eta_2 = 0.05$  is the rate of increase for the claim amounts over consecutive accident years.

#### 5.1.2 Schiegl method

This method is proposed in Schiegl (2004) and the simulation of the run-off triangles (containing 10 accident and development years) is based on the collective model, where a different distribution is used for the number of claims and for a single claim amount. The number of claims  $N_{ij}$  (for  $i = 1, \dots, 10$  and  $j = 1, \dots, 10$ ) follows a Poisson distribution with mean



$\lambda_{ij}$  as defined by equation (11) of the benchmark model. Individual claim amounts are then generated using the Gamma distribution with mean  $r = 2,000$  such that the final data are obtained as follows

$$Y_{ij} \sim \sum_{k=1}^{N_{ij}} Z_k \text{ with } N_{ij} \sim \text{Poisson}(\lambda_{ij}) \text{ and } Z_k \sim \text{Gamma}(r, 1).$$

### 5.1.3 Kaishev method

Cowell (2009) noted that, given the same set of parameter values, the distribution of reserves in the Schiegl method is the same regardless of the sample drawn for the upper triangle (since each  $C_{ij}$  value is sampled independently from the others). Therefore, Cowell (2009) presents a slightly more complex model (proposed by Kaishev through personal communication) so that, even if all parameters are kept fixed, different simulations will lead to different distributions of reserves.

As before, the method is based on the collective model and we again generate individual claim amounts using a Gamma distribution with mean  $r = 2,000$ . The total number of claims  $N_i$  for accident year  $i = 1, \dots, 10$  is generated from a Poisson distribution with mean  $\lambda_i$ . Each claim (for  $k = 1, \dots, N_i$ ) follows a Gamma distribution (yielding a certain claim amount  $Z_k$ ) and this amount is then assigned randomly to a certain development year  $j = 1, \dots, 10$ . This random allocation is done using a Beta( $\alpha = 1, \beta = 1.6$ ) distribution (as suggested in Cowell (2009)). Similar as in the Schiegl method we defined  $\lambda_i$  as

$$\lambda_i = \lambda_0(1 + (i - 1)\eta_2)$$

where  $\lambda_0$  and  $\eta_2$  were also chosen as before.

Note that plugging in a Pareto or a Lognormal distribution for the individual claim size (instead of a Gamma distribution) in the Schiegl or Kaishev model yielded similar conclusions.

## 5.2 Simulation setup

During the simulation study we consider the following bootstrap procedures:

- TCL: standard approach using the classical chain-ladder method and standard bootstrap
- CB: robust chain-ladder method combined with the classical bootstrap
- WB: winsorized bootstrap
- FRB: fast and robust bootstrap procedure
- IFB: influence function bootstrap

Note that for IFB, the weight function to determine the resampling probability of each residual contains three tuning parameters, see equation (10). These values need to be chosen to guarantee a sufficiently high efficiency at the uncontaminated case whilst offering robustness when contamination is present. Taking into account this tradeoff between efficiency and robustness, the parameter  $c$  is calculated as the 90% quantile of the  $\text{RESIF}_{ij}$  values, whereas the parameters  $d$  and  $\gamma$  were set to 30 and 10 respectively. For the winsorized bootstrap the 10% highest and 10% smallest residuals are winsorized.

The bootstrap procedures are tested for both clean and contaminated data. Two contamination settings are shown. Firstly, contamination was generated by multiplying observations  $Y_{23}$  and  $Y_{32}$  in the original clean data by a factor of five. Secondly, observations  $Y_{23}$  and  $Y_{41}$  in the original clean data are multiplied by two and a half. We also studied other locations and multiplication factors to create outliers, but these results led to similar conclusions. The corresponding tables are not included in the paper, since they do not add new insights in the proposed methodology.

## 5.3 Simulation results

### 5.3.1 Benchmark model

For each bootstrap procedure, we investigated the quantiles obtained from 10,000 bootstrap runs. Since we simulate complete data using the benchmark model, the true total reserve can be calculated from the lower triangle according to equation (1). Hence, for 10,000 generated data sets we recorded whether its true total reserve is smaller than the calculated bootstrap quantile, for different probability levels. These coverage percentages are shown in Table 6 for different bootstrap procedures and confidence levels. If the bootstrap procedure is able to approximate the  $(1 - \alpha)$ -quantiles with  $\alpha \in (0, 0.5)$  well, one may expect the true total reserve to be smaller than the bootstrap quantile in  $(1 - \alpha)\%$  of the cases.

For the uncontaminated data, it is clear that the WB procedure leads to significantly lower coverage percentages than the other methods. The TCL, CB and IFB methodologies give comparable results whereas the FRB procedure yields slightly lower coverage probabilities (but still much better than the WB method).

To study the effect of outliers, we repeated the simulation study using the first (i.e. multiplying  $Y_{23}$  and  $Y_{32}$  by five) and second (i.e. multiplying  $Y_{23}$  and  $Y_{41}$  by two and a half) contamination setting. In both situations, the FRB and IFB clearly outperform the other bootstrap procedures. We immediately see that the TCL and CB procedures are highly influenced by the outliers as almost all the coverage percentages equal 100% (this is because the obtained bootstrap quantiles became very large due to the outliers), whereas the WB bootstrap quantiles are too small leading to the smallest coverage percentages. It can be seen that in contamination scenario two the estimates of the 95% and 99.5% quantiles obtained by the IFB bootstrap procedure are still affected by the outliers. In this setting the outliers are closer to the real data and therefore it is more difficult to detect them.

To study the effect of the residual adjustment on the coverage percentages, Table 7 shows

		75%	90%	95%	99.5%
Clean	TCL	71.50	86.60	92.00	98.50
	CB	72.60	88.00	91.80	98.70
	WB	66.50	78.60	84.80	93.60
	FRB	69.60	83.50	89.20	96.70
	IFB	72.00	86.80	92.40	98.70
Cont 1	TCL	100.00	100.00	100.00	100.00
	CB	87.00	100.00	100.00	100.00
	WB	65.50	79.30	83.80	94.40
	FRB	70.00	85.70	91.30	97.90
	IFB	63.10	80.80	88.40	97.60
Cont 2	TCL	100.00	100.00	100.00	100.00
	CB	92.00	100.00	100.00	100.00
	WB	74.90	85.70	90.00	96.60
	FRB	77.40	89.40	93.50	98.50
	IFB	76.20	91.80	98.30	100.00

Table 6: Coverage percentages for the benchmark model using the Cordeiro adjusted residuals  $\mathbf{r}^C$  in the bootstrap process.

		75%	90%	95%	99.5%
Clean	TCL	67.80	81.30	87.50	96.00
	CB	69.00	82.50	88.60	96.80
	WB	63.00	73.90	79.00	89.40
	FRB	67.70	80.70	86.80	95.20
	IFB	67.90	81.50	87.60	96.40
Cont 1	TCL	100.00	100.00	100.00	100.00
	CB	81.40	100.00	100.00	100.00
	WB	63.20	73.30	79.90	89.50
	FRB	68.30	83.20	89.00	96.90
	IFB	60.80	76.00	83.20	94.70
Cont 2	TCL	100.00	100.00	100.00	100.00
	CB	86.40	100.00	100.00	100.00
	WB	71.90	80.70	85.30	93.20
	FRB	75.90	87.50	91.60	97.70
	IFB	75.50	93.60	100.00	100.00

Table 7: Coverage percentages for the benchmark model using the original Pearson residuals  $\mathbf{r}$  in the bootstrap process.

the percentages obtained by using the original Pearson residuals. These coverage percentages are lower and further away from the target than those obtained with the adjusted Pearson residuals. The coverage percentages based on the adjusted residuals using the England and Verrall (1999) adjustment factor ( $\mathbf{r}^E$ ) and the correction proposed by Pinheiro et al. (2003) ( $\mathbf{r}^P$ ) were comparable to the results found in Table 6 (using  $\mathbf{r}^C$ ) and are therefore omitted.

