Actuarial Models For Estimating Non Life Risks

By

BANTOUNAS IOANNIS

Department of Statistics and Insurance Science
UNIVERSITY OF PIRAEUS

A Thesis submitted to the University of Piraeus in accordance with the requirements for the degree of *Doctor of Philosophy*.

2019

Abstract

Loss reserving is one of the most critical actuarial procedures in non-life insurance. This procedure projects losses to their ultimate value and estimates the total reserves. The actual amount of the insurers' liability is initially unknown until all claims are finally settled. Inappropriate actuarial methods may lead to misestimation of the total reserve, which has a significant impact on the insurers' solvency. Each reserving method gives a different estimate for the required reserves which means that the appropriate method will be selected according to the judgement of the actuary. In non-life insurance, the insurer should have reserves, for his future obligations concerning with incurred but not reported claims and incurred but not enough reported. In this thesis, we present new methods for estimating the ultimate claims and the total reserves, according to insurance regulations and the market's needs. Using the data in a log-linear way, robust estimators are applied to the chain ladder procedure. We incorporate robust random coefficients regression models and robust cross-section models for the estimation of the total reserves. These models provide a solution to the problem of outlier claims, which have an effect to the pattern of outstanding claims and lead to misreserving. We present an application of the recursive Kalman filter algorithm, in order to estimate the reserves of an insurance company. A robustified version of this Kalman filter algorithm is also provided. Using quantile regression, which offers a more thorough description of the distribution than the classical least squares estimation, we construct methods for loss reserves estimation. In addition, we propose a loss reserving method for a non-life insurance portfolio consisting by several correlated run-off sub-portfolios that can be embedded within the quantile regression model for longitudinal data. Our numerical results indicate that our proposed loss reserving methods provide more reliable results than the existing ones.

Περιληψη

Η Αποθεματοποίηση ζημιών είναι μία από τις πιο σημαντικές διαδικασίες στις ασφαλίσεις κατά ζημιών. Η διαδικασία αυτή εκτιμά την τελική τιμή των ζημιών καθώς και τα αποθέματα. Το πραγματικό ποσο που ευθύνονται να πληρώσουν οι ασφαλιστικές εταιριες είναι αρχικά άγνωστο μέχρι να διευθετηθουν οι ζημιές. Μη κατάλληλες αναλογιστικές μέθοδοι αποθεματοποιησης μπορούν να οδηγήσουν σε λανθασμένη εκτίμηση των αποθεμάτων, το οποίο έχει σημαντική επίδραση στη φερεγγυότητα μίας ασφαλιστικής εταιρίας. Κάθε μέθοδος αποθεματοποίησης δίνει διαφορετική εκτίμηση για τα απαιτούμενα αποθεματα, το οποίο σημαίνει οτι η κατάλληλη μέθοδος θα επιλεγεί σύμφωνα με την κρίση του αναλογιστή. Στις ασφαλίσεις κατά ζημιών ο ασφαλιστής θα πρέπει να δημιουργήσει αποθέματα ώστε να μπορεί να καλύψει τις μελλοντικές του υποχρεώσεις που βασίζονται σε ζημιές που έγιναν και δεν έχουν αναφερθεί ή έγιναν και δεν έχουν αποθεματοποιηθεί αρκετά. Σε αυτή τη διατριβή, παρουσιάζουμε νέες μεθόδους για την εκτίμηση των τελικών ζημιών και αποθεμάτων που βασίζονται στο νομοθέτικο πλαίσιο και στις ανάγκες της αγοράς.

Μια κλάση ανθεκτικών εκτιμητών εφαρμόζεται με μορφή λογαριθμογραμμικών μοντέλων. Ενσωματώνουμε ανθεκτική παλινδρόμηση με τυχαίους παράγοντες και $cross - section$ μοντέλα για την εκτίμηση των τελικών αποθεμάτων. Αυτα τα μοντέλα δίνουν λύση στο πρόβλημα των ακραίων ζημιών, οι οποίες επηρεάζουν το μοτίβο πληρωμών και οδηγούν σε λάθος αποθεματοποίηση.

Παρουσιάζουμε μια εφαρμογή του Φίλτρου Κάλμαν με αλγορίθμους χώρου καταστάσεων, με σκοπό να εκτιμηθούν τα αποθεματα από μία ασφαλιστική εταιρία. Επίσης δίνεται ο ανθεκτικός αλγόριθμος του Φίλτρου Κάλμαν. Χρησιμοποιώντας παλινδρόμηση ποσοστημορίων, η οποία παρέχει ολοκληρωμένη εικόνα της κατανομής σε σχέση με την κλασσική παλινδρόμηση, κατασκευάζουμε μεθόδους για την εκτίμηση των αποθεμέτων. Επιπλέον, προτείνουμε μέσω παλινδρόμησης ποσοστημορίων μια μέθοδο αποθεματοποίησης η οποία συνδυάζει διαφορα συσχετισμένα χαρτοφυλάκια, ως διαχρονικά δεδομένα. Τα αριθμητικά αποτελέσματα υποδηλώνουν οτι οι προτεινόμενες μέθοδοι παρέχουν πιο αξιόπιστα αποτελέσματα από τις υπάρχουσες μεθόδους αποθεματοποίησης.

Dedication and acknowledgements

First and foremost, I offer my sincerest gratitude to my supervisor Assistant Professor Mr. Pitselis Georgios for the guidance, encouragement, friendship, inspiration and advice he has provided throughout my doctoral studies. I especially want to thank him for supporting me in my choices and for believing in me. He answered the many (sometimes not well-defined) questions that I had. I also want to express my gratitude for giving me the chance to teach in the university as his assistant. I would like to acknowledge the support, both financial and emotional, of him, who in his own special way, kept me going till the end. It has been a privilege to study under him and I will always respect him in both personal and professional level.

I would also like to thank the other members of the doctoral committee (Papaioannou Takis, Antzoulakos Demetrios, Politis Konstantinos, Zimbidis Alexandros, Tzavelas Georgios, Psarrakos Georgios) for their valuable contributions as committee members.

After the professionals, a word of thanks is addressed to all my friends (especially to George) for their support. Of course, a big thanks to a friend of mine who is a wonderful cook as well as a good actuary for his support.

I would also like to thank my parents Georgios and Maria and my brothers Andrew and Nick for their love, guidance, and support.

A special thanks to my angel, Theodora, who constantly reminded me of her confidence to me, supported and encouraged me to pursue my scientific vocation, especially in moments of doubt. She had always believed in me and that was a great moral support.

I would also like to thank all these little heroes who gave and continue giving life to my life.

In loving memory of my late grandmother Anastasia and my uncle Nick.

Finally, I thank God for letting me through all the difficulties.

Table of Contents

List of Tables

Table **Page**

List of Figures

Chapter 1

Non Life Risks

In the modern world of globalization that we live in, it is more than obvious that risk and uncertainty preponderate. The main object of non life insurance is the risk which is the uncertain future damage. What separates the risk from any other losses or expenses is the uncertainty of the claim amount and the time that will occur. In order to limit this uncertainty, insurance companies have been formed, which for a specified amount of money (the premium), transfer the possible loss between two contracting parties. When working with many of independent risks, the error of estimating the average of these losses is reduced and therefore it is easier to manage these policies.

The insurance company is exposed to a large number of risks and many of these risks are general in the sense that they have an impact on a wide range of companies. Some examples are operational risk, market risk, insurance risk, and others. When a client buys an insurance contract, the risk is being assinged to the insurance company. In this way, the insurer is obliged to cover the losses that will come from a particular risk for a predetermined period of time. Obviously, the insurer does not offer his services free of charge and the insured person has to pay the premium. Then, in the event of damage, the insurer is obliged to pay the indemnity agreed in the insurance contract to the insured. Depending on the insurance contract type, determining the amount of final compensation is in many cases quite difficult and time-consuming. Factors such as the delay between the settlement and the payment of a claim make the calculation of the necessary reserves a difficult task.

1.1 Insurance Risks

Life is inherently risky which means that people are not able to protect themselves against every potential risk they face. It is obvious that people can not identify where they might be vulnerable to any loss but they can protect themselves from the impact of damage by insurance. In insurance terms, the risk is the possibility of something unexpected to happen. This might involve the damage, loss or theft of valuable property, or it may involve a bodily injury. Insurance companies assess the risk they want to take and decide the premium if a policyholder suffered a loss for something covered by the policy. For that reason, insurers calculate the probability of an accident at the insured property and how much it would cost to repair or replace. For that reason, insurance companies need to hold the necessary loss reserves.

One of the most fundamental aspects of the insurance company is to handle the loss reserves that are required to cover future payments, which arose from the incurred claims. Insurance companies make promises to policyholders to compensate them if they have a loss or sustain damage that is caused by a peril (claim). They may face the possibility of paying a claim, which is too large. A large claim can be due to scarce events, such as a car accident in which the policyholder's car will be totally broken down or a fire which may burn completely a house and may threaten other houses in the nearby area. The company must have sufficient reserves to cover such demands and for that reason, it is necessary to predict the future values of those claims. The magnitude of the reserves that an insurer should hold requires great consideration and is an exquisitely subtle point. The case of holding too low reserves may lead the insurance company to bankruptcy, while on the other hand, holding too high reserves may reduce the profitability and the competitiveness of the company in the insurance industry.

The prediction of future claims is generally made according to the insurer's portfolio and the line of business (LoBs), which are generally classified by the insurance type, e.g. Property Insurance, Auto Insurance, Health Insurance, etc. The most well-known framework which demonstrates historical payments from a single LoB involves the visualization of the payments in a triangular manner. Then, future claims are estimated using the triangular way, which is the run-off triangle. This is comprised of monetary amounts due to determinate claims and dates concerning with the event, settlement, and report of each one. It should be mentioned that using the run-off triangles for the estimation of the future claims, makes sense only when the loss development model is the same for all accident years (Schmidt, 2006). Thus, using this assumption, the patterns of claims occurring in the past are supposed to be the same in the future (known as a homogeneous development pattern). It does not make sense to forecast future claims if this assumption is not valid. To ensure a development pattern like that, LoBs should be separated in a way that each run-off triangle consists of homogeneous observed claims (Wiendorfer, 2012).

The most popular of all known methods which are used to make estimations for future obligations is primarily the Chain Ladder method and secondarily the method of Bornhuetter-Ferguson (Bornhuetter and Ferguson, 1972). The purpose is, therefore, to make estimations for the incurred but not reported claims (IBNR) and project ultimate loss amounts. IBNR is referring to claims that have a time lag between the incident of the accident ant the actual payment. There are many possible reasons for this:

- Firstly, the policyholders often delay to report the occurrence of a claim.
- Secondly, it may take time to settle the real cost of a claim. For example, in Motor Insurance, the establishment of the effects of an incident and the payments may last many years after the contract expiration.

Mainly, the IBNR claims are referred to as the difference between the ultimate (total) claims (claims that have been closed or matured) and the incurred claims. The basic parts of this (Schlemmer and Tarkowski, 2013) are:

- Pure IBNyR which is referred to the reserves that have not yet been reported.
- IBNeR or development on known claims which is referred to the estimate of ultimate losses for known claims which have not been enough reported.

$$\begin{bmatrix} \\ \\ Ultimate \\ \\ \\ \end{bmatrix} = \begin{bmatrix} \\ Claim\ reserves \\ \\ Payments \\ \\ \end{bmatrix} = \begin{bmatrix} IBNyR \\ IBNeR \\ Case\ reserves \\ Payments \\ \\ \end{bmatrix} = \begin{bmatrix} IBNR \\ \\ Case\ reserves \\ Payments \\ \\ \end{bmatrix} = \begin{bmatrix} IBNR \\ \\ \\ Incurred \\ \\ \end{bmatrix}$$

Remark 1.1.

- Payments are often called paid losses.
- Some terms are not consistent within the actuarial world. For instance, actuaries often understand under IBNR only the IBNyR claims.

- The definitions depend on the accounting standard. For instance, under IFRS 4, the cashflows should be discounted. Moreover, the begin of the coverage period is used instead of the accident date.

Nevertheless, it is obvious how important it is to use better methods not only to make estimations for the reserves but also to take measures in order to compute their variability. Chain-ladder is the benchmark in the reserving industry because of its simplicity and its global recognition. It is the duty of the actuary to estimate the accurate amount of variation in the reserves.

1.2 Solvency Regulations

1.2.1 Solvency I

The protection of the policyholders and the financial stableness of the insurance industry is a crucial aspect and the regulatory authorities intervene to ensure it. Each insurance and reinsurance company is obliged to hold an extra reserve, which is the solvency margin and is used to meet the underwriting liabilities for the future.

Many books are written about the solvency of an insurance company. The books of Pentikainen (1982) and Rantala (1982) provide much information and defined the term solvency margin which shows how the assets differ from the liabilities of a company. Starting with simple formulas, the researchers calculated the solvency risk margin based on the risk to invest and the technical risk (the risk of not adopting the correct claim rates). The simplicity of the calculations was the key to understand the term but the complexity of the market imposed the use of more sophisticated methods and paved the way for the estamblishment of Solvency I in Europe in 2002.

The plan behind Solvency I was the development of insurance services throughout the European single market. With the new legislation, Solvency I made improvements to the previous regulations importing robust methods to regulate the solvency of insurance companies. Nevertheless, Solvency I maintained its simplicity. The most important accomplishment of Solvency I was that the protection to the policyholders had been increased while the requirements should be met constantly (see ChandraShekhar et al., 2007).

Solvency I was primarily focused on capital adequacy for insurers and was widely recognized as being calibrated at a really too low level of capital. Hence, most regulators had informally expected companies to hold twice the amount of capital. Moreover, Solvency I lacked the inclusion of risk management within companies. Thus, the introduction of an improved regulatory system was essential.

1.2.2 Solvency II

Solvency II (SII) was the state-of-the-art European Union (EU) legislation risk-sensitive system which came into effect on January 1^{st} in 2016. It has been designed in order to make policyholders feel more secure and at the same time to create stable financial markets. If an insurance company satisfies the criteria of SII, it obtains the right to run its business in every EU country. Solvency II is a giant progress in the regulation of EU insurers and reinsurers because compared to Solvency I, it provides a more transparent risk based on the regulatory framework. It has a more complex system than the one that Solvency I used but covers a multitude of actuarial issues. The key characteristics of Solvency II system are:

- It is a three-pillar system of supervision. This is analogous to the Basel II system, which was started for banks and dealt with quantitative capital adequacy (Pillar I), makes the requirements for the risk management and governance of the insurance company (Pillar II) and finally reporting and publication which ensures more transparency (Pillar III).

- Pillar I is a two level pillar which requires the computation of Minimum Capital requirement (MCR) and the Solvency Capital Requirement (SCR). The SCR is the minimum amount of capital that an insurance undertaking must have so as to be authorized to conduct its business in the EU. If the capital is smaller than its SCR then the insurance undertaking can lose its authorization of conducting its business. Note that the SCR computation is based on a Value-at-Risk (VaR). Based on the Directive 2009/138/EC and article 101.4, the condition for solvency at time $t = 0$ is

  $$\text{Basic Own Funds at time } 0 = BoF_{t=0} = Assets_{t=0} - Liabilities_{t=0} \geq SCR.$$

  Moreover, at the same Directive and article 101.3, the SCR is defined as an amount which will be the Value at Risk (VaR) of the basic own funds of the insurer or the reinsurer using a confidence level of 99.5% for one-year. Specifically,

  $$SCR = E[BoF_{t=1}] - VaR_{99.5\%}[BoF_{t=1}],$$

  where $VaR_{99.5\%}[X] = \sup\{x : P(X \geq x) \geq 99.5\%\}$.

- Solvency II allows the usage of internal models. Then, the SCR is computed by stochastic simulation methods, taking into account the necessary risk categories. Nevertheless, the aforementioned models have to be approved by the supervisory authority. The standard approach of SII is the default method and can be used without requirements.

Various approaches have been considered when analyzing a claim portfolio. Many sophisticated stochastic models have been widely applied so as to to investigate the structure and the uncertainty of the data (Taylor 2000, Klugman et al., 2008). Moreover, generalized linear models which basically are based on loss distributions have been used in order to evaluate the claims liability of a company using the past claims data (de Jone and Heller, 2008). Nevertheless, the existence of uncertainty and non robustness of the mean, makes this measure sensitive to outlier claims and their liability is different from the central estimates. Practically, the quantiles approach leads to a provision in which a specified probability, say 80%, makes the provision sufficient to cover the run-off claims. Adding the necessary margin to the central estimate, the evaluation of the claims liability provides and makes the provision sufficient to cover the future liabilities. It should be mentioned that high claims are a concerning scenario for the insurance companies because dangerous situations may be cause at the liquidity of the company. Finally, a more volatile portfolio requires a higher risk margin in comparison with stable portfolios (de Jong and Heller, 2008).

1.3 Premium and Reserve Risk

In non-life market, risk is separated into the reserve and premium risk. The reserve risk deals with the liabilities for insurance policies. On the other hand, premium risk focuses on the future dangers such as the provision for unearned premium. These two risks are contained in the computation of risk margins especially using the Cost-of-Capital (CoC) approach. According to Solvency II, the time period is one year and defined by the EU Commission (2007) as all possible claims, as well as the adverse recalculation of assets and liabilities for a period of 12 months which are to be estimated.

The target of the reserve risk is the capability of an insurance company to pay for the claims up to their full run-off. If R_0 is the estimation of the reserves when the year begins and C_n are the total payments during the run-off period, the reserve risk is the analysis of the distribution function of $R_0 - C_n$. This is a stochastic claims reserving process and it has been studied a lot over the last years by Mack (1993), England and Verall (2002) and many others.

If the next year the payments are C_1, R_0 is the reserve when the year begins and R_1 the estimation of the reserve when the year ends, the technical result is calculated as

$$T = R_0 - R_1 - C_1$$

and the 1 year reserve risk is measured using the distribution of T, and the data at time zero.

In non-life insurance where the portfolio is divided into lines of business (LoB), the computation of the reserve risk is the combination of all LoB's calculation which means that we focus on the probability distribution of the run-off result T for all LOB's (Wutrich et al., 2008).

Moreover, the premium risk is related to the contracts to be written in a specific period, and to the risk of the unexpired contracts (CEIOPS, 2007). If $\tilde{C}_1$ are first year payments and $\tilde{R}_1$ is the claims reserve when the year ends, then the premium risk is the risk in the cost $(\tilde{C}_1 + \tilde{R}_1)$. Moreover, if $\tilde{P}$ denotes the earned premium for that period and E are the corresponding expenses, the quantity (see Ohlsson and Lauzeningks, 2008)

$$\tilde{T} = \tilde{P} - E - (\tilde{C}_1 + \tilde{R}_1)$$

measures the premium risk.

1.4 Review on Loss Reserving

In order to assess the reserve liability of the insurer, the source, the form, the quantity as well as the quality of the data which are used in the methods, should be taken into serious consideration (Carrato, 2016).

When analyzing data, an actuarial analyst should be informed about the origin of the triangulated data. The specialist should be informed about the relationship with the business process and what they represent (Source). The Form is an important feature because the data should be captured correctly at any cell in the triangle. Moreover, the reliability of the results is also very important and is connected with the quantity of the data. Actuaries should decide how much information is needed to make the necessary estimations so as to include basic characteristics such as the inflation over the period or the market conditions etc. A data triangle should include information up to the ultimate claims and then tail factors are not necessary to be applied at the procedures. Moreover, in this case, there is more uncertainty of the total results. Finally, the quality of the data is clearly an important factor because data without good quality result to corruption and inappropriate estimations.

1.4.1 Chain Ladder Notations

Run-off triangles are used in non-life insurance in order to estimate the future obligations of an insurance company which can be the claim numbers or the claim amounts. They are usually used in non-life insurance because it may take some period from when a loss occurred up to be paid. It's very critical for the insurer to know the amount of its liabilities for the future claims or the possible reopenings so that it can calculate the surplus to be made. Obviously, the insurer does not know the explicit amount of ultimate claims for each period but it should be estimated with accuracy.

The run-off triangles (see Table 1.1) are divided into cells where each cell is the corresponding payment arising from a speciffic accident year $i \in \{1, \dots, n\}$ (rows) and a development year $j \in \{1, \dots, n\}$ (columns). The accident year shows the losses appeared during a specific period while the development year denotes the total years that passed after the accident when a compensation has been made. It should be mentioned that sometimes instead of the accident year, the underwriting year can be used which is the year commencing with the effective date of a policy (Wutrich et al., 2008).

The calendar year k is the diagonal element of the triangle and is defined as $k = i + j$, with $k \in \{1, \dots, n\}$. Then $C_{i,j}$ is defined as the total incremental payments in accident year i with development j, where $i+j \leq n$ because the calendar $i+j > n$ has not occurred yet. Some reserving methods use the cumulative payments $S_{i,j}$ where $S_{i,n} = \sum_{j=1}^{n} C_{i,j}$, with $S_{i,0} = C_{i,0}$. We will denote with D_n the data up to time n (which is the upper section of the run-off triangle)

$$D_n = \{C_{i,j}, i + j \leq n\} = \{S_{i,j}, i + j \leq n\}.$$

For any accident year i, we need to find the best estimate for the ultimate amount of a compensation, i.e.

$$\widehat{S}_{i,\infty}^{n-i} = \lim_{j \to \infty} E[S_{i,j} | D_n] = E[S_{i,\infty} | D_n],$$

and the deviation with the payment made at year n will be the appropriate reserve

$$\widehat{R}_i = \widehat{S}_{i,\infty}^{n-i} - S_{i,n-i+1}.$$

Finally, one important quantity is the total uncertainty, measured by the variance of the total year amount of payment

$$Var(S_{i,\infty}|D_n) \text{ or } Var(\widehat{S}_{i,\infty}^{n-i}).$$

For our examples, we make the assumption that all claims do not evolve after n years. After n years they are closed, which means that $C_{i,j>n} = 0$.

Table 1.1: Run-off Triangle

	Development Year j				
Accident Year i	Year 1	Year 2	$\cdots$		Year n
Year 1	$C_{1,1}$	$\cdots$			$C_{1,n}$
Year 2		$\ddots$		$\ddots$	
$\vdots$	$\vdots$		$C_{i,j}$		
		$\ddots$			
Year n	$C_{n,1}$				

Remark 1.2. An important point that should be mentioned is the existence of non positive amounts such as recoveries, internal errors or cancellation of outstanding claims which is the result of overestimating them (see Marco et al., 2005). Moreover, large settlement amounts should be investigated because such amounts may lead to unreliable estimations.

Remark 1.3. Usually run-off triangles (paid or paid and outstanding triangles) contain loss adjustment expenses (LAE). Each claim may have LAE and they can be found in the run-off triangle. Expenses which are associated to the process of a compensation are called allocated loss adjustment expenses (ALAE). They are mainly expenses for external partners, accountants etc. On the other side, internal loss adjustment expenses (income of claims handling commissions, production fees, etc.) are not contained have to be computed separated because they are not included in the claims.

1.4.2 Chain-Ladder Model

Chain Ladderis the most famous method used in loss reserving. There are different derivations for this model. A distribution-free one is presented below (Mack, 1991). In the actuarial bibliography the Chain Ladder model is usually explained as a non-parametric model using an algorithm which can easily estimate the claims reserve. In the latest years, actuaries expressed the CL algorithm as a stochastic model (Mack, 1991). Specifically, in 1993, Mack published an article based on the computation of the variances in the Chain Ladder method.

Proposition 1.1 (Model Assumptions)**.**
Supposing that the development factors $f_0,\ f_1,\dots,f_{n-1}$ *exist, then for all* $0 \leq i \leq n$ *and* $1 \leq j \leq n$ *the following equation holds:*

$$E[S_{i,j}|S_{i,0},\dots,S_{i,j-1}] = E[S_{i,j}|S_{i,j-1}] = f_{j-1}\cdot S_{i,j-1} \tag{1.1}$$

where the accident years are independent.

Remark 1.4.

- We make the independence assumption between the accident years. This assumption is common in almost all of the loss reserving methods. This means that the accounting year effects have been eliminated in the triangle data.

- A stronger assumption for the cumulative claims $S_{i,0}, S_{i,1},\dots$ is that they form Markov chains. In addition, the quantity

$$S_{i,j}\prod_{k=1}^{j-1} f_k^{-1}$$

  forms a martingale for $j \geq 0$.

- The estimated quantities f_j are usually called loss development factors (LDF's), CL factors or age-to-age factors and they are one of the most important points of concern in the Chain-Ladder model.

Using the model assumptions (Proposition 1.1) and the upper trapezoid set of observations D_n we take that

$$E[S_{i,n}|D_n] = E[S_{i,n}|S_{i,n-i}] = S_{i,n-i}\cdot f_{n-i}\cdots f_{n-1}. \tag{1.2}$$

Using (1.2), a recursive algorithm can be constructed for estimating the total claims given by the observations D_n. For known CL factors f_j, the estimated outstanding claims for accident year i based on the upper triangle D_n are given by

$$E[S_{i,n}|D_n] - S_{i,n-i} = S_{i,n-i}[f_{n-i} \cdots f_{n-1} - 1]. \tag{1.3}$$

It should be mentioned that CL factors are not known and it is necessary to estimate them. A common way to estimate f_{j-1} is the following:

$$\hat{f}_{j-1} = \frac{\sum_{k=0}^{n-j} S_{k,j}}{\sum_{k=0}^{n-j} S_{k,j-1}}. \tag{1.4}$$

Remark 1.5. The result of (1.3) is well known as the Best Estimate (BE) reserves for accident year i.

Proposition 1.2.
The CL estimator for $E[S_{i,n}|D_n]$ is

$$\widehat{S_{i,j}}^{CL} = \hat{E}[S_{i,j}|D_n] = S_{i,n-i} \cdot f_{n-i} \cdots f_{j-1}, \tag{1.5}$$

for $i + j > n$.

Lemma 1.1. *We define the subset*

$$B_k = \{C_{i,j},\ i + j \leq n,\ 0 \leq j \leq k\} \subseteq D_n.$$

Then, under the model assumptions (Proposition 1.1) the following results are taken:

1. *$\hat{f}_j|B_j$ is unbiased, $E[\hat{f}_j|B_j] = f_j$.*
2. *$\hat{f}_j$ is unbiased, $E[\hat{f}_j] = f_j$.*
3. *$\hat{f}_0, \ldots, \hat{f}_{n-1}$ are not correlated. Mathematically, $E[\hat{f}_0, \ldots, \hat{f}_{n-1}] = E[\hat{f}_0] \cdots E[\hat{f}_{n-1}]$.*
4. *$\widehat{S_{i,n}}^{CL}|S_{i,n-i}$ is unbiased, $E[\widehat{S_{i,n}}^{CL}|S_{i,n-i}] = E[S_{i,n}|D_n]$.*
5. *$\widehat{S_{i,n}}^{CL}$ is unbiased estimator, $E[\widehat{S_{i,n}}^{CL}] = E[S_{i,n}]$.*

Proof. 1. First of all

$$E[\hat{f}_{j-1}|B_{j-1}] = \frac{\sum_{k=0}^{n-j} E[S_{k,j}|B_{j-1}]}{\sum_{k=0}^{n-j} S_{k,j-1}} = \frac{\sum_{k=0}^{n-j} S_{k,j-1} \cdot f_{j-1}}{\sum_{k=0}^{n-j} S_{k,j-1}} = f_{j-1},$$

which means that there is conditional unbiasedness.

2. Similar as in the previous part.

3. For $j < k$, it holds

$$E[\hat{f}_j \cdot \hat{f}_k] = E\Big[E[\hat{f}_j \cdot \hat{f}_k | B_k]\Big] = E\Big[\hat{f}_j \cdot E[\hat{f}_k | B_k]\Big] = E\Big[\hat{f}_j \cdot f_k\Big] = f_j \cdot f_k.$$

4. In order to prove the Chain-Ladder estimator we have

$$\begin{aligned} E[\widehat{S_{i,n}}^{CL} | S_{i,n-i}] &= E[S_{i,n-i} \cdot \hat{f}_{n-i} \cdots \hat{f}_{n-1} | S_{i,n-i}] \\ &= E[S_{i,n-i} \cdot \hat{f}_{n-i} \cdots \hat{f}_{n-2} \cdot E[\hat{f}_{n-1} | B_{n-1}] | S_{i,n-i}] \\ &= f_{n-1} \cdot E[\widehat{S_{i,n-1}}^{CL} | S_{i,n-i}]. \end{aligned}$$

Iterating this procedure many times we take the results.

5. The same as the previous part.

■

Remark 1.6. According to the Lemma 1.1, the Chain Ladder factors $\hat{f}_j$ are not correlated. But, it should be taken into account that they are dependent. Moreover, unbiased estimators for the BE of the reserves $E[S_{i,n}|D_n]$ have been produced.

1.4.3 Mack Model

In order to estimate the mean square error (MSE) of the forecast for the CL estimator, it is important to make extensions to the Chain-Ladder assumptions and include the second moments.

Proposition 1.3 (Model Assumptions)**.**
Suppose that accident years are independent. Then, $C_{i,j}$ forms a Markov Chain with LDF's $f_0,\ f_1, \dots, f_{n-1}$ and variance parameters $\sigma_0^2,\ \sigma_1^2, \dots, \sigma_{n-1}^2$ for all $0 \le i \le n$ and $1 \le j \le n$. Moreover (see Wuthrich and Merz, 2008),

$$\begin{aligned} E[S_{i,j}|S_{i,j-1}] &= f_{j-1} \cdot S_{i,j-1}, \\ V[S_{i,j}|S_{i,j-1}] &= \sigma_{j-1}^2 \cdot S_{i,j-1}. \end{aligned}$$

Supposing that the development factors $f_0,\ f_1, \dots, f_{n-1}$ exist, then for all $0 \le i \le n$ and $1 \le j \le n$ the following holds

$$E[S_{i,j}|S_{i,0}, \dots, S_{i,j-1}] = E[S_{i,j}|S_{i,j-1}] = f_{j-1} \cdot S_{i,j-1},$$

while the accident years i are independent.

Similarly to the results of the previous section and Lemma (1.1), the following hold

- The estimators for f_j and σ_{j-1}^2 are
$$\hat{f}_j = \frac{\sum_{i=0}^{n-j-1} S_{i,j+1}}{\sum_{i=0}^{n-j-1} S_{i,j}},$$
$$\hat{\sigma}_j^2 = \frac{1}{n-j-1} \sum_{i=0}^{n-j-1} S_{i,j} \left(\frac{S_{i,j+1}}{S_{i,j}} - \hat{f}_j \right)^2.$$
- $\hat{f}_j | B_j$ are (un)conditionally unbiased estimators for f_j.
- $\hat{f}_0, \hat{f}_1, \ldots, \hat{f}_{n-1}$ are uncorrelated.

Using the individual development factors $F_{i,j+1} = \frac{S_{i,j+1}}{S_{i,j}}$ it is observed that $\hat{f}_j$ are averages of $F_{i,j+1}$ with weights, which means that
$$\hat{f}_j = \sum_{i=0}^{n-j-1} \frac{S_{i,j}}{\sum_{k=0}^{n-j-1} S_{k,j}} F_{i,j+1}.$$

Lemma 1.2. *Under Assumptions 1.3, the following hold:*

1. $\hat{\sigma}_j^2 | B_j$ *is unbiased,* $E[\hat{\sigma}_j^2 | B_j] = \sigma_j^2$.
2. $\hat{\sigma}_j^2$ *is unbiased,* $E[\hat{\sigma}_j^2] = \sigma_j^2$.

Proof. 1. We have that
$$E\left[\left(\frac{S_{i,k+1}}{S_{i,k}} - \hat{f}_k \right)^2 \middle| B_k \right] = E\left[\left(\frac{S_{i,k+1}}{S_{i,k}} - f_k \right)^2 \middle| B_k \right] -$$
$$- 2E\left[\left(\frac{S_{i,k+1}}{S_{i,k}} - f_k \right) (\hat{f}_k - f_k) \middle| B_k \right] + E\left[(\hat{f}_k - f_k)^2 \middle| B_k \right].$$

Each quantity on the right of the above equation will be separately calculated.
$$E\left[\left(\frac{S_{i,k+1}}{S_{i,k}} - f_k \right)^2 \middle| B_k \right] = Var\left[\frac{S_{i,k+1}}{S_{i,k}} \middle| B_k \right] = \frac{1}{S_{i,k}} \sigma_k^2.$$

Using the independence of the accident years we take
$$E\left[\left(\frac{S_{i,k+1}}{S_{i,k}} - f_k \right) (\hat{f}_k - f_k) \middle| B_k \right] = Cov\left[\frac{S_{i,k+1}}{S_{i,k}}, \hat{f}_k \middle| B_k \right]$$
$$= \frac{S_{i,k}}{\sum_{i=0}^{n-k-1} S_{i,k}} Var\left[\frac{S_{i,k+1}}{S_{i,k}} \middle| B_k \right]$$
$$= \frac{\sigma_k^2}{\sum_{i=0}^{n-k-1} S_{i,k}}.$$

For the last term of the equation we have

$$E\left[(\hat{f}_k - f_k)^2 \middle| B_k\right] = Var[\hat{f}_k|B_k] = \frac{\sigma_k^2}{\sum_{i=0}^{n-k-1} S_{i,k}}.$$

Putting all these terms together the following hold:

$$E\left[\left(\frac{S_{i,k+1}}{S_{i,k}} - \hat{f}_k\right)^2 \middle| B_k\right] = \sigma_k^2\left(\frac{1}{S_{i,k}} - \frac{1}{\sum_{i=0}^{n-k-1} S_{i,k}}\right)$$

and hence

$$E[\hat{\sigma}_k^2|B_k] = \frac{1}{n-k-1}\sum_{i=0}^{n-k-1} S_{i,k} E\left[\left(\frac{S_{i,k+1}}{S_{i,k}} - \hat{f}_k\right)^2 \middle| B_k\right] = \sigma_k^2.$$

2. The proof is similar to the proof of 1.

■

Moreover, the second moment of the $\hat{f}_k$ plays a significant role in order to derive an estimator for $E[\hat{f}_k^2|B_k]$

$$E[\hat{f}_k^2|B_k] = Var[\hat{f}_k|B_k] + f_k^2 = \frac{\sigma_k^2}{\sum_{i=0}^{n-k-1} S_{i,k}} + f_k^2.$$

So, in order to compute the prediction MSE of $\widehat{S_{i,n}}^{CL}$ we obtain

$$\begin{aligned} msep_{S_{i,n}|D_n}(\widehat{S_{i,n}}^{CL}) &= E\left[\left(\widehat{S_{i,n}}^{CL} - S_{i,n}{}^{CL}\right)^2 |D_n\right] \\ &= Var[S_{i,n}|D_n] + \left(\widehat{S_{i,n}}^{CL} - ES_{i,n}|D_n\right)^2. \end{aligned}$$

while for all accident years we have

$$msep_{\sum_i S_{i,n}|D_n}(\sum_{i=1}^{n} \widehat{S_{i,n}}^{CL}) = E\left[\left(\sum_{i=1}^{n} \widehat{S_{i,n}}^{CL} - \sum_{i=1}^{n} S_{i,n}{}^{CL}\right)^2 |D_n\right].$$

1.4.4 Bootstrap Reserving

Almost each loss reserving method make point estimates for the total reserve but the precision of these estimates is very important. The bootstrap methodology is a useful method which enables us to derive this precision for the point estimates. In non life reserving procedures, bootstrap techniques not only make estimations for the adequate reserve but also helps us to assess the variability of the predictions.