To compare the speed of the different bootstrap procedures, the average computation time for a bootstrap run consisting of 10,000 bootstrap samples for a run-off triangle, simulated using the benchmark model, is shown in Table 8. These computations were performed on an Intel Ivy Bridge Xeon E5-2680V2 2,8 GHz CPU and averaged across 100 runs.

	TCL	CB	WB	IFB	FRB
mean	0.98	6.75	19.47	1.03	0.016
st. dev.	0.06	2.90	14.44	0.05	0.002

Table 8: Computation times expressed in minutes.

The FRB method clearly outperforms the other alternatives in terms of computational efficiency. Given the high time complexity and bad performance of the CB and WB methods in the (contaminated) benchmark model, we discard these procedures in further studies. The other procedures are now compared using the more advanced simulation models.

### 5.3.2 Schiegl and Kaishev model

We have applied the remaining bootstrap techniques on clean and contaminated data sets simulated by the Schiegl and Kaishev models. Boxplots of the obtained 99.5% quantiles based on 10,000 bootstrapped estimates for the total reserve are shown in Figures 4 and 5. The boxes in the boxplots are white for the clean data, orange for the first contamination setting (i.e. multiplying  $y_{23}$  and  $y_{32}$  by 5) and red for the second contamination setting (i.e. multiplying  $y_{23}$  and  $y_{41}$  by 2.5).

It is immediately clear from Figures 4 and 5 that TCL breaks down as soon as the data is contaminated, whereas the other methods give results that are almost the same as those obtained by the classical method on the initial clean data without outliers. The presence of the outliers thus only has a minor effect on these methods. In order to better compare the robustness properties of IFB and FRB, we have removed the results for TCL on contaminated data in Figures 6 and 7. When no outliers are included, the boxplots for FRB are nearly identical to the ones obtained with TCL on the clean data, whereas the results for IFB are clearly lower. These differences are even more obvious when looking at the contaminated settings: the boxplots for FRB hardly change when outliers are added, whereas the boxplots for IFB are getting further away from the target (i.e. the results obtained with TCL on the clean data).

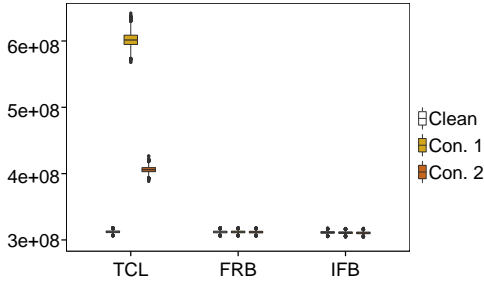


Figure 4: Boxplots of the obtained 99.5% quantiles of the 10,000 bootstrapped estimates using Schiegl model.

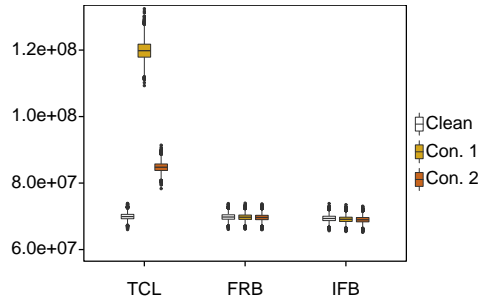


Figure 5: Boxplots of the obtained 99.5% quantiles of the 10,000 bootstrapped estimates using Kaishev model.

The same conclusions are obtained from the results in Tables 9 and 10, where the 10,000 bootstrapped estimates for the total reserve are compared with the true reserve (obtained by summing up the lower triangle of the simulated data set). The percentages in the tables again represent in how many of the 10,000 runs the true reserve was covered by the corresponding confidence interval for different confidence levels. For the uncontaminated data the coverage percentages for TCL are fairly close to their expected values with FRB percentages being only slightly smaller. The IFB method achieves smaller coverage percentages. When outliers are added, TCL clearly breaks down: the bootstrap quantiles become very large, which explains the high coverage rates in almost all situations. However, for data simulated from the Schiegl model and the second contamination setting, the 75% bootstrap quantiles never succeed in covering the real future reserve. Similar observations can be made for the Kaishev model. On the other hand, the coverage percentages for the robust methods always remain close to the

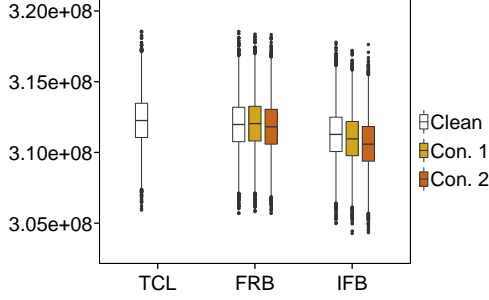


Figure 6: Detailed version of Figure 4 (discarding large values on  $y$ -axis).

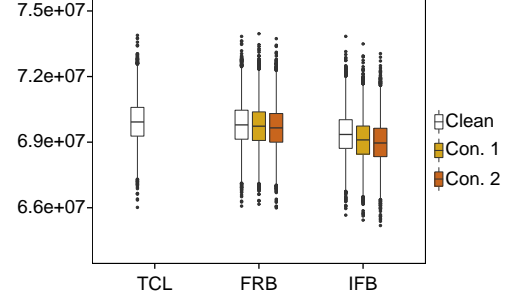


Figure 7: Detailed version of Figure 5 (discarding large values on  $y$ -axis).

target percentages. The results for FRB are again more stable than for IFB.

		75%	90%	95%	99.5%
Clean	TCL	71.84	87.09	92.66	98.62
	FRB	71.02	85.75	91.34	97.90
	IFB	67.13	80.50	86.67	95.51
Cont 1	TCL	100.00	100.00	100.00	100.00
	FRB	65.37	83.18	90.31	98.14
	IFB	55.43	72.31	80.89	93.79
Cont 2	TCL	0.00	100.00	100.00	100.00
	FRB	59.23	79.12	87.47	97.35
	IFB	47.68	65.64	74.91	90.89

Table 9: Coverage results for Schiegl model.

		75%	90%	95%	99.5%
Clean	TCL	72.95	88.02	93.01	98.75
	FRB	72.64	87.06	91.84	98.21
	IFB	67.52	80.86	86.93	95.23
Cont 1	TCL	2.99	100.00	100.00	100.00
	FRB	62.71	81.98	89.41	97.97
	IFB	52.06	69.43	78.40	92.50
Cont 2	TCL	0.01	97.59	100.00	100.00
	FRB	58.70	79.41	87.80	97.55
	IFB	46.67	65.20	74.63	90.77

Table 10: Coverage results for Kaishev model.