The bootstrap technique (Michael and Robert, 2011) is a nonparametric method which uses resampling methods. Bootstrapping is based on Monte Carlo approximation with the usage of computers and goes back to the beginning of the early 1940s.

However, 1979, Efron published a paper in the Annals of Statistics (Efron, 1979) where he defined a resampling procedure that he coined as bootstrap and approximated the jackknife technique (a resampling method which was developed by John Tukey). The basic idea of the bootstrap technique is to take several Monte Carlo samples with replacement from the original data and make estimations. If you randomly generate $K = 10000$ or 100000 bootstrap samples, then the distribution of the bootstrap estimates will approximate the bootstrap distribution for the estimate. The basic steps of the Monte Carlo approximation are:

Step 1: Generate a random sample with replacement from the empirical distribution for the data (bootstrap sample).

Step 2: Compute $T(F_n^*)$ the bootstrap estimate of $T(F)$ by replacing the original sample with a bootstrap sample.

Step 3: Repeat the first two Steps K times for large K.

If $T(F_n^*)$ converges to $T(F)$ then the bootstrap technique works but it is not guaranteed that it will work.

Kaas et al. (2009) and England and Verrall (1999, 2002) suggested a procedure in order to create bootstrap samples for the estimation of loss reserve. First of all, the Pearson residuals are

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

where $\mu_{i,j}$ should be replaced by its estimation $\hat{\mu}_{i,j}$. For run-off triangles where the data are not big enough, the residuals produce big sample bias. For that reason, the residuals have been adjusted (see England and Verrall (1999) and England (2002)) with a multiplication with the correction factor

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

where $N = \binom{n}{2}$ is the size of the data and $p = 2n - 1$ is the number of fitted coefficients. Another option is the usage of a factor to adjust each residual individually (see Pinheiro et al. (2003)) and standardize the Pearson residuals likewise linear regression does, as follows

$$r_{i,j}^P = \frac{r_{i,j}}{\sqrt{1 - h_{kk}}},$$

where h_{kk} is the diagonal of a hat matrix $\boldsymbol{H}$. In order to use the bootstrap method, the procedure can be applied by using the original Pearson residuals or any transformed residuals. The general notation r^* will be used for the selected residuals. Afterwards, the following bootstrap steps are performed many times, e.g. $m = 10000$ times.

Step 1: Take a sample with replacement using the selected residuals r^* and create a new sample of residuals $r(b)$. In case of fitting a GLM model, the fact that the last accident and development years have only one observation, will lead to zero residual.

Step 2: Use these new pseudo-residuals $r(b)$ to go back to the real data, which will be pseudo-data:

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

Step 3: Estimate the new total future reserve $R(b)$.

1.4.5 The Multiplicative Model and the CL Method

At this section the multiplicative model is represented and its relationship with the Chain Ladder method (Kremer, 1982). We will use the symbols x_i and y_j as the parameters in the multiplicative model. For that model we assume that the mean of the claims can be presented as

$$E(C_{i,j}) = x_i\, y_j, \tag{1.6}$$

where $C_{i,j}$ is the stochastic variable of the incremental claims, x_i, y_j which are unknown parameters and the following assumption holds

$$y_1 + y_2 + \ldots + y_n = 1. \tag{1.7}$$

By the definition in (1.6) and the assumption (1.7), which means that the sum of y_j equals one, the parameter x_i is

$$x_i = E(S_{in}),$$

where S_{in} are the cumulative claims. Making it more explicit, (1.6) suggests that the forecast of the incremental payments can be computed as a product of an accident year coefficient x_i and a development year coefficient y_j. The fact that x_i is the estimated total claim, we can assume that the sum over j of y_j is 1. In the case of $C_{i,j}$ to be number of payments, then y_j is the empirical frequency of a claim appeared in accident year i and developed in development year j.

(1.1) shows a stochastic model and is similar to the multiplicative model (Mack, 1994). It can be proved by finding the parameters x_i and y_j. We take:

$$\begin{aligned} E(C_{i,j}) &= E(S_{i,j}) - E(S_{i,j-1}) \\ &= (f_{j+1} f_{j+2} \cdots f_n)^{-1} E(S_{in}) - (f_j f_{j+1} \cdots f_n)^{-1} E(S_{in}) \\ &= E(S_{in}) \Big((f_{j+1} f_{j+2} \cdots f_n)^{-1} - (f_j f_{j+1} \cdots f_n)^{-1} \Big). \end{aligned} \tag{1.8}$$

The next step is to find the variables y_j in order to make (1.8) be equal to $x_i\ y_j$. The variable x_i has already been recognized as $x_i = E(S_{in})$. So it is clear that

$$y_j = (f_{j+1} f_{j+2} \cdots f_n)^{-1} - (f_j f_{j+1} \cdots f_n)^{-1}. \tag{1.9}$$

For the development years $2 \leq j < n$ the variables y_j are

$$\begin{aligned} y_1 &= (f_2 f_3 \cdots f_n)^{-1}, \\ y_j &= (f_{j+1} f_{j+2} \cdots f_n)^{-1} - (f_j f_{j+1} \cdots f_n)^{-1}, \\ y_n &= 1 - (f_n)^{-1}. \end{aligned}$$

We observe that $y_j \geq 0$ if $f_j \geq 1$ for $j = 1, \ldots, n$. Also, if in (1.9), we sum up its terms, then $\sum_{j=1}^{n} y_j = 1$. So, the definition of y_j seems to be a good choice. The cumulative claim for the accident year i and the development year j can be written as a sum of the incremental claims, and using y_j makes it easy to see that for the accident year $i = 1, \ldots, n$

$$\begin{aligned} E(S_{in}) &= x_i(y_1 + \ldots + y_n) = x_i y_1 + x_i y_2 + \ldots + x_i y_n \\ &= E(C_{i1}) + E(C_{i2}) + \ldots + E(C_{in}). \end{aligned} \tag{1.10}$$

Hence, by appropriately choosing x_i and y_j, it is clear that the simple stochastic model is equivalent to the multiplicative model (Mack 1994b).

The development factor can be derived by using the identities from the multiplicative model. For $2 \leq j \leq n$,

$$f_j = \frac{E(S_{i,j})}{E(S_{i,j-1})} = \frac{x_i(y_1 + y_2 + \ldots + y_j)}{x_i(y_1 + y_2 + \ldots + y_{j-1})} = \frac{y_1 + y_2 + \ldots + y_j}{y_1 + y_2 + \ldots + y_{j-1}}. \tag{1.11}$$

This development factor does not have the same appearance as the Chain Ladder development factor, but it is actually the same. This can be proved by induction.

1.4.6 The Poisson Model and the CL Method

The Poisson model can be presented as a particular case of the multiplicative method. It has an equivalent format but it is also supposed that the incremental claims $C_{i,j}$ have the Poisson distribution. According to Verral (2000), the Poisson model gives the same estimations for the reserves as the CL method. This is true when maximum likelihood estimators (MLE) are used.

Lets suppose that $C_{i,j}$ are independent and follow the Poisson distribution with $E(C_{i,j}) = x_i y_j$ and $\sum_{j=1}^{n} y_j = 1$. From the multiplicative model, the parameter x_i has already been decided. It has been mentioned that $x_i = E(S_{in})$. The parameter x_i is the expected value of the cumulative claims. We have

$$E(C_{i,j}) = x_i\ y_j = E(S_{i,j})y_j = \frac{E(S_{i,n-i+1})y_j}{\sum_{j=1}^{n-i+1} y_j} = \frac{z_i y_j}{s_{n-i+1}}, \tag{1.12}$$

where $z_i = E(S_{i,n-i+1})$ and $s_k = \sum_{j=1}^{k} y_j$. Since y_j can be interpreted as the percentage of the total claims in development year j, it is logical that $E(S_{i,n-i+1})$ divided by the proportion of claims until $j = n - i + 1$ equals $E(S_{i,n})$.

Equation (1.12) can be written as a formula for predicting the expectation of the total claim $E(S_{i,n})$. Approximating $E(S_{i,n})$ with $\widehat{S}_{i,n}$, the equation is:

$$\widehat{S}_{i,n} = E(S_{i,n}) = x_i = \frac{z_i}{\sum_{k=1}^{n-i+1} y_k} = \frac{z_i}{1 - \sum_{k=n-i+2}^{n} y_k}. \tag{1.13}$$

Verral (2000) claimed that this equation is equivalent to the CL estimator:

$$\widehat{S}_{n-j+1,n} = d_{n-j+1,j}\widehat{f}_{j+1}\widehat{f}_{j+2}\cdots\widehat{f}_n \quad \text{where} \quad \widehat{f}_j = \frac{\sum_{i=1}^{n-j+1} d_{i,j}}{\sum_{i=1}^{n-j+1} d_{i,j-1}}. \tag{1.14}$$

To make it clear that (1.13) and (1.14) are in fact equivalent, it is natural to look for the estimators of the unknown coefficients in (1.13). The maximum likelihood function will be used to find the estimators. In this case the observations $c_{i,j}$ are considered to be known and the parameters are considered as variables. Since $c_{i,j} = x_i y_j = z_i y_j / s_{n-i+1}$, the ML function is:

$$L = \prod_{i=1}^{n} \prod_{j=1}^{n-i+1} \frac{(z_i y_j / s_{n-i+1})^{c_{i,j}} e^{-z_i y_j / s_{n-i+1}}}{c_{i,j}!} \tag{1.15}$$

Further calculations show that

$$L = \prod_{i=1}^{n} \left[\frac{z_i^{d_{i,n-i+1}} e^{-z_i}}{d_{i,n-i+1}!} \left(\frac{d_{i,n-i+1}!}{\prod_{j=1}^{n-i+1} c_{i,j}!} \prod_{j=1}^{n-i+1} \left(\frac{y_j}{s_{n-i+1}} \right)^{c_{i,j}} \right) \right]. \tag{1.16}$$

It should be marked that in order to take the final equation, the below calculations were used:

- $\prod_{j=1}^{n-i+1}\left[e^{(\frac{y_j}{s_{n-i+1}})}\right] = e^{(\frac{\sum_{j=1}^{n-i+1} y_j}{s_{n-i+1}})} = e^{\frac{s_{n-i+1}}{s_{n-i+1}}} = e$
- $\prod_{j=1}^{n-i+1}\left[z_i^{c_{i,j}}\right] = z_i^{c_{i1}} z_i^{c_{i2}} \cdots z_i^{c_{i,n-i+1}} = z_i^{\sum_{j=1}^{n-i+1} c_{i,j}} = z_i^{d_{i,n-i+1}}$
- We multiplied and divided with the quantity $d_{i,n-i+1}!$

However, the equation (1.16) can be written as a product of two quantities

$$L = L_c L_d, \tag{1.17}$$

where

$$L_c = \prod_{i=1}^{n}\left[\left(\frac{d_{i,n-i+1}!}{\prod_{j=1}^{n-i+1} c_{i,j}!}\prod_{j=1}^{n-i+1}\left(\frac{y_j}{s_{n-i+1}}\right)^{c_{i,j}}\right)\right] \tag{1.18}$$

while

$$L_d = \prod_{i=1}^{n}\left[\frac{z_i^{d_{i,n-i+1}} e^{-z_i}}{d_{i,n-i+1}!}\right]. \tag{1.19}$$

Theorem 1.1. *The quantity L_c is the conditioned maximum likelihood function where $C_{i,j}$ conditioning $d_{i,n-i+1}$ follow the Multinomial distribution with probabilities*

$$y_{i(j)} = \frac{y_j}{s_{n-i+1}}.$$

Proof. The probability $y_{i(j)}$ is that of a payment that occurred at the year i to be mentioned at the year j. By the equation (1.16) we take

$$L = \prod_{i=1}^{n}\left[\frac{z_i^{d_{i,n-i+1}} e^{-z_i}}{d_{i,n-i+1}!}\left(\frac{d_{i,n-i+1}!}{\prod_{j=1}^{n-i+1} c_{ij}!}\prod_{j=1}^{n-i+1}\left(y_{i(j)}\right)^{c_{i,j}}\right)\right] \tag{1.20}$$

Let's suppose that $C_{i,j}$ for $i = 1,\dots,n$ and $j = 1,\dots,n-i+1$, are independent random variables which follow a Poisson($y_{i(j)}$) with $y_{i(j)} = \frac{y_j}{\sum_{k=1}^{n-i+1} y_k}$. Since this is a parameter, the notation changes so as to make it more tangible. So, $y_{i(j)} = p_{i(j)} = \frac{p_j}{\sum_{k=1}^{n-i+1} p_k}$. Then $S_{i,j}$, for $i = 1,\dots,n$ and $j = 1,\dots,n-i+1$, are independent random variables which follow a Poisson distribution. This stands because $S_{i,j}$ is a descrete sum of independent Poisson random variables.

The conditioned probability distribution function is

$$\begin{aligned} f_{C_{i,j}|S_{i,n-i+1}}(c_{i,j}|d_{i,n-i+1}) &= \left(\prod_{j=1}^{n-i+1} \frac{p_j^{c_{i,j}} e^{-p_j}}{c_{i,j}!}\right)\left(\frac{(p_1+\ldots+p_{n-i+1})^{d_{i,n-i+1}} e^{-(p_1+\ldots+p_{n-i+1})}}{p_{i,n-i+1}!}\right) \\ &= \frac{d_{i,n-i+1}!}{\prod_{j=1}^{n-i+1} c_{i,j}!}\left(\frac{p_1}{\sum_{j=1}^{n-i+1} p_j}\right)^{c_{i1}} \cdots \left(\frac{p_{n-i+1}}{\sum_{j=1}^{n-i+1} p_j}\right)^{c_{i,n-i+1}} \\ &= \frac{d_{i,n-i+1}!}{\prod_{j=1}^{n-i+1} c_{i,j}!}(p_{i(1)})^{c_{i1}}(p_{i(2)})^{c_{i2}} \cdots (p_{i(n-i+1)})^{c_{i,n-i+1}}. \end{aligned}$$

From the last expression it is easy to recognize that the distribution of $C_{i,j}$ conditioned $d_{i,n-i+1}$ is a Polynomial distribution. ■

Correspondingly, for the quantity L_d it is easier to see that it is the maximum likelihood function when $S_{i,n-i+1}$ follows the Poisson distribution(z_i). So, it is feasible to find the MLE for z_i. Specifically the MLE of z_i is $d_{i,n-i+1}$ because $S_{i,n-i+1}$ follows the Poisson distribution.
Using the MLE of z_i, the MLE for the final ultimate claims is

$$\widehat{S}_{in} = \frac{d_{i,n-i+1}}{1-\sum_{k=n-i+2}^{n} y_k}. \tag{1.21}$$

For the accident year $n-j+1$ this expression is

$$\widehat{S}_{n-j+1,n} = \frac{d_{n-j+1,j}}{1-\sum_{k=j+1}^{n} y_k}. \tag{1.22}$$

In the equation (1.22), y_k is the only unknown parameter. It can be estimated by using the MLE based on the function L, but the quantity L_c can also be used. The logarithm of L_c is calculated and the final expression is derived by y_k for $k = 1, \ldots, n$. Nevertheless, it does not have an analytic solution, so Renshaw (1998) suggested a procedure. A value for the parameter $\widehat{y}_n$ is given, afterwards the value $\widehat{y}_{n-1}$ is decided and so on. The calculations that are used in order to find $\widehat{y}_n$ and the general formula about $\widehat{y}_j$ is shown below:

$$\begin{aligned} \log(L_c) = l_c &\propto \sum_{i=1}^{n}\sum_{j=1}^{n-i+1} \log\left(\frac{y_j}{\sum_{k=1}^{n-i+1} y_k}\right)^{c_{i,j}} \\ &= \sum_{i=1}^{n}\sum_{j=1}^{n-i+1} c_{i,j}\Big(\log(y_j) - \log\Big(\sum_{k=1}^{n-i+1} y_k\Big)\Big) \\ \frac{\partial l_c}{\partial y_n} &= 0 \Rightarrow \frac{c_{1n}}{\widehat{y}_n} - \sum_{j=1}^{n} \frac{c_{1j}}{\sum_{k=1}^{n} \widehat{y}_k} = \frac{c_{1n}}{\widehat{y}_n} - \sum_{j=1}^{n} \frac{c_{1j}}{1} = 0. \end{aligned}$$

Using the last equation we obtain the mle

$$\widehat{y}_n = \frac{c_{1n}}{\sum_{j=1}^{n} c_{1j}} = \frac{c_{1n}}{d_{1n}}. \tag{1.23}$$

Correspondingly,

$$\begin{aligned}\frac{\partial l_c}{\partial y_n} = 0 \Rightarrow \sum_{i=1}^{n-j+1} \left(\frac{c_{i,j}}{\widehat{y}_j} - \sum_{j=1}^{n-j+1} \frac{c_{i,j}}{\sum_{k=1}^{n-i+1} \widehat{y}_k} \right) \\ = \frac{\sum_{k=1}^{n-i+1} c_{i,j}}{\widehat{y}_j} - \sum_{i=1}^{n-j+1} \left(\frac{d_{i,n-i+1}}{\sum_{k=1}^{n-i+1} \widehat{y}_k} \right) = 0.\end{aligned}$$

Finally, from the last expression it holds that

$$\widehat{y}_j = \frac{\sum_{i=1}^{n-j+1} c_{i,j}}{\sum_{i=1}^{n-j+1} \left(\frac{d_{i,n-i+1}}{\sum_{k=1}^{n-i+1} \widehat{y}_k} \right)} = \frac{c_{1j} + \ldots + c_{n-j+1,j}}{d_{1,n} + \frac{d_{2,n-1}}{1-\widehat{y}_n} + \ldots + \frac{d_{n-j+1,j}}{1-\widehat{y}_{j+1}-\ldots-\widehat{y}_n}}. \tag{1.24}$$

Observing the equation (1.24) it is obvious that with this iterative procedure all estimations for the quantities y_j can be found. The next step is to find an expression for the development factor $\widehat{f}_j$ using the MLE of $\widehat{y}_j$. According to (1.14),

$$\widehat{f}_{j+1}\widehat{f}_{j+2} \cdots \widehat{f}_n = \frac{\widehat{S}_{n-j+1,n}}{d_{n-j+1,j}}.$$

Inserting the expression for the quantity $\widehat{S}_{n-j+1,n}$ from the equation (1.22) and using the estimator $\widehat{y}_j$ instead of y_j, the product of the development factors is:

$$\widehat{f}_{j+1}\widehat{f}_{j+2} \cdots \widehat{f}_n = \frac{1}{1 - \widehat{y}_{j+1} - \widehat{y}_{j+2} - \ldots - \widehat{y}_n}, \tag{1.25}$$

and

$$\widehat{f}_j\widehat{f}_{j+1} \cdots \widehat{f}_n = \frac{1}{1 - \widehat{y}_j - \widehat{y}_{j+1} - \ldots - \widehat{y}_n}. \tag{1.26}$$

From the equation (1.25), an expression for $1 - \widehat{y}_{j+1} - \widehat{y}_{j+2} - \ldots - \widehat{y}_n$ can be found and used in (1.26). So,

$$\widehat{f}_j\widehat{f}_{j+1} \cdots \widehat{f}_n = \frac{1}{\frac{1}{\widehat{f}_{j+1}\widehat{f}_{j+2} \cdots \widehat{f}_n} - \widehat{y}_j}. \tag{1.27}$$

Finally, an estimator for the development factor can be found according to the next equation

$$\widehat{f}_j = \frac{1}{1 - \widehat{y}_j \widehat{f}_{j+1} \widehat{f}_{j+2} \cdots \widehat{f}_n}. \tag{1.28}$$

Using the MLE of y_n from the equation (1.23), the development factors are

$$\widehat{f}_n = \frac{1}{1 - \widehat{y}_n} = \frac{1}{1 - \frac{c_{1n}}{d_{1n}}} = \frac{d_{1n}}{d_{1n} - c_{1n}} = \frac{d_{1n}}{d_{1,n-1}}. \tag{1.29}$$

The estimator in (1.29) is the same with the one that was found at CL method when $j = n$. In order to prove that the other LDF's of the Poisson model are the same with them of the CL method, induction can be used. Since it is proved for $j = n$, the first step of the induction has already been done. The next step is to find a general formula for the development factors $\widehat{f}_j$. In order to do this, the expression for $\widehat{y}_j$ must become simpler. The equation (1.24) gives an expression for $\widehat{y}_j$ but the fractions at the denominators can be modified using the equations (1.25) and (1.26). So,

$$\widehat{y}_j = \frac{c_{1j} + c_{2j} + \ldots + c_{n-j+1,j}}{d_{1n} + d_{1,n-1}\widehat{f}_n + d_{n-j+1}\widehat{f}_{j+1}\widehat{f}_{j+2} \ldots \widehat{f}_n}. \tag{1.30}$$

If the equation of $\widehat{y}_j$ is examined, it is clear that it is a percentage of the final claims. The numerator counts the claims during the whole accident years for the development year j but the decominator counts the estimated values of the final claims during the same accident years. The equation (1.28) is a general expression for $\widehat{f}_j$. The final expression for $\widehat{y}_j$ is used in (1.28). So,

$$\widehat{f}_j = \frac{1}{1 - \frac{c_{1j} + c_{2j} + \ldots + c_{n-j+1,j}}{d_{1n} + d_{1,n-1}\widehat{f}_n + d_{n-j+1}\widehat{f}_{j+1}\widehat{f}_{j+2} \ldots \widehat{f}_n} \widehat{f}_{j+1}\widehat{f}_{j+2} \ldots \widehat{f}_n}. \tag{1.31}$$

This is the general formula. It has already been proved that $\widehat{f}_n$ is the development factor of the CL method.

As a part of the induction it is supposed that for $k = j+1, \ldots, n$, $\widehat{f}_k$ is equal to the corresponding development factor of the CL method. At the last step we must prove that $\widehat{f}_k$ is equal to the corresponding development factor of the CL method but for $k = j$. The denominator in (1.31) must be simplified showing that

$$d_{1n} + d_{1,n-1}\widehat{f}_n + d_{n-j+1}\widehat{f}_{j+1}\widehat{f}_{j+2} \ldots \widehat{f}_n = \widehat{f}_{j+1}\widehat{f}_{j+2} \ldots \widehat{f}_n \sum_{i=1}^{n-j+1} d_{i,j}. \tag{1.32}$$

But this is true for $j = n - 1$ because

$$\begin{aligned} d_{1n} + d_{2,n-1}\widehat{f}_n = d_{1n} + d_{2,n-1}\frac{d_{1n}}{d_{1,n-1}} &= \frac{d_{1n}}{d_{1,n-1}}\Big(d_{1,n-1} + d_{2,n-1}\Big) \\ &= \widehat{f}_n\Big(d_{1,n-1} + d_{2,n-1}\Big) \end{aligned}$$

Similarly for $j = n - 2$, a similar equation emerges

$$\begin{aligned} d_{1n} + d_{2,n-1}\widehat{f}_n + d_{3,n-2}\widehat{f}_{n-1}\widehat{f}_n &= \widehat{f}_n\Big(d_{1,n-1} + d_{2,n-1} + d_{3,n-2}\widehat{f}_{n-1}\Big) \\ &= \widehat{f}_n\Big(d_{1,n-1} + d_{2,n-1}\Big)\frac{d_{1,n-1} + d_{2,n-1}}{d_{1,n-2} + d_{2,n-2}} + d_{3,n-2}\widehat{f}_{n-1} \\ &= \widehat{f}_{n-1}\widehat{f}_n\Big(d_{1,n-2} + d_{2,n-2} + d_{3,n-2}\Big). \end{aligned}$$

Doing this procedure $n - j$ times, we can prove the equation (1.32), and the equation for $\widehat{f}_j$ in (1.31) can be shortened as follows

$$\widehat{f}_j = \frac{1}{1 - \dfrac{c_{1j} + c_{2j} + \ldots + c_{n-j+1,j}}{\widehat{f}_{j+1}\widehat{f}_{j+2}\ldots\widehat{f}_n \sum_{i=1}^{n-j+1} d_{i,j}}\widehat{f}_{j+1}\widehat{f}_{j+2}\ldots\widehat{f}_n}.$$

From all the above, it is shown that

$$\widehat{f}_j = \frac{\sum_{i=1}^{n-j+1} d_{i,j}}{\sum_{i=1}^{n-j+1} d_{i,j-1}} = \frac{\sum_{i=1}^{n-j+1} d_{i,j}}{\sum_{i=1}^{n-j+1} d_{i,j} - \sum_{i=1}^{n-j+1} c_{i,j}}.$$

So, the induction method has been completed because $\widehat{f}_j$ is the same with the one that was calculated at the CL method. This means that if the maximum likelihood estimators are used for the Poisson model, then the same estimations will be used at the CL method.

1.4.7 Run-off Triangle Interdependence and Cross Dependence

Within a run-off triangle there are three kinds of certain dependencies that can arise.

Horizontal Dependence: The development time of an incurred accident may create horizontal trends in a run-off triangle. This happens because the payments logically decrease as the years pass. Of course, the payments' development depends on the policy and the Line of Business while the horizontal trend is apparent.

Vertical Dependence: On the other hand, the existence of vertical trends may appear between development years. Within the insurance industry, a cyclical pattern, may be obvious and can create a vertical seasonal trend (Doherty et al., 1995).

Calendar Year Dependence: Payments of the same calendar year may have effects at a run-off triangle because they are mainly macroeconomic effects such as inflation. The closing claims of a certain year may also effect the development of the triangle.

Table 1.2: Lines of Business in Non-Life insurance

LoB number	**Category name**
1	Motor, vehicle liability
2	Motor, other classes
3	Marine, aviation, transport (MAT)
4	Fire and other property damage
5	Third-party liability
6	Credit and suretyship
7	Legal expenses
8	Assistance
9	Miscellaneous
10	Non-proportional reinsurance - property
11	Non-proportional reinsurance - casualty
12	Non-proportional reinsurance - MAT

Source: CEIOP'S Advice for Level 2 Implementing Measures on Solvency II: SCR Standard Formula Calibration of Non-life Underwriting Risk.

In general insurance, the triangles are classified by LoB (see Table 1.2). So, cross dependence indicates the dependence between the triangles. In general, a large incident may effect many LoB at the same time. For that reason, this kind of dependency should be taken into consideration during the aggregated reserving procedure. The existence of the interdependence makes more obstacles to capture the cross-dependence between the triangles. In cases that the interdependence is not computed properly, the cross-correlation is overestimated.

CEIOPS has published a Quantitative Impact Study (QIS4) with a description on the derivation of the correlations between the different LoBs. This study is available on CEIOPS' website (Table 1.3).

Table 1.3: Correlation of LoBs

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
1	1											
2	0.5	1										
3	0.5	0.25	1									
4	0.25	0.25	0.25	1								
5	0.5	0.25	0.25	0.25	1							
6	0.25	0.25	0.25	0.25	0.5	1						
7	0.5	0.5	0.25	0.25	0.5	0.5	1					
8	0.25	0.5	0.5	0.5	0.25	0.25	0.25	1				
9	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	1			
10	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.5	0.25	1		
11	0.25	0.25	0.25	0.25	0.5	0.5	0.5	0.25	0.25	0.25	1	
12	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.5	0.25	0.25	1

Source: CEIOP'S Advice for Level 2 Implementing Measures on Solvency II: SCR Standard Formula Calibration of Non-life Underwriting Risk.

CHAPTER 2

Robust Loss Reserving for Log-linear Regression Model

It is well known that the presence of outlier events may overestimate or underestimate the overall reserve in case of using the chain-ladder method. The lack of robustness of loss reserving estimators is a very important matter because the appearance of outlier events (including large claims or catastrophic events) can offset the result of the ordinary chain ladder technique and perturb the reserving estimation. A solution to this, is to apply robust statistical procedures to the loss reserving estimation, which are insensitive to the occurrence of outlier events in the data. Robust log-linear and ANOVA models can be used to the analysis of loss reserving by using different type of robust estimators, such as LAD-estimators, M-estimators, LMS-estimators, LTS-estimators, MM-estimators (with initial S-estimate) and Adaptive-estimators.

2.1 Review on Robust Estimation

The presence of outliers due to large claims or catastrophic events is a special problem in loss reserving calculation. Outliers can be described as points which do not follow the trend of the majority of the data. The problem appears if a trend (due to an outlier event) that appeared in one of the development years in a chain ladder setting carried on for the next years resulting in an overestimation or underestimation of claim reserves. In particular, excess claims (large claims) lead to an unsatisfactory behavior of chain ladder methodology. The purpose is to robustify the claim reserves calculations using robust estimators. In simple regression (two-dimensional case), it is easy to detect outlier events just by plotting the observations. This is no longer possible in the log-linear multiple regression. So, in practice, one needs a procedure that is able to lessen the impact of outliers. Kremer (1997) incorporated the ideas of robust statistics into loss reserving techniques by using the lagfactor-method (or link-ratio method). Verdonck et al. (2009) created a technique for detecting outlying observations in a run-off triangle of claim amounts and solved the problem of non-robustness of the chain ladder by replacing the mean by the median. Verdonck and Debruyne (2011) based on the influence function approach presented a diagnostic tool for highlighting the influence of every individual claim on the classical chain-ladder estimates. They considered the chain ladder method as generalized linear models (GLM) and obtained robust estimates of GLM in a chain ladder framework. Busse et al. (2010) designed a filter for outliers and large jumps, and presented a robust version of Mack's variance estimator. They verified the reliability of their methods with several loss triangles. Venter and Tampubolon (2010) presented an introduction of robust methods for loss reserving and compared development triangle based on the sensitivity of the reserve estimates. At this work a class of robust estimators is applied to a chain ladder procedure where the data is in a log linear form and that was transformed into a two-way analysis of variance. This class of estimators includes robust estimators that simultaneously attain maximum breakdown point (BP) and full asymptotic efficiency under error normality. At this robust loss reserving estimation we initially ignore the bias present due to the robustification of the large claims, but add in a second stage, a share of the excess (correction term) to ultimate claims, to obtain a final unbiasedness. This robust log-linear regression estimation can provide quite good claim reserves estimates by guaranteeing the recovery of ultimate claims. Of course, these robust estimators can be embedded within several loss reserving techniques providing reliable claims reserves estimation. Very often, assumptions, made in statistics, i.e. normality, linearity, independence are at most approximations of reality. Robust regression models are useful for filtering linear relationships when the random variation in the data is not normal or when the data contain significant outliers (see Hampel et al.,1986).

Let F_n be the empirical distribution function of a sample $X_1, \ldots, X_n$. Formally F_n is given by $(1/n)\sum_{i=1}^{n}\Delta_{x_i}$, where Δ_x has point mass 1 at x. As an estimator of a parameter θ we consider real-valued statistics $T_n = T(F_n)$. Sometimes also referred to as $T_n(X_1, ..., X_n)$ by extension of the notation. In a broader sense, an estimator can be viewed as a sequence of statistics $\{T_n; n \geq 1\}$, one for each possible sample size n.

Definition 1: The influence function (IF) of T at F is given by

$$IF(x; T, F) = \lim_{t \to 0} \frac{T(F_t) - T(F)}{t}, \quad (2.1)$$

with $T(F_t) = T[(1-t)F + t\Delta_x]$. If we replace F by F_{n-1}, where F_n is the empirical distribution estimate of F generated by the sample $(x_1, ..., x_n)$, and put $t = \frac{1}{n}$, we realize that the $IF(x; T, F_{n-1})$ measures approximately n times the change in T caused by an additional observation in x when T is applied to a large sample of size $n-1$ (see Hampel et al., 1986). The breakdown point (BP) provides us with a rough upper bound on the fraction of outliers for which such a linear approximation might be useful, (see Rousseeuw and Leroy, 1987).