### 5.3.3 Conclusion

From this simulation study we can conclude that TCL and CB are very sensitive to outliers and that WB consistently underestimates the total reserve quantile. The IFB and FRB procedures yield efficient and robust results and FRB is generally less influenced by outliers than IFB. Taking computational speed into account as well, we advice to combine the robust chain-ladder method with the FRB approach for prediction purposes in the claims reserving framework.

## 6 Real data application

We now apply our optimal robust methodology (i.e. robust chain-ladder method combined with the FRB approach) to a run-off triangle from practice, which is shown in Table 11.

This real data set is extracted from the schedule P - Analysis of Losses and Loss Expenses in

	1	2	3	4	5	6	7	8	9	10
1988	794	569	497	409	221	55	62	9	0	0
1989	696	499	429	437	227	154	10	57	0	0
1990	799	690	391	529	326	34	4	37	45	0
1991	796	405	234	287	131	263	438	0	1	4
1992	742	498	368	199	346	49	1	0	0	1
1993	637	518	140	231	79	34	19	89	5	0
1994	703	427	364	194	98	40	15	6	61	3
1995	580	375	227	132	121	18	10	52	45	6
1996	563	266	106	203	71	30	23	4	3	19
1997	324	193	173	30	36	49	44	19	3	0

Table 11: Real run-off triangle from Schedule P data.

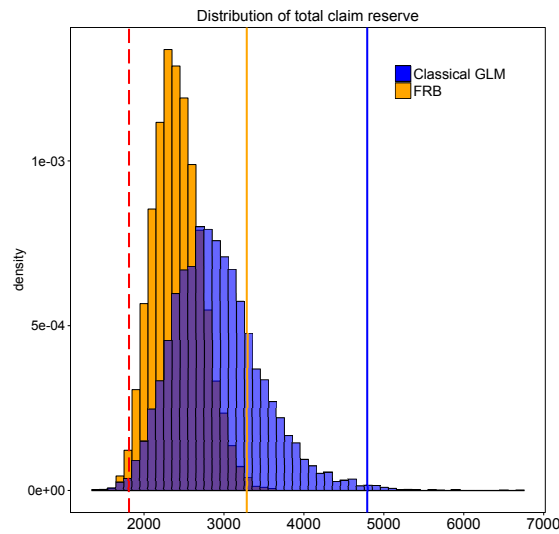


Figure 8: Bootstrapped overall reserve estimate for the Rockford Mutual Insurance company.

the National Association of Insurance Commissioners (NAIC) database and is available on the website of the Casualty Actuarial Society <sup>2</sup>. This database contains data on several lines of business for a variety of businesses in the United States and here we focus on the *other liability line of business* for the Rockford Mutual Insurance Company. The true total reserve is 1812 (the sum of the observations in the lower triangle of Table 11).

We recommend to always apply both the classical and robust procedures to a given dataset and compare their results. For this example, the classical chain-ladder method leads to a total reserve estimate of 2837 whereas the robust alternative leads to a total claim reserve estimate of 2304, which is a difference of roughly 20%. The obtained bootstrap 99.5% quantile for the total reserve is 4781 for the classical methodology and 3285 for the robust procedure based on FRB. This clear discrepancy may indicate the presence of possible outlier(s) in the data. The bootstrap distributions are shown in figure 8. The true total reserve is marked with a red dashed line.

<sup>2</sup>[http://www.casact.org/research/index.cfm?fa=loss`reserves`data](http://www.casact.org/research/index.cfm?fa=loss%27reserves%27data)

An advantage of the robust chain-ladder method is that the most influential points can easily be detected using the weights of the robust chain-ladder method defined in equation (7). From Table 12 it is clear that the sixth and seventh development year of accident year 1991 have a very low weight and should be examined in more detail. If these observations are actually errors, then the robust estimate is a reliable total reserve estimate, otherwise the total reserve estimate should be adjusted accordingly (taking into account the possibility of observing such atypical claims in future).

	1	2	3	4	5	6	7	8	9	10
1988	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1989	1.00	1.00	1.00	1.00	1.00	0.63	1.00	1.00	1.00	
1990	1.00	1.00	1.00	0.93	1.00	1.00	1.00	1.00		
1991	1.00	1.00	1.00	1.00	1.00	0.22	0.07			
1992	1.00	1.00	1.00	0.83	0.57	1.00				
1993	1.00	0.80	0.76	1.00	0.91					
1994	1.00	1.00	1.00	1.00						
1995	1.00	1.00	1.00							
1996	1.00	1.00								
1997	1.00									

Table 12: Weights from the robust chain-ladder method for the Schedule P data.

	1	2	3	4	5	6	7	8	9	10
1988	-0.94	0.70	6.35	2.76	-1.06	-5.44	-5.82	-4.28	-0.02	0.00
1989	-3.25	-1.34	3.71	5.13	-0.04	4.40	-10.17	4.28	0.02	
1990	-2.78	3.96	-0.54	7.45	4.45	-7.92	-11.37	0.09		
1991	-0.56	-6.05	-7.07	-3.51	-6.77	14.33	27.72			
1992	0.20	0.10	1.77	-6.77	9.12	-5.19				
1993	3.13	7.16	-7.27	-0.87	-6.56					
1994	1.73	-0.77	3.69	-5.45						
1995	1.05	0.07	-1.64							
1996	3.03	-3.70								
1997	-0.00									

Table 13: Pearson residuals obtained from the classical chain-ladder algorithm for the Schedule P data.

For comparison, the Pearson residuals obtained using the classical chain-ladder method are shown in Table 13. Development year six and seven of accident year 1991 have the largest residuals, but other observations also have large residuals (e.g. above 5 or even above 10 in absolute value). It is thus not obvious to determine how many observations have an unusually large residual and it is also not clear what adjustments should be made to these observations or how they should be treated in the bootstrap process. On the other hand, our robust techniques yield an automatic, data driven and objective methodology to treat and detect possible outliers.

Under the new Solvency II regulations insurers are also compelled to investigate one year risk measures (Dreksler et al., 2015). Wüthrich et al. (2009) proposed to use the Claims Development Result (CDR) for this purpose. Let  $D_k = \{Y_{ij} \mid i + j \leq k + 1\}$  denote the upper triangle of stochastic variables up to calendar year  $k$  with  $n \leq k \leq 2n - 1$ . Note that for  $k = n$  this corresponds to the past claims data and for  $k = 2n - 1$  the full square of data is available. The true CDR for claims in accident year  $i \in \{1, \dots, n\}$  during the next calendar year  $(k, k + 1]$  is defined as

$$\text{CDR}_i(k + 1) = E[\text{TR}_i \mid D_k] - (Y_{i,k-i+1} + E[\text{TR}_i \mid D_{k+1}]),$$

where  $TR_i = \sum_{j=k-i+1}^n Y_{ij}$  denotes the total claims amount for accident year  $i$ . The aggregated CDR is then defined as

$$CDR(k+1) = \sum_{i=1}^n CDR_i(k+1).$$

To estimate the CDR the stochastic quantities are replaced by their observations and estimates respectively. Several authors already considered this problem including Diers (2008) and Ohlsson and Lauzenings (2009). Lacoume (2008), Boisseau (2006) and Boumezoued et al. (2012) complement these papers and propose to add the following steps to the bootstrap procedure (explained in Section 2.2) in every run  $b = 1, \dots, B$ :

4. Denote the fitted values obtained from the GLM in step 3 of the bootstrap loop as  $\hat{Y}_{ij}^{(b)}$ , see Table 14(b). The bootstrap reserve  $R^{(b)}$  then corresponds to  $\sum_{i=2}^n \sum_{j=n-i+2}^n \hat{Y}_{ij}^{(b)}$ . An extended bootstrap history is constructed by adding the first fitted anti-diagonal  $\hat{Y}_{ij}^{(b)}$  with  $i+j = n+2$  to the bootstrapped historic observations. Based on this extended bootstrap history the GLM is refitted leading to the fitted values  $\tilde{Y}_{ij}^{(b)}$ , see Table 14(c). This leads to a new total future reserve estimate  $R_{ext}^{(b)} = \sum_{i=3}^n \sum_{j=n-i+3}^n \tilde{Y}_{ij}^{(b)}$ .
5. A bootstrapped value for the CDR may then be obtained by computing :

$$CDR^{(b)}(k+1) = R^{(b)} - R_{ext}^{(b)} = \sum_{i=2}^n \sum_{j=n-i+2}^n \hat{Y}_{ij}^{(b)} - \sum_{i=3}^n \sum_{j=n-i+3}^n \tilde{Y}_{ij}^{(b)}.$$

The bootstrapped CDR is shown in Figure 9. The 99.5% quantile corresponds to 1,677 for the classical procedure and 1,160 for the robust procedure. As the Value at Risk under the Solvency II regulations is directly related to this quantile, this example again shows the impact of outliers on standard claims reserving results and the good performance of the proposed robust alternative.

	1	2	...	J
1	$Y_{11}^{(b)}$	$Y_{12}^{(b)}$	...	$Y_{1J}^{(b)}$
2	$Y_{21}^{(b)}$	$Y_{22}^{(b)}$	...	
...	...	...		
I	$Y_{I1}^{(b)}$			

(a) Obtain a bootstrap triangle

	1	2	...	J
1	$Y_{11}^{(b)}$	$Y_{12}^{(b)}$	...	$Y_{1J}^{(b)}$
2	$Y_{21}^{(b)}$	$Y_{22}^{(b)}$	...	$Y_{2J}^{(b)}$
...	...	...	...	...
I	$Y_{I1}^{(b)}$	$\hat{Y}_{I2}^{(b)}$	...	$\hat{Y}_{IJ}^{(b)}$

(b) Fit a GLM to the bootstrap history completing the triangle.

	1	2	...	J
1	$Y_{11}^{(b)}$	$Y_{12}^{(b)}$	...	$Y_{1J}^{(b)}$
2	$Y_{21}^{(b)}$	$Y_{22}^{(b)}$	...	$Y_{2J}^{(b)}$
...	...	...	...	...
I	$Y_{I1}^{(b)}$	$\tilde{Y}_{I2}^{(b)}$	...	$\tilde{Y}_{IJ}^{(b)}$

(c) Add the first fitted anti-diagonal to the bootstrap history and refit the GLM.

Table 14: Graphical representation of the CDR calculation for bootstrap loop  $b$ .

## 7 Conclusions

The bootstrap procedure is a powerful and popular technique to obtain information from a single sample of data, which would usually be obtained using analytic techniques. It



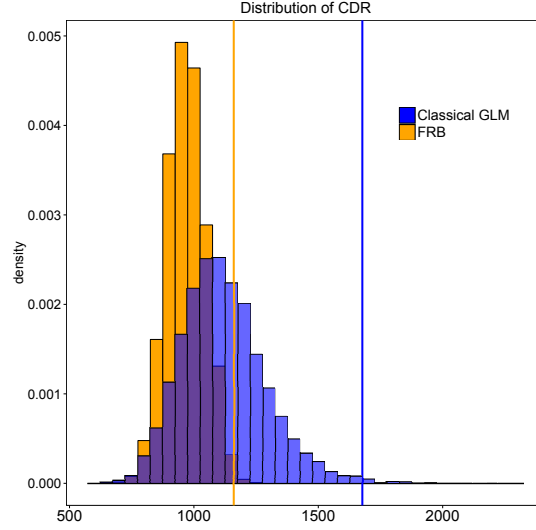


Figure 9: Estimated distribution of the CDR for the Rockford Mutual Insurance Company using the classical GLM approach in blue (dark color) and the FRB approach in orange (light color).

complements sample estimates with measures of accuracy (e.g. standard error, confidence interval, prediction error). In claims reserving, the bootstrap procedure is typically used to produce a standard error of prediction and an approximate predictive distribution for the random future losses. In this paper we studied the effect of outliers on these procedures and it is shown that it is not sufficient to plug a robust estimator into the classical bootstrap procedure. Therefore, we implemented and compared all existing (up to our knowledge) robust bootstrap methods that could be adapted to the claims reserving framework. The Fast and Robust Bootstrap (FRB) approach appears to be a good choice in the sense that it yields very reliable results and is computationally fast.

The presence of outliers may reveal that the data are more heterogeneous than assumed and cannot be handled by the original statistical model. Outliers can be isolated or may come in clusters, indicating that there are subgroups in the population that behave differently. We believe that a robust analysis together with some exploratory data analysis techniques can reveal structures in the data that would remain hidden in a traditional analysis. Robust methods try to fit the model followed by the majority of the data: if the data contain no outliers, they give approximately the same results as the classical method, while if the data is contaminated they give approximately the same results as the classical method applied to the outlier-free data. The best estimates obtained by the robust methodology is hence the result that one would obtain when applying the traditional methodology on the data without outliers. Note that we do not know the proportion of outliers in advance.

Using our robust methodology, insurers obtain estimates that are similar to the results that would have been obtained from outlier-free data. Moreover, the procedure also detects outlying observations (if any). Of course, it is important to examine the detected outliers

and to understand the reasons for their atypical behaviour. In this way the proposed robust methodology is helpful to gain insight in the data. If the outliers are errors or due to causes that are not expected to re-occur in the future, then the robust estimates allow to build up more realistic reserve estimates and confidence intervals. However, if such atypical observations are likely to happen again in future, it is necessary to model also their process (outside the scope of this paper) and to predict how much extra reserve (besides the robust total reserve estimate) is needed to cope with such outlying observations in future years. In such a case, the final estimate may for instance be equal to the robust total reserve estimate plus a safe margin for the outliers.

## Acknowledgments

This work was supported by the Internal Funds KU Leuven under Grants *C16/15/068* and *C24/15/001*; the Flemish Science Foundation (FWO) under Grant *1523915N*. The computational resources and services used in this work were provided by the VSC (Flemish Supercomputer Center), funded by the Research Foundation - Flanders (FWO) and the Flemish Government – department EWI.