Definition 2: The finite-sample breakdown point of the estimator T_n at an observed sample $\boldsymbol{X} = (X_1, ..., X_n)$ is defined as

$$\varepsilon_n^\star(T, \boldsymbol{X}) = \min[\frac{m}{n}; \text{bias}(m; T_n, x) < \infty] \ ,$$

where bias$(m; T_n, x)$ is the maximum bias that can be caused by a contamination (presence of outliers) and m is the number of original points replaced by arbitrary values. In other words $\varepsilon_n^\star(T_n, x)$ is the smallest fraction of contamination that can cause the estimator T to take values arbitrarily far from T_n. The breakdown point usually does not depend on the sample value x, but depends only slightly on the sample size n.

Definition 3: The gross-error sensitivity of T at F is defined by

$$\gamma^\star = \sup_x |IF(x; T, F)| \ ,$$

where the supremum is taken over all x, where $IF(x; T, F)$ exists. The $\gamma^\star$ describes the maximal effect on T induced by a small contamination of the data set.

2.1.1 LAD and M-Estimators

The idea of least absolute deviation (LAD) also known as L_1 regression is actually older than that of least squares. It is clear that outliers have a very large influence on Ordirary Least Squares (OLS) because the residuals r_i are squared. Estimates are found by minimizing the sum of the absolute values of the residuals

$$\min_{\hat{\beta}} \sum_{i=1}^{n} |r_i|, \quad (2.2)$$

where $r_i = y_i - \boldsymbol{x}_i^T \boldsymbol{\beta}$ is the i^{th} residual.

The LAD is a case of the general quantile regression. Unfortunately the BP of L_1 regression is no better than 0%. The BP of a regression estimator is the largest proportion of the data which can be replaced by large values (outlier events) before the estimator breaks down. As its name implies, L_1 regression finds the coefficients estimate that minimizes the sum of the absolute values of the residuals.

Moreover, M-estimators are generalization of maximum likelihood estimator proposed by Huber (1973), who suggested that we obtain M-estimators as solutions of the following minimization problem,

$$\min_{\hat{\beta}} \sum_{i=1}^{n} \rho\left(r_i\right), \tag{2.3}$$

where r_i is the i^{th} residual, ρ is a symmetric function with unique minimum at zero. Differentiating this expression with respect to the regression coefficients $\hat{\boldsymbol{\beta}}$ yields, $\sum_{i=1}^{n} \psi\left(r_i\right) \boldsymbol{x}_i = 0$, where ψ is the derivative of ρ and $\boldsymbol{x}_i$ is the row vector of explanatory variables of the i^{th} case. In practice one has to standardize the residuals by means of some estimate of S, yielding

$$\sum_{i=1}^{n} \psi\left(\frac{r_i}{S}\right) \boldsymbol{x}_i = 0, \tag{2.4}$$

where S is a scale parameter and must be estimated simultaneously. In practice, it is advisable to use $S = med\{|r_i|\}$ as an initial value. The advantage of M-estimates is that they can be computed in much less time than other robust estimates. The disadvantage is that they are sensitive to high leverage points and they don't enjoy high breakdown point (BP). The BP of M-estimators are 0% (see Rousseeuew and Leroy 1987, p. 145).

The location-scale M-estimators of $\boldsymbol{\beta}$, with an appropriate choice of ψ, may attain a high efficiency and at the same time be robust against large residuals. But these estimators are not robust to outliers in the design matrix space, i.e. if the explanatory variables are random or otherwise subject to errors the classical M-estimators may be unreliable. In this case the domain of the ψ function has been enlarged to include the design points, as well as the residuals. The influence function of the Huber M-estimator (ignoring the scale) is defined as

$$IF(\boldsymbol{x}^T, y; T, F) = \frac{\psi_c(r)}{E\psi_c'(r)}(E\boldsymbol{x}\boldsymbol{x}^T)^{-1}\boldsymbol{x}. \tag{2.5}$$

The first part of the influence function in (2.5) is called the influence of the residuals and is bounded, but the second part that is called the influence of position in factor space is unbounded. Thus, a single x_i, which is an outlier in the factor space, will almost completely determine the fit. In this case the Huber estimator and all estimators defined through (2.4), including L_1, are only the first step in the robustification of the regression estimator (see Hampel et al. 1986, p. 313).

In order to find M-estimates for a regression model, an iterative algorithm is necessary. A closed form (such as OLS) is impossible because the residuals cannot be found until the model is fitted, and moreover the coefficient estimates cannot be found without knowing the values of the residuals. For that reason, iteratively reweighted least squares (IRLS) will be used:

Step 1: Before starting the iterations, set $i = 0$ and an OLS regression is fitted to the data. Then, the initial estimates of the regression coefficients $\hat{b}^{(0)}$ are found.

Step 2: The residuals $r_i^{(0)}$ of the model are extracted from the preliminary OLS regression and then used to calculate initial estimates for the weights of the regression.

Step 3: Then, a weight function is selected and applied to the initial OLS residuals so as to create some preliminary weights $w[r_i^{(0)}]$.

Step 4: If $\mathbf{W}$ represents the $n \times n$ diagonal matrix of individual weights, at the first iteration, $i = 1$, the solution of the regression coefficients is

$$\hat{b}^{(1)} = \left(\mathbf{X}'\mathbf{W}\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{W}\mathbf{y}. \tag{2.6}$$

Step 5: The procedure continues by using the residuals from the initial WLS in order to calculate the new weights, $w[r_i^{(2)}]$.

Step 6: The new weights $w[r_i^{(2)}]$ are used in the next iteration, $i = 2$ and using (2.6) in order to produce the new regression coefficient estimates $\hat{b}^{(2)}$.

Step 7: Steps 4-6 are repeated until the estimate of the coefficients converges.

2.1.2 LMS Estimators and LTS Estimators

A robust equivariant regression estimator that first attained the maximum asymptotic BP=0.5 is the least median of squares (LMS) estimator proposed by Hampel in 1975 and further developed by Rousseeuw in 1984 (see Rousseeuw and Leroy 1987). The LMS is given by

$$\min_{\hat{\beta}} med_i \left(y_i - \sum x_{ij}\beta_j\right)^2 = \min_{\hat{\beta}} med_i r_i^2. \tag{2.7}$$

Apart from the regression coefficients, also the scale parameter S (the dispersion of errors e_i) has to be estimated in a robust way. The idea is that by replacing the sum with the more robust median, the resulting estimator will be resistant to outliers. The above estimator is very robust with respect to outliers in y as well as outliers in x. The LMS is equivariant with respect to linear transformation on the explanatory variables.

Unfortunately, the LMS performs poorly from the point of view of asymptotic efficiency (at best a relative efficiency of 37%) and it does not have a well-defined influence function because of its convergence rate of $n^{-1/3}$ (Rousseeuw, 1984). Rousseeuw in order to improve the asymptotic efficiency of LMS introduced the least trimmed squares (LTS) estimators given by (see Rousseeuw and Leroy, 1987)

$$\min \sum_{i=1}^{h} (r^2)_{i:n}, \tag{2.8}$$

where $(r^2)_{1:n} \leq ... \leq (r^2)_{n:n}$ are the ordered squared residuals (note that the residuals are first squared and then ordered). The best robustness properties are achieved when h is approximately $n/2$, in which case the BP attains 50%. Although highly resistant, LTS suffers badly in terms of relative efficiency at about 8%.

2.1.3 S-Estimators and MM-Estimators

The S-Estimators introduced by Rousseeuw and Yohai (1984) are a generalization of LMS and LTS and have the same asymptotic properties as corresponding to M-estimators and also can handle 50% of the outliers present in the data and are the first high BP regression to achieve the usual $n^{1/2}$- consistency under appropriate regularity conditions. They are defined by minimization of the dispersion of the residuals

$$\min_{\theta} s(r_1(\theta), ..., r_n(\theta)), \tag{2.9}$$

with final scale estimate $\widehat{\sigma} = s(r_1(\widehat{\theta}), ..., r_n(\widehat{\theta}))$. However, Hössjer (1992) showed that S-estimators cannot achieve simultaneously high BP and high efficiency under the normal model. The scale estimator can be obtained through the following dispersion minimization problem

$$\hat{\beta}_j = \arg\min_{\hat{\beta}} S\{r_1(\beta), ..., r_n(\beta)\}, \tag{2.10}$$

subject to

$$\frac{1}{n}\sum_{i=1}^{n} \rho\left(\frac{r_i(\beta)}{S(\beta)}\right) = K, \tag{2.11}$$

with $r_i(\beta) = y_i - \boldsymbol{x}_i^T \boldsymbol{\beta}$ and $K = E_{\Phi}(\rho)$, which assures the consistency of S at the normal distribution Φ. For the initial S-estimate Tukey's bisquare ρ function has been used. This is defined as (Tukey, 1977)

$$\rho(t) = \begin{cases} \left(\frac{t}{c}\right)^6 - 3\left(\frac{t}{c}\right)^4 + 3\left(\frac{t}{c}\right)^2 & \text{if} \quad |t| \leq c \\ 1 & \text{if} \quad |t| > c. \end{cases} \tag{2.12}$$

Another class of robust estimators for the linear model, so called MM-estimators, introduced by Yohai (1987) and has become very popular in the statistical literature. MM-estimators have simultaneously the following properties: a) they are highly efficient when the errors have normal distribution and b) their BP is 0.5. The MM-estimate denoted as $\boldsymbol{T}^1$ is defined in a three-stage procedures as follows:

Step 1: Compute an initial estimate $\boldsymbol{T}^0$ of the regression coefficient $\boldsymbol{\beta}$. This initial regression estimate is consistent robust and with high BP but not necessarily efficient. The S-estimator defined above will be used as the initial part of an overall MM-estimate computational strategy proposed by Yohai, Stahel and Zamar (1991).

Step 2: Compute the M-scale of the residuals $r_i(\boldsymbol{T}^0)$, $S(\beta)$, using the function ρ_0, i.e. in this second stage an M-estimate of the errors scale is computed using residuals based on the initial estimate.

Step 3: Define $\boldsymbol{T}^1$ as any solution to the equation $\sum_{i=1}^{n} \psi_1 \left(\frac{r_i(\boldsymbol{T}^1)}{S(\beta)} \right) \boldsymbol{x}_i = 0$, where $\psi_1 = \rho_1^{'}$ that also satisfies, $\sum_{i=1}^{T} \rho_1 (\frac{r_i(\boldsymbol{T}^1)}{S(\hat{\beta})}) \boldsymbol{x}_i \leq \sum_{i=1}^{T} \rho_1 (\frac{r_i(\boldsymbol{T}^o)}{S(\hat{\beta})}) \boldsymbol{x}_i$.

Yohai and Zamar (1997) defined some optimal functions ρ and ψ in the sense that, the final M-estimate has a BP equal to 0.5 thus minimizing the maximum bias under contamination distributions, subject to achieving a desired efficiency when the data is Gaussian (by an appropriate choice of constant c). The influence function of MM-estimates is given by

$$IF(\boldsymbol{x}^T, y; T, H) = \psi_1(y - \boldsymbol{x}^T \boldsymbol{\beta}) \boldsymbol{x} \sigma^2 (B(\psi_1, F) \boldsymbol{V})^{-1}, \tag{2.13}$$

where $B(\psi_1, F) = E_F \left(\psi_1^{'} \left(\frac{\epsilon}{\sigma} \right) \right)$, $\boldsymbol{V} = E_G(\boldsymbol{x}_i \boldsymbol{x}_i^T)$ and $S(\beta)$ is an estimate of scale estimate which converges to σ see Yohai (1987).

2.1.4 Adaptive one-step Robust Estimator

However tuning up the above estimators for high efficiency will be accompanied by an increase in bias as an unpleasant side-effect and will never achieve maximum asymptotic efficiency and positive breakdown point simultaneously. A new class of robust estimators that simultaneously attain maximum BP and full asymptotic efficiency under normal errors introduced by Gervini and Yohai (2002) so called robust and efficient weighted least squares estimators (REWLSE). The model considers a pair of initial robust estimators of regression and scale $\boldsymbol{T}_{0n}$ and S_n, respectively. If $S_n > 0$, the standardized residuals are defined as $r_i = \frac{y_i - \boldsymbol{x}_i^T \boldsymbol{T}_{0n}}{S_n}$. Assuming a normal-error model, it seems reasonable to consider outliers those points with $|r_i| \geq 2.5$. The REWLS estimator uses adaptive cut-off values which are constructed in a way that the resulting estimator is asymptotically efficient under the normal-error model and is robust under some deviations from the linear model. For detecting outliers the empirical distribution function of standardized absolute residuals $F_n^+(t)$ is compared with the distribution of the absolute errors under the hypothetical distribution $F_n(t)$, which is never known. In practice F^+ denotes the distribution of $|X|$ when $X \sim F$ and η is some large quantile of F^+. The measure of proportion of outliers is defined as $d_n = \sup_{t\geq\eta}\left\{F^+(t)-F_n^+(t)\right\}^+$, where $\{.\}^+$ denotes the positive part. Note that if $|r|_{(1)} \leq ... \leq |r|_{(n)}$ are the order statistics of the standardized absolute residuals and if $i_0 = \max\left\{i : |r|_{(1)} \leq \eta\right\}$, then $d_n = \sup_{i\geq i_0}\left\{F^+(|r|_{(i)})-\frac{(i-1)}{n}\right\}^+$. These $[nd_n]$ observations are liminated with the largest standardized absolute residuals ($[x]$ is the larger integer $\leq x$). The resulting cut-off value is

$$t_n = \min\left\{t : F_n^+(t) \geq 1 - d_n\right\}, \tag{2.14}$$

i.e. $t_n = |r|_{(i_n)} = n - [nd_n]$. Observe that $i_n \leq i_0$ and $t_n > \eta$. Therefore based on the above, weights are defined of the form $w_i = w\left(|r_i|/t_n\right)$ and REWLS estimator is

$$\boldsymbol{T}_{1n} = \begin{cases} (\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{Y}, & \text{if} \quad S_n \geq 1, \\ \boldsymbol{T}_{0n}, & \text{if} \quad S_n = 0. \end{cases} \tag{2.15}$$

The weight function $w[0,\infty) \to [0,1]$ is non-increasing continuous in the neighborhood of 0, $w(0) = 1$, $w(u) > 0$, for $0 < u < 1$ and $w(u) = 0$ for $u \geq 1$. The influence function of REWLS-estimators is given by

$$IF_{T_1}(\boldsymbol{x}^T, y, T_1, F) = \tau_1^{-1}\left\{w\left(\frac{|y-\boldsymbol{x}^T\boldsymbol{\beta}|}{\sigma t_0}\right)\boldsymbol{\Sigma}^{-1}\boldsymbol{x}(y-\boldsymbol{x}^T\boldsymbol{\beta}) + \tau_2 IF_{T_0}(\boldsymbol{x}^T, y),\right.$$

where IF_{T_0} is the initial influence of the initial estimator. The $\tau_1 = h_1\left(\frac{1}{\sigma t_0}, 0\right)$ where $h_1(s,t) = \int \omega(s|\epsilon - t|)dF(\epsilon/\sigma)$ and is continuous at $(s, 0)$ for every $s \geq 0$, and $\tau_2 = \frac{\partial h_2}{\partial t}\left(\frac{1}{\sigma t_0}, 0\right)$, where $h_2(s,t) = \int \omega(s|\epsilon - t|)dF(\epsilon/\sigma)$ is differentiable in the variable t for every $s \geq 0$. The $\partial h_2/\partial t$ is continuous and bounded in both variables. For details the reader may be referred to Gervini and Yohai (2002, p. 591).

2.2 Robust Log-linear Loss Reserving Model Estimation

In this section, we incorporate the robust estimators into loss reserving techniques. A robust algorithm for the robustification of the log-linear model of Verrall (1991) is derived as well as a robust estimation in the ANOVA setup of Kremer (1982).

2.2.1 Robust estimation for claims run-off triangles in the Log-linear Model

In the following we illustrate the steps required for obtaining robust loss reserving estimation, based on regression M-estimators. The rest of estimators, can be applied in a similar manner by substituting the M estimator by LAD, LMS, LTS, MM (with S), and Adaptive estimator, respectively.

ALGORITHM:

Step 1: Obtain a robust estimator $\widehat{\boldsymbol{\beta}}^M$ of $\boldsymbol{\beta}$ by the following minimization problem (see Huber and Dutter (1974)). In regression analysis M-estimators can be obtained as solutions of the following minimization problem, i.e.

$$\min_{\hat{\beta}} \sum_{t=1}^{T} \rho\left(\frac{r_t}{S}\right), \tag{2.16}$$

which is used to find $\hat{\boldsymbol{\beta}}^M$, a robust estimator of $\boldsymbol{\beta} = (\beta_1, ..., \beta_p)'$ and S, with Tukey's bisquare ρ function defined as

$$\rho(t) = \begin{cases} \left(\frac{t}{c}\right)^6 - 3\left(\frac{t}{c}\right)^4 + 3\left(\frac{t}{c}\right)^2 & \text{if} \quad |t| \leq c \\ 1 & \text{if} \quad |t| > c. \end{cases} \tag{2.17}$$

The solution of the minimization problem (2.16) is equivalent to simultaneously solving the following equations;

$$\sum_{t=1}^{T} \psi\left(\frac{r_t}{S}\right) \sum_{t=1}^{T} x_{kt} = 0, \quad k = 1, 2, ..., p, \quad \text{and} \quad \sum_{t=1}^{T} \chi\left(\frac{r_t}{S}\right) = a, \tag{2.18}$$

where $\psi(u) = \rho'(u)$ and $\chi(u) = u\psi(u) - \rho(u)$. If we want S to be asymptotically unbiased for normal errors we take $a = [(n-p)/n]E_\Phi(\chi)$ with Φ being the normal distribution.

Step 2: Calculate the robust covariance matrix of $\hat{\beta}$ as

$$\widehat{\text{Cov}(\boldsymbol{\beta}^M)} = \widehat{\sigma_M}^2 (\boldsymbol{X}^T \boldsymbol{X})^{-1}, \tag{2.19}$$

with

$$\widehat{\sigma_M}^2 = L^2 \frac{1}{(t-p)} \frac{\sum_{i=1}^{t} \psi\left(\frac{r_i}{S}\right)^2 S^2}{\left[(1/t)\sum_{i=1}^{t} \psi'\left(\frac{r_i}{S}\right)\right]^2}, \tag{2.20}$$

where L is the correction factor for unbiasedness (see Huber (1973)), defined as

$$L = 1 + \frac{p}{T} \frac{\mathrm{Var}[\psi'(\frac{r_t}{S})]}{E[\psi'(\frac{r_t}{S})]^2}. \tag{2.21}$$

In practice $E[\psi'(\frac{r_i}{S})]$ and $\mathrm{Var}[\psi'(\frac{r_i}{S})]$ are unknown and will be estimated by

$$E[\psi'(\frac{r_t}{S})] \simeq \frac{1}{t} \sum_{i=1}^{T} \psi'(\frac{r_i}{S}) = m, \tag{2.22}$$

$$\mathrm{Var}[\psi'(\frac{r_i}{S})] \simeq \frac{1}{t} \sum_{i=1}^{T} [\psi(\frac{r_i}{S}) - m]^2. \tag{2.23}$$

In the special case where

$$\psi(u) = \min[c, \max(-c, u)], \tag{2.24}$$

we obtain

$$L = 1 + \frac{1}{t} \frac{(1-m)}{m}, \tag{2.25}$$

where m is the relative frequency of the residuals satisfying $-c < \frac{r_i}{S} < c$.

Step 3: Calculate the robust M-estimate of $\tilde{\theta}_{ij}$

$$\tilde{\theta}_{ij}^M = \exp(\boldsymbol{x}_{ij}^T \widehat{\boldsymbol{\beta}^M}) g_m[\frac{1}{2}(1 - \boldsymbol{x}_{ij}^T (\boldsymbol{X}^T \boldsymbol{X})^{-1} \boldsymbol{x}_{ij}) \sigma_M^2]. \tag{2.26}$$

Step 4: The robust variance of the asymptotically unbiased estimator of τ_{ij}^2 is obtained as

$$\tilde{\tau}_{M_{ij}}^2 = \exp(2\boldsymbol{x}_{ij}' \widehat{\boldsymbol{\beta}^M})[g_m(\frac{1}{2}(1 - \boldsymbol{x}_{ij}' (\boldsymbol{X}' \boldsymbol{X})^{-1} \boldsymbol{x}_{ij}) s_M^2) - g_m((1 - 2\boldsymbol{x}_{ij}^T (\boldsymbol{X}^T \boldsymbol{X})^{-1} \boldsymbol{x}_{ij}) s_M^2].$$

Remark 2.1. The robust parameter estimation of the log-linear model and the standard errors are obtained in a straightforward way by substituting the parameters by their robust ones.

2.2.2 Robust Loss Reserving in the ANOVA Model

The robust ANOVA model can be implemented similarly as in the regression case in two-stage procedures. Since in the ANOVA setup we have only factor variables, in the first stage procedure the least absolute deviation (LAD) estimate can be used as an initial estimate. LAD estimate is also a robust M-estimate with $\rho(x) = |x|$. In the second stage, by using the initial estimates, we compute robust sums-of-squares by using robust estimates in place of the usual least squares estimates of the main effects. In the following we obtain parameter estimation in a more general model, where more than one observations is appeared in each cell, determined by the two factors in a two-way ANOVA setup.

Armstrong and Frome (1979) demonstrated how LAD estimates can efficiently be obtained for the two-way ANOVA. LAD estimates of the parameters are obtained by solution of the following problem:

$$\min \quad Q = \sum_{i=1}^{r}\sum_{j=1}^{c}\sum_{k=1}^{n_{ij}} |Y_{ijk} - (\mu + \alpha_i + \beta_j)|. \tag{2.27}$$

Thus, Y_{ijk} is the k-th observation at the $i-th$ level of the first factor and the $j-th$ level of the second factor, α_i represents the effect of the $i-th$ level of the first factor (row effect), β_j represents the effect of the $j-th$ level of the second factor (column effect), and μ is a typical value. There are two degrees of freedom in the assignment of values to μ, α_i and β_j, thus restrictions should be added to the problem. Similarly as in the least squares estimates analysis, in the LAD analysis, estimates that minimize

$$\min \sum_{i=1}^{r} |\alpha_i| + \sum_{j=1}^{c} |\beta_j| \tag{2.28}$$

are provided subject to the optimal value for Q in (2.27) being maintained. This additional criterion does not necessarily provide a unique solution. Further restrictions, or a completely different set of criteria, may determine a unique solution [see Armstrong and Frome (1979)]. By letting $\tau_i = \mu + \alpha_i$, the problem (2.28) can be restated and written as a linear programming problem:

$$\min \sum_{i=1}^{r}\sum_{j=1}^{c}\sum_{k=1}^{n_{ij}} (d_{ijk}^{+} - d_{ijk}^{-}), \tag{2.29}$$

subject to

$$\begin{aligned} \tau_i + \beta_j - Y_{ijk} + d_{ijk}^{+} - d_{ijk}^{-} &= 0, \\ d_{ijk}^{+} \geq 0, \quad d_{ijk}^{-} &\geq 0, \\ i = 1, ..., r, \quad j = 1, ..., c, \quad k &= 1, ..., n_{ij}, \end{aligned} \tag{2.30}$$

where d_{ijk}^{+} and d_{ijk}^{-},; are the positive and negative deviations of the regression equation from the observation Y_{ijk}, respectively.

Remark 2.2. For obtaining LAD estimates for Kremer's (1982) model we drop the third summation in (2.27) and (2.29) which indicates that we have only one observation in each cell determined by the two factors in a two-way ANOVA setup. The summation of k observations (in the case that we have more that one value in each cell) is useful when more than one run-off triangles of similar line of business are involved for the estimation of claims reserves.

Remark 2.3. Very often in a two way ANOVA, not a single cell but rather a bigger substructure of the design such as a row or several rows together behaves like an outlier showing an unusual behavior of parameter estimation. As a result of the above the absolute residuals in the substructure tend to be large especially residuals from the robust fit. This large variability can be detected by applying robust techniques to absolute residuals. In some cases a robust least squares analysis of the absolute residuals might be more informative, usually when a small substructure, but no single cells, show outlying behavior, (see Huber, 1981).

Apart from factors with no effects we can define main effects after determination of the grand mean. If the majority of the main effects between pairs of rows is constant then we have a well defined main effects difference, but only for part of the data, and we know that the remaining pairs of data contain outliers, but we don't know which value in each pair is an outlier. This can be done by a comparison with other rows.

Remark 2.4. Hampel et al. (1986, p. 426) discusses various ways of identifying outlier evens where the role of prior information is very crucial which tell us whether some effect or error is quite possible or is very surprising and unlikely. In two way tables, the median provides a robust estimator of row and column effects. Writing the two way ANOVA as a linear model regression, it is not always possible to apply the LMS to it because the number of cases (cells) must be larger than twice the number of parameters. On the other hand, the LMS becomes more useful when also interval-scaled covariates occur, which may contain outliers (see Rousseeuw and Leroy, 1987, p. 285).

Remark 2.5. Tukey (1977) suggested that with the presence of outlier events in the data we can use an iterative procedure called median polish that would produce better estimates than using the mean in the two-way ANOVA. The median polish procedure subtracts the median from each row from each observation in that row, and then subtracts the median from each column from the updated table. This continuous until the median of each row and column is zero. Sposito (1987) has shown the equivalence of median polish and L_1 estimator.

Remark 2.6. Transforming the ANOVA model to the regression one, we have the advantage that a whole variety of robust estimators that appears in the literature for the regression case can be applied. Although, the two-way ANOVA can be transferred to an equivalent regression model, this does not imply equivalence on the results of robust regression and robust ANOVA models. Robust results that are obtained by this transformation are slightly different, due to the different way the robustification applied to ANOVA and regression models.

2.3 Example on Loss Reserving for Robust Log-linear Regression Model

In what it follows we will illustrate the LS and our robust procedures using the data given in Taylor and Ashe (1983) (see Table 2.1).

Table 2.1: Taylor and Ashe (1983) data

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
1	357848	766940	610542	482940	527326	574398	146342	139950	**227229**	**67948**
2	352118	884021	933894	1183289	445745	320996	527804	266172	**425046**	
3	290507	1001799	926219	1016654	750816	146923	495992	280405		
4	310608	1108250	776189	1562400	272482	352053	206286			
5	443160	693190	991983	769488	504851	470639				
6	396132	937085	847498	805037	705960					
7	440832	847631	1131398	1063269						
8	359480	1061648	1443370							
9	**376686**	**986608**								
10	**344014**									

The advantage of using this data set is that is well known and it has been widely used in the actuarial literature (see for example Mack 1993; England and Verrall 1999, Verdonc et al., 2010). Our first step is to identify the existence of outlier events (if there are any) in the original data and apply a log-linear regression model, a two way ANOVA, as well as robust estimators. In the second step, in order to investigate the effect of outlier events in the loss reserving estimation, we create one artificial outlier by multiplying by ten (10) and in the sequel two artificial outliers by multiplying two specific claims by ten (10).

Following Verrall (1991, p. 78) we apply robust estimation on the following two-way analysis of variance in a log-normal regression setting, i.e.,

$$\ln(C_{ij}) = Y_{ij} = \mu + a_i + b_j + \epsilon_{ij}, \quad i = 2, ..., 10; \quad j = 2, ..., 10, \tag{2.31}$$

where μ is the overall mean, a_i is the effect of the $i-th$ accident year and b_j is the effect of the $j-th$ development year, ϵ_{ij} is the error term and the design matrix is of dimension 55×18.

Most robust estimators are calculated based on (2.31). The calculation of LTS and LMS estimators failed to be implemented based on (2.31) model and for that reason we have tried alternative models with different design matrices.

A model that works properly for the calculation of the LTS estimator is the model where there is a unique level for each accident year and a unique value for the zero development period. The parameters for development periods 1 to 10 are assumed to follow some linear relationship (straight line) with the same slope or parameters s (see Christofides, 1990, p. 19). Then the model is defined as

$$\ln(C_{ij}) = Y_{ij} = \mu + a_i + d_j + \epsilon_{ij}, \quad \text{for} \quad i, j \quad \text{from 1 to 10}, \tag{2.32}$$

where $d_0 = d$, $d_j = s \times j$, for $j > 1$, ϵ_{ij} is the error term and the design matrix is of dimension 55×12.

Unfortunately, the LMS estimator for the model in (2.32) failed to be implemented. An appropriate model for the calculation of the LMS estimator is the following,

$$\ln(C_{ij}) = Y_{ij} = a.i + b.j + \epsilon_{ij}, \quad i = 1, ..., 10; \quad j = 1, ..., 10, \tag{2.33}$$

where ϵ_{ij} is the error term and the design matrix is of dimension 55×2.

2.3.1 Diagnostics and Robust Loss Reserving Estimation

Outlier diagnostics are statistical methods that focus attention on observations having a large influence on the least squares estimates, which are known to be nonrobust. Many diagnostics are based on the residuals (see Rousseeuw and Leroy, 1987). In connection with diagnostic statistics, there is a variety of plots used for diagnostic purposes. These plots investigate the appearance of outlier events and compare the results of the classical and robust fits.

2.3.1.1 Diagnostics and Robust Loss Reserving Estimation based on the Original Data

In Figure 2.1, the Normal QQ-Plot of residuals shows that the classical least squares residuals are approximately normally distributed, except for one moderately sized outlier in the left hand part of the left panel (25th observation). However, the normal QQ-Plot of residuals from the robust fit in the left hand panel shows clearly that the 25th observation corresponds to an outlier event, and that all the other residuals conform quite well to a normal distribution.

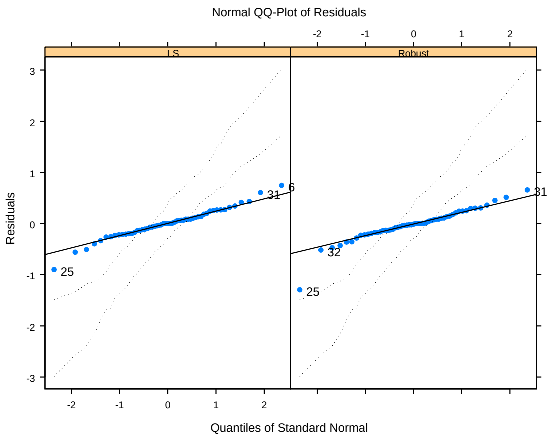

Figure 2.1: Normal QQ-Plots for LS and Robust Residuals with original data

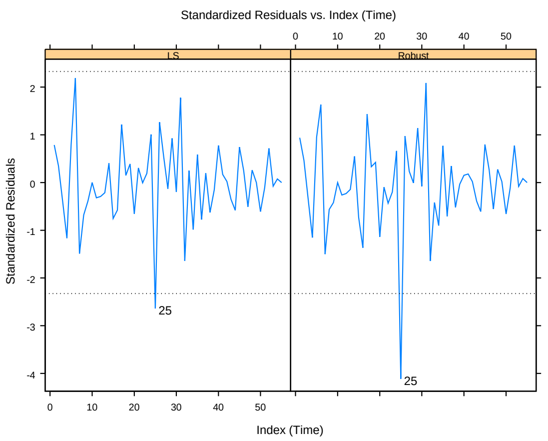

Figure 2.2: Standardized Residuals vs Index for the LS and Robust fits with Original data

The plots of Standardized Residuals versus Index (in this case observation number) for the LS and Robust fits are shown in Figure 2.2. The horizontal reference lines at ± 2.5 correspond to tail probabilities of 0.006 for a standard normal random variable. The LS residuals barely hint the presence of an outlier, corresponding to the 25th observation, while the robust residuals clearly identify observation 25th as an outlier. These plots are consistent with the behavior of the normal QQ-Plots in Figure 2.1.

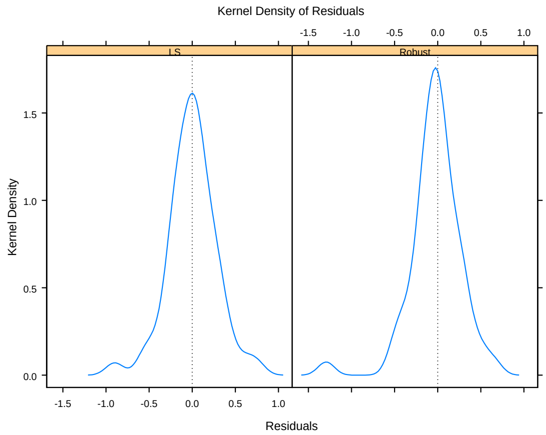

Figure 2.3: LS and Robust Residuals Density Estimates with Original data

The density estimate in Figure 2.3 provides a much more accurate picture of the distribution of the error term in the model. The main mode of the density estimate for both the LS and robust residuals is well centered on zero, and has a single bump in the left-hand tail reflecting the presence of a single large outlier. Consequently, our original data has an outlier and this has to be taken into account.