## References

- Amado, C., Bianco, A. M., Boente, G., and Piers, A. M. (2014). Robust bootstrap: An alternative to bootstrapping robust estimators. *REVSTAT - Statistical Journal*, 12(2):169–197.
- Amado, C. and Pires, A. M. (2004). Robust bootstrap with non random weights based on the influence function. *Communications in Statistics - Simulation and Computation*, 33(2):377–396.
- Antonio, K. and Plat, R. (2014). Micro-level stochastic loss reserving for general insurance. *Scandinavian Actuarial Journal*, 2014(7):649–669.
- Ashe, F. (1986). An essay at measuring the variance of estimates of outstanding claim payments. *ASTIN Bulletin*, 16:S99–S113.
- Barnett, G. and Zehnirith, B. (2008). Modeling with the multivariate probabilistic trend family. In *Casualty Actuarial Society E-forum*.
- Boisseau, J.-P. (2006). Solvabilité 2 et mesure de volatilité dans les provisions pour sinistres à payer. *Mémoire Centre D’études Actuarielles*.
- Boumezoued, A., Angoua, Y., Devineau, L., and Boisseau, J.-P. (2012). One-year reserve risk including a tail factor : closed formula and bootstrap approaches. *arXiv*, 1107.0164.
- Brazauskas, V. (2009). Robust and efficient fitting of loss models: diagnostic tools and insights. *North American Actuarial Journal*, 13(3):356–369.

- Brazauskas, V., Jones, B. L., and Zitikis, R. (2009). Robust fitting of claim severity distributions and the method of trimmed moments. *Journal of Statistical Planning and Inference*, 139(6):2028–2043.
- Camponovo, L., Scaillet, O., and Trojani, F. (2012). Robust subsampling. *Journal of Econometrics*, 167(1):197 – 210.
- Cantoni, E. and Ronchetti, E. (2001). Robust inference for generalized linear models. *Journal of the American Statistical Association*, 96:1022–1030.
- Chatterjee, A. and Lahiri, S. N. (2011). Bootstrapping lasso estimators. *Journal of the American Statistical Association*, 106(494):608–625.
- Cordeiro, G. M. (2004). On pearson’s residuals in generalized linear models. *Statistics & Probability Letters*, 66(3):213 – 219.
- Cowell, R. (2009). Exploration of a novel bootstrap technique for estimating the distribution of outstanding claims reserves in general insurance. Working Paper Actuarial Research Paper No. 192, City University London, London, UK.
- Dell’Aquila, R. and Embrechts, P. (2006). Extremes and robustness: a contradiction? *Financial Markets and Portfolio Management*, 20:103–118.
- Diers, D. (2008). Stochastic re-reserving in multi-year internal models. An approach based on simulations. *ASTIN Colloquium, Helsinki*.
- Dreksler, S., Allen, C., Akoh-Arrey, A., Courchene, J. A., Junaid, B., Kirk, J., Lowe, W., O’Dea, S., Piper, J., Shah, M., Shaw, G., Storman, D., Thaper, S., Thomas, L., Wheatley, M., and Wilson, M. (2015). Solvency II technical provisions for general insurers. *British Actuarial Journal*, 20:7–129.
- Efron, B. (1981). Nonparametric estimates of standard error: The jackknife, the bootstrap and other methods. *Biometrika*, 68(3):pp. 589–599.
- England, P. (2002). Addendum to analytic and bootstrap estimates of prediction errors in claims reserving? *Insurance: Mathematics and Economics*, 31(3):461 – 466.
- England, P. and Verrall, R. (1999). Analytic and bootstrap estimates of prediction errors in claims reserving. *Insurance: mathematics and economics*, 25(3):281–293.
- England, P. and Verrall, R. (2002). Stochastic claims reserving in general insurance. *British Actuarial Journal*, 8:443–518.
- England, P. D. and Verrall, R. J. (2006). Predictive distributions of outstanding liabilities in general insurance. *Annals of Actuarial Science*, 1:221–270.
- Hall, P., Horowitz, J. L., and Jing, B.-Y. (1995). On blocking rules for the bootstrap with dependent data. *Biometrika*, 82(3):561–574.
- Hampel, F. (1974). The influence curve and its role in robust estimation. *Journal of the American Statistical Association*, 69:383–393.

- Hampel, F., Ronchetti, E., Rousseeuw, P., and Stahel, W. (1986). *Robust Statistics: The Approach Based on Influence Functions*. Wiley, New York.
- Hoedemakers, T., Beirlant, J., Goovaerts, M. J., and Dhaene, J. (2005). On the distribution of discounted loss reserves using generalized linear models. *Scandinavian Actuarial Journal*, 2005(1):25–45.
- Hubert, M., Dierckx, G., and Vanpaemel, D. (2013). Detecting influential data points for the Hill estimator in Pareto-type distributions. *Computational Statistics & Data Analysis*, 65:13–28.
- Johnson, N. L., Kotz, S., and Kemp, A. W. (1993). *Univariate Discrete Distributions*. John Wiley & Sons, 2nd edition.
- Kaas, R., Goovaerts, M., Dhaene, J., and Denuit, M. (2009). *Modern Actuarial Risk Theory: Using R*. Springer, Berlin.
- Kremer, E. (1982). Ibm-claims and the two-way model of anova. *Scandinavian Actuarial Journal*, 1982(1):47–55.
- Lacoume, A. (2008). Mesure du risque de réserve sur un horizon de un an. *Mémoire ISFA*.
- Lowe, J. (1994). A practical guide to measuring reserve variability using: Bootstrapping, operational time and a distribution free approach. In *Proceedings of the 1994 General Insurance Convention, Institue if Actuaries and Faculty if Actuaries*.
- Mack, T. (1993). Distribution-free calculation of the standard error of chain ladder reserve estimates. *ASTIN Bulletin*, 23:213–225.
- Mack, T. (1994). Measuring the variability of chain ladder reserve estimates. *Casualty Actuarial Society Forum*, pages 101–182.
- Mack, T., Quarg, G., and Braun, C. (2006). The mean square error of prediction in the chain ladder reserving method – a comment. *ASTIN Bulletin*, 36:543–552.
- McCullagh, P. and Nelder, J. A. (1983). *Generalized Linear Models*. Chapman & Hall, London.
- Norberg, R. (1993). Prediction of outstanding liabilities in non-life insurance. *Astin Bulletin*, 23(01):95–115.
- Ohlsson, E. and Lauzeningsks, J. (2009). The one-year non-life insurance risk. *Insurance: Mathematics and Economics*, 45(2):203 – 208.
- Pinheiro, P. J. R., Andrade e Silva, J. M., and De Lourdes Centeno, M. (2003). Bootstrap methodology in claim reserving. *Journal of Risk and Insurance*, 70(4):701–714.
- Pison, G. and Van Aelst, S. (2004). Diagnostic plots for robust multivariate methods. *Journal of Computational and Graphical Statistics*, 13:310–329.
- Pitselis, G., Grigoriadou, V., and Badounas, I. (2015). Robust loss reserving in a log-linear model. *Insurance: Mathematics and Economics*, 64:14–27.