Table 2.2: Outstanding Reserves based on Robust Estimation - Original Data

Diagonal	**LS**	**M**	**LAD**	**LTS**	**LMS**	**MM**	**Adaptive**	**Robust Anova**
3901463	0	0	0	0	0	0	0	0
5339085	110586	103441	115638	278009	317029	107555	124637	108809
4909315	481398	508024	608036	590827	713704	519563	524978	557983
4588268	659908	612192	718506	569505	1207232	827843	829983	629323
3873311	1089378	1031823	1456540	1717002	1818476	1180906	1061998	1033773
3691712	1528635	1492761	1801564	2229653	2572737	1433696	1360182	1632835
3483130	2308095	2298118	3065129	3501744	3500711	2339766	2256340	2415429
2864498	3802283	3624321	3997726	3584548	4639652	3800707	3709255	3887952
1363294	4444529	4214877	4798948	4415824	6034796	4402712	4322244	4502493
344014	5043474	4673040	5338526	5179459	7741096	3775148	4862427	5073682
34358090	**19468286**	**18558601**	**21900617**	**22066571**	**28545434**	**18387899**	**19052048**	**19842279**

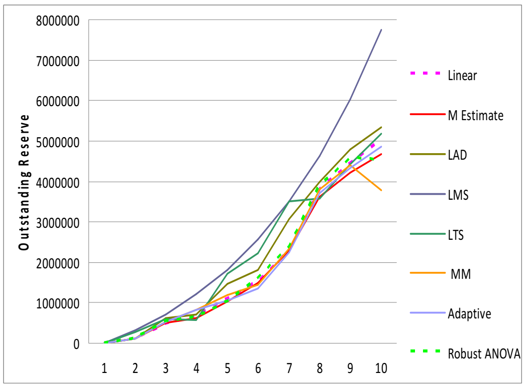

Figure 2.4: Outstanding Reserves based on Robust Estimation - Original Data

The results based on LS and robust chain ladder estimation are shown in Tables 2.2 and 2.3. In Table 2.3, we see a difference on the value of the overall reserves based on the log-linear model (LS value-19468286) and the value of the overall reserves based on the robust M-estimate (value-18558601), which is due to the robustification of the outlier event $X_{3,6}$ with M-estimation. We have similar behavior to MM-estimation producing a value of the overall reserves (value-18387899). Robust ANOVA estimation produces almost the same results with LS estimation, which means that Robust ANOVA procedure does not robustify the outlier event, $X_{3,6}$, of the original data. Adaptive estimators produce slightly higher estimations for the ultimate claims and the total reserves compared with M-estimation. The rest of the Robust LAD, LMS and LTS estimators produce different values of ultimate claims and total reserves in comparison with LS and M-estimators. Figure 2.4 shows a summary of the outstanding reserves in a diagram based on LS and all robust estimators.

Table 2.3: Ultimate Claims based on Robust Estimation - Original Data

Diagonal	**LS**	**M**	**LAD**	**LTS**	**LMS**	**MM**	**Adaptive**	**Robust Anova**
3901463	3901463	3901463	3901463	3901463	3901463	3901463	3901463	3901463
5339085	5449671	5442526	5454723	5656114	5617094	5446640,039	5463722	5447895
4909315	5390715	5417339	5517351	5623019	5500142	5428878,836	5434294	5467298
4588268	5248177	5200460	5306774	5795500	5157773	5416111,088	5418251	5217590
3873311	4962691	4905135	5329852	5691787	5590313	5054217,039	4935310	4907084
3691712	5220351	5184473	5493276	6264449	5921365	5125408,547	5051894	5324547
3483130	5791230	5328571	6095581	6983841	6984874	5370218,277	5286793	5898559
2864498	6666790	5962831	6336235	7504150	6449046	6139216,048	6047764	6752450
1363294	5807834	5578172	6162243	7398090	5779118	5766006,544	5685539	5865787
344014	5387504	5017054	5682541	8085110	5523473	4119162,748	5206442	5417696
34358090	**53826426**	**51938024**	**55280040**	**62903524**	**56424661**	**51767322**	**52431471**	**54200369**

Thus, regression model is equivalent to ANOVA model, robust M-regression produces different results from those of robust ANOVA (see Remark 2.3) and this is due to the appearance of outlier events in the original data. This is also evident looking at the normal QQ-Plots (Figure 2.1).

Remark 2.7. In general, LTS, LMS, MM and Adaptive robust methods are minimally influenced by outliers in the independent variables space, in the response (dependent variables) space, or in both. But these methods provide unreliable results in cases where the independent variables are only dummy variables (there are not leverage points in the data), as it is in our loss reserving estimation. This is due to the fact that the robustification of the design matrix with dummy variables may affect the estimation of the regression coefficients, (see Rousseuew, 1987).

2.3.1.2 Diagnostics and Robust Loss Reserving Estimation with Data with 1 Artificial Outlier

We multiply the $X_{4,4} = 1562400$ value by 10 in order to create an artificial outlier.

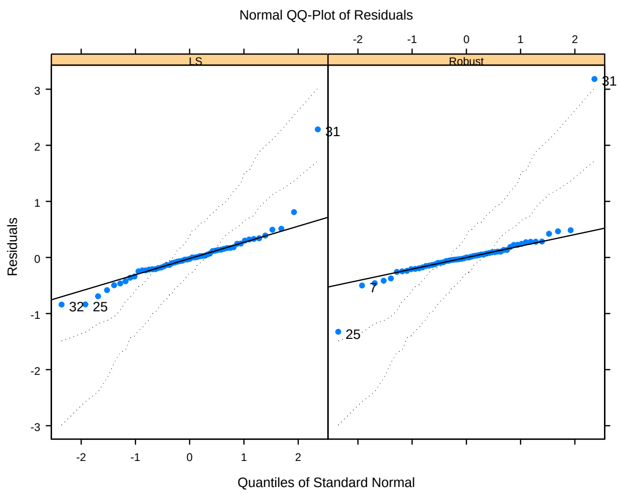

Figure 2.5: Normal QQ-Plots for LS and Robust Residuals with 1 artificial outlier

According to Figure 2.5 we see that for the classical LS the residuals are approximately normally distributed, except from one moderately sized outlier, corresponding to the 31st observation which is tenfold the initial loss of the (4,4) cell of the original run-off triangle and two marginal outliers in the left hand part of the left panel. However, the residuals QQ-Plot from the robust fit in the right hand panel confirms that there are two residual outliers, corresponding to the 31th and 25th observations and that all the other residuals conform quite well to a normal distribution. Similar conclusions we may obtain if we plot Standardized Residuals versus Index or density estimate plot.

Table 2.4: Outstanding Reserves based on Robust Estimation with 1 artificial outlier

Diagonal	**LS**	**M**	**LAD**	**LTS**	**LMS**	**MM**	**Adaptive**	**Robust Anova**
3901463	0	0	0	0	0	0	0	0
5339085	135125	103436	137247	315003	235655	120598	120598	108809
4909315	567567	508005	699543	709222	568192	504477	504477	560252
18649868	1073178	612170	821910	1199751	566930	12953580	12953580	558916
3873311	1219207	1031787	1656235	1807307	1227695	1033734	1033734	1047718
3691712	1664314	1492709	2045232	2557002	1808461	1333650	1333650	1676244
3483130	2455135	2298040	3477598	3479283	2367687	2207391	2207391	2497626
2864498	4977126	3624194	4550623	4611084	2724235	3462214	3462214	3760173
1363294	5734526	4214717	5524428	5997223	3738782	4124256	4124256	4389823
344014	6691870	4672813	6383944	7692107	4495872	4491231	4491231	4970616
48419690	**24518048**	**18557870**	**25296760**	**28367982**	**17733508**	**30231131**	**30231131**	**19570177**

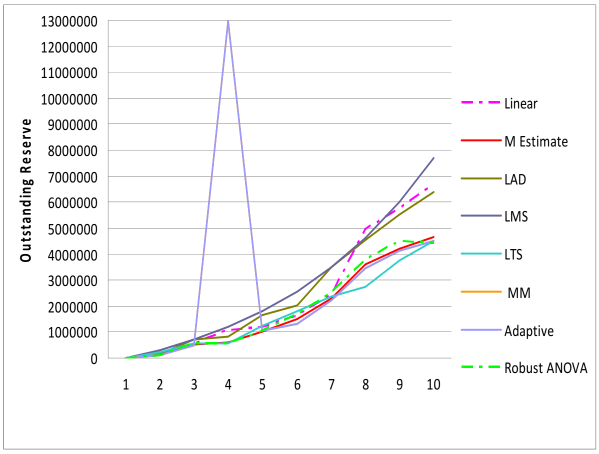

Figure 2.6: Outstanding Reserves based on Robust Estimation with 1 artificial outlier

The outstanding reserves and ultimate claims are illustrated in Tables 2.4 and 2.5. There is a significant difference between the LS and robust estimation. With the appearance of the artificial outlier $X_{4,4} = 15624000$ the estimated reserves (Table 2.4) and ultimate claims (Table 2.5) calculated using robust M-estimation and robust ANOVA are similar to the estimated ultimate claims and reserves based on the original data. LTS-estimators produce outstanding reserves 17733508, while the rest of robust estimators (LAD, LMS, MM and adaptive) produce much higher ultimate claims and total reserves (see Table 2.4). Figure 2.6 shows a summary of the outstanding reserves based on LS and all robust estimators.

Table 2.5: Ultimate Claims based on Robust Estimation with 1 artificial outlier

Diagonal	**LS**	**M**	**LAD**	**LTS**	**LMS**	**MM**	**Adaptive**	**Robust Anova**
3901463	3901463	3901463	3901463	3901463	3901463	3901463	3901463	3901463
5339085	5474210	5442521	5476332	5654088	5574740	5459683	5459683	5447895
4909315	5476882	5417320	5608858	5618537	5477507	5413792	5413792	5469567
18649868	19723046	5200438	5410178	19849619	19216798	17541848	17541848	5147184
3873311	5092518	4905098	5529546	5680618	5101006	4907045	4907045	4921029
3691712	5356026	5184421	5736944	6248714	5500173	5025362	5025362	5367956
3483130	5938265	5328492	6508050	6962413	5850817	5237843	5237843	5980757
2864498	7841624	5962703	6889132	7475582	5588733	5800723	5800723	6624670
1363294	7097820	5578011	6887722	7360517	5102076	5487550	5487550	5753116
344014	7035884	5016827	6727958	8036121	4839886	4835245	4835245	5314630
48419690	**72937738**	**51937293**	**58676183**	**76787672**	**66153198**	**63610554**	**63610554**	**53928267**

2.3.1.3 Diagnostics and Robust Loss Reserving Estimation with 2 Artificial Outliers

By multiplying the value of $X_{2,2} = 766940$ by 10 we create the second outlier event. Figure 2.7 shows that the residuals are approximately normally distributed, except for two moderately sized outliers, corresponding to the 31st and the 2nd observation which is now tenfold the initial losses of the (4, 4) and (1, 2) cells of the original run-off triangle, in the right hand part of the left panel. The normal QQ-Plot of residuals from the robust fit in the right hand panel confirms that there are two residual outliers, corresponding to the 31th and 2nd observations, and one marginal outlier in the left hand part of the right panel corresponding to the 25th observation. All the other residuals conform quite well to a normal distribution.

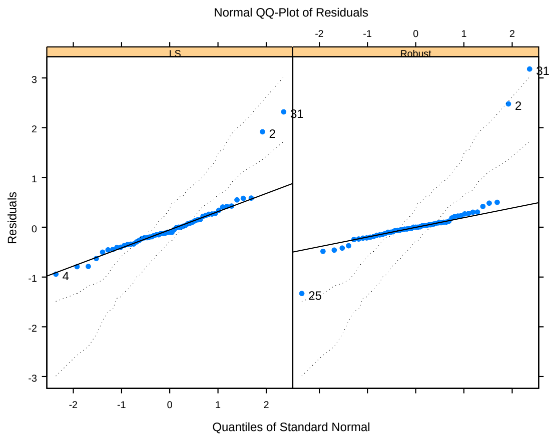

Figure 2.7: Normal QQ-Plots for LS and Robust Residuals with 2 artificial outliers

Table 2.6: Outstanding Reserves based on Robust Estimation with 2 artificial outliers

Diagonal	**LS**	**M**	**LAD**	**LTS**	**LMS**	**MM**	**Adaptive**	**Robust Anova**
10803923	0	0	0	0	0	0	0	0
5339085	120816	99836	155056	284415	522100	133442	133441	111958
4909315	538249	501073	773113	644780	1057266	544647	544644	572162
18649868	1025028	601180	904656	1098828	1634639	13828866	13828803	569959
3873311	1171249	1013311	1458875	1668391	2266398	1110943	1110938	1066561
3691712	1601241	1470983	2238977	2380349	2963473	1428205	1428199	1705895
3483130	2361095	2280379	3805330	3267824	3740338	2361984	2361973	2541942
2864498	4784998	3598542	4991035	4371641	4622958	3712952	3712936	3831037
1363294	5367285	4171748	5933694	5742149	5678468	4464882	4464861	4486581
344014	7824554	4648719	7251634	7441472	7226432	4759120	4759092	5018421
55322150	**24794515**	**18385771**	**27512369**	**26899850**	**29712071**	**32345040**	**32344888**	**19904516**

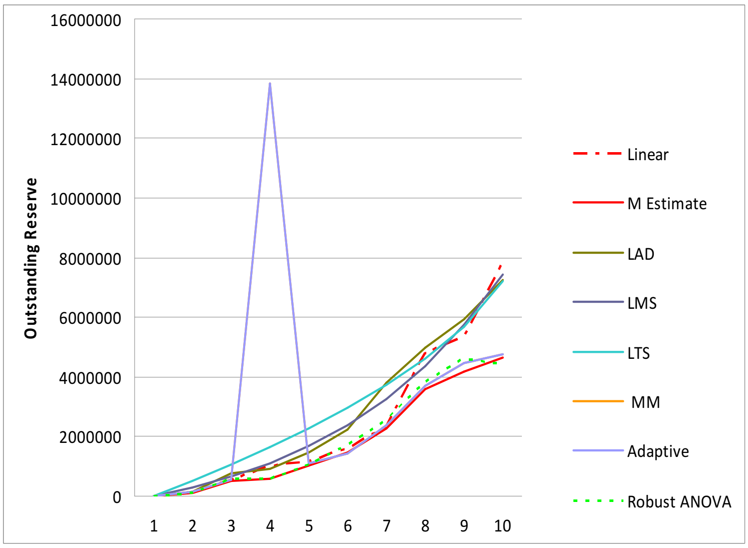

Figure 2.8: Outstanding Reserves based on Robust Estimation with 2 artificial outliers

In Tables 2.6 and 2.7 we observe that in the presence of the 2 artificial outliers in the data, the estimated ultimate claims and reserves calculated using robust M-estimation and robust ANOVA, produce values similar with values those obtained with 1 artificial outlier (Tables 2.4, 2.5). The rest of robust estimators, in comparison with robust M-estimation and robust ANOVA, produce much higher values for the ultimate claims and the total reserves. Figure 2.8 shows a summary of the outstanding reserves in a diagram based on LS and robust estimators.

Table 2.7: Ultimate Claims based on Robust Estimation with 2 artificial outliers

Diagonal	**LS**	**M**	**LAD**	**LTS**	**LMS**	**MM**	**Adaptive**	**Robust Anova**
10803923	10803923	3901463	3901463	10803923	10803923	3901463	3901463	3901463
5339085	5459901	5438921	5494141	5623500	5861185	5472527	5472526	5451043
4909315	5447564	5410388	5682428	5554095	5966581	5453962	5453959	5481477
18649868	19674896	5189448	5492924	19748696	20284507	18417134	18417071	5158227
3873311	5044560	4886622	5332186	5541702	6139709	4984254	4984249	4939872
3691712	5292953	5162695	5930689	6072061	6655185	5119917	5119911	5397607
3483130	5844225	5310831	6835782	6750954	7223468	5392436	5392425	6025072
2864498	7649496	5937051	7329544	7236139	7487456	6051461	6051445	6695535
1363294	6730579	5535042	7296988	7105443	7041762	5828176	5828155	5849875
344014	8168568	4992733	7595648	7785486	7570446	5103134	5103106	5362435
55322150	**80116665**	**51765194**	**60891792**	**82222000**	**85034221**	**65724463**	**65724311**	**54262607**

At this point it is appropriate to estimate the prediction errors for the LSE and the most stable robust estimation procedures, the M-estimation and robust ANOVA. Table 2.8 shows the individual standard errors for the LS-estimation, M-estimation and robust ANOVA, with the presence of no-outliers, 1-outlier and 2-outliers. The overall standard errors (and % error) are presented at the end of the Table 2.8. We see that the M-estimation method has a lower standard error than LSE and robust ANOVA.

Since the influence of an outlier depends on its location on the triangle, it is appropriate to see the behavior of the run-off triangle in the presence of an outlier in each location. Therefore, following the idea of Verdonc et al. (2009), we contaminate the data by multiplying each observation (cell) by 10 and observe the outstanding reserve and standard errors at each outlying observation, separately.

In the following we compare the values of reserves and standard errors of LS-estimation with the values of M-estimation and robust ANOVA. The first line of Table 2.10 illustrates the values of reserves and total standard errors for the 3.2.1 and M-estimation and robust ANOVA based on the original data. From the second line and on, the first column shows which observation (cell) was contaminated (multiplied by 10) and can be considered as an outlier.

As we can observe in Table 2.8, the ordinary LS-estimation is very sensitive to the presence of outlier events in the data producing very large reserves and standard errors in comparison with M-estimation. In the presence of large claims (outliers) most of the times we have an overestimation of the reserves, but there are cases where we have an underestimation of the reserves.

Table 2.8: Standard Errors for future values $\widehat{C}_{i,j}$, for LS, M and Robust ANOVA

	No outlier			One outlier			Two outliers		
Cell	**LS**	**M**	**R-ANOVA**	**LS**	**M**	**R-ANOVA**	**LS**	**M**	**R-ANOVA**
$\widehat{C}_{10,2}$	370295	211105	955014	592553	210995	955014	1437222	206868	945019
$\widehat{C}_{9,3}$	310075	183660	1074261	535059	183567	1074261	714244	177395	1093571
$\widehat{C}_{10,3}$	383365	220147	982683	614711	220032	982683	1137745	213173	992862
$\widehat{C}_{8,4}$	314663	177638	1242469	724450	177549	1104180	1003373	172666	1120888
$\widehat{C}_{9,4}$	323738	181306	1095556	778431	181214	973619	1027121	175667	992730
$\widehat{C}_{10,4}$	397117	215652	1002163	886967	215539	890621	1622290	209471	901307
$\widehat{C}_{7,5}$	149949	95642	640151	303725	95594	662566	312177	93904	671294
$\widehat{C}_{8,5}$	180881	113327	701122	296765	113270	704577	406547	111177	715715
$\widehat{C}_{9,5}$	184743	114831	618219	316514	114773	621266	413048	112292	633883
$\widehat{C}_{10,5}$	224545	135354	565518	357218	135283	568305	646034	132696	575508
$\widehat{C}_{6,6}$	91859	59267	518555	167016	59238	539466	191889	57599	546952
$\widehat{C}_{7,6}$	106201	69236	563792	212949	69201	590578	216359	67383	599775
$\widehat{C}_{8,6}$	127205	81461	617489	206582	81420	628025	279735	79217	639464
$\widehat{C}_{9,6}$	128788	81828	544476	218377	81786	553766	281662	79319	566350
$\widehat{C}_{10,6}$	154760	95374	498061	243570	95324	506559	435257	92684	514194
$\widehat{C}_{5,7}$	81385	48214	329211	139013	48190	340278	169435	46519	344685
$\widehat{C}_{6,7}$	87517	52415	354849	157314	52388	369203	178302	50541	374541
$\widehat{C}_{7,7}$	100473	60804	385805	199161	60773	404183	199607	58714	410713
$\widehat{C}_{8,7}$	119312	70930	422551	191531	70894	429811	255821	68438	437891
$\widehat{C}_{9,7}$	119473	70474	372588	200210	70438	378989	254684	67780	387824
$\widehat{C}_{10,7}$	141424	80931	340826	219870	80888	346681	387389	78035	352109
$\widehat{C}_{4,8}$	56719	32366	210590	149484	32350	186862	180872	31167	189642
$\widehat{C}_{5,8}$	65186	38024	235768	122428	38005	236518	146522	36504	240189
$\widehat{C}_{6,8}$	69602	41046	254128	137555	41026	256622	153079	39381	260993
$\widehat{C}_{7,8}$	79221	47210	276298	172635	47186	280936	169873	45360	286200
$\widehat{C}_{8,8}$	93052	54477	302613	164192	54449	298750	215294	52300	305138
$\widehat{C}_{9,8}$	91821	53344	266832	169092	53316	263425	211130	51049	270250
$\widehat{C}_{10,8}$	106407	59988	244085	181669	59956	240969	314044	57553	245362
$\widehat{C}_{3,9}$	107944	71435	439640	209305	71399	441428	248657	69060	449659
$\widehat{C}_{4,9}$	98242	57707	329923	265440	57677	293145	311705	55412	298890
$\widehat{C}_{5,9}$	112086	67305	369367	215785	67271	371044	250614	64432	378556
$\widehat{C}_{6,9}$	118637	72026	398133	240300	71989	402584	259484	68909	411345
$\widehat{C}_{7,9}$	133552	81939	432864	298222	81897	440727	284709	78505	451072
$\widehat{C}_{8,9}$	154585	93185	474092	279437	93137	468672	355438	89209	480921
$\widehat{C}_{9,9}$	149385	89376	418034	281700	89329	413256	341114	85289	425934
$\widehat{C}_{10,9}$	167477	97282	382398	292457	97230	378027	489941	93072	386709
$\widehat{C}_{2,10}$	42191	24459	108810	87786	24446	108810	98165	23034	111958
$\widehat{C}_{3,10}$	39275	25409	118343	81885	25396	118824	90246	24061	122502
$\widehat{C}_{4,10}$	35464	20368	88809	102994	20357	78909	112171	19157	81428
$\widehat{C}_{5,10}$	40092	23542	99427	82932	23530	99878	89308	22075	103131
$\widehat{C}_{6,10}$	41954	24911	107170	91270	24898	108368	91360	23345	112064
$\widehat{C}_{7,10}$	46524	27923	116519	111526	27908	118635	98668	26204	122887
$\widehat{C}_{8,10}$	52719	31096	127617	102235	31080	126158	120458	29160	131019
$\widehat{C}_{9,10}$	49286	28868	112527	99591	28852	111241	111628	26984	116039
$\widehat{C}_{10,10}$	52004	29611	102934	97009	29595	101758	150142	27751	105353

Table 2.9: Standard Errors for LS, M-estimation and Robust ANOVA

	Least Squares			**M Huber Estimator**			**Robust ANOVA**		
Outlier	**Reserves**	**SE**	**% SE**	**Reserves**	**SE**	**% SE**	**Reserves**	**SE**	**% SE**
-	19468286	6101444	31.34%	18558601	3612493	19.47%	19842279	2964004	14.94%
$C_{1,1}$	18308221	8826406	48.21%	18597639	3619389	19.46%	20719948	3268648	15.78%
$C_{2,1}$	17787293	7691080	43.24%	17744100	3477427	19.60%	20223446	3457358	17.10%
$C_{3,1}$	18115239	7519667	41.51%	17568151	3079402	17.53%	19454268	3025417	15.55%
$C_{4,1}$	18817623	8247422	43.83%	18141553	3763249	20.74%	20542392	3641893	17.73%
$C_{5,1}$	19878945	9301157	46.79%	18380006	3462032	18.84%	20387511	3225948	15.82%
$C_{6,1}$	20191754	8931295	44.23%	18493494	4039739	21.84%	20303095	3364249	16.57%
$C_{7,1}$	21494476	9377440	43.63%	18309534	3176384	17.35%	20309336	3396320	16.72%
$C_{8,1}$	23848448	9026257	37.85%	19158871	3596848	18.77%	20615596	3445356	16.71%
$C_{9,1}$	30005476	11647462	38.82%	27389972	5274105	19.26%	20386257	3389146	16.62%
$C_{10,1}$	64862426	24078797	37.12%	60614855	13918539	22.96%	66598545	20895598	31.38%
$C_{1,2}$	19577281	9252449	47.26%	18392957	3576225	19.44%	20214902	3144311	15.55%
$C_{2,2}$	19345924	8574995	44.32%	18096253	3337792	18.44%	20130808	3310026	16.44%
$C_{3,2}$	20413645	9626407	47.16%	18144658	3269378	18.02%	20264828	3384797	16.70%
$C_{4,2}$	21275583	10606917	49.85%	18552876	3610718	19.46%	19910049	2957072	14.85%
$C_{5,2}$	20364569	8362497	41.06%	18132834	2903904	16.01%	19703360	2925098	14.85%
$C_{6,2}$	21668208	9632939	44.46%	18515817	3385192	18.28%	20264707	3365793	16.61%
$C_{7,2}$	22430981	9148969	40.79%	18736623	3103675	16.56%	19986587	3004514	15.03%
$C_{8,2}$	25937515	10668600	41.13%	19499438	3250120	16.67%	20190244	3223008	15.96%
$C_{9,2}$	31855834	12800182	40.18%	28041595	5593105	19.95%	19957068	3186062	15.96%
$C_{1,3}$	20117650	8733421	43.41%	18612355	4385652	23.56%	20156303	3335526	16.55%
$C_{2,3}$	20542651	9147554	44.53%	18541678	3971289	21.42%	20108583	3238736	16.11%
$C_{3,3}$	21310177	9686295	45.45%	18462838	3812473	20.65%	20145591	3268911	16.23%
$C_{4,3}$	21477590	9504400	44.25%	18837637	4135333	21.95%	19948358	3061940	15.35%
$C_{5,3}$	22343862	10223909	45.76%	18747839	3990133	21.28%	20176181	3292139	16.32%
$C_{6,3}$	22343862	10223909	45.76%	19047612	4157201	21.83%	20235529	3272080	16.17%
$C_{7,3}$	24310768	10807611	44.46%	19076085	4097259	21.48%	20035417	3242712	16.18%
$C_{8,3}$	28014262	12603185	44.99%	18761047	3656633	19.49%	19142946	2882014	15.06%
$C_{1,4}$	20630611	8046460	39.00%	18687793	3909517	20.92%	20267021	3436755	16.96%
$C_{2,4}$	22249695	10461429	47.02%	18684856	3937601	21.07%	19876576	3036550	15.28%
$C_{3,4}$	22667283	10406643	45.91%	18705510	3832001	20.49%	20093135	3172822	15.79%
$C_{4,4}$	24518048	13000682	53.02%	18557870	3609162	19.46%	19570177	2908006	14.86%
$C_{5,4}$	22779253	9385193	41.20%	18901527	3378486	17.87%	20061531	3052453	15.22%
$C_{6,4}$	23456382	9576789	40.83%	19155170	3505763	18.30%	20461500	3338119	16.31%
$C_{7,4}$	25259838	10803101	42.77%	19510726	4429032	22.70%	20223866	3326220	16.45%
$C_{1,5}$	22429788	10811508	48.20%	18540276	3580452	19.31%	20065393	3138556	15.64%
$C_{2,5}$	21357813	8724737	40.85%	18721174	3222521	17.21%	20129514	3217831	15.99%
$C_{3,5}$	23550396	11611796	49.31%	18536008	3386596	18.27%	20026766	3237246	16.16%
$C_{4,5}$	21508813	7689022	35.75%	19086752	3294073	17.26%	20200353	2998589	14.84%
$C_{5,5}$	23346284	10065776	43.12%	18874771	2646329	14.02%	19780274	2904911	14.69%
$C_{6,5}$	24869408	11647766	46.84%	18784258	3702816	19.71%	19744418	3167619	16.04%
$C_{1,6}$	23951471	12916930	53.93%	18581943	3615560	19.46%	19641973	2936141	14.95%
$C_{2,6}$	21733141	8955657	41.21%	19289723	3924210	20.34%	20369335	3186426	15.64%
$C_{3,6}$	20575678	6049331	29.40%	18938587	3358215	17.73%	20086275	3192376	15.89%
$C_{4,6}$	23437188	10544310	44.99%	19215998	3494730	18.19%	20400990	3339052	16.37%
$C_{5,6}$	24484074	11528165	47.08%	18748141	3590374	19.15%	20023384	3201374	15.99%
$C_{1,7}$	21313505	7539055	35.37%	19721845	4310623	21.86%	20524783	3420288	16.66%
$C_{2,7}$	24033738	11792284	49.07%	18675868	3632869	19.45%	19885923	3201931	16.10%
$C_{3,7}$	24738817	12177617	49.22%	18898817	3672506	19.43%	19862547	3273394	16.48%
$C_{4,7}$	22958708	8709664	37.94%	20252613	4240824	20.94%	21069997	3198709	15.18%
$C_{1,8}$	22430392	8651428	38.57%	19540360	4759587	24.36%	20419567	3403837	16.67%
$C_{2,8}$	23474250	10085995	42.97%	18843755	3796999	20.15%	20296462	3472717	17.11%
$C_{3,8}$	24485861	10962561	44.77%	18793893	3587348	19.09%	20217493	3416093	16.90%
$C_{1,9}$	27410145	10466799	38.19%	25692797	6379276	24.83%	20900218	3566671	17.07%
$C_{2,9}$	28587845	12135176	42.45%	25482099	5366916	21.06%	55609323	13197632	23.73%
$C_{1,10}$	28171197	9697075	34.42%	26899234	5736970	21.33%	29119964	5440289	18.68%

A similar phenomenon appears if we divide each observation by 10. For example, if we divide the value of the cell $C_{4,4} = 1562400$ by 10 the value of the outstanding reserves becomes lower (18351108), but if we divide the value of the cell $C_{2,2} = 766940$ by 10, the value of the outstanding reserves becomes higher (19292133).

Huber M-estimation robustifies very well most of the outliers except in the upper right corner that contains the cells (observations), $C_{10,1}$, $C_{9,2}$ and $C_{9,1}$ observations, and the lower left corner that contains the cells, $C_{1,9}$, $C_{2,9}$ and $C_{1,10}$ (see observations in red color in Table 2.1). Verdonck et al. (2009) faced similar situations applying robust generalized linear model techniques and explained why robust methods can fail in some borderline cases and presented some ad-hoc solutions. Venter and Tampubolon (2010) illustrated that these observations on the upper right corner and on the lower left corner tend to have high impact on the estimation of outstanding reserves and suggested models to limit the impact of these cells.

Similar behavior appears with the application of robust Anova that produces slightly different values than M-estimation, but with similar behavior in the upper right corner and the lower left corner. The values of standard errors based on robust Anova are a bit lower than the standard errors based on M-estimation, in comparison with LS estimation that produces much higher values of standard errors than M-estimation and robust Anova.

Remark 2.8. At this point, for comparison reasons, we may suggest as the true value of outstanding reserves a) the reserves with the original data (no artificial outlier) based on M-estimation adding the value of the bias term due to the robustification (18558601+1073464=19632065), or b) the corresponding value based on robust ANOVA (19842279+459483=20301762), or c) the corresponding value based of LS - estimation (19468286), see Table 2.2. Of course, experts' opinion is necessary for the ultimate decision on the true value of outstanding reserves.

2.4 Concluding Remarks

In this chapter we have shown how robust estimation techniques can be incorporated in a loss reserving framework, providing a fair value for the estimation of outstanding reserves. An overview of the main robust estimators has been given in the first part of the chapter. Least squares estimators and robust estimators were applied to Ashe and Taylor data. Table 2.10 presents an overview of the robust loss reserving estimation in a log-linear model with the original data (without artificial outliers) and the presence of 1 and 2 additional (artificial) outliers, based on each individual robust estimator presented in Section 2.2. The data was tested in detail and few diagnostic plots were presented for the investigation of outlier events. The values of claim reserves were presented and the sensitivity of log-linear loss reserving model was shown by embedding one or two artificial outliers to the data set. Implementing the data set we have seen the superiority of robust M-estimator and Robust ANOVA in comparison with the least squares estimator and the rest of robust estimators for claims reserves estimation. With the original data and the presence of 1 and 2 outliers, the values of outstanding reserves estimated based on robust M-estimates are almost the same to the value of outstanding reserves estimated based on non-robust estimation techniques with the original data (pointing out again the existence of 1 outlier in the original data). In general, M-estimation and robust ANOVA show a consistency on the expected ultimate losses and reserves with the existence of these two artificial extreme events. Robust estimators that are robust to outliers in the design matrix space, such as the LMS, LTS, MM and adaptive estimators may be unreliable for the loss reserving estimation with the presence of outlier events, which means that these estimators robustify also the independent variables (in our case dummy variables) and are not appropriate for the robust loss reserving estimation. Since the influence of an outlier depends on its location on the triangle, it is appropriate to see the behavior of the run-off triangle in the presence of an outlier in each location. As expected, the Log-linear LS estimation cannot handle a single outlier and most of the times produces an overestimation of the outstanding reserves, although some times we have an underestimation of outstanding reserves, depending on the location of that outlier. The loss reserving estimation based on M-estimation and robust ANOVA works well independently of the location of the outlier (large or catastrophic event) in the triangle, except in the observations in the upper right corner and in the lower left corner of the triangle. Special attention must be paid to these sensitive observations, which must be treated (diagnostic tests) separately, in order to obtain a complete robustness of the loss reserving estimation. The same behavior is observed if artificial outliers are multiplied by 100. M-estimation and robust ANOVA work also very well for the loss reserving estimation if we apply an ordinary linear regression model instead of a log-linear regression model.

The advantage of applying M-estimators and robust ANOVA for the estimation of the outstanding reserves is that they are simple, can be computed in much less time than other robust estimators and there is a wide literature on M-estimation for regression and diagnostic tools as well. The disadvantage of M-estimators is that they don't enjoy high breakdown point (BP) and we have to take into consideration the number of outliers in the data.

Table 2.10: Overview of Loss Reserving Estimation

Behavior of LS and Robust Estimators in the presence of 0,1,2 outliers			
Models	**Number of outliers**	**Overview**	**Explanation**
LS-Estimates	0	∨ ∨ ∨∨	OK Reserves (R)
LS-Estimates	1	∨	High values of R
LS-Estimates	2	×	Very High values of R
M-Estimates	0	∨ ∨ ∨∨	OK Robust R
M-Estimates	1	∨ ∨ ∨∨	OK Robust R
M-Estimates	2	∨ ∨ ∨∨	OK Robust R
LAD-estimates	0	∨	High values of R
LAD-estimates	1	∨	High values of R
LAD-estimates	2	∨	High values of R
Robust ANOVA	0	∨ ∨ ∨∨	OK Robust R
Robust ANOVA	1	∨ ∨ ∨∨	OK Robust R
Robust ANOVA	2	∨ ∨ ∨∨	OK Robust R
MM-Estimates	0	∨ ∨ ∨	OK Robust R
MM-Estimates	1	×	Very High values of R
MM-Estimates	2	×	Very High values of R
Robust Adaptive	0	∨ ∨ ∨	OK Robust R
Robust Adaptive	1	×	Very High values of R
Robust Adaptive	2	×	High values of R
LTS-estimates	0	∨	High values of R
LTS-estimates	1	×	Very High values of R
LTS-estimates	2	×	Very High values of R
LMS-estimates	0	×	Very low values of R
LMS-estimates	1	×	Very High values of R
LMS-estimates	2	×	Very High values of R

Another advantage involved in the use of robust regression is the availability of R packages for the performance of different robust regression procedures. The R-project has the additional advantage of being interactive with other spreadsheet programs (excel) for better organizing and presenting data and results. It should also be emphasized that the robust statistical procedure should be followed by the practical knowledge of the actuary for estimating the outstanding reserves. Of course, the bias term which is due to the robustification of the original values of the triangle, must be added at the end to the value of the ultimate claims to obtain total unbiasedness.

CHAPTER 3

Random Coefficients Regression Models

In insurance loss reserving applications, the concept of constancy of regression coefficients (RC) in consecutive observations is a matter of discussion. In some situations in which the coefficients are random, this happens due to the fact that the development of claims varies with accident years. The model of RC frequently implies to various decision making problems. As far as the usual regression model is concerned, the decision will have an effect only in the average outcome, while in the RC model not only the average but also the variance of the outcome will be affected. In this chapter, we illustrate two models, the individual purely random coefficients model and the RCR cross effect model.

3.1 Individual Random Coefficients Model

In general, in loss reserving estimation, we make the assumption that the development factors (DFs) remain the same for every accident period. Here, we relax that assumption in order to permit the DFs be subject of the random variation. Verrall (1994) suggested a model in order to permit the evolution of the DFs to evolve in a recursive way. Next, we use the model of Hildreth and Houck (1968) where the response parameters in a GLM are supposed to be random variables and we can estimate the distribution's mean. Swamy (1974), Raj and Ullah (1981), and Hsiao (1986, 2003) have studied RC models in more details.

The linear regression model with many explanatory variables and RC coefficients is given by

$$Y_i = \beta_{1i} + \sum_{k=2}^{p}(\beta_k + \upsilon_{ki})X_{ki} + \upsilon_{0i} = \beta_1 + \sum_{k=2}^{p}\beta_k X_{ki} + u_i,\ i = 1, ..., n, \tag{3.1}$$

where Y is the dependent variable and the Xs are the independent variables. We make the assumption that υ_{0i} are independent variables with zero mean and variance σ_0^2, while β_k and υ_{ki} are the the two parts of the parameters, with β_{ki}'s be the RC with $E(\beta_{ki}) = \beta_k$, for all i, $\mathrm{Var}(\beta_{ki}) = \mathrm{Var}(\upsilon_{ki}) = \sigma_k^2$, for all i, or $\mathrm{Var}(\upsilon_{ki}) = E(\upsilon_{ki}^2) = \sigma_k^2$ for all i, $\mathrm{Cov}(\beta_{ki}, \beta_{k'i'}) = 0$, or $\mathrm{Cov}(\upsilon_{ki}, \upsilon_{k'i'}) = 0$ for $i \neq i'$ and $k \neq k'$, where $i' = 1, ..., n$ and $k' = 1, ..., p$.

In (3.1) we also have $u_i = \upsilon_{0i} + \upsilon_{1i} + \sum_{k=2}^{p} X_{ki}\upsilon_{ki}$, with $E(u_i) = 0$, $\mathrm{Cov}(u_i, u_{i'}) = 0\ for\ i \neq i'$ and $\mathrm{Var}(u_i) = (\sigma_0^2 + \sigma_1^2) + \sum_{k=2}^{p}\sigma_k^2 X_{ik}^2 = \lambda_{ii}$. For n observations (3.1) is written as

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{u}, \tag{3.2}$$

where $E(\boldsymbol{u}) = 0$ and $\boldsymbol{\Lambda} = E(\boldsymbol{uu}') = \mathrm{diag}(\lambda_{11}, ..., \lambda_{nn})$. The regression coefficients $\boldsymbol{\beta} = (\beta_1, \beta_2, ..., \beta_p)'$ and their variance are estimated by

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{\Lambda}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Lambda}^{-1}\boldsymbol{Y}, \mathrm{Var}(\widehat{\boldsymbol{\beta}}) = (\boldsymbol{X}'\boldsymbol{\Lambda}^{-1}\boldsymbol{X})^{-1}. \tag{3.3}$$

The fact that σ_k^2's are not known, leads to an unknown $\boldsymbol{\Lambda}$. So, in (3.3), we use an estimator of $\boldsymbol{\Lambda}$. Hildreth and Houck (1968) suggested a method which is shown below.

Let $\boldsymbol{r} = \boldsymbol{Y} - \boldsymbol{Xb} = \boldsymbol{MY} = \boldsymbol{Mu}$ be the least square residuals and $\boldsymbol{b}$ is the least squares estimator of $\boldsymbol{\beta}$, i.e., $\boldsymbol{b} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}$ and $\boldsymbol{M} = \boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$. Then we have $Er_i = 0$ and

$$E(r_i^2) = \mathrm{Var}(r_i) = \sum_{j=1}^{n} m_{ij}\lambda_{ij} \quad i.e. \quad E(\dot{\boldsymbol{r}}) = \dot{\boldsymbol{M}}\dot{\boldsymbol{X}}\dot{\sigma}, \tag{3.4}$$

where $\dot{\boldsymbol{r}} = (r_1^2, ..., r_n^2)'$, $\dot{\boldsymbol{M}} = \boldsymbol{M} * \boldsymbol{M} = \{m_{ij}^2\}_{i,j=1,...,n}$ and $\dot{\boldsymbol{\sigma}} = (\sigma_1^2, ..., \sigma_p^2)'$. We use $*$ in order to symbolize the Hadamard matrix product, i.e. for two matrices $\boldsymbol{A} = \{a_{ij}\}_{i,j=1,..n}$ and $\boldsymbol{B} = \{b_{ij}\}_{i,j=1,..n}$, $\boldsymbol{A} * \boldsymbol{B} = \{a_{ij}b_{ij}\}_{i,j=1,..n}$. Then (3.4) can be written as

$$\dot{\boldsymbol{r}} = \dot{\boldsymbol{M}}\dot{\boldsymbol{X}}\ \boldsymbol{\sigma} + \boldsymbol{w} = \boldsymbol{G}.\dot{\boldsymbol{\sigma}} + \boldsymbol{w}. \tag{3.5}$$

The formulation of (3.5) is similar to a regression LSE model with dependent variable $\dot{\boldsymbol{r}}$ and independent variable $\dot{\boldsymbol{\sigma}}$, with $\boldsymbol{G} = \dot{\boldsymbol{M}}\dot{\boldsymbol{X}}$, $E(\boldsymbol{w}) = 0$ and the variance-covariance matrix is $\mathrm{Cov}(\boldsymbol{w}) = E(\boldsymbol{ww}') = 2\dot{\boldsymbol{\Omega}}$, with $\boldsymbol{\Omega} = E(\boldsymbol{rr}') = E(\boldsymbol{Muu}'\boldsymbol{M}) = \boldsymbol{M\Lambda M}$. Then the generalized LSE model of $\dot{\sigma}$ in (3.5) is given by (see Hildreth and Houck, 1968)

$$\widehat{\dot{\boldsymbol{\sigma}}} = (\boldsymbol{G}^T\dot{\boldsymbol{\Omega}}^{-1}\boldsymbol{G})^{-1}\boldsymbol{G}^T\dot{\boldsymbol{\Omega}}^{-1}\dot{\boldsymbol{r}}, \tag{3.6}$$

where

$$\hat{\dot{\boldsymbol{\Omega}}} = \widehat{\boldsymbol{\Omega}} * \widehat{\boldsymbol{\Omega}} \quad \text{with} \quad \widehat{\boldsymbol{\Omega}} = \boldsymbol{M}\widehat{\boldsymbol{\Lambda}}\boldsymbol{M} \tag{3.7}$$

and $\widehat{\boldsymbol{\Lambda}} = \mathrm{diag}(\hat{\lambda}_{11}, ..., \hat{\lambda}_{11})$, with $\hat{\lambda}_{ii} = \hat{\sigma}_1^2 + \sum_k^p \hat{\sigma}_k^2 X_{ki}^2$.

3.1.1 Actual Response Coefficients in a Purely Random Coefficients Model

Sometimes it is important to make predictions about the individual part of the actual response coefficients $\boldsymbol{\beta}_i$ in order to derive all needed information about the the behavior of all claims separately. Griffiths (1972) studied the actual response coefficients and showed that the best estimator for them is not the same as the best estimator of the mean response coefficients.

The vector $\boldsymbol{\beta}$ in (3.3) contains the mean response coefficients and the response coefficients are given by

$$\boldsymbol{b} = \boldsymbol{L}\boldsymbol{\beta} + \boldsymbol{v}, \tag{3.8}$$

where $\boldsymbol{L}$ is a $p \times np$ matrix, $\boldsymbol{b}^{'}$ a $np \times 1$ vector defined respectively as

$$\boldsymbol{L}^{'} = \begin{pmatrix} 1\ 1..\ 1 & & & \\ & 1\ 1...\ 1 & & \\ & & & \\ & & & 1\ 1...\ 1 \end{pmatrix}, \tag{3.9}$$

$$\boldsymbol{b}^{'} = (\boldsymbol{b}_1^{'}, \boldsymbol{b}_2^{'}, ..., \boldsymbol{b}_p^{'})^{'}, \text{ with } \boldsymbol{b}_k^{'} = (b_{1k}, b_{2k}, ..., b_{nk})$$

and

$$\boldsymbol{v}^{'} = (\boldsymbol{v}_1^{'}, \boldsymbol{v}_2^{'}, ..., \boldsymbol{v}_p^{'}), \text{ with } \boldsymbol{v}_k^{'} = (v_{1k}, v_{2k}, ..., v_{nk}). \tag{3.10}$$

Then we obtain

$$\boldsymbol{A} = E(\boldsymbol{v}\boldsymbol{v}^{'}) = \begin{pmatrix} \sigma_{11}\boldsymbol{I} & & & \\ & \sigma_{22}\boldsymbol{I} & & \\ & & . & \\ & & & \sigma_{pp}\boldsymbol{I} \end{pmatrix}. \tag{3.11}$$

If $\boldsymbol{X}_k = diag(X_{1k}, X_{2k}, ..., X_{nk})$ and $\boldsymbol{Z}$ a $n \times np$ matrix defined as

$$\boldsymbol{Z} = (\boldsymbol{X}_1, \boldsymbol{X}_2, ..., \boldsymbol{X}_p) \tag{3.12}$$

then

$$\boldsymbol{u} = \boldsymbol{Z}\boldsymbol{v} \text{ and } \boldsymbol{\Lambda} = E(\boldsymbol{u}\boldsymbol{u}^{'}) = \boldsymbol{Z}\boldsymbol{A}\boldsymbol{Z}^{'}. \tag{3.13}$$

Following Griffiths (1972), the kth diagonal element of $\boldsymbol{V}$, estimates the errors derived by the kth coefficient according to the matrix form

$$\hat{\boldsymbol{v}}_k = \sigma_{kk}\boldsymbol{X}_k\boldsymbol{\Lambda}^{-1}\hat{\boldsymbol{u}}. \tag{3.14}$$

When estimations of the disturbances are included in (3.14), it becomes

$$\hat{\boldsymbol{v}} = \boldsymbol{A}\boldsymbol{Z}_k\boldsymbol{\Lambda}^{-1}\hat{\boldsymbol{u}}. \tag{3.15}$$

Then, we obtain an estimator of $\boldsymbol{b}$ as follows

$$\begin{aligned} \hat{\boldsymbol{b}} &= \boldsymbol{L}\hat{\boldsymbol{\beta}} + \hat{\boldsymbol{v}} \\ &= \boldsymbol{L}(\boldsymbol{X}^{'}\boldsymbol{\Lambda}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}\boldsymbol{\Lambda}^{-1}\boldsymbol{Y} + \boldsymbol{A}\boldsymbol{Z}^{'}\boldsymbol{\Lambda}^{-1}\hat{\boldsymbol{u}}, \end{aligned} \tag{3.16}$$

with $\hat{\boldsymbol{u}} = [\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}^{'}\boldsymbol{\Lambda}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^{'}\boldsymbol{\Lambda}^{-1}]\boldsymbol{Y}$, $\hat{\boldsymbol{b}}$ is

$$\hat{\boldsymbol{b}} = \boldsymbol{L}(\boldsymbol{X}^{'}\boldsymbol{\Lambda}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}\boldsymbol{\Lambda}^{-1} + \boldsymbol{A}\boldsymbol{Z}^{'}\boldsymbol{\Lambda}^{-1}[\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}^{'}\boldsymbol{\Lambda}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^{'}\boldsymbol{\Lambda}^{-1}]\boldsymbol{Y}. \tag{3.17}$$

3.2 Random Coefficients Model in Loss Reserving Procedure

The run-off triangles (see Table 3.1) are divided into cells where each cell is the corresponding payment arising from a specific accident year $i \in \{1, \cdots, n\}$ (rows) and a development year $j \in \{1, \cdots, n\}$ (columns). The accident year shows the losses that occurred during a specific period while the development year specifies after how many years of the claim reported it is getting settled.

The calendar year k is the diagonal element of the triangle and is defined as $k = i + j$, with $k \in \{1, \cdots, n\}$. Then $Y_{i,j}$ is defined as the total incremental payments in accident year i with development j, where $i+j \leq n$ because the calendar $i+j > n$ has not occurred yet. Some reserving methods use the cumulative payments $S_{i,j}$ where $S_{i,n} = \sum_{j=1}^{n} Y_{i,j}$, with $S_{i,0} = Y_{i,0}$. We will denote with D_n the available data up to period n (the known section of the triangle)

$$D_n = \{Y_{i,j}, i + j \leq n\} = \{S_{ij}, i + j \leq n\}.$$

Table 3.1: Run-off Triangle of Paid Claims

Accident Year i	Development Year j: 1	2	$\cdots$	j	$\cdots$	n-1	n
1	$Y_{1,1}$	$Y_{1,2}$	$\cdots$	$Y_{1,j}$	$\cdots$	$Y_{1,n-1}$	$Y_{1,n}$
2	$Y_{2,1}$	$Y_{2,2}$	$\cdots$	$Y_{2,j}$	$\cdots$	$Y_{2,n-1}$	
$\vdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$		
i	$Y_{i,1}$	$\cdots$	$\cdots$	$Y_{i,n+1-i}$			
$\vdots$	$\cdots$	$\cdots$	$\cdots$				
n	$Y_{n,1}$						

For accident year i, we want to get the best estimate for the ultimate amount of payment, i.e.

$$\widehat{S}_{i,\infty}^{n-i} = \lim_{j\to\infty} E[S_{ij}|D_n] = E[S_{i,\infty}|D_n]$$

and the required reserve will be calculated as the total payment amount minus the payments done at the specific year

$$\widehat{R}_i = \widehat{S}_{i,\infty}^{n-i} - S_{i,n-i+1}.$$

Finally, a critical quantity is the (total) uncertainty, which is the variance of the total year amount of payment $Var(S_{i,\infty}|D_n)$ or $Var(\widehat{S}_{i,\infty}^{n-i})$. In our examples, we make the assumption that all claims have a lifetime of n years. After n years all claims are closed, which means that $Y_{i,j>n} = 0$.

The RCR model defined in (3.1) in connection with Table 3.1 is

$$\begin{pmatrix} Y_{1,\leq j} \\ \\ \\ Y_{n,\leq j} \end{pmatrix} = \begin{pmatrix} 1 & X_{11} & ... & X_{p-1,1} \\ \\ \\ 1 & X_{1n} & ... & X_{p-1,n} \end{pmatrix} \begin{pmatrix} \beta_1 \\ \\ \\ \beta_p \end{pmatrix} + \begin{pmatrix} \boldsymbol{X}_1' & 0 & . & . & . & 0 \\ 0 & . & & & & 0 \\ & . & & & & . \\ & & & & & . \\ 0 & 0 & . & . & . & \boldsymbol{X}_n' \end{pmatrix} \begin{pmatrix} \boldsymbol{v}1 \\ \\ \\ \boldsymbol{v}_n \end{pmatrix} \tag{3.18}$$

or compactly

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{D}_x\boldsymbol{v}, \tag{3.19}$$

where

(*i*) $\boldsymbol{Y} = (Y_{1,\leq j}, ..., Y_{n,\leq j})'$ is a $n(n+1)/2 \times 1$ vector of the incremental claims at the accident year i and development year j,

(*ii*) $\boldsymbol{X} = (\boldsymbol{X}'_1, ..., \boldsymbol{X}'_n)'$ is a $n(n+1)/2 \times p$ design matrix, with $p = 2n - 1$, reflecting the position of incremental claims in Table 3.1,

(*iii*) $\boldsymbol{\beta} = (\beta_1, ..., \beta_p)'$, with $p = 2n - 1$, are unknown parameters,

(*iv*) $\boldsymbol{D}_x$ is a diagonal block of $\boldsymbol{X}'_i s$, in the right hand side of (3.18) and $\boldsymbol{X}'_i$ is the i-th row of the design matrix $\boldsymbol{X}$,

(*v*) $\boldsymbol{\upsilon} = (\boldsymbol{\upsilon}'_1, ..., \boldsymbol{\upsilon}'_n)'$ is a $n(n+1)/2 \times 1$ vector, with $\boldsymbol{\upsilon}_i = (\upsilon_{1,i}, .., \upsilon_{n,i})'$.

Remark 3.1. *The magnitude of p depends on the choice of the appropriate design matrix that reflects the position of the claims in the run-off triangle (see Christofides, 1990).*

3.3 RCR Cross Effect Model

In the following, we present a multivariate regression model with random coefficients that handles multiple run-off triangles from different lines of business. Here, we have a generalization of the chain ladder model to a multivariate settings where multiple run-off triangles are estimated simultaneously. The model is defined for the i^{th} cross-section unit (i^{th} triangle) different years of observations. Hence, we consider a RCR model with cross effects as developed by Swamy (1971, 1974), it allows for differences in behavior over multiple run-off triangles (units) and within each unit (triangle) the data are varying over time. The dependent variable is the claim incurred experience for a certain unit (triangle) and the time period and the explanatory variables are dummy variables that show the position of incurred (or paid) claims in a chain ladder representation. Swamy (1971) studied the regression coefficient vector supposing that it is a random quantity and showed that they are similar to the Bayesian ones. The model is defined as

$$Y_{it} = \sum_{k=1}^{p} \beta_{kit} X_{kit} + u_{it}, \tag{3.20}$$

with

$$\beta_{kit} = \beta_{ki} = \beta_k + \epsilon_{ki}, \tag{3.21}$$

where $i = 1, ..., N$ are the cross-sectional units, $t = 1, ..., T$ represents the time period, Y_{it} is the dependent variable for the i^{th} individual at time t, X_{kit} is the value of the k^{th}

independent variable for the i^{th} individual at time t and β_{kit} shows the response of Y_{it} to a unit change in X_{kit}. Then, n cross-sectional groups with T known data are used. The disturbances u_{it} in (3.20) are all independent such that $E(u_{it}) = 0$ and $V(u_{it}) = \sigma_i^2$. The β_{ki} are random parameters with mean β_k and variances σ_{kk}^2, i.e. $E(\epsilon_{ki}) = 0$ and $V(\epsilon_{ki}) = \sigma_{kk}^2$. Also $\text{Cov}(\epsilon_{ki}, \epsilon_{kl}) = 0$ for $l \neq i$, $l = 1, ..., N$. The ϵ_{ki} and u_{it} are jointly independent. Then, the model (3.20) is (Swamy, 1974)

$$Y_{it} = \sum_{k=1}^{p} \beta_k X_{kit} + w_{it}, \tag{3.22}$$

where

$$w_{it} = \sum_{k=1}^{p} \epsilon_{ki} X_{kit} + u_{it}. \tag{3.23}$$

If we now write (3.22) in matrix notation for the i^{th} cross-section unit as

$$\boldsymbol{Y}_i = \boldsymbol{X}_i \boldsymbol{\beta} + \boldsymbol{w}_i, \tag{3.24}$$

where $\boldsymbol{Y}_i = (Y_{i1}, ..., Y_{iT})'$, $\boldsymbol{\beta} = (\beta_1, ..., \beta_p)'$, $\boldsymbol{X}_i = (\boldsymbol{X}_{i1}', ..., \boldsymbol{X}_{iT}')'$ a $T \times p$ matrix and $\boldsymbol{w}_i = (w_{i1}, ..., w_{iT})'$, then $E(w_i) = 0$ and

$$\boldsymbol{\Omega}_i = E(\boldsymbol{w}_i\, \boldsymbol{w}_l') = \begin{cases} \sigma_i^2 I + \boldsymbol{X}_i \boldsymbol{\Sigma} \boldsymbol{X}_i', & \text{if} \quad l = i, \\ 0, & \text{if} \quad l \neq i, \end{cases} \tag{3.25}$$

where $\boldsymbol{\Sigma}$ is a diagonal matrix with elements $\sigma_{11}, ..., \sigma_{kk}$. For the full sample of observations, $\boldsymbol{\Omega}$ is a block-diagonal matrix with blocks $\boldsymbol{\Omega}_1, ..., \boldsymbol{\Omega}_n$. Now let $\hat{\boldsymbol{\beta}}_i$ be the i^{th} ordinary least squares given by $\widehat{\boldsymbol{\beta}}_i = (\boldsymbol{X}_i'\boldsymbol{X}_i)^{-1}\boldsymbol{X}_i'\boldsymbol{Y}_i$, then the estimator of $\boldsymbol{\beta}$, is known as the GLS estimator and is calculated by

$$\widehat{\boldsymbol{\beta}}_{GLS} = \sum_{i=1}^{n} \boldsymbol{\Lambda}_i \hat{\boldsymbol{\beta}}_i, \tag{3.26}$$

which is a matrix of weighted combination of ordinary least squares estimator $\widehat{\boldsymbol{\beta}}_i$ of $\boldsymbol{\beta}_i$. Following Greene (2000, p. 610) the matrix of weights is

$$\begin{aligned} \boldsymbol{\Lambda}_i &= \left[\sum_{l=1}^{N} (\boldsymbol{\Sigma} + \sigma_l^2 (\boldsymbol{X}_l'\boldsymbol{X}_l)^{-1})^{-1}\right]^{-1} [\boldsymbol{\Sigma} + \sigma_i^2 (\boldsymbol{X}_i'\boldsymbol{X}_i)^{-1}]^{-1} \\ &= \left[\sum_{i=1}^{N} \boldsymbol{X}_i'\boldsymbol{\Omega}_i^{-1}\boldsymbol{X}_i\right]^{-1} \boldsymbol{X}_i'\boldsymbol{\Omega}_i^{-1}\boldsymbol{X}_i, \end{aligned} \tag{3.27}$$

such that $\sum_{i=1}^{N} \boldsymbol{\Lambda}_i = I$. The GLS estimator of $\boldsymbol{\beta}$ is given by

$$\widehat{\boldsymbol{\beta}}_{GLS} = (\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{Y} \tag{3.28}$$

and the variance-covariance matrix of $\widehat{\boldsymbol{\beta}}$ then is given by

$$\text{Cov}(\widehat{\boldsymbol{\beta}}_{GLS}) = \left[\sum_{i=1}^{N}(\boldsymbol{\Sigma} + \sigma_i^2(\boldsymbol{X}_i'\boldsymbol{X}_i)^{-1})^{-1}\right]^{-1} = (\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{X})^{-1}. \tag{3.29}$$

The unknown matrix $\boldsymbol{\Sigma}$ can be estimated as

$$\hat{\boldsymbol{\Sigma}} = \frac{1}{n-1}\left[\sum_{i=1}^{N}\hat{\boldsymbol{\beta}}_i\hat{\boldsymbol{\beta}}_i' - \frac{1}{N}\sum_{i=1}^{N}\hat{\boldsymbol{\beta}}_i\sum_{i=1}^{N}\hat{\boldsymbol{\beta}}_i'\right] - \frac{1}{N}\sum_{i=1}^{N}(\hat{\sigma}_i)^2(\boldsymbol{X}_i'\boldsymbol{X}_i)^{-1}, \tag{3.30}$$

where

$$\hat{\sigma}_i^2 = \frac{\boldsymbol{w}_i'\boldsymbol{w}_i}{T-n} = \frac{1}{T-n}\boldsymbol{Y}_i'[\boldsymbol{I} - \boldsymbol{X}_i(\boldsymbol{X}_i'\boldsymbol{X}_i)^{-1}\boldsymbol{X}_i]\boldsymbol{Y}_i. \tag{3.31}$$

Sometimes it is useful to predict the individual component $\boldsymbol{\beta}_i$, because it provides not only information on the behavior of each individual, but also a basis for predicting future values of the dependent variable for a given individual. Swamy (1970, 1971) has shown that the best linear unbiased predictor, conditional on given $\boldsymbol{\beta}_i$, is the least-squares estimator $\widehat{\boldsymbol{\beta}}_i$. On the other hand, if the sampling properties of the class of predictors are considered in terms of repeated sampling over both time and individuals, Lee and Griffiths (1979) (see also Hsiao, 2003) suggested estimating $\boldsymbol{\beta}_i$ by

$$\widehat{\boldsymbol{\beta}}_i^{\star} = \widehat{\boldsymbol{\beta}}_{GLS} + \boldsymbol{\Sigma}\boldsymbol{X}_i'(\boldsymbol{X}_i\boldsymbol{\Sigma}\boldsymbol{X}_i' + \sigma_i^2\boldsymbol{I}_T)^{-1}(\boldsymbol{Y}_i - \boldsymbol{X}_i\widehat{\boldsymbol{\beta}}_{GLS}). \tag{3.32}$$

This predictor is the best linear unbiased estimator in the sense of $E(\widehat{\boldsymbol{\beta}}_i^{\star} - \boldsymbol{\beta}_i) = 0$, where the expectation is an unconditional one.

3.3.1 RCR Cross Effect Model in Loss Reserving

Suppose that N same sized run-off triangles are available and $i \in \{1, 2, \cdots, N\}$ refers to the i^{th} triangle while $r \in \{1, 2, \cdots, I\}$ refers to the accident year and $j \in \{1, 2, \cdots, I\}$ refers to the development year. Denote, $\boldsymbol{Y}_{r,j} = (Y_{r,j}^{(1)}, \cdots, Y_{r,j}^{(N)})'$ the $N \times 1$ vector with the incremental losses at accident year r and development year j for all triangles N. The data for the N run-off triangles is displayed in Table 3.2.

The overall outstanding reserve R that needs to be paid in future, is defined as

$$R = \sum_{i=1}^{N}\sum_{r=2}^{I}\left(S_{r,I}^{(i)} - S_{r,I-r+1}^{(i)}\right) \tag{3.33}$$

where for run-off triangle i, the cumulative claims of accident year r and development year j are given by

$$S_{r,j}^{(i)} = \sum_{k=1}^{j} Y_{r,k}^{(i)}$$

and $Y_{r,k}^{(i)}$ are the incremental claims of run-off triangle i, the cumulative claims of accident year r and development year j. Our purpose is to estimate the future claims in the bottom right corner of the run-off triangles and estimate the overall outstanding reserves.

Table 3.2: Representation of N run-off triangles

Accident	Development Year j						
Year r	1	2	$\cdots$	j	$\cdots$	I-1	I
1	$\boldsymbol{Y}_{1,1}$	$\boldsymbol{Y}_{1,2}$	$\cdots$	$\boldsymbol{Y}_{1,j}$	$\cdots$	$\boldsymbol{Y}_{1,I-1}$	$\boldsymbol{Y}_{1,I}$
2	$\boldsymbol{Y}_{2,1}$	$\boldsymbol{Y}_{2,2}$	$\cdots$	$\boldsymbol{Y}_{2,j}$	$\cdots$	$\boldsymbol{Y}_{2,I-1}$	
$\vdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$		
r	$\boldsymbol{Y}_{r,1}$	$\cdots$	$\cdots$	$\boldsymbol{Y}_{r,I+1-r}$			
$\vdots$	$\cdots$	$\cdots$	$\cdots$				
I	$\boldsymbol{Y}_{I,1}$						

Denote $D = \{\boldsymbol{Y}_{r,j},\ r+j \leq I+1,\ 1 \leq r \leq I,\ 1 \leq j \leq I\}$ as the observed losses, $D_{\cdot,j} = \{\boldsymbol{Y}_{r,j}\ 1 \leq r \leq I,\ j \leq k\}$ as the losses up to development year k (including it) and $D_{r,j} = \{\boldsymbol{Y}_{r,j},\ k \leq j\}$ as the losses for accident year r up to development year j (including it). According to the data, the sets $D_{\cdot,j}$ and $D_{r,j}$ are the observed values and should be used to estimate the adequate reserve to fund losses that have been incurred but not yet developed.

Similarly as in Zhang (2010) and Peremans et al. (2018), the cross section effects model defined in (3.22) and (3.23) in connection with Table 3.2 can be written in a multivariate form as

$$\begin{pmatrix} \boldsymbol{Y}_{,\leq j}^{(1)} \\ \\ \\ \boldsymbol{Y}_{,\leq j}^{(N)} \end{pmatrix} = \begin{pmatrix} \boldsymbol{X}_{j}^{(1)} \\ \\ \\ \boldsymbol{X}_{j}^{(N)} \end{pmatrix} \begin{pmatrix} \boldsymbol{\beta}_{j}^{(1)} \\ \\ \\ \boldsymbol{\beta}_{j}^{(N)} \end{pmatrix} + \begin{pmatrix} \boldsymbol{X}_{j}^{(1)} & 0 & . & . & . & 0 \\ 0 & . & & & & 0 \\ & . & & & & . \\ & & & & & . \\ 0 & 0 & . & . & . & \boldsymbol{X}_{j}^{(N)} \end{pmatrix} \begin{pmatrix} \boldsymbol{\epsilon}_{j}^{(1)} \\ \\ \\ \boldsymbol{\epsilon}_{j}^{(N)} \end{pmatrix} + \begin{pmatrix} \boldsymbol{u}_{j}^{(1)} \\ \\ \\ \boldsymbol{u}_{j}^{(N)} \end{pmatrix}, \tag{3.34}$$

which can be written in a matrix notation as

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{D}(\boldsymbol{X})\boldsymbol{\epsilon} + \boldsymbol{u}, \tag{3.35}$$

where

(*i*) $\boldsymbol{Y} = (\boldsymbol{Y}_{,\leq j}^{(1)'}, ..., \boldsymbol{Y}_{,\leq j}^{(N)'})'$, with $\boldsymbol{Y}_{,\leq j}^{(i)} = (Y_{1,j}^{(i)}, ..., Y_{I,j}^{(i)})'$ a $I(I+1)/2 \times 1$ vector of the incremental claims at the accident year r and the development year j from the i^{th} triangle, (*ii*) $\boldsymbol{X} = (\boldsymbol{X}_j^{(1)'}, ..., \boldsymbol{X}_j^{(N)'})'$, where $\boldsymbol{X}_j^{(i)}$ is a $I(I+1)/2 \times p$ design matrix, with $p = 2I-1$, at development year j from each run-off triangle reflecting the position of claims in each of the triangles $i = 1, ..., n$. In our case we consider the same design matrix $\boldsymbol{X}_j^{(i)}$ for all triangles.

(*iii*) $\boldsymbol{\beta} = (\boldsymbol{\beta}_j^{(1)}, ..., \boldsymbol{\beta}_j^{(N)})'$, with $\boldsymbol{\beta}_j^{(i)} = (\beta_{1j}^{(i)}, ..., \beta_{pj}^{(i)})'$ and $p = 2I - 1$

(*iv*) $\boldsymbol{\epsilon} = (\boldsymbol{\epsilon}_j^{(1)'}, ..., \boldsymbol{\epsilon}_j^{(N)'})'$, with $\boldsymbol{\epsilon}_j^{(i)} = (\epsilon_{1,j}^{(i)}, .., \epsilon_{I,j}^{(i)})'$

(*v*) $\boldsymbol{u} = (\boldsymbol{u}_j^{(1)'}, ..., \boldsymbol{u}_j^{(N)'})'$, with $\boldsymbol{u}_j^{(i)} = (u_{1,j}^{(i)}, .., u_{I,j}^{(i)})'$

(*vi*) $\boldsymbol{D}(\boldsymbol{X})$ denotes the block-diagonal matrix in the right hand side of (3.34).

3.4 Robust Estimation and Identification of Outliers for the Insurance Application

A lot of insurance data cannot be studied by short tail distributions. Robust regression methods are used as auxiliary tools for the ordinary least squares estimation in cases of errors which are not normally distributed because of large deviations or outlier claims.

An unusual event such as a catastrophe or data that belongs to different populations, can cause an outlier. Therefore, these extreme values are data that diverge from the pattern of the whole data. A major problem is being faced when an outlier produces a pattern in a year which affects the following years and leads to the misestimation of the reserves. A solution to the outliers problem is the usage of robust statistics which are not sensitive to the existence of data errors. If we generalize the Maximum Likelihood (ML) estimators we get a class of estimators called M-estimators which minimize a function $\rho(x, \theta)$ of the errors (see Huber, 1981).