- Renshaw, A. and Verrall, R. (1998). A stochastic model underlying the chain-ladder technique. *British Actuarial Journal*, 4:903–923.
- Salibian-Barrera, M., Van Aelst, S., and Willems, G. (2008). Fast and robust bootstrap. *Statistical Methods and Applications*, 17(1):41–71.
- Salibian-Barrera, M., n-Barrera, M. and Zamar, R. H. (2002). Bootstrapping robust estimates of regression. *The Annals of Statistics*, 30(2):pp. 556–582.
- Schiegl, M. (2004). Simulation. In *Handbuch zur Schadenreservierung*. Verlag Versicherungswirtschaft.
- Schiegl, M. (2015). A model study about the applicability of the chain ladder method. *Scandinavian Actuarial Journal*, 2015(6):482–499.
- Singh, K. (1998). Breakdown theory for bootstrap quantiles. *The Annals of Statistics*, 26(5):pp. 1719–1732.
- Stromberg, A. J. (1997). Robust covariance estimates based on resampling. *Journal of Statistical Planning and Inference*, 57(2):321 – 334. Robust Statistics and Data Analysis, Part II.
- Taylor, G. C. and Ashe, F. (1983). Second moments of estimates of outstanding claims. *Journal of Econometrics*, 23(1):37–61.
- Van Aelst, S. and Willems, G. (2013). Fast and robust bootstrap for multivariate inference: The R package FRB. *Journal of Statistical Software*, 53(3):1–32.
- Venter, G. G. (2006). Discussion of the mean square error of prediction in the chain ladder reserving method. *ASTIN Bulletin*, 36:566–571.
- Verdonck, T. and Debruyne, M. (2011). The influence of individual claims on the chain-ladder estimates: Analysis and diagnostic tool. *Insurance: Mathematics and Economics*, 48(1):85 – 98.
- Verdonck, T. and Van Wouwe, M. (2011). Detection and correction of outliers in the bivariate chain-ladder method. *Insurance: Mathematics and Economics*, 49(2):188–193.
- Verdonck, T., Wouwe, M. V., and Dhaene, J. (2009). A robustification of the chain-ladder method. *Nort American Actuarial Journal*, 13(2):280 – 298.
- Verrall, R. (1991a). On the estimation of reserves from loglinear models. *Insurance: Mathematics and Economics*, 10(1):75 – 80.
- Verrall, R. (2000). An investigation into stochastic claims reserving models and the chain-ladder technique. *Insurance: Mathematics and Economics*, 26(1):91 – 99.
- Verrall, R. J. (1991b). Chain ladder and maximum likelihood. *Journal of the Institute of Actuaries*, 118(03):489–499.
- Wüthrich, M. V., Merz, M., and Lysenko, N. (2009). Uncertainty of the claims development result in the chain ladder method. *Scandinavian Actuarial Journal*, 2009(1):63–84.

Wüthrich, M. V. and Merz, M. (2008). *Stochastic claims reserving methods in insurance*, volume 435. John Wiley & Sons.

Wüthrich, M. V., Merz, M., and Bühlmann, H. (2008). Bounds on the estimation error in the chain ladder method. *Scandinavian Actuarial Journal*, 2008(4):283–300.

## 8 Appendix

### 8.1 Hat matrix of the robust GLM estimator

To calculate the adjusted Pearson residuals  $\mathbf{r}^C$  an expression for the hat matrix corresponding to the robust estimator of Cantoni and Ronchetti (2001) is needed. To this end we rewrite the estimator as an iteratively weighted least squares problem. Using a first order Taylor expansion of equation (4), we obtain

$$\boldsymbol{\theta}^* = \left[ -\frac{\partial}{\partial \boldsymbol{\theta}} \bar{\boldsymbol{\psi}}(\mathbf{y}, \boldsymbol{\mu}) \right]^{-1} \bar{\boldsymbol{\psi}}(\mathbf{y}, \boldsymbol{\mu}) + \boldsymbol{\theta} \quad (12)$$

where  $\bar{\boldsymbol{\psi}}(\mathbf{y}, \boldsymbol{\mu}) = \sum_{i=1}^n \sum_{j=1}^{n-i+1} \boldsymbol{\psi}(y_{ij}, \mu_{ij}) = \mathbf{X}^t \mathbf{D} \mathbf{1}_N$  with  $\mathbf{1}_N$  a unit vector of length  $N$  and  $\mathbf{D}$  the diagonal matrix with elements

$$\psi_c(r_{ij}) w(x_{ij}) \frac{1}{V^{1/2}(\mu_{ij})} \mu_{ij} - \frac{2}{n(n+1)} \sum_{i=1}^n \sum_{j=1}^{n-i+1} E[\psi_c(r_{ij})] \omega(x_{ij}) \frac{1}{V^{1/2}(\mu_{ij})} \mu_{ij}.$$

We then replace  $-\frac{\partial}{\partial \boldsymbol{\theta}} \bar{\boldsymbol{\psi}}(\mathbf{y}, \boldsymbol{\mu})$  by  $-\mathbf{E} \left[ -\frac{\partial}{\partial \boldsymbol{\theta}} \bar{\boldsymbol{\psi}}(\mathbf{y}, \boldsymbol{\mu}) \right]$ . Cantoni and Ronchetti (2001) computed this to be equal to

$$\mathbf{X}^t \mathbf{B} \mathbf{X}$$

where  $\mathbf{B}$  is the diagonal matrix with elements  $\mathbf{E} \left[ \psi_c(r_{ij}) \frac{Y_{ij} - \mu_{ij}}{\mu_{ij}} \right] \frac{w(\mathbf{x}_{ij})}{\mu_{ij}^{3/2}}$  and

$$\begin{aligned} \mathbf{E} \left[ \psi_c(r_{ij}) \frac{Y_{ij} - \mu_{ij}}{\mu_{ij}} \right] &= c(\mathbf{P}(Y_{ij} = j_1) + \mathbf{P}(Y_{ij} = j_2)) \\ &\quad + \sqrt{\mu_{ij}} (\mathbf{P}(Y_{ij} = j_1 - 1) - \mathbf{P}(Y_{ij} = j_1) - \mathbf{P}(Y_{ij} = j_2 - 1) + \mathbf{P}(Y_{ij} = j_2)) \\ &\quad + \mu_{ij} \mathbf{P}(j_1 \leq Y_{ij} \leq j_2 - 1). \end{aligned}$$

Equation (12) then reduces to

$$\boldsymbol{\theta}^* = (\mathbf{X}^t \mathbf{B} \mathbf{X})^{-1} \mathbf{X}^t \mathbf{D} \mathbf{1}_N + \boldsymbol{\theta}$$

which can be rewritten as

$$\boldsymbol{\theta}^* = (\mathbf{X}^t \mathbf{B} \mathbf{X})^{-1} \mathbf{X}^t \mathbf{B} [\mathbf{B}^{-1} \mathbf{D} \mathbf{1}_N + \mathbf{X} \boldsymbol{\theta}],$$

assuming  $\mathbf{B}$  and  $(\mathbf{X}^t \mathbf{B} \mathbf{X})$  are invertible. This expression may be recognized as that of a weighted least squares regression where  $[\mathbf{B}^{-1} \mathbf{D} \mathbf{1}_N + \mathbf{X} \boldsymbol{\theta}]$  is regressed onto  $\mathbf{X}$  with weight matrix  $\mathbf{B}$ . The hat matrix for the robust estimator of Cantoni and Ronchetti (2001) is thus

$$\mathbf{H} = \mathbf{X} (\mathbf{X}^t \mathbf{B} \mathbf{X})^{-1} \mathbf{X}^t \mathbf{B}.$$