Not only the M-estimators can be used in the RCR model. We can use many robust estimators and decide which one is the most appropriate according to the contamination of the data, the efficiency of these estimators that we want to achieve and the degree of influence of the outliers on the design matrix of the model.

For more details on these estimators see Hampel et al. (1986), Rousseeuw and Leroy (1987), Yohai and Zamar (1997) and Gervini and Yohai (2002), Pitselis (2005). The algorithm for obtaining robust M-estimation of the RC model is due to Pitselis (2005).

Remark 3.2. In general, M-estimation shows that if outrageous events appear, the total expected losses and reserves demonstrate consistency. If we use estimators (such as the LMS, LTS, MM and adaptive estimators) that are robust to outliers in the design matrix, produce more reliable estimations for the reserve even if outlier claims exist. , This means that these estimators robustify in addition to dependent variables the independent variables (here the variables only take the values 0 and 1) and it seems that they are not appropriate for the robust loss reserving estimation.(Pitselis et al., 2015).

3.4.1 Robust M-Estimation of the RCR model

Next, we will present how to achieve the robust M estimation of the RC model as was shown in Section 2.1. In addition to the algorithm presented in Pitselis (2005) an additional step is required (Step(6)) to obtain robust M-estimation of the actual response RCR model.

ALGORITHM:

Step 1: Compute the residuals $\boldsymbol{r} = \boldsymbol{y} - \boldsymbol{X}\boldsymbol{b}$, with $\boldsymbol{b}$ as a least squares estimator of $\boldsymbol{\beta}$ in the model (3.2).

Step 2: Similarly as in Huber and Dutter (1974) we can obtain a robust estimator of $\dot{\boldsymbol{\sigma}} = (\sigma_1^2, \sigma_2^2, ..., \sigma_p^2)^T$ in (3.5) by minimizing

$$Q(\dot{\boldsymbol{\sigma}}, S_1) = (1/n)\sum_{j=1}^{n}\left\{\rho\left[\sum_{i=1}^{n} q_{ij}(\dot{r}_i - G_{(i)}^T\dot{\sigma})/S_1\right] + a_1\right\} S_1, \tag{3.36}$$

with respect to $\dot{\boldsymbol{\sigma}}$ and S_1, where $\boldsymbol{G}_{(i)}^T$ is the i-th row of G. The quantity S_1 is an estimate of scale and q_{ij} are elements of the matrix $\boldsymbol{Q} = \{q_{ij}\}_{i,j=1,...,n}$ of the factorization with $(\dot{\boldsymbol{M}})^{g+1} = \boldsymbol{Q}^T\boldsymbol{Q}$. The matrix $(\dot{\boldsymbol{M}})^{g+1}$ is symmetric with full rank.

Step 3: Calculate the robust M-estimate of $\boldsymbol{\Omega}$ in (3.7), i.e.

$$\widehat{\boldsymbol{\Omega}^M} = \boldsymbol{M}\widehat{\boldsymbol{\Lambda}^M}\boldsymbol{M} = \boldsymbol{M}\text{diag}(\hat{\lambda}_{11}^M, ..., \hat{\lambda}_{nn}^M)\boldsymbol{M}, \tag{3.37}$$

where $\hat{\lambda}_{ii}^M = \hat{\sigma}_1^{2^M} + \sum_k^p \hat{\sigma}_k^{2^M} X_{ki}^2$ and $\hat{\sigma}_k^{2^M}$'s corresponds to the robust estimators of σ_k^2's calculated in Step 2.

Step 4: Let $f_i(\boldsymbol{\beta}) = \beta_1 + \sum_{k=2}^p \beta_k X_{ki}$. Then the robust estimator of $\boldsymbol{\beta}$ and scale S_2 can be obtained by the following minimization problem

$$Q(\beta, S_2) = (1/n)\sum_{i=1}^{n}\left\{\rho\left[\widehat{\lambda}_{ii}^{1/2}[Y_i - f_i(\boldsymbol{\beta})]/S_2\right] + a_2\right\} S_2. \tag{3.38}$$

ρ as mentioned before and $a_2 = [(n-p)/n]E_\Phi(\chi)$ with Φ being the normal distribution.

Step 5: Steps 3 and 4 are repeated until the values converge. One iterate may often be enough which can be obtained by minimizing the following function

$$Q(\dot{\boldsymbol{\sigma}}, S_3) = (1/n)\sum_{j=1}^{n}\rho\left\{\left[\sum_{i=1}^{n} w_{ji}(\dot{r}_i - \boldsymbol{G}_{(i)}^T\dot{\boldsymbol{\sigma}})/S_3\right] + a_3\right\} S_3, \tag{3.39}$$

where $\boldsymbol{G}_{(i)}^T$ is the i-th row of G. S_3 is an estimate of scale and w_{ij} are elements of the matrix $\boldsymbol{W} = \{w_{ij}\}_{i,j=1,...,n}$ of the factorization with $(2\hat{\dot{\boldsymbol{\Omega}}})^{-1} = (2\widehat{\boldsymbol{\Omega}} * \widehat{\boldsymbol{\Omega}})^{-1} = \boldsymbol{W}^T\boldsymbol{W}$. The matrix $2\hat{\dot{\boldsymbol{\Omega}}}^{-1}$ is symmetric with full rank, with $a_3 = [(n-p)/n]E_\Phi(\chi)$ and χ and ψ as mentioned before.

Step 6: To obtain a robust estimator $\boldsymbol{b}^R$, of the actual response $\boldsymbol{b}$ in (3.8) we substitute the non robust estimators by the robust ones, i.e.,

$$\hat{\boldsymbol{b}}^R = \boldsymbol{L}\hat{\boldsymbol{\beta}}^R + \boldsymbol{\Lambda}^R\boldsymbol{Z}'(\boldsymbol{\Lambda}^R)^{-1}\hat{\boldsymbol{u}}^R \tag{3.40}$$

where the elements of $\boldsymbol{u}^R = (\hat{u}_i^R, \hat{u}_2^R, ..., \hat{u}_n^R)'$ are the winsdorized residuals based on M-estimation given by (see Huber, 1981)

$$\hat{u}_i^R = \frac{S\psi\left(\frac{u_i}{S}\right)}{\frac{1}{n}\sum_{i=1}^{n}\psi'\left(\frac{u_i}{S}\right)}, \qquad i = 1, 2, ..., n, \tag{3.41}$$

where ψ' is the derivative of ψ function and S is a scale parameter defined similarly as in Step 2.

Some of the most well known robust functions are:

1. **Huber function**:

   The family of Huber functions is defined as,

$$\rho_k(x) = \begin{cases} \frac{1}{2}x^2, & \text{if } |x| \le k \\ k(|x| - \frac{k}{2}), & \text{if } |x| > k \end{cases}$$

$$\psi_k(x) = \begin{cases} x, & \text{if } |x| \le k \\ k\ sign(x), & \text{if } |x| > k \end{cases}.$$

   When $k = 1.345$, there is 95% efficiency of the regression estimator.

2. **Hampel function**:

   The Hampel family of functions (see Hampel et al., 1986) is defined as,

$$\rho_{a,b,r}(x) = \begin{cases} \frac{1}{2}x^2/C, & \text{if } |x| \le a \\ (\frac{1}{2}a^2 + a(|x| - a))\ C, & \text{if } a < |x| \le b \\ \frac{a}{2}\Big(2b - a + (|x| - b)(1 + \frac{r-|x|}{r-b})\Big)\ C, & \text{if } b < |x| \le r \\ 1, & \text{if } r < |x| \end{cases},$$

   where $C := \rho(\infty) = \rho(r) = \frac{a}{2}(2b - a + (r - b) = \frac{a}{2}(b - a + r)$. When $k = 0.901608$ the efficiency of the regression estimator is 95%.

3. **Bisquare function**:

   Tukey's bisquare (or biweight) family of functions is defined as,

$$\rho_k(x) = \begin{cases} 1 - (1 - (\frac{x}{k})^2)^3, & \text{if } |x| \le k \\ 1, & \text{if } |x| > k \end{cases}$$

   with derivative $\rho_k'(x) = 6\psi_k(x)/k^2$, where

$$\psi_k(x) = x\bigg[1 - \Big(\frac{x}{k}\Big)^2\bigg]^2 I_{\{|x| \le k\}}.$$

   When $k = 4.685$, the regression estimator has 95% efficiency.

3.4.2 Robust M-Estimation of RCR Cross Effect Model

An outlier claim may also be appeared in a cross-section model (maybe in every run-off triangle) and lead to differences not only in the results of the least squares estimators but also to the generalized least squares estimator of the cross-section model. These outliers could dramatically change the estimations and their variances in the cross-section model. What we suggest is the robustification of the cross-section regression by specific robust techniques.

Krishnakumar (1995) constructed a method in order to produce robust estimators for the regression parameters and their covariance adopting an econometric GLM. The reader can see more details at the papers of Caroni (1987), Koenker and Portnoy (1990), and Bilodeau and Duchesne (2000).

Using the RCR cross effect model which was fully studied in Section 2.2 and the outcomes of the robust estimation techniques, we can obtain a robust M-estimation of random coefficient regression with cross effect. For further information on the algorithm steps and their properties like the efficiency of the estimator, the reader may be referred to Pitselis (2005). For more on robust estimation techniques see Huber and Dutter (1974), Hampel et al. (1986), Rousseeuw and Leroy (1987) and Maronna et al. (2006).

Remark 3.3. *The excess of loss reserves (bias) which occured according to the robustification of our random coefficient regression model can be distributed equally to all accident periods ultimate reserves, or according to practicient proofs. Estimators for the parameters which arised from robust methods are asymptotically biased if the distributions of the errors are asymmetric. It is crusial to treat the bias appears in the robustified model. Wang et al. (2005), created a distribution-free method in order to correct this bias.*

3.4.3 Robust M-Estimation of the RCR Cross Effect Model

M-Estimation of RCR cross sectional model can be evaluated by following the steps of the algorithm below. In addition to the algorithm presented in Pitselis (2005) an additional step is required (Step 6) to obtain a robust predictor of the individual robust regression component.

ALGORITHM:

Step 1: Minimize [Huber and Dutter (1974)]

$$(1/T)\sum_{t=1}^{T}\rho\left(\frac{y_{it}-\sum_{k=1}^{p}\beta_{ki}x_{kit})}{S_i}\right)S_i+a_iS_i \tag{3.1}$$

to find $\hat{\boldsymbol{\beta}}_i^R$ a robust estimator of $\boldsymbol{\beta}_i=(\beta_{1i},...,\beta_{pi})'$ and S_i, for $i=1,...,n$. In practice we have to calculate $\hat{\boldsymbol{\beta}}_i^R$ and scale parameter S_i by simultaneous iterations.

Step 2: Calculate the robust estimate of σ_i^2 in (3.8), i.e.

$$(\hat{\sigma}_i^R)^2=\frac{1}{(T-p)}\frac{\sum_{t=1}^{T}\psi\left(\frac{r_{it}}{S_i}\right)^2 S_i^2}{\left[\frac{1}{T}\sum_{t=1}^{T}\psi'\left(\frac{r_{it}}{S_i}\right)\right]^2}, \tag{3.2}$$

where $r_{it}=y_{it}-\sum_{k=1}^{p}\beta_{ki}x_{kit}$ and $\psi'(u)=\frac{\partial}{\partial u}\psi(u)$.

Step 3: Calculate the robust estimate of $\boldsymbol{\Sigma}$ by substituting the parameters of (3.3) by their robust ones, i.e.

$$\boldsymbol{\Sigma}^R=\frac{1}{n-1}\left[\sum_{i=1}^{N}\hat{\boldsymbol{\beta}}_i^R(\hat{\boldsymbol{\beta}}_i^R)'-\frac{1}{N}\sum_{i=1}^{N}\hat{\boldsymbol{\beta}}_i^R\sum_{i=1}^{N}(\hat{\boldsymbol{\beta}}_i^R)'\right]-\frac{1}{N}\sum_{i=1}^{N}(\hat{\sigma}_i^R)^2(\boldsymbol{X}_i'\boldsymbol{X}_i)^{-1}, \tag{3.3}$$

where $(\hat{\sigma}_i^R)^2(\boldsymbol{X}_i^{'}\boldsymbol{X}_i)^{-1}$ is the robust covariance matrix of $\boldsymbol{\beta}_i$.

In the case of $\boldsymbol{\Sigma}^R$ not being positive definite we can proceed the same way as of proceeding when $\boldsymbol{\Sigma}$ in (3.3) is not positive definite. Having obtained $\boldsymbol{\Sigma}^R$ in (3.3), by using (3.8) a robust estimate $\boldsymbol{\Omega}^R$, of $\boldsymbol{\Omega}$ can obtained.

Step 4: Substitute the values of (3.2) and (3.3) into (3.27) to obtain a robust estimate

$$\boldsymbol{\Lambda}_i^R = \left[\sum_{j=1}^{N}(\boldsymbol{\Sigma}^R + (\hat{\sigma}_i^R)^2(\boldsymbol{X}_j^{'}\boldsymbol{X}_j)^{-1})^{-1}\right]^{-1}(\boldsymbol{\Sigma}^R + (\hat{\sigma}_i^R)^2(\boldsymbol{X}_i^{'}\boldsymbol{X}_i)^{-1})^{-1}. \tag{3.4}$$

Step 5: Calculate the robust estimate of $\boldsymbol{\beta} = (\beta_1, ..., \beta_p)^{'}$

$$\hat{\boldsymbol{\beta}}_{GLS}^R = \sum_{i=1}^{N}\boldsymbol{\Lambda}_i^R\hat{\boldsymbol{\beta}}_i^R, \tag{3.5}$$

which is a matrix weighted combination of $\hat{\boldsymbol{\beta}}_i^R$ obtained in (3.1). Then the robust variance-covariance matrix can be obtained in an analogous way of obtaining (3.29), substituting the $\hat{\sigma}_i$ and $\boldsymbol{\Sigma}$ with the robust $\hat{\sigma}_i^R$ and $\boldsymbol{\Sigma}^R$, respectively

$$\begin{aligned}\mathrm{Cov}(\widehat{\boldsymbol{\beta}_{GLS}^R}) &= \left[\sum_{i=1}^{N}(\boldsymbol{\Sigma}^R + (\hat{\sigma}_i^R)^2(\boldsymbol{X}_i^{'}\boldsymbol{X}_i)^{-1})^{-1}\right]^{-1} \\ &= (\boldsymbol{X}^{'}(\boldsymbol{\Omega}^R)^{-1}\boldsymbol{X})^{-1}.\end{aligned} \tag{3.6}$$

Step 6: In order to obtain a robust predictor of the individual component $\boldsymbol{\beta}_i$, we substitute the parameters in (3.32) by their robust ones, i.e.,

$$\widehat{\boldsymbol{\beta}}_i^{R\star} = \widehat{\boldsymbol{\beta}}_{GLS}^R + \boldsymbol{\Sigma}^R\boldsymbol{X}_i^{'}(\boldsymbol{X}_i\boldsymbol{\Sigma}^R\boldsymbol{X}_i^{'} + (\hat{\sigma}_i^R)^2\boldsymbol{I}_T)^{-1}(\boldsymbol{Y}_i^{\star} - \boldsymbol{X}_i\widehat{\boldsymbol{\beta}}_{GLS}^R), \tag{3.7}$$

where $\boldsymbol{Y}_i^{\star} = (Y_{1i}^{\star}, ..., Y_{Ti}^{\star})^{'}$ are pseudo-observations derived from Winsorizing the observations $\boldsymbol{Y}_i = (Y_{1i}, ..., Y_{Ti})^{'}$, with

$$Y_{ti}^{\star} = \begin{cases} Y_{ti}, & \text{if } |r_{ti}| \leq cS_{ti}, \\ \hat{Y}_{ti} - cS_{ti} & \text{if } r_{ti} < -cS_{ti}, \\ \hat{Y}_{ti} + cS_{ti} & \text{if } r_{ti} > cS_{ti}. \end{cases} \tag{3.8}$$

The constant c regulates the amount of robustness; good choices are in the range between 1 and 2, say $c = 1.5$. Then use the pseudo-observations $Y_{ti}^{\star}$ to calculate new fitted values $\hat{Y}_{ti}$ (and new S_{ti}), and iterate to convergence. For details, see Huber (1981, p. 18).

3.5 Numerical Example based on RCR Model for One LOB

In what it follows we apply the non random and random linear regression models to the Taylor and Ashe (1983) data set. We provide estimations of the ultimate reserves for each accident year, as well as of the total reserves. The results of these estimations are presented in Table 3.3, where we observe that the random regression coefficients model provides almost identical values of total reserves (21522631) from the non random model (21522633). This provides an indication that there is no true randomness in the regression coefficients (see Table 3.3).

Similar results for random and non random models are obtained by applying robust estimations. In each of Tables 3.3, 3.4 and 3.5, three different robust estimators are given depending on three different ρ functions (Humber, Hambel, Bisqure). In order to study the influence of outliers we create two fake (artificial) outliers, just by multiplying two incremental claims by **10**.

Tables 3.4 and 3.5 show the total reserves in the presence of one and two outliers. In the presence of one outlier ($Y_{1,3} = 610542\mathbf{0}$), the LS regression model provides an underestimation of the total reserves, while in the presence of two outliers ($Y_{1,3} = 610542\mathbf{0}$ and $Y_{8,3} = 1443370\mathbf{0}$), the LS regression model overestimates the total reserves.

Table 3.3: Reserve Estimation without Outliers

	Non Random Estimation				**Random Estimation**			
	Non Robust	**Robust**			**Non Robust**	**Robust**		
Year/Model	**LS**	**Huber**	**Hampel**	**Bisquare**	**LS**	**Huber**	**Hampel**	**Bisquare**
1	0	0	0	0	0	0	0	0
2	235234	230237	236131	222024	235240	230237	236123	222025
3	640016	671678	641679	711057	640007	671681	641706	711057
4	906623	676215	709952	490433	906607	676201	709823	490444
5	1064309	1091327	1094211	1137952	1064325	1091331	1094184	1137957
6	1597933	1650607	1643383	1767579	1597913	1650622	1643420	1767575
7	2662660	2777956	2744028	2960095	2662652	2777967	2744079	2960093
8	4788626	4401235	4711443	4087469	4788641	4401253	4710944	4087367
9	4614418	4544003	4541028	4501191	4614432	4544007	4540932	4501167
10	5012813	4891731	4939423	4814302	5012813	4891731	4939327	4814289
Total Reserves	**21522633**	**20934989**	**21261278**	**20692103**	**21522631**	**20935029**	**21260538**	**20691975**

In contrast to the previous indication of the non-randomness of regression coefficients, we observe the following (see Table 3.5). In the presence of two outliers the total values of reserves under the robust non-random regression model are different than the values of reserves under the robust random regression model for the three ρ functions (Humber, Hambel, Bisqure). For example, using the Huber function the total reserve is 21408539 under the non random model, while the total reserve is 20603298 under the random model. This gives us a new indication of any randomness of the coefficients. Of course, in

Table 3.4: Reserve Estimation with 1 Outlier

	Non Random Estimation				Random Estimation			
	Non Robust	Robust			Non Robust	Robust		
Year/Model	**LS**	**Huber**	**Hampel**	**Bisquare**	**LS**	**Huber**	**Hampel**	**Bisquare**
1	0	0	0	0	0	0	0	0
2	235234	227553	236516	210478	235215	227555	236515	210485
3	1293657	787881	676612	811298	1293659	787876	676608	811263
4	750994	622259	690500	400900	751006	622277	690504	400844
5	753052	1051983	1080595	1080334	753042	1051976	1080590	1080240
6	1112371	1617363	1631108	1810722	1112366	1617348	1631099	1810654
7	1959218	2721823	2724706	3004209	1959201	2721800	2724695	3004110
8	3773926	4312027	4623247	3662468	3773914	4311826	4623263	3662554
9	3055018	4328576	4467628	4393663	3055036	4328559	4467631	4393704
10	2000878	4429053	4804316	4552209	2000869	4429079	4804321	4552228
Total Reserves	**14934349**	**20098519**	**20935228**	**19926282**	**14934309**	**20098296**	**20935226**	**19926083**

Table 3.5: Reserve Estimation with 2 Outliers

	Non Random Estimation				Random Estimation			
	Non Robust	Robust			Non Robust	Robust		
Year/Model	**LS**	**Huber**	**Hampel**	**Bisquare**	**LS**	**Huber**	**Hampel**	**Bisquare**
1	0	0	0	0	0	0	0	0
2	235234	231120	234901	222469	235251	234480	238501	212258
3	1924568	930519	670459	782825	1924562	775427	673919	813196
4	1802513	804192	753764	514589	1802513	892598	880261	579029
5	1804571	1231266	1080262	1139540	1804572	1040086	1063199	1090611
6	2163890	1809307	1629757	1775562	2163884	1523536	1570021	1765795
7	3010737	2906680	2715234	2942002	3010742	2812194	2772655	3044115
8	4825445	4477382	4702353	4290423	4825457	4348292	4667584	3648566
9	4106537	4472048	4506306	4464424	4106546	4406185	4546342	4380001
10	3052397	4546025	4846462	4675612	3052390	4570498	4889212	4612842
Total Reserves	**22925891**	**21408539**	**21139498**	**20807448**	**22925916**	**20603298**	**21301694**	**20146412**

the statistical literature there are tests of verifying the randomness of the coefficients, but this is out of the scope of our research since this does not affect the reserve estimation.

These results that illustrate the behaviour of ultimate reserves by accident year reflecting the results of Tables 3.3, 3.4 and 3.5 are shown in Figures 3.1, 3.2 and 3.3.

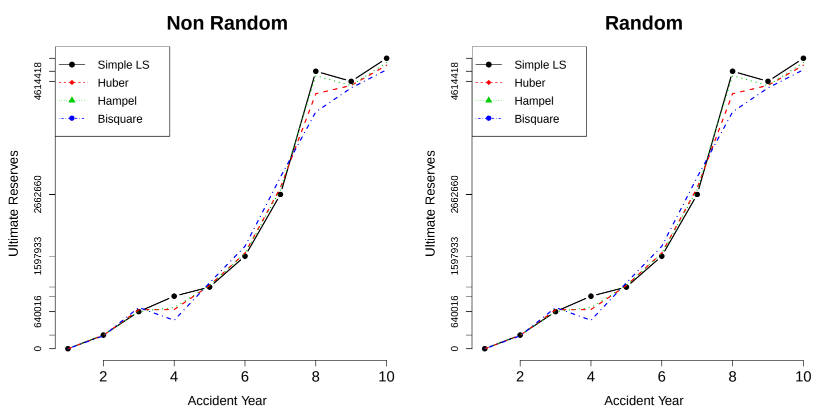

Figure 3.1: Reserves estimation using Random and Non Random regression models (no outliers)

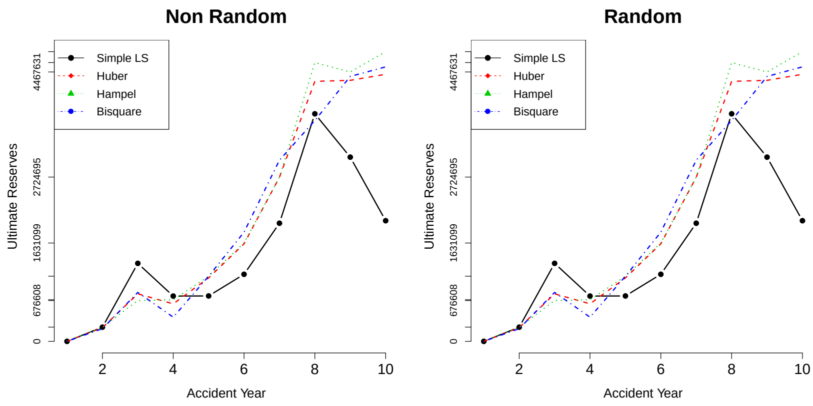

Figure 3.2: Reserves estimation using Random and Non Random regression (1 outlier)

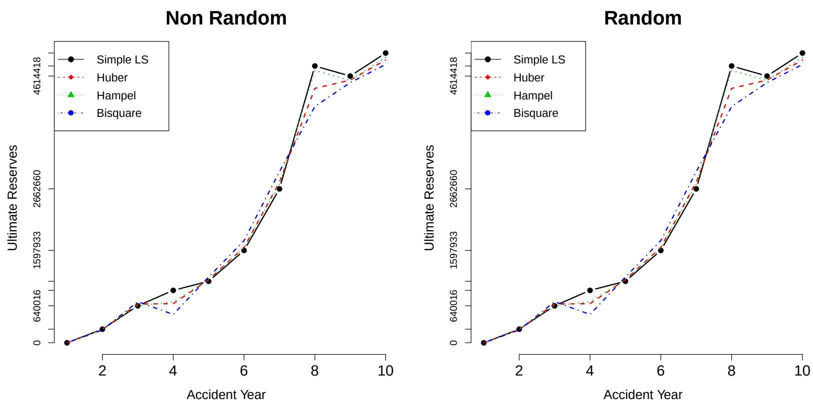

Figure 3.3: Reserves estimation using Random and Non Random regression (two outliers)

Figures 3.4-3.11 display the residuals versus fitted values and the corresponding histograms, respectively, based on the random coefficient regression model for the three different ρ functions (Humber, Hambel, Bisqure).

Residual scatter plots show that residuals are scattered around zero almost arbitrary, with some exemptions in cases where outliers are appeared (see Figures 3.4, 3.6, 3.8 and 3.10). The random pattern of the residuals suggests that there is no heteroscedasticity, which could make the estimators inefficient.

The histogram of the residuals in Figures 3.5, 3.7, 3.9 and 3.11 shows a symmetric bell - shaped histogram of the residuals, which is distributed around zero and indicates that the normality assumption is likely to be true.

In general, both LS model and RCR model are very sensitive to outlier events (very large values) and most of the times result to an overestimation of the total reserves.

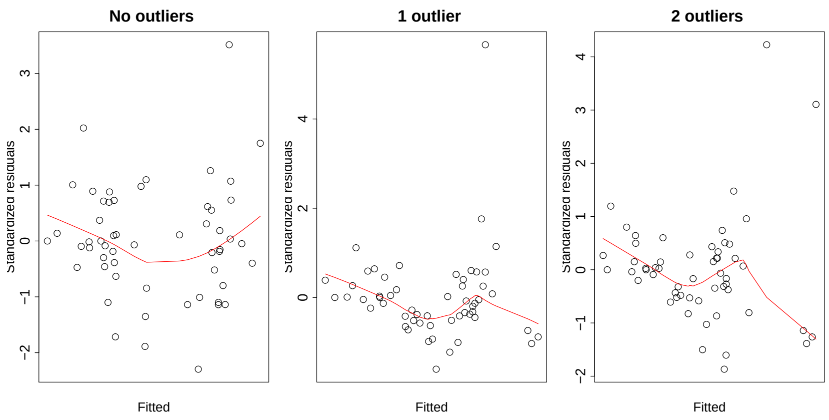

Figure 3.4: Residuals versus fitted - Random coefficients

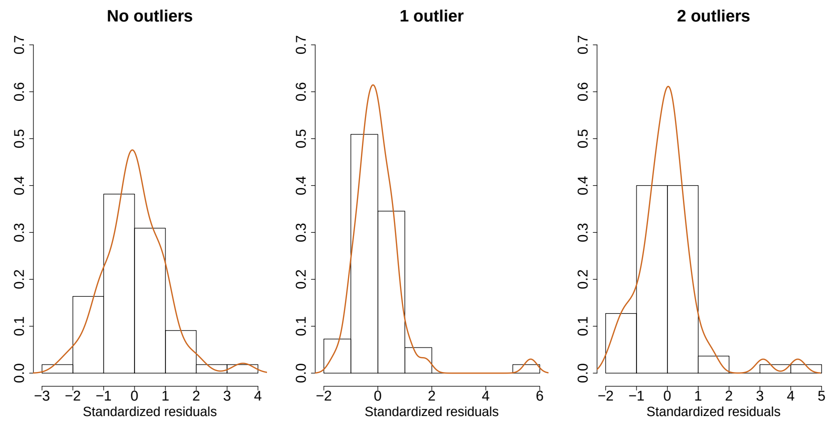

Figure 3.5: Histogram of residuals - Random coefficients

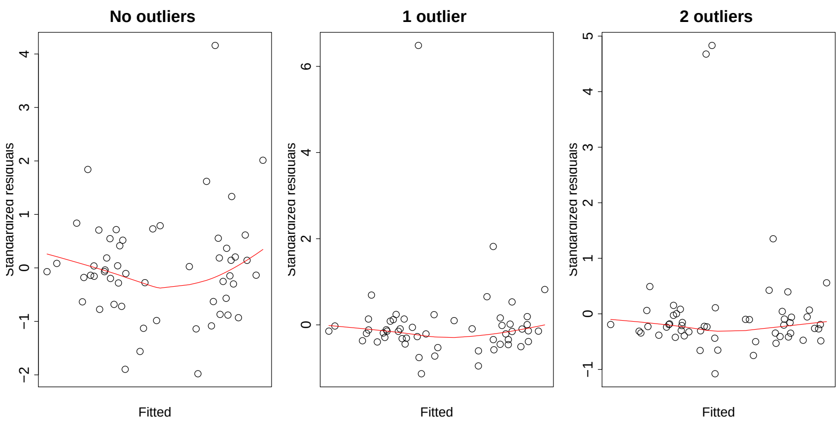

Figure 3.6: Residuals versus fitted - Random coefficients (Huber)

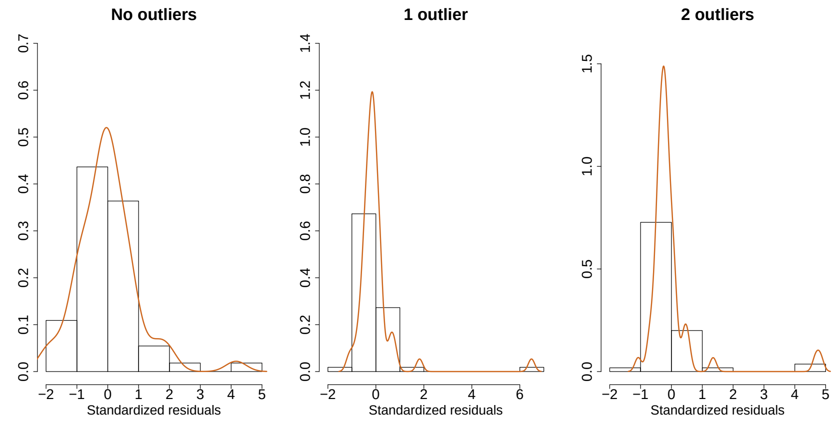

Figure 3.7: Histogram of residuals - Random coefficients (Huber)

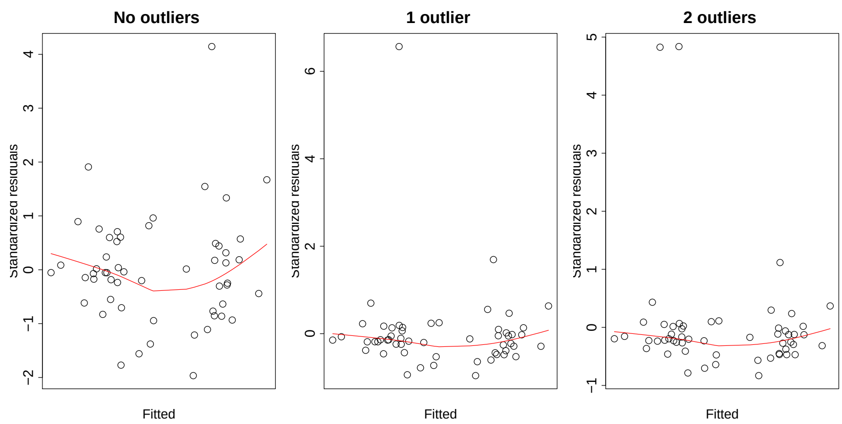

Figure 3.8: Residuals versus fitted - Random coefficients (Hampel)

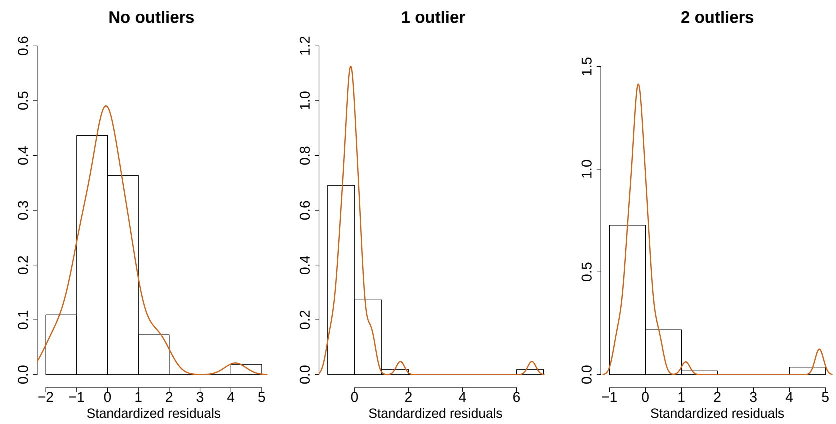

Figure 3.9: Histogram of residuals - Random coefficients (Hampel)

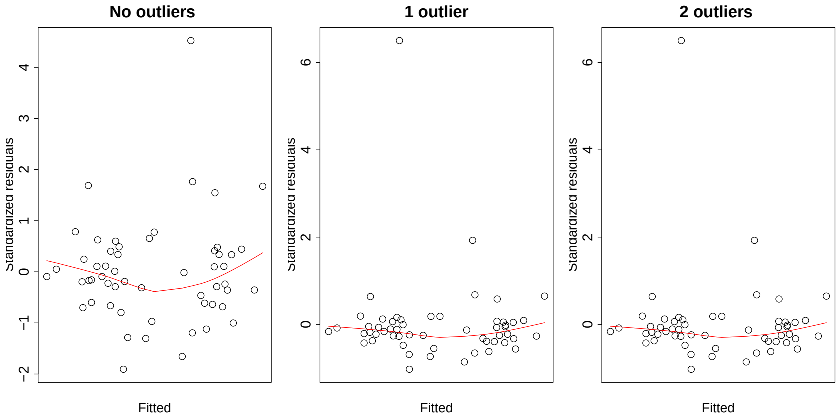

Figure 3.10: Residuals versus fitted -Random (Bisquare)

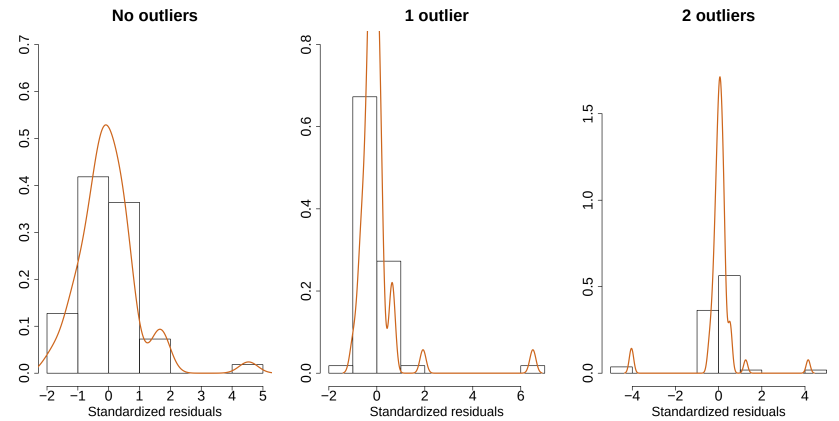

Figure 3.11: Histogram of residuals Random (Bisquare)

However, there are cases where we obtain an underestimation of the total reserves depending on the position and the size of the outliers (for details see Pitselis et al., 2015). Implementing the data set we see the superiority of robust M-estimator in comparison with the least squares estimator (with fixed and random coefficients) for claims reserves estimation.