## 8.2 Calculation of the correction terms proposed by Cordeiro (2004)

Based on the general expression given by Cordeiro (2004), the following vector of expected Pearson residuals is obtained for the Poisson GLM model

$$E[\mathbf{r}] = -\frac{1}{2}(I - H)Jz,$$

whereas the vector of variances of the Pearson residuals is given by

$$\text{Var}[\mathbf{r}] = \mathbf{1} + \frac{1}{2}(QHJ - T)z.$$

For these expressions it holds that

$$\begin{aligned} W &= \text{diag}(\mu_{ij}) \\ J &= \text{diag}(\mu_{ij}^{1/2}) = W^{1/2} \\ Q &= \text{diag}(\mu_{ij}^{-1/2}) = W^{-1/2} \\ T &= 2W + I \\ H &= W^{1/2}X(X^TWX)^{-1}X^TW^{1/2} \\ Z &= X(X^TWX)^{-1}X^T, \end{aligned}$$

where  $z$  is the diagonal of  $Z$  and  $\mathbf{1}$  is a unit vector. Note that  $(QHJ - T)z$  corresponds to the diagonal of  $(QHJ - T)Z$  and since

$$QHJ = W^{-1/2}W^{1/2}X(X^TWX)^{-1}X^TW^{1/2}W^{1/2} = X(X^TWX)^{-1}X^TW,$$

we obtain

$$\begin{aligned} (QHJ - T)Z &= (X(X^TWX)^{-1}X^TW)(X(X^TWX)^{-1}X^T) - (2W + I)(X(X^TWX)^{-1}X^T) \\ &= (X(X^TWX)^{-1}X^T) - (2W + I)(X(X^TWX)^{-1}X^T) \\ &= -2W(X(X^TWX)^{-1}X^T) \\ &= -2WX(X^TWX)^{-1}X^T. \end{aligned}$$

Therefore we obtain

$$\text{Var}[\mathbf{r}] = \text{diag}(I) - \text{diag}(WX(X^TWX)^{-1}X^T)$$

Since the diagonal of  $W(X(X^TWX)^{-1}X^T)$  equals the diagonal of the hat matrix  $H = W^{1/2}(X(X^TWX)^{-1}X^T)W^{1/2}$ , we finally obtain that

$$\text{Var}[\mathbf{r}] = \text{diag}(I - H). \quad (13)$$

Rewriting the expression for  $E[\mathbf{r}]$  we obtain that

$$E[\mathbf{r}] = \text{diag}\left(-\frac{1}{2}(I - H)HW^{-1/2}\right). \quad (14)$$

In practice the matrices  $H$  and  $W$  may be replaced by their sample estimates.

### 8.3 Calculations for Fast and Robust bootstrap for GLM

Since the term  $E[\psi_c(r_{ij})]$  is not differentiable using its exact expression, we calculate this term by approximating the Poisson distribution with a  $\chi^2$  distribution and then using the Wilson-Hilferty approximation, see (Johnson et al., 1993, chapter 4, p. 162). This leads to the following lemma:

**Lemma 1.** *If  $Y_{ij} \sim \text{Poisson}(\mu_{ij})$  it holds that*

$$P(Y_{ij} \leq y) \approx (2\pi)^{-1/2} \int_z^\infty \exp\left(-\frac{u^2}{2}\right) du$$

where

$$z = 3 \left[ \left( \frac{\mu_{ij}}{y+1} \right)^{1/3} - 1 + \frac{1}{9(y+1)} \right] (y+1)^{1/2}.$$

Using Lemma (1) we may rewrite equation (6) as follows:

$$\begin{aligned} E \left[ \psi_c \left( \frac{y_{ij} - \mu_{ij}}{\mu_{ij}^{1/2}} \right) \right] &\approx c \left[ 1 - (2\pi)^{-1/2} \left( \int_{\hat{z}_2}^\infty \exp\left(-\frac{u^2}{2}\right) du + \int_{\hat{z}_1}^\infty \exp\left(-\frac{u^2}{2}\right) du \right) \right] \\ &\quad + (2\pi)^{(-1/2)} \mu_{ij}^{1/2} \left[ - \int_{\hat{z}_1}^{\hat{z}_1} \exp\left(-\frac{u^2}{2}\right) du + \int_{\hat{z}_2}^{\hat{z}_2} \exp\left(-\frac{u^2}{2}\right) du \right] \end{aligned}$$

with

$$\begin{aligned} \tilde{z}_l &= 3 \left[ \left( \frac{\mu_{ij}}{j_l + 0.5} \right)^{1/3} - 1 + \frac{1}{9(j_l + 0.5)} \right] (j_l + 0.5)^{1/2}, \\ \hat{z}_l &= 3 \left[ \left( \frac{\mu_{ij}}{j_l + 1.5} \right)^{1/3} - 1 + \frac{1}{9(j_l + 1.5)} \right] (j_l + 1.5)^{1/2}, \end{aligned}$$

and  $j_1 = \mu_{ij} - c\mu_{ij}^{1/2}$  and  $j_2 = \mu_{ij} + c\mu_{ij}^{1/2}$ . Note that in turn  $j_1$  and  $j_2$  are approximated by considering their non-truncated versions. However  $j_1$  must be bounded from below by zero to ensure all square roots may be calculated without problems. Clearly we need to calculate the terms

$$\frac{\partial}{\partial \boldsymbol{\theta}} \int_{\hat{z}_i}^\infty \exp\left(-\frac{u^2}{2}\right) du \text{ and } \frac{\partial}{\partial \boldsymbol{\theta}} \int_{\hat{z}_i}^{\hat{z}_i} \exp\left(-\frac{u^2}{2}\right) du.$$

To calculate the first term note that

$$\frac{\partial}{\partial \boldsymbol{\theta}} \int_{\hat{z}_i}^\infty \exp\left(-\frac{u^2}{2}\right) du = -\frac{\partial}{\partial \boldsymbol{\theta}} \int_0^{\hat{z}_i} \exp\left(-\frac{u^2}{2}\right) du = -\exp\left(-\frac{\hat{z}_i^2}{2}\right) \frac{\partial}{\partial \boldsymbol{\theta}} \hat{z}_i$$

where we used the generalized Leibniz integration rule. Moreover,

$$\frac{\partial}{\partial \boldsymbol{\theta}} \int_{\hat{z}_i}^{\hat{z}_i} \exp\left(-\frac{u^2}{2}\right) du = \exp\left(-\frac{(\hat{z}_i)^2}{2}\right) \frac{\partial}{\partial \boldsymbol{\theta}} \hat{z}_i - \exp\left(-\frac{(\hat{z}_i)^2}{2}\right) \frac{\partial}{\partial \boldsymbol{\theta}} \hat{z}_i.$$



This leads to

$$\begin{aligned} \frac{\partial}{\partial \boldsymbol{\theta}} E \left[ \psi_c \left( \frac{y_{ij} - \mu_{ij}}{\mu_{ij}^{1/2}} \right) \right] &\approx c \left[ (2\pi)^{-1/2} \left( \exp(-\frac{\hat{z}_2^2}{2}) \frac{\partial}{\partial \boldsymbol{\theta}} \hat{z}_2 + \exp(-\frac{\hat{z}_1^2}{2}) \frac{\partial}{\partial \boldsymbol{\theta}} \hat{z}_1 \right) \right] \\ &+ \frac{1}{2} (2\pi)^{(-1/2)} \mu_{ij}^{1/2} x_i \left[ \int_{\hat{z}_2}^{\hat{z}_2} \exp(-\frac{u^2}{2}) du - \int_{\hat{z}_1}^{\hat{z}_1} \exp(-\frac{u^2}{2}) du \right] \\ &+ (2\pi)^{(-1/2)} \mu_{ij}^{1/2} \left[ \exp(-\frac{(\hat{z}_2)^2}{2}) \frac{\partial}{\partial \boldsymbol{\theta}} \hat{z}_2 - \exp(-\frac{(\hat{z}_2)^2}{2}) \frac{\partial}{\partial \boldsymbol{\theta}} \hat{z}_2 \right. \\ &\left. - \exp(-\frac{(\hat{z}_1)^2}{2}) \frac{\partial}{\partial \boldsymbol{\theta}} \hat{z}_1 + \exp(-\frac{(\hat{z}_1)^2}{2}) \frac{\partial}{\partial \boldsymbol{\theta}} \hat{z}_1 \right]. \end{aligned}$$

Differentiating  $\hat{z}_l$  and  $\hat{z}_l$  yields

$$\begin{aligned} \frac{\partial}{\partial \boldsymbol{\theta}} \hat{z}_l &= 3 \left[ \frac{\partial}{\partial \boldsymbol{\theta}} \left( \frac{\mu_{ij}}{j_l + 0.5} \right)^{1/3} + \frac{\partial}{\partial \boldsymbol{\theta}} \frac{1}{9(j_l + 0.5)} \right] (j_l + 0.5)^{1/2} \\ &+ 3 \left[ \left( \frac{\mu_{ij}}{j_l + 0.5} \right)^{1/3} - 1 + \frac{1}{9(j_l + 0.5)} \right] \frac{\partial}{\partial \boldsymbol{\theta}} \left( (j_l + 0.5)^{1/2} \right), \\ \frac{\partial}{\partial \boldsymbol{\theta}} \hat{z}_l &= 3 \left[ \frac{\partial}{\partial \boldsymbol{\theta}} \left( \frac{\mu_{ij}}{j_l + 1.5} \right)^{1/3} + \frac{\partial}{\partial \boldsymbol{\theta}} \frac{1}{9(j_l + 1.5)} \right] (j_l + 1.5)^{1/2} \\ &+ 3 \left[ \left( \frac{\mu_{ij}}{j_l + 1.5} \right)^{1/3} - 1 + \frac{1}{9(j_l + 1.5)} \right] \frac{\partial}{\partial \boldsymbol{\theta}} \left( (j_l + 1.5)^{1/2} \right). \end{aligned}$$

Moreover,

$$\begin{aligned} \frac{\partial}{\partial \boldsymbol{\theta}} \left( \frac{\mu_{ij}}{j_l + m} \right)^{1/3} &= \frac{1}{3} \frac{\mu_{ij}^{1/3}}{(j_l + m)^{4/3}} \left( x_{ij}(j_l + m) - \frac{\partial}{\partial \boldsymbol{\theta}} j_l \right), \\ \frac{\partial}{\partial \boldsymbol{\theta}} \frac{1}{9(j_l + m)} &= \frac{-\frac{\partial}{\partial \boldsymbol{\theta}} j_l}{9(j_l + m)^2}, \\ \frac{\partial}{\partial \boldsymbol{\theta}} \left( (j_l + m)^{1/2} \right) &= \frac{1}{2} \frac{1}{(j_l + m)^{1/2}} \frac{\partial}{\partial \boldsymbol{\theta}} j_l, \end{aligned}$$

with  $\frac{\partial}{\partial \boldsymbol{\theta}} j_1 = x_{ij}(\mu_{ij} - \frac{c}{2}\mu_{ij}^{1/2})$  and  $\frac{\partial}{\partial \boldsymbol{\theta}} j_2 = x_{ij}(\mu_{ij} + \frac{c}{2}\mu_{ij}^{1/2})$ .

## 8.4 Proof of theorem 1

The estimator by Cantoni and Ronchetti (2001) may be generally written as

$$\sum_{i=1}^n \sum_{j=1}^{n-i+1} \psi(y_{ij}, \mu_{ij}) = 0,$$

where

$$\psi(y_{ij}, \mu_{ij}) = \nu(y_{ij}, \mu_{ij}) w(x_{ij}) \mu'_{ij} - a(\boldsymbol{\theta})$$

can be seen as the score function and  $a(\boldsymbol{\theta}) = \frac{1}{N} \sum_{i=1}^n \sum_{j=1}^{n-i+1} E[\nu(y_{ij}, \mu_{ij})] w(x_{ij}) \mu'_{ij}$ . This is the structure of an  $M$ -estimator which leads to the following influence function:

$$\text{IF}([z, l, m], T_{\boldsymbol{\theta}}, F) = M(\psi, F)^{-1} \psi(z, \mu_{lm}),$$

with  $M(\psi, F) = -E\left[\frac{\partial}{\partial \boldsymbol{\theta}} \psi(z, \mu_{lm})\right]$ . More details may be found in e.g. Hampel et al. (1986). The asymptotic variance can then be found as

$$V(\psi, F) = M(\psi, F)^{-1} Q(\psi, F) M(\psi, F)^{-t},$$

where  $Q(\psi, F) = E[\psi(z, \mu_{lm}) \psi(z, \mu_{lm})^t]$ . Using the above expressions, the standardised influence function of the  $M$ -estimator becomes

$$(\text{IF}^t V^{-1} \text{IF})^{\frac{1}{2}} = (\psi^t M^{-t} M^t Q^{-1} M M^{-1} \psi)^{\frac{1}{2}} = (\psi^t Q^{-1} \psi)^{\frac{1}{2}},$$

where the arguments of the function have been dropped for lightning the notation.

Choosing  $\nu(z, \mu_{lm}) = \psi_c(r_{lm}) \frac{1}{V^{1/2}(\mu_{lm})}$  leads to the robust estimator of Cantoni and Ronchetti (2001) as detailed in section 3.1. However, for the influence function bootstrap procedure we need the expression for the influence function of the classical estimator. Therefore, we choose  $\nu(z, \mu_{lm}) = \frac{r_{lm}}{V^{1/2}(\mu_{lm})}$  and set  $w(\cdot) = 1$  which are the functions corresponding to the Poisson GLM yielding the same estimates as the classical chain-ladder. Thus (9) reduces to

$$\text{RESIF}([z, l, m], T_{\boldsymbol{\theta}}^{nr}, F_{\hat{\boldsymbol{\theta}}^r}) = \left[ \frac{\psi(z, \hat{\mu}_{lm})^t \psi(z, \hat{\mu}_{lm})}{Q(\psi, F_{\hat{\boldsymbol{\theta}}^r})} \right]^{\frac{1}{2}}. \quad (15)$$

It is easily checked that  $E[r_{lm}] = 0$ , which leads to  $a(\boldsymbol{\theta}) = 0$ . Hence,  $\psi(z, \hat{\mu}_{lm}) = r_{lm} \hat{\mu}_{lm}^{1/2}$  and thus (15) becomes

$$\text{RESIF}([z, l, m], T_{\boldsymbol{\theta}}^{nr}, F_{\hat{\boldsymbol{\theta}}^r}) = \frac{|r_{lm}| \mu_{lm}^{1/2}}{(E[r_{lm}^2 \mu_{lm}])^{1/2}}.$$

This immediately simplifies to

$$\text{RESIF}([z, l, m], T_{\boldsymbol{\theta}}^{nr}, F_{\hat{\boldsymbol{\theta}}^r}) = \frac{|z - \mu_{lm}|}{\mu_{lm}^{1/2}}$$

which proves the theorem.