3.6 Numerical Example based on RCR Cross Effects Model for Two LOB

For the Random Coefficient Regression Cross Effects model we consider a data set of two run-off triangles obtained from a private insurance company that operates in Greece.

Tables 3.6 and 3.7 show the incremental incurred losses (paid and outstanding claims) for two lines of business (Company A and company B) which both operate in Greece. Company A mainly focuses on Motor Business and underwrites all vehicle categories apart from taxis and trucks while company B underwrites all vehicle categories for Motor Business. Moreover, the premiums of the companies by accident year are presented next to the run-off triangles. For the implementation of the data set we divided the losses by the premium in order to put the accident years on a more nearly equal basis.

Table 3.6: Motor Triangle and Premiums for Company A

Accident	**Development Year**										
Year	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**Premium**
2007	58134	162688	101105	100964	61591	71009	34024	2746	646	10190	1051637
2008	51437	197139	120641	74807	76771	77276	39070	4396	13809		1190965
2009	57906	116191	143953	103883	70760	177194	35341	6088			1327568
2010	40352	121837	88389	320429	75127	70190	63723				1418348
2011	82227	279591	151260	230293	82378	47315					1504056
2012	196417	119755	228499	99894	44266						1580233
2013	67161	107098	198252	75172							1619382
2014	78293	141865	106150								1727540
2015	74472	118886									1820104
2016	43281										1883017

Table 3.7: Motor Triangle and Premiums for Company B

Accident	**Development Year**										
Year	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**Premium**
2007	63078	143002	144235	75007	60775	70804	27508	4757	3172	6385	1633833
2008	65567	177292	107870	137305	72741	68708	102864	4335	6107		1675707
2009	87394	146346	158876	199846	53161	72764	42915	10898			1636855
2010	70017	153893	119028	93771	49600	185689	28331				1689715
2011	104638	186326	335477	136857	87941	69248					1649386
2012	76390	190629	192606	121704	66297						1712587
2013	58620	184557	135174	118180							2105361
2014	87845	166511	145385								2265432
2015	53616	152751									1976188
2016	62904										1351719

It is obvious that for Company A for accident year 1, a big claim has been paid 10 years after the accident date (the amount of this claim is embedded at the total incremental amount of $X_{1,10}$ =10190 and could be represented as an outlier claim. This claim will dramatically change the development pattern of the payments in the case of using the Chain Ladder method because the Loss Development Factor for this year increases from 1.01 to 1.02 (see Table 5.4).

Table 3.8: Loss Development Factors

Company	0-1	1-2	2-3	3-4	4-5	5-6	6-7	7-8	8-9
A	2.93	1.61	1.37	1.13	1.15	1.07	1.01	1.01	1.02
B	3.25	1.68	1.30	1.12	1.15	1.08	1.01	1.01	1.01

This is commonly observed in Motor Business because there are many cases where claims are settled many years later especially when accidents with large compensations (such as partial or total disability, deaths, etc.) are observed. This can also be observed in accident year 2 where a large amount is observed at the last known development year ($Y_{29} = 13809$). For company B the run-off triangle data seems to be more stable without big fluctuations.

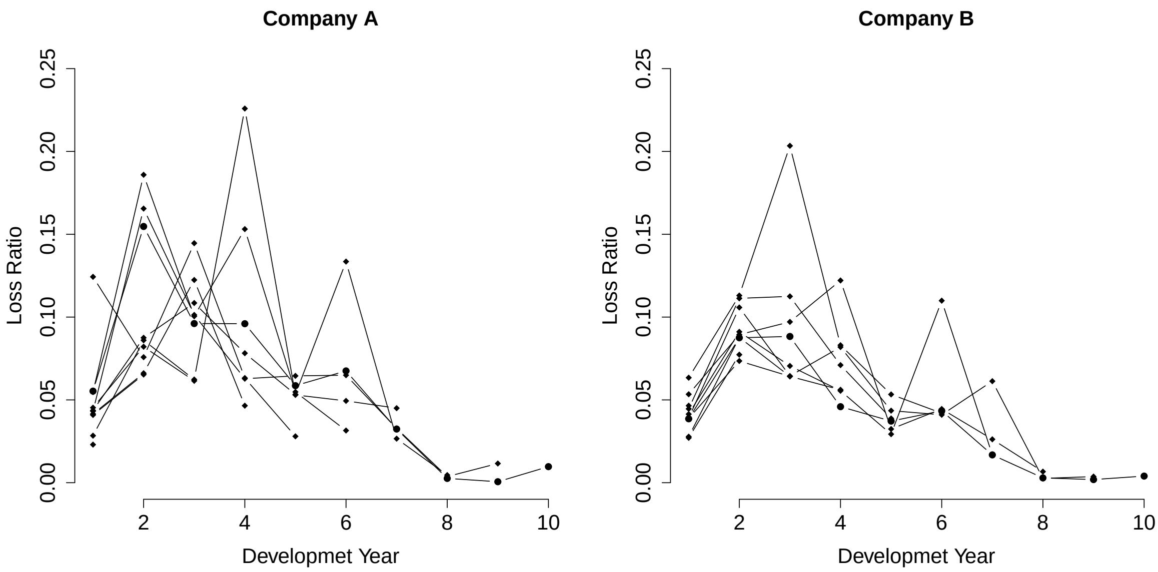

Figure 3.12: Plots of incremental loss ratios for companies A and B

Figure 3.12 exhibits scatter plots for the incremental loss ratios for companies A and B. A decreasing trend for both companies confirms that claims will close within ten years. Figure 3.13 exhibit scatterplots for the accumulated loss ratios for companies A and B. A comparison of the two panels shows that company A seems to be more stable than company B.

Figure 3.14 illustrates loss ratios of company A versus company B line of business. The plots suggest a positive (somewhat linear) relationship between company A and company B line of business (the Pearson correlation coefficient is equal with 0.944 while the Spearman correlation coefficient is equal with 0.916). This positive correlation reflects

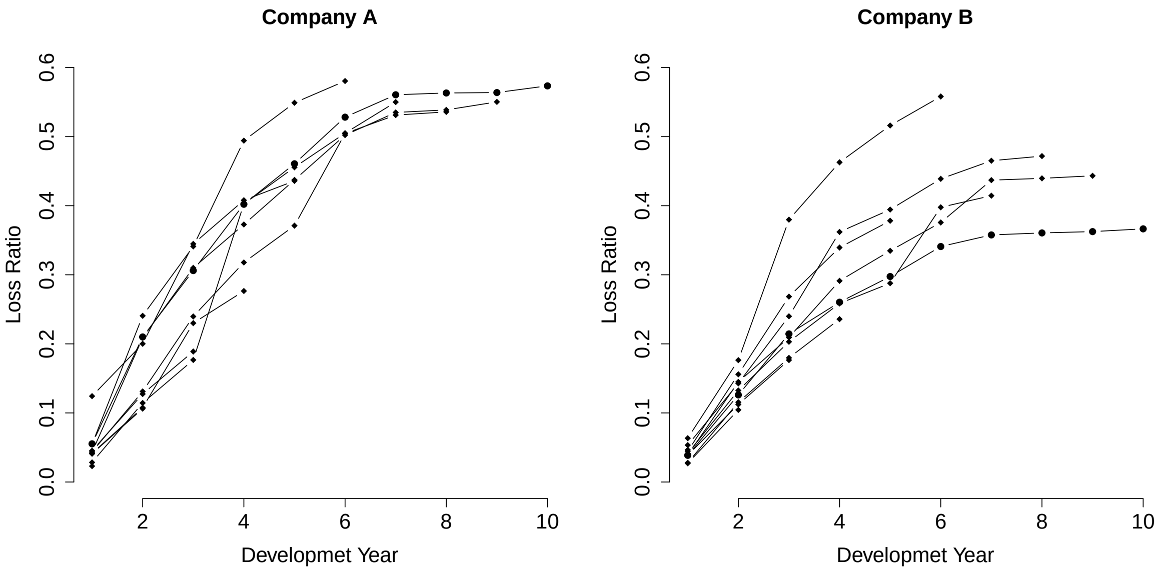

Figure 3.13: Plots of accumulated loss ratios for companies A and B

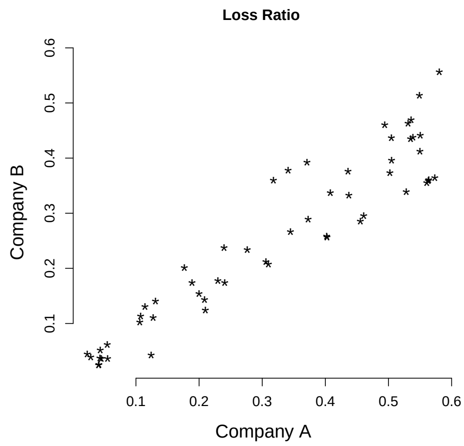

Figure 3.14: Loss Ratios of company A versus company B

the effect of some economic components (new policy of the insurance company for faster settlement of claims, inflation rates, etc.), to all outstanding claims. Naturally, loss ratios at the first development years are much higher than those at the future years. Moreover, new government rules could lead to higher payments and consequently higher loss ratios across all lines of business. Figure 3.15 presents the QQ-Plots for the Loss Ratios of the two companies. It is obvious that the normality assumption is not the best choice for the distribution of these quantities.

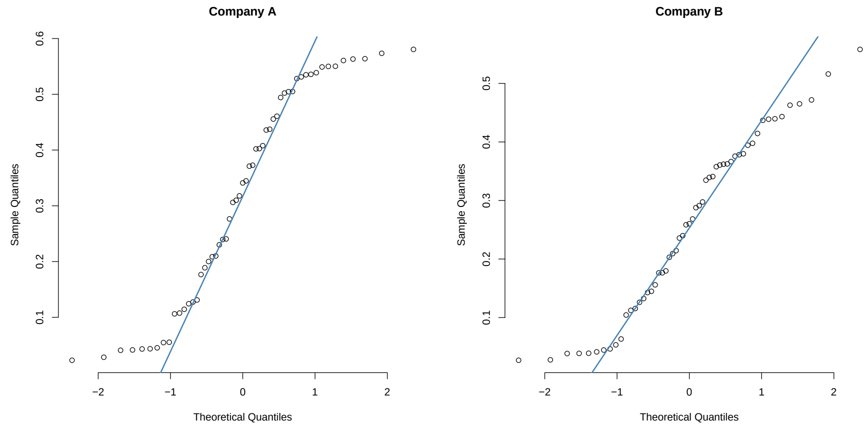

Figure 3.15: Loss Ratio's QQ-Plot for companies A and B

Applying individual LS estimation for each line of business and cross section effects model we get the coefficients that are shown in Table 3.9.

Table 3.9: Coefficients based on LS and Cross Effects

	Least Squares (LS)		**Cross Effect**
Coefficients	**Company A**	**Company B**	**A & B**
(Intercept)	63186	54954	62098
b1	6938	16717	47616
b2	15271	22018	54784
b3	26844	14004	56320
b4	51950	59234	93184
b5	39173	30718	71680
b6	-41	10950	40960
b7	692	17052	40960
b8	-3099	1888	38400
b9	-19905	7950	86016
b10	73183	92682	47616
b11	61492	91041	42496
b12	60430	49193	19456
b13	-18066	-13650	-48640
b14	5211	16094	-22528
b15	-32409	-17734	-59904
b16	-66179	-61202	-97280
b17	-59427	-58673	-92160
b18	-52996	-48570	-139264

For comparison reasons Table 3.10 illustrates the ultimate claims and reserves for each line of business based on the individual LS estimation and on the cross section effects model.

Table 3.10: Ult. Claims and Reserves based on LS and Cross Effects Model

	Company A				**Company B**			
Accident	**Individual LS Model**		**Cross Effects Model**		**Individual LS Model**		**Cross Effects Model**	
Year	**Reserves**	**Ult. Claims**	**Reserves**	**Ult. Claims**	**Reserves**	**Ult. Claims**	**Reserves**	**Ult. Claims**
2007	0	603097	0	603097	0	598722	0	598722
2008	156097	811444	116464	771811	175658	918448	155586	898376
2009	279977	991293	226352	937667	317716	1089916	301828	1074029
2010	379285	1159332	348534	1128581	374685	1075015	339333	1039663
2011	209145	1082210	158125	1031190	231611	1152098	203057	1123543
2012	417332	1106163	358801	1047632	449831	1097457	423799	1071425
2013	318894	766577	253428	701110	367737	864268	351112	847643
2014	127952	454260	74662	400970	139650	539391	98057	497798
2015	191378	384735	140951	334309	187933	394300	134061	340428
2016	22478	65758	209534	252814	36706	99610	237991	300895
Total	**2102538**	**7424869**	**1886849**	**7209180**	**2281526**	**7829225**	**2244823**	**7792523**
Loss Ratio	**49,10%**		**47,67%**		**44,24%**		**44,03%**	

We observe that the sum 2102538+ 2281526 =4384064 of individual triangle reserves based on the LS estimation is greater than the sum 1886849+2244823=4131672 of the claims reserves based on cross section effects model. This gives substance to the additivity property, emphasized by Anje (1994), where a simple aggregation of the reserves of individual triangle ignores the correlation among triangle.

The corresponding loss ratios are also provided at the end of the table, showing a decrease (from 49.10% to 44.24% for company A and from 47.67% to 44.03% for company B) when we apply the cross section effects model than the individual LS estimation. In this case, insurance companies are collecting more premiums than the amount paid in claims. Loss ratio is considered as one of the tools with which explains a company's suitability for coverage.

In order to test the validity of the RCR cross effects model, we display the residuals in Figure 3.16 up to Figure 3.19. There are some outliers but the majority of the residuals are located randomly and close to 0.

In order to study the influence of outliers we create two fake (artificial) outliers just by multiplying two incremental claims by **10**. Tables 3.11 and 3.12 show the total reserves

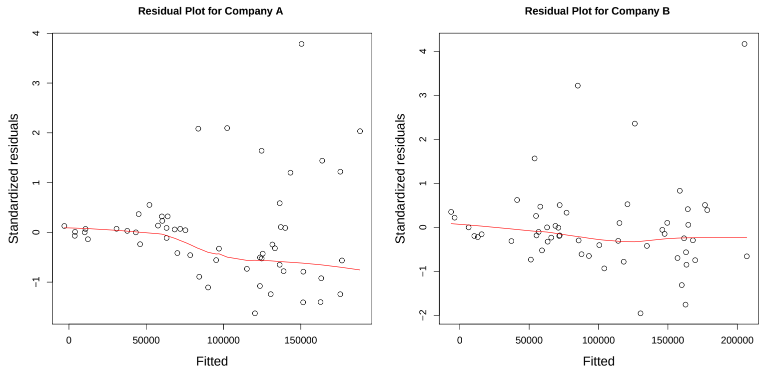

Figure 3.16: Residual Plots (LS RCR Cross Effect Model)

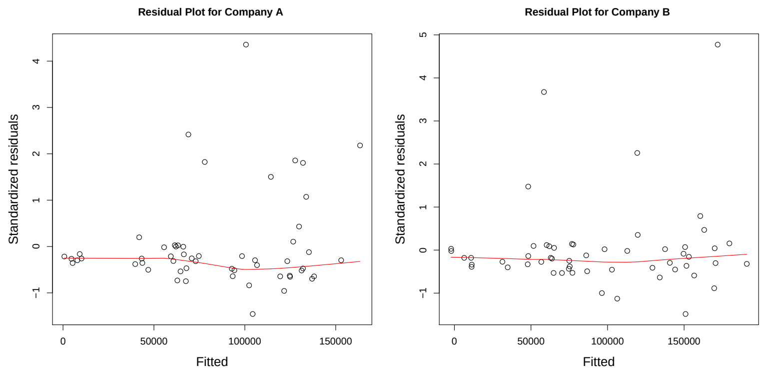

Figure 3.17: Residual Plots (Huber RCR Cross Effect Model)

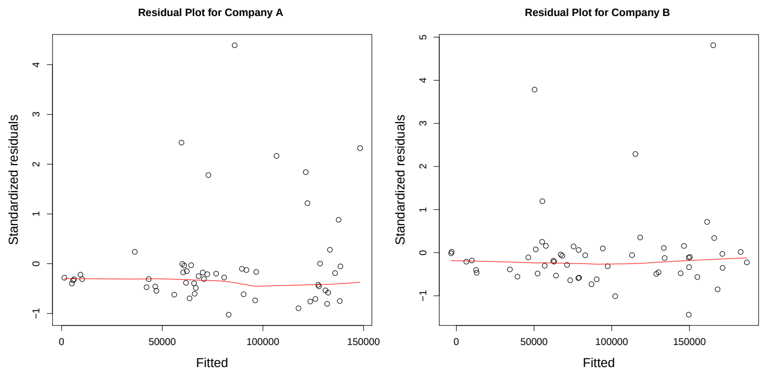

Figure 3.18: Residual Plots (Hampel RCR Cross Effect Model)

without the presence of outliers while Tables 3.13 up to 3.16 show the total reserves in the presence of one and two outliers, respectively.

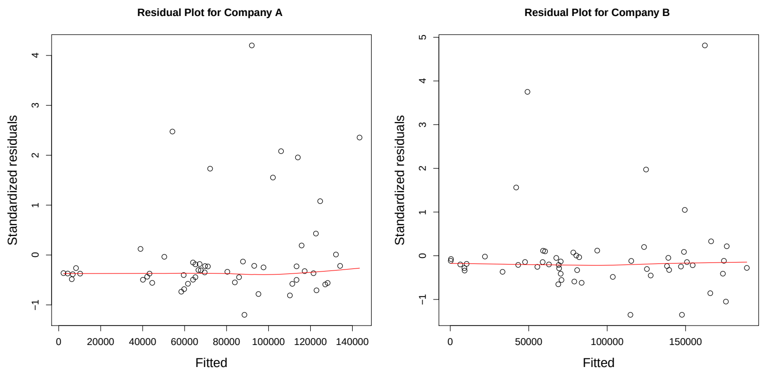

Figure 3.19: Residual Plots (Bisquare RCR Cross Effect Model)

Table 3.11: RCR Cross Effect Model with no Outlier (Company A)

Accident	**Least Squares**		**Robust (Huber)**		**Robust (Hampel)**		**Robust (Bisquare)**	
Year	**Reserves**	**Ultimate**	**Reserves**	**Ultimate**	**Reserves**	**Ultimate**	**Reserves**	**Ultimate**
2007	0	603097	0	603097	0	603097	0	603097
2008	116464	771811	112404	767750	114525	769872	106311	761658
2009	226352	937667	222886	934202	226795	938110	211998	923313
2010	348534	1128581	284790	1064838	259273	1039320	280397	1060444
2011	158125	1031190	221525	1094589	247589	1120653	252180	1125245
2012	358801	1047632	346529	1035360	316119	1004950	324862	1013693
2013	253428	701110	264450	712133	245105	692788	247840	695523
2014	74662	400970	33274	359582	-3728	322580	21694	348002
2015	140951	334309	77566	270923	30895	224253	60055	253412
2016	209534	252814	131043	174323	85199	128480	104048	147328
Total Reserve	**1886849**	**7209180**	**1694466**	**7016796**	**1521772**	**6844103**	**1609385**	**6931716**
Loss Ratio	**47,67%**		**46,40%**		**45,26%**		**45,84%**	

Table 3.12: RCR Cross Effect Model with no Outlier (Company B)

Accident	**Least Squares**		**Robust (Huber)**		**Robust (Hampel)**		**Robust (Bisquare)**	
Year	**Reserves**	**Ultimate**	**Reserves**	**Ultimate**	**Reserves**	**Ultimate**	**Reserves**	**Ultimate**
2007	0	598722	0	598722	0	598722	0	598722
2008	155586	898376	155797	898587	155586	898376	158264	901053
2009	301828	1074029	267395	1039595	261890	1034091	255636	1027836
2010	339333	1039663	325215	1025546	317766	1018097	330194	1030525
2011	203057	1123543	219536	1140023	226391	1146877	221757	1142243
2012	423799	1071425	345467	993094	325655	973281	298874	946500
2013	351112	847643	278977	775508	317264	813795	219948	716480
2014	98057	497798	51576	451317	55639	455380	12004	411745
2015	134061	340428	63331	269698	62647	269014	21117	227484
2016	237991	300895	150632	213536	163335	226239	84888	147792
Total Reserve	**2244823**	**7792523**	**1857925**	**7405624**	**1886172**	**7433872**	**1602682**	**7150382**
Loss Ratio	**44,03%**		**41,85%**		**42,01%**		**40,40%**	

Table 3.13: RCR Cross Effect Model with one outlier (Company A)

Accident	**Least Squares**		**Robust (Huber)**		**Robust (Hampel)**		**Robust (Bisquare)**	
Year	**Reserves**	**Ultimate**	**Reserves**	**Ultimate**	**Reserves**	**Ultimate**	**Reserves**	**Ultimate**
2007	0	603097	0	603097	0	603097	0	603097
2008	49303	704650	108952	764298	115996	771343	111032	766379
2009	117215	828530	220694	932009	231458	942773	214272	925587
2010	203418	983465	273414	1053461	258501	1038548	293151	1073199
2011	228884	1101949	273391	1146456	233763	1106828	351848	1224912
2012	429560	1118390	401985	1090815	330059	1018889	422276	1111107
2013	324186	1376318	318847	1370979	256018	1308150	346801	1398933
2014	145421	471729	86987	413295	4718	331026	120269	446577
2015	211710	405067	132673	326031	43860	237217	156646	350003
2016	280293	323573	184057	227337	101856	145136	200639	243920
Total Reserve	**1989989**	**7916769**	**2000999**	**7927779**	**1576229**	**7503008**	**2216933**	**8143713**
Loss Ratio	**52,35%**		**52,42%**		**49,61%**		**53,85%**	

Table 3.14: RCR Cross Effect Model with one outlier (Company B)

Accident	**Least Squares**		**Robust (Huber)**		**Robust (Hampel)**		**Robust (Bisquare)**	
Year	**Reserves**	**Ultimate**	**Reserves**	**Ultimate**	**Reserves**	**Ultimate**	**Reserves**	**Ultimate**
2007	0	598722	0	598722	0	598722	0	598722
2008	155586	898376	155465	898255	155586	898376	157214	900004
2009	301828	1074029	272738	1044939	263230	1035431	263455	1035656
2010	339333	1039663	317868	1018199	314831	1015161	310429	1010760
2011	203057	1746771	220661	1764375	226116	1769830	226814	1770529
2012	1047027	1694653	435995	1083621	362385	1010011	994622	1642248
2013	558855	1055386	318518	815049	326288	822819	444918	941450
2014	305799	705540	93439	493180	70361	470102	238408	638149
2015	341804	548171	102842	309209	75074	281442	245027	451394
2016	445733	508637	196566	259470	175098	238002	335340	398245
Total Reserve	**3699022**	**9869949**	**2114092**	**8285019**	**1968969**	**8139896**	**3216228**	**9387155**
Loss Ratio	**55,77%**		**46,82%**		**46,00%**		**53,04%**	

Table 3.15: RCR Cross Effect Model with two outliers (Company A)

Accident	**Least Squares**		**Robust (Huber)**		**Robust (Hampel)**		**Robust (Bisquare)**	
Year	**Reserves**	**Ultimate**	**Reserves**	**Ultimate**	**Reserves**	**Ultimate**	**Reserves**	**Ultimate**
2007	0	603097	0	603097	0	603097	0	603097
2008	49303	704650	109041	764387	119080	774427	112278	767625
2009	117215	828530	212434	923749	231410	942726	213653	924969
2010	203418	983465	267642	1047690	266534	1046581	293239	1073286
2011	228884	1527785	196492	1495393	179705	1478606	243269	1542170
2012	855396	1544227	566203	1255033	504270	1193100	344234	1033065
2013	466132	1518264	359703	1411835	335584	1387716	250203	1302335
2014	287366	613674	121382	447690	82781	409089	22250	348558
2015	353656	547013	186284	379642	149071	342428	61109	254466
2016	422238	465519	248392	291673	217653	260934	104315	147595
Total Reserve	**2983608**	**9336223**	**2267573**	**8620189**	**2086088**	**8438704**	**1644549**	**7997165**
Loss Ratio	**61,74%**		**57,00%**		**55,80%**		**52,88%**	

Table 3.16: RCR Cross Effect Model with two outliers (Company B)

Accident	Least Squares		Robust (Huber)		Robust (Hampel)		Robust (Bisquare)	
Year	Reserves	Ultimate	Reserves	Ultimate	Reserves	Ultimate	Reserves	Ultimate
2007	0	598722	0	598722	0	598722	0	598722
2008	96966	839756	150504	893294	153078	895868	154381	897171
2009	206570	978771	264157	1036357	258223	1030423	251156	1023357
2010	212671	913002	298226	998557	307960	1008291	313399	1013729
2011	264817	1808531	232038	1775752	229434	1773149	229442	1773156
2012	1108787	1756413	431637	1079263	365761	1013387	344214	991840
2013	620615	1644729	314522	1338636	327904	1352018	246091	1270204
2014	367560	767301	100036	499777	74092	473833	49006	448747
2015	403564	609932	109425	315792	77479	283846	55375	261742
2016	507494	570398	195389	258293	176702	239606	122047	184951
Total Reserve	**3789045**	**10487555**	**2095935**	**8794444**	**1970635**	**8669144**	**1765111**	**8463620**
Loss Ratio	**59,26%**		**49,70%**		**48,99%**		**47,83%**	

In the presence of one outlier ($Y_{1,3}^A = 101105\mathbf{0}$, $Y_{1,3}^B = 144235\mathbf{0}$), the LS regression model provides an overestimation of the total reserves for both companies (1989989 for company A and 3699022 for company B), which leads to an increase of 5.68% of the loss ratio for the Company A (the loss ratio without outliers is 47.6%, while with one outlier is 52.35%) and in an increase of 11.74% for the Company B (the loss ratio without outliers is 44.03%, while with one outlier is 55.77%). The robust regression model (e.g Hampel) identifies this outlier, resulting to an increase of 1.94% of the loss ratio for the Company A and an increase of 1.97% for the Company B. This increase due to the robust estimation is significantly less than the increase due to the LS estimation.

In the presence of two outliers for both companies ($Y_{13}^A = 101105\mathbf{0}$, $Y_{8,3}^A = 106150\mathbf{0}$ and $Y_{13}^B = 144235\mathbf{0}$, $Y_{8,3}^B = 145385\mathbf{0}$), the LS regression model provides a significantly large overestimation of the total reserves for both companies (the estimated total reserve for Company A is 2983608, while the estimated total reserve for Company B is 3789045), which leads to an increase of 14.07% of the loss ratio for Company A and of 15.23% for Company B. This increase shows that in the presence of outlier events the LS model does not provide robust estimation of the reserves and cannot be used. The robust regression model (e.g Bisquare) gives and increase of 5.21% for Company A and 3.8% for Company B. Clearly, the robust regression model identifies outliers and provides reliable estimation of the total reserves in contrast to the LS model that provides unreliable reserve estimation. In general, the underestimation or overestimation of reserves depend on the position of the outliers in the run of triangle (see Pitselis et al., 2015)

According to the residuals of the least squares RCR cross effect model of both companies, we check the dependence between the two companies. Recall that we have observed a positive correlation between the loss ratios which might reflect the accident and/or development year effects (see Figure 3.14).

Figure 3.20 indicates a negative relationship between the residuals of the two triangles ($r_{A,B} = -0.12$). This is opposed to the correlation of the loss ratios of the companies which was computed 0.944 (and 0.916 the Spearman correlation coefficient). A possible reason may be the effect of accident and development years, which play an important role to the estimations (see Shi and Frees, 2011).

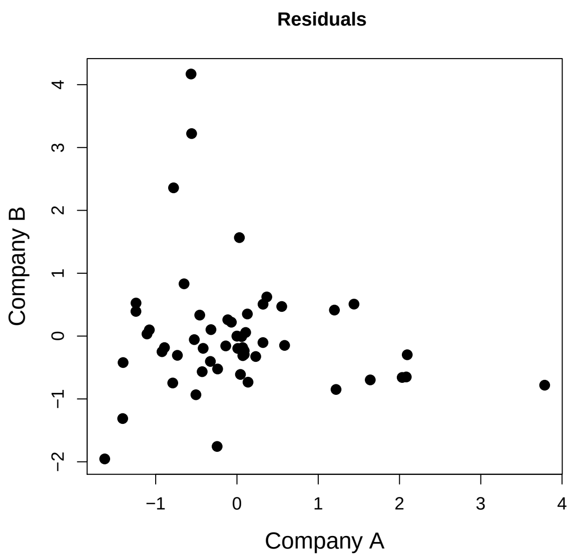

Figure 3.20: Residuals Scatter Plot for companies A and B

Figures 3.21, 3.22 and 3.23 present the cumulative reserve estimation by accident year for both companies in the presence of no outliers (Figure 3.21), of one outlier (Figure 3.22) and two outliers (Figure 3.23).

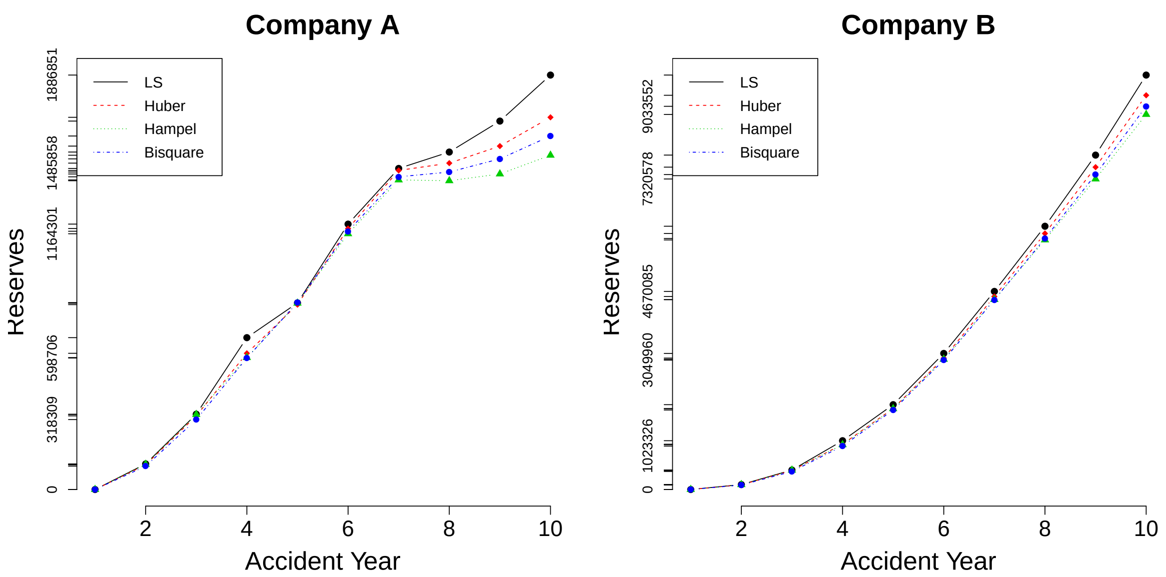

Figure 3.21: RCR cross section cumulative reserve estimation by accident year without outliers for companies A and B

As we can see, there are cases where the LS method underestimates or overestimates the total reserve. The robust functions seem to be more stable and can be used to identify outliers and provide robust estimations of the total reserves.

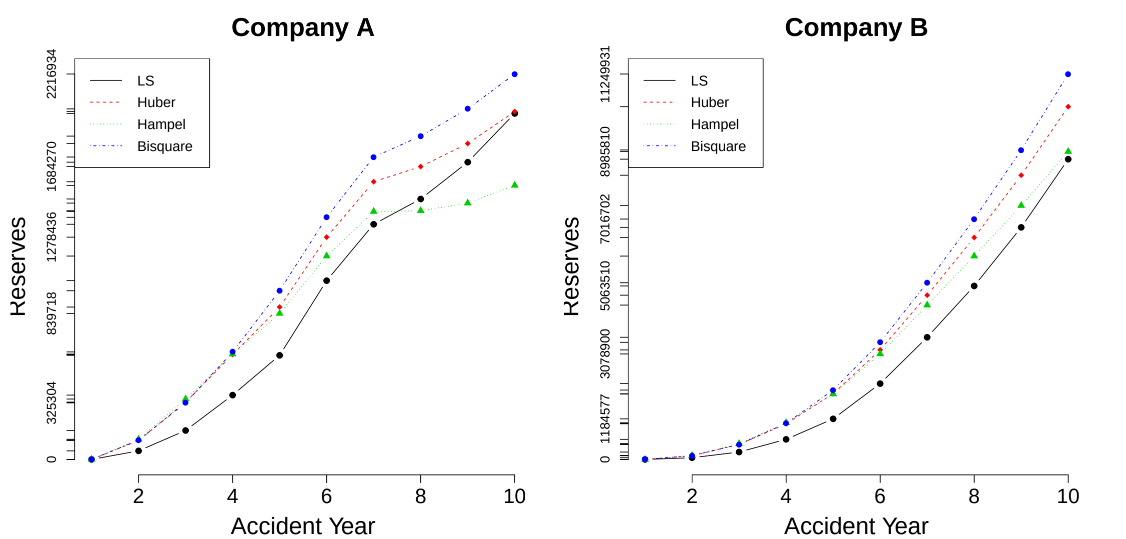

Figure 3.22: RCR cross section cumulative reserve estimation by accident year with one outlier for companies A and B

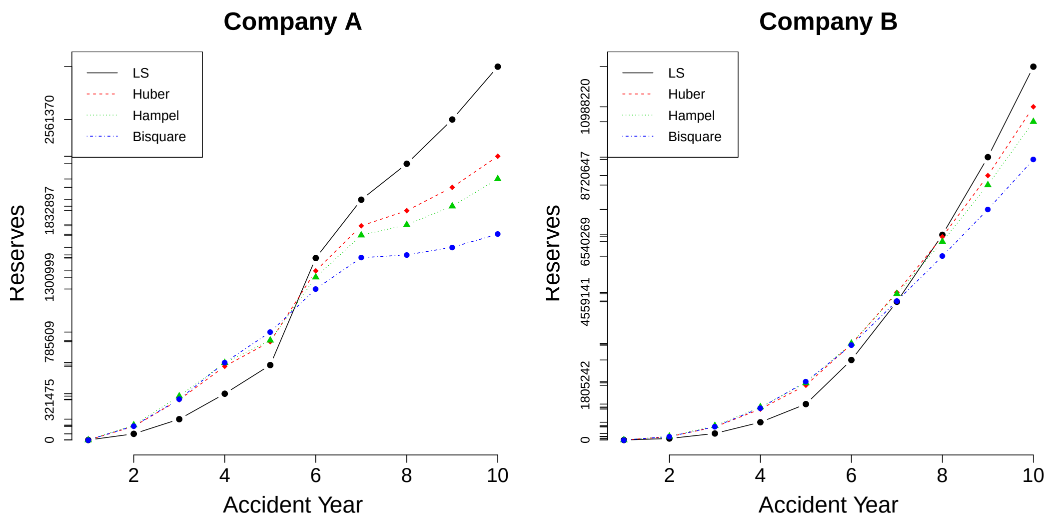

Figure 3.23: RCR cross section cumulative reserve estimation by accident year with two outliers for companies A and B

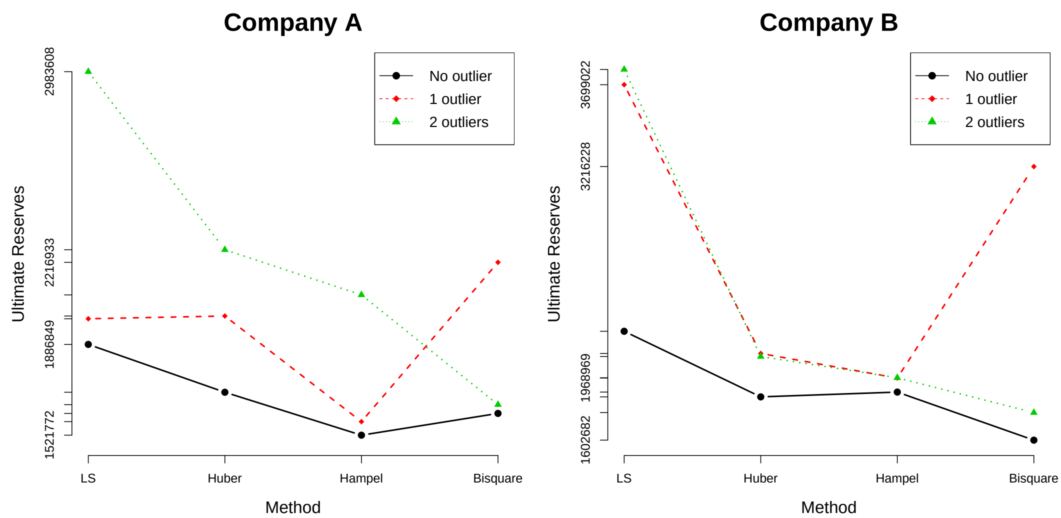

Figure 3.24: Total Reserves for non-robust (LS) and robust RCR Cross Effect Models for companies A and B

Based on the results displayed in Tables 3.13 up to 3.16, the values of total reserves are shown in Figure (3.24), in the presence of no outlier, one artificial outlier and two artificial outliers, for the RCR cross effects model and for the robustified versions of it, based on the three robust functions (Huber, Hampel, Bisquare). In the presence of one or two outliers, the value of reserves for Company A and Company B based on the RCR cross effects model become very large in comparison to reserves based on robust estimations. Even in the presence of no artificial outlier the value of the reserves under the RCR cross is bigger that the corresponding reserves under robust estimation. This means that the robust functions identify and other observations as outliers in addition to artificial outliers we set. Therefore the superiority of robust estimation against non-robust estimation of the RCR cross effects model is evident.

3.7 Concluding Remarks

In this chapter we showed how random coefficient regression models can be incorporated in loss reserving techniques. These models provide a fair value for the estimation of outstanding reserves in cases we have indications that the run-off patterns are changing. First, we proposed a univariate reserving model. We relax the assumptions that the development factors are constant and we propose a linear regression model with random coefficients. Second, we proposed a multivariate reserving model based on the cross sectional regression model. Nevertheless as mentioned above RCR models are very sensitive to outlier events (very large values) and most of the times resulting to an overestimation of the total reserves. Although there are cases where we obtain an underestimation of the total reserves depending on the position and the size of the outliers.

In order to remediate the effect of outliers to the estimation of the total reserves, robust versions of the above two random coefficient models are applied. Implementing the data sets for both models we showed the superiority of robust M-estimator in comparison with the non-robust estimators. The advantage of applying M-estimators for the estimation of reserves are: a) they are simple, b) robust procedures are available in R packages and c) these packages are interactive with other programs (excel) for better implementation of the sets. Of course at the final step of reserves estimation, the bias term due to robustification of the outliers (if there are true large values and not fake outliers) must be added to the value of total reserve.

CHAPTER 4

Robust Kalman Filter in a State Space Model

In this chapter, a robust application of the Kalman Filter (KF) to a state-space model recursive algorithm is going to be used, in order to estimate the reserves of the insurer. The application of the KF algorithm to a state-space model leads to a flexible framework, which uses the past claims data and produce future reserve estimations. The most important point of this algorithm is the model's parameters which can be assumed to evolve over time in contrast with other models which assume constant parameters (see De Jong and Zeinwirth, 1983 and Verrall, 1994).

4.1 State Space Models

The development of the state space models has been made the last 20 years. They are important models because they mainly give a rich family of models which are easily interpretable for data and simultaneously they lead to highly efficient estimations and forecasting algorithms via the Kalman Filter recursions. Moreover, state space models allow the creation of models with meaningful dynamic parameters. In non life insurance, they can be used as a framework for automating the reserving procedure. A state space model is regularly specified by 2 sets of equations and characterizes the dependency of the observations and the state variable (see Koller and Friedman, 2009):

- An observation equation

$$\boldsymbol{y}_t = \boldsymbol{F}_t\boldsymbol{\theta}_t + \boldsymbol{w}_t,\ t \geq 0, \tag{4.1}$$

where $\boldsymbol{y}_t$ is a vector of observed data, $\boldsymbol{\theta}_t$ is the unknown state-vector, $\boldsymbol{F}_t$ is a known matrix which connects the state vector $\boldsymbol{\theta}_t$ to the observations and finally $\boldsymbol{w}_t$ is a noise process.

- A state equation

$$\boldsymbol{\theta}_{t+1} = \boldsymbol{G}_{t+1}\boldsymbol{\theta}_t + \boldsymbol{v}_{t+1},\ t \geq 0, \tag{4.2}$$

which characterizes the evolution of the vector $\boldsymbol{\theta}_t$ in Markovian way using a transition matrix $\boldsymbol{G}_{t+1}$. Likewise in the observation equation, the state equation incorporates a random error vector $\boldsymbol{v}_t$.

The main point of this approach is to make inference about $\boldsymbol{\theta}_t$ given the vector of the observations $\boldsymbol{y}_t$. Generally,

- state space models represent dynamic systems,
- the states $\boldsymbol{\theta}_t$ of the system are not directly observable,
- these states drive the observable set of values $\boldsymbol{y}_t$,
- the state space model has conditional independence structure.

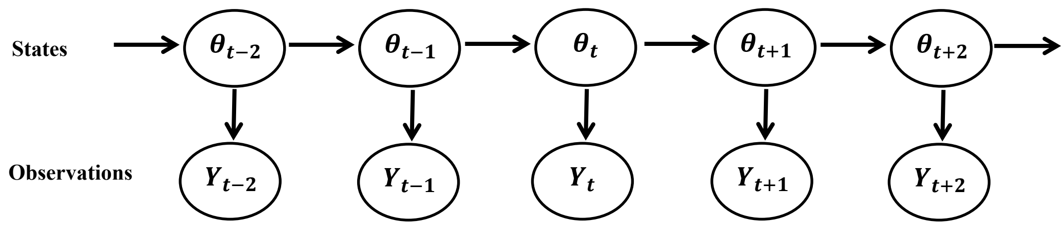

Figure 4.1: State space models

It is supposed that $\boldsymbol{y}_t$ is a vector of n observations, $\boldsymbol{\theta}_t$ is the r dimensional state vector, $\boldsymbol{F}_t$ is a matrix with dimensions $r \times n$ and $\boldsymbol{G}_t$ an $r \times r$ matrix. The nuisance vectors $\boldsymbol{w}_t$ and $\boldsymbol{v}_t$ are n and r dimensional vectors with variance-covariance matrices

$$E[\boldsymbol{v}_t\boldsymbol{v}'_{t+k}] = \begin{cases} \boldsymbol{Q}_t, & k = 0, \\ 0, & \text{otherwise}, \end{cases} \qquad E[\boldsymbol{w}_t\boldsymbol{w}'_{t+k}] = \begin{cases} \boldsymbol{R}_t, & k = 0, \\ 0, & \text{otherwise}, \end{cases}$$

where $\boldsymbol{Q}_t$ is an $r \times r$ matrix and $\boldsymbol{R}_t$ is an $n \times n$ matrice. Moreover, it is supposed that $E[\boldsymbol{v}_t\boldsymbol{w}'_{t+k}] = 0\ \forall t, k$.

Remark 4.1. The fact that the variance-covariance matrices $\boldsymbol{Q}_t$ and $\boldsymbol{R}_t$ depend on t means that this model assumes heteroscedasticity. This is not observed when using common methods.

4.2 The Kalman Filter Algorithm

The estimation of $\boldsymbol{\theta}_t$ is based on the sequence $\boldsymbol{y}_1, \boldsymbol{y}_2, \dots \boldsymbol{y}_t$ and the model assumptions (see section 4.1) should be satisfied. The KF algorithm is comrised of some recursive steps which are linear. The algorithm estimates the state vector $\boldsymbol{\theta}_t$ based on $\boldsymbol{\theta}_{t-1}$ and the new data vector $\boldsymbol{y}_t$. The Steps of this algorithm are the following (see Kailath et al., 2000, Li, 2006)

Step 1: $\hat{\boldsymbol{y}}_t = \boldsymbol{F}_t \hat{\boldsymbol{\theta}}_{t|t-1}$,

Step 2: $\boldsymbol{\epsilon}_t = \boldsymbol{y}_t - \hat{\boldsymbol{y}}_t$,

Step 3: $\hat{\boldsymbol{\theta}}_{t|t} = \hat{\boldsymbol{\theta}}_{t|t-1} + \boldsymbol{P}_{t|t-1}\boldsymbol{F}_t(\boldsymbol{F}'_t\boldsymbol{P}_{t|t-1}\boldsymbol{F}_t + \boldsymbol{R}_t)^{-1}\boldsymbol{\epsilon}_t$,

Step 4: $\hat{\boldsymbol{\theta}}_{t+1|t} = \boldsymbol{G}_{t+1}\hat{\boldsymbol{\theta}}_{t|t}$,

Step 5: $\boldsymbol{P}_{t|t} = \boldsymbol{P}_{t|t-1} - \boldsymbol{P}_{t|t-1}\boldsymbol{F}_t(\boldsymbol{F}'_t\boldsymbol{P}_{t|t-1}\boldsymbol{F}_t + \boldsymbol{R}_t)^{-1}\boldsymbol{F}'_t\boldsymbol{P}_{t|t-1}$,

Step 6: $\boldsymbol{P}_{t+1|t} = \boldsymbol{G}_{t+1}\boldsymbol{P}_{t|t}\boldsymbol{G}'_{t+1} + \boldsymbol{Q}_{t+1}$.

Here $\hat{\boldsymbol{\theta}}_{t|t}$ is the estimate of $\boldsymbol{\theta}_t$ given the data vector $\boldsymbol{y}_1, \boldsymbol{y}_2, \dots \boldsymbol{y}_t$ till time t while $\boldsymbol{P}_{t|t}$ is the variance-covariance matrix. The vector $\hat{\boldsymbol{\theta}}_{t+1|t}$ is the estimate of the vector $\boldsymbol{\theta}_{t+1}$ at time $t+1$ given the data vector $\boldsymbol{y}_1, \boldsymbol{y}_2, \dots \boldsymbol{y}_t$ up to time t ("again") while the matrix $\boldsymbol{P}_{t+1|t}$ is the variance of that estimate. This means that the KF algorithm not only updates the estimates $\hat{\boldsymbol{\theta}}_{t|t}$ and $\hat{\boldsymbol{\theta}}_{t+1|t}$ but also their variance-covariance matrices. Figure 4.2 shows the Steps of the Kalman filter procedure.

In cases of noise parameters being normally distributed, the estimates are also normally distributed which permits the construction of a confidence interval for the parameters

$$\hat{\theta}_{k,t|t} \pm t_a\sqrt{p_{kk,t|t}} = \hat{\theta}_{k,t+1|t} \pm t_a\sqrt{p_{kk,t+1|t}}, \tag{4.3}$$

where $\hat{\theta}_{k,t|t}$ is the k^{th} component of $\hat{\boldsymbol{\theta}}_{t|t}$, $p_{kk,t|t}$ is the diagonal element of the matrix $\boldsymbol{P}_{t|t}$ and t_a is the upper quantile at confidence level a.

Let's suppose that $\hat{\boldsymbol{\theta}}_{t|t-1}$ is a forecast for $\boldsymbol{\theta}_t$ based on the data $\boldsymbol{y}_1, \boldsymbol{y}_2, \dots \boldsymbol{y}_{t-1}$ up to time $t-1$. Then, using Step 1 in the Kalman Filter equation, a good prediction for $\boldsymbol{y}_t$

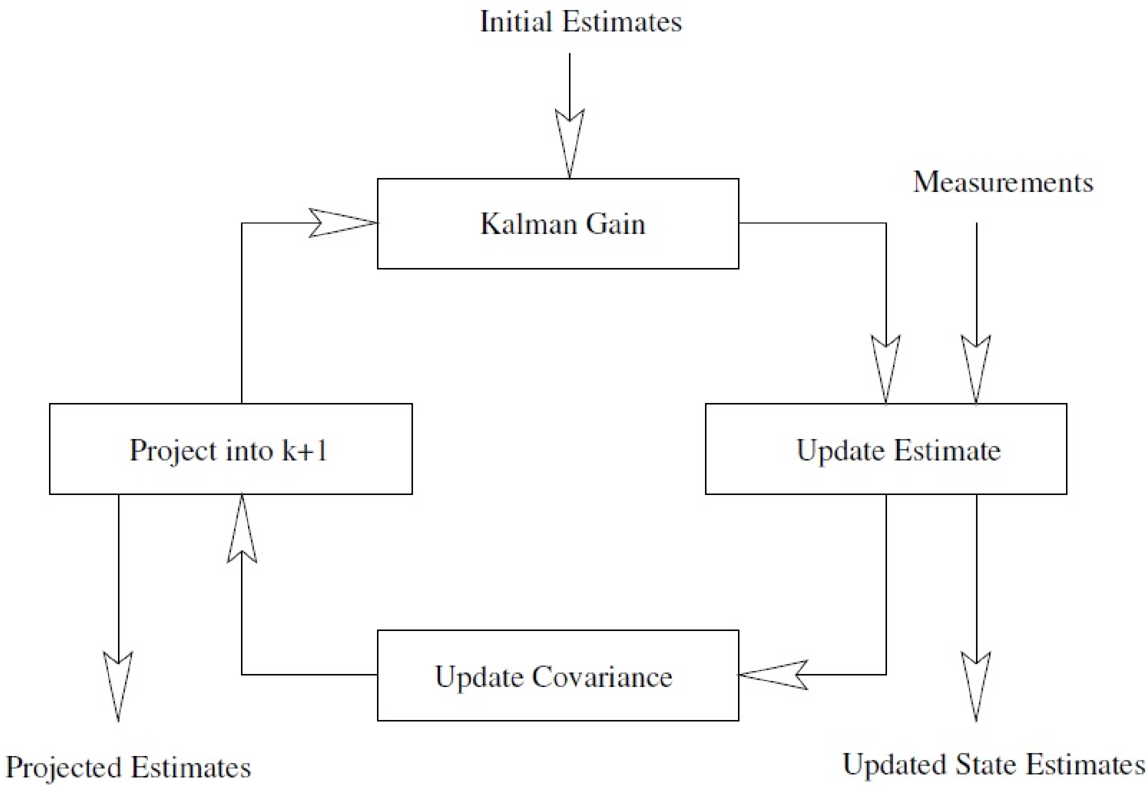

Figure 4.2: Kalman Filter Algorithm

can be made

$$\boldsymbol{y}_t = \boldsymbol{F}_t\boldsymbol{\theta}_t + \boldsymbol{w}_t,$$
$$\hat{\boldsymbol{y}}_t = \boldsymbol{F}_t\hat{\boldsymbol{\theta}}_t + 0 = \boldsymbol{F}_t\hat{\boldsymbol{\theta}}_{t|t-1}.$$

It should be mentioned that the best prediction for $\boldsymbol{w}_t$ based on the data $\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots \boldsymbol{y}_{t-1}$ is zero because of the property of the white noise model. If $\boldsymbol{\theta}_{t|t-1}$ is a good estimate of $\boldsymbol{\theta}_t$, then $\hat{\boldsymbol{y}}_t$ is a good estimate of $\boldsymbol{y}_t$, which means that $\boldsymbol{w}_t$ should be zero.

In Step 2, the quality of the forecast at Step 1 can be checked by calculating the forecast error as far as the new observations point $\boldsymbol{y}_t$ is available. This means that the quantity $\boldsymbol{\epsilon}_t^{'}\boldsymbol{\epsilon}_t$ should be small[1] in case of a good prediction at Step 1.

Step 3 is more complicated than the previous two Steps. The estimation of $\hat{\boldsymbol{\theta}}_{t|t}$ based on the new observation $\boldsymbol{y}_t$ is a combination of the forecast $\hat{\boldsymbol{\theta}}_{t|t-1}$ and the forecast error $\boldsymbol{\epsilon}_t$. When the forecast error is small, then it is deduced that $\hat{\boldsymbol{\theta}}_{t|t-1}$ may be a good estimate of $\boldsymbol{\theta}_t$ which means that there is no need to revise the old forecast. When the forecast error is surprisingly large, then $\hat{\boldsymbol{\theta}}_{t|t-1}$ is not a good estimate of $\boldsymbol{\theta}_t$ and the old estimate should

[1]This means that a criterion to estimate the unknown parameter can be constructed.

be recalculated. The term $\boldsymbol{P}_{t|t-1}\boldsymbol{F}_t(\boldsymbol{F}_t'\boldsymbol{P}_{t|t-1}\boldsymbol{F}_t + \boldsymbol{R}_t)^{-1}$ is called the *Kalman Gain* and is the most important element in this Step as this is an optimal mean square weighting quantity.

In order to derive the Kalman-Gain, we will use the simple MSE of the not known weight $\boldsymbol{\beta}$ in the regression

$$\boldsymbol{\theta}_t = \hat{\boldsymbol{\theta}}_{t|t-1} + \boldsymbol{\beta}\boldsymbol{\epsilon}_t + \boldsymbol{z}_t.$$

The least squares estimate for the parameter $\boldsymbol{\beta}$ is then

$$Cov[\boldsymbol{\theta}_t, \boldsymbol{\epsilon}_t|\boldsymbol{y}_{t-1}, \boldsymbol{y}_{t-2}, \dots, \boldsymbol{y}_1]Var[\boldsymbol{\epsilon}_t|\boldsymbol{y}_{t-1}, \boldsymbol{y}_{t-2}, \dots, \boldsymbol{y}_1]^{-1}. \tag{4.4}$$

Taking each part separately we have

$$\begin{aligned} Cov[\boldsymbol{\theta}_t, \boldsymbol{\epsilon}_t|\boldsymbol{y}_{t-1}, \boldsymbol{y}_{t-2}, \dots, \boldsymbol{y}_1] &= E[(\boldsymbol{\theta}_t - \boldsymbol{\theta}_{t|t-1})(\boldsymbol{y}_t - \hat{\boldsymbol{y}}_{t|t-1})'] \\ &= E[(\boldsymbol{\theta}_t - \boldsymbol{\theta}_{t|t-1})(\boldsymbol{F}'(\boldsymbol{\theta}_t - \boldsymbol{\theta}_{t|t-1}) + \boldsymbol{w}_t)'] \\ &= \boldsymbol{P}_{t|t-1}\boldsymbol{F}. \end{aligned}$$

In the same way we have

$$\begin{aligned} Var[\boldsymbol{\epsilon}_t|\boldsymbol{y}_{t-1}, \boldsymbol{y}_{t-2}, \dots, \boldsymbol{y}_1]^{-1} &= Var[\boldsymbol{y}_t - \hat{\boldsymbol{y}}_{t|t-1}] \\ &= Var[\boldsymbol{F}'(\boldsymbol{\theta}_t - \boldsymbol{\theta}_{t|t-1}) + \boldsymbol{w}_t] \\ &= \boldsymbol{F}'\boldsymbol{P}_{t|t-1}\boldsymbol{F} + \boldsymbol{R}. \end{aligned}$$

Inserting the above two expressions into the least squares estimation in (4.4) we obtain the Kalman-Gain.

For the next Step, the state space equation is firstly considered

$$\boldsymbol{\theta}_{t+1} = \boldsymbol{G}_{t+1}\boldsymbol{\theta}_t + \boldsymbol{v}_{t+1}.$$

In the case of predicting $\boldsymbol{\theta}_{t+1}$ based on data $\boldsymbol{y}_1, \boldsymbol{y}_2, \dots \boldsymbol{y}_t$ (up to time t), the new updated quantity $\hat{\boldsymbol{\theta}}_{t|t}$ should be used into this equation

$$\hat{\boldsymbol{\theta}}_{t+1} = \boldsymbol{G}_{t+1}\hat{\boldsymbol{\theta}}_{t|t}$$

and $\hat{\boldsymbol{v}}_{t+1} = 0$ because of the assumption of the white noise model. This means that a full circle which started from $\boldsymbol{\theta}_{t|t-1}$ and ended to $\boldsymbol{\theta}_{t+1|t}$, has been completed.

In Step 4 the variances should also be updated and in Step 5 the updated variance $\boldsymbol{P}_{t|t-1}$ is taken in cases when we have new data $\boldsymbol{y}_t$, which affect the parameters and possibly decrease the uncertainty of the estimation.

For the derivation of the expression $\boldsymbol{P}_{t|t}$ we have

$$\begin{aligned}
\boldsymbol{P}_{t|t} &= Var[\hat{\boldsymbol{\theta}}_{t|t}] \\
&= Var[\hat{\boldsymbol{\theta}}_{t|t-1} + \boldsymbol{P}_{t|t-1}\boldsymbol{F}_t(\boldsymbol{F}_t'\boldsymbol{P}_{t|t-1}\boldsymbol{F}_t + \boldsymbol{R}_t)^{-1}\boldsymbol{\epsilon}_t] \\
&= Var[\hat{\boldsymbol{\theta}}_{t|t-1}] + Var[\boldsymbol{P}_{t|t-1}\boldsymbol{F}_t(\boldsymbol{F}_t'\boldsymbol{P}_{t|t-1}\boldsymbol{F}_t + \boldsymbol{R}_t)^{-1}\boldsymbol{\epsilon}_t] \\
&+ 2Cov[\hat{\boldsymbol{\theta}}_{t|t-1}, \boldsymbol{P}_{t|t-1}\boldsymbol{F}_t(\boldsymbol{F}_t'\boldsymbol{P}_{t|t-1}\boldsymbol{F}_t + \boldsymbol{R}_t)^{-1}\boldsymbol{\epsilon}_t].
\end{aligned}$$

But,

$$\begin{aligned}
&Var[\boldsymbol{P}_{t|t-1}\boldsymbol{F}_t(\boldsymbol{F}_t'\boldsymbol{P}_{t|t-1}\boldsymbol{F}_t + \boldsymbol{R}_t)^{-1}\boldsymbol{\epsilon}_t] \\
&= Var[\boldsymbol{P}_{t|t-1}\boldsymbol{F}_t(\boldsymbol{F}_t'\boldsymbol{P}_{t|t-1}\boldsymbol{F}_t + \boldsymbol{R}_t)^{-1}(\boldsymbol{F}_t'(\boldsymbol{\theta}_t - \hat{\boldsymbol{\theta}}_{t|t-1} + \boldsymbol{w}_t))] \\
&= \boldsymbol{P}_{t|t-1}\boldsymbol{F}_t(\boldsymbol{F}_t'\boldsymbol{P}_{t|t-1}\boldsymbol{F}_t + \boldsymbol{R}_t)^{-1}(\boldsymbol{F}_t'\boldsymbol{P}_{t|t-1}\boldsymbol{F}_t + \boldsymbol{R}_t)(\boldsymbol{P}_{t|t-1}\boldsymbol{F}_t(\boldsymbol{F}_t'\boldsymbol{P}_{t|t-1}\boldsymbol{F}_t + \boldsymbol{R}_t)^{-1})' \\
&= \boldsymbol{P}_{t|t-1}\boldsymbol{F}_t(\boldsymbol{F}_t'\boldsymbol{P}_{t|t-1}\boldsymbol{F}_t + \boldsymbol{R}_t)^{-1}(\boldsymbol{F}_t'\boldsymbol{P}_{t|t-1}\boldsymbol{F}_t + \boldsymbol{R}_t)(\boldsymbol{F}_t'\boldsymbol{P}_{t|t-1}\boldsymbol{F}_t + \boldsymbol{R}_t)^{-1}\boldsymbol{F}_t'\boldsymbol{P}_{t|t-1}' \\
&= \boldsymbol{P}_{t|t-1}\boldsymbol{F}_t(\boldsymbol{F}_t'\boldsymbol{P}_{t|t-1}\boldsymbol{F}_t + \boldsymbol{R}_t)^{-1}\boldsymbol{F}_t'\boldsymbol{P}_{t|t-1}.
\end{aligned}$$

Similarly,

$$Cov[\hat{\boldsymbol{\theta}}_{t|t-1}, \boldsymbol{P}_{t|t-1}\boldsymbol{F}_t(\boldsymbol{F}_t'\boldsymbol{P}_{t|t-1}\boldsymbol{F}_t + \boldsymbol{R}_t)^{-1}\boldsymbol{\epsilon}_t] = -\boldsymbol{P}_{t|t-1}\boldsymbol{F}_t(\boldsymbol{F}_t'\boldsymbol{P}_{t|t-1}\boldsymbol{F}_t + \boldsymbol{R}_t)^{-1}\boldsymbol{F}_t'\boldsymbol{P}_{t|t-1}.$$

Using these two expressions we take

$$\boldsymbol{P}_{t|t} = \boldsymbol{P}_{t|t-1} - \boldsymbol{P}_{t|t-1}\boldsymbol{F}_t(\boldsymbol{F}_t'\boldsymbol{P}_{t|t-1}\boldsymbol{F}_t + \boldsymbol{R}_t)^{-1}\boldsymbol{F}_t'\boldsymbol{P}_{t|t-1}.$$

In the final Step, the new variance $\boldsymbol{P}_{t+1|t}$ of the new forecast of $\boldsymbol{\theta}_{t+1}$ is taken:

$$\begin{aligned}
Var[\boldsymbol{\theta}_{t+1}|\boldsymbol{y}_1, \ldots \boldsymbol{y}_t] &= Var[\boldsymbol{G}_{t+1}\boldsymbol{\theta}_t|\boldsymbol{y}_1, \ldots \boldsymbol{y}_t] + Var[\boldsymbol{v}_{t+1}|\boldsymbol{y}_1, \ldots \boldsymbol{y}_t] \\
&= \boldsymbol{G}_{t+1}\boldsymbol{P}_{t|t}\boldsymbol{G}'_{t+1} + \boldsymbol{Q}_{t+1}.
\end{aligned}$$

Remark 4.2. Given the data vector $\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots \boldsymbol{y}_{t-1}$ till time $t-1$, under the model assumptions, the estimate $\hat{\boldsymbol{\theta}}_{t|t-1}$ is the conditional mean of $\boldsymbol{\theta}_t$ and the conditional variance $\boldsymbol{P}_{t|t-1}$. This means that in case of Gaussian noise errors the distribution of $\hat{\boldsymbol{\theta}}_{t|t-1}$ is completely specified $\forall t$. In different circumstances, the Kalman Filter algorithm gives the best linear unbiased estimate of $\boldsymbol{\theta}_t$ (see Cazan, 2011).

Remark 4.3. In order to start the Kalman Filter algorithm it is crucial to have the starting values $\hat{\boldsymbol{\theta}}_{1|0}$ and $\boldsymbol{P}_{1|0}$. For $\hat{\boldsymbol{\theta}}_{1|0}$ we can use the unconditional mean $E(\boldsymbol{\theta}_1)$ and for $\boldsymbol{P}_{1|0}$ we can use the unconditional variance $Var(\boldsymbol{\theta}_1)$.

4.3 Regression Analysis into State Space Representation

Let y_t be a dependent variable and $x_{1t}, x_{2t}, \ldots, x_{pt}$ be a set of $p > 0$ explanatory variables. The regression equation can be defined as

$$y_t = \beta_0 + \beta_1 x_{1t} + \ldots + \beta_p x_{pt} + \epsilon_t, \tag{4.5}$$

where ϵ_t denotes the sequence of random errors which follow the normal distribution with constant variance $Var(\epsilon_t) = \sigma^2$ and $\beta_0, \beta_1, \ldots, \beta_p$ are unknown parameters which must be estimated. Then the state space representation model in (4.1) and (4.2) in connection with the regression model in (4.5) is defined as

$$\begin{aligned} y_t &= \boldsymbol{F}_t \boldsymbol{\theta}_t + w_t, \\ \boldsymbol{\theta}_{t+1} &= \boldsymbol{G} \boldsymbol{\theta}_t. \end{aligned} \tag{4.6}$$

where

- y_t is the value of the incremental payment at time t,
- $\boldsymbol{\theta}_t' = (\beta_0, \beta_1, \ldots, \beta_p)$ is the vector of the unknown coefficients at time t,
- $w_t \equiv \epsilon_t$ is the random normal error,
- $\boldsymbol{F}_t = (1, x_{1t}, \ldots, x_{pt})$ is a $1 \times (p+1)$ vector of the covariates,
- $R = \sigma^2 > 0$ the known variance of the errors w_t,
- $\boldsymbol{G}_t$ is a $(p+1) \times (p+1)$ identity matrix,
- $\boldsymbol{Q}_t = \mathbf{0}$ is a $(p+1) \times (p+1)$ matrix with elements 0.

Remark 4.4. We do not have to know the variance σ^2 so as to estimate the regression coefficients because $\boldsymbol{Q} = 0$. We can give an arbitrary value for σ^2 without affecting the final result ($\boldsymbol{Q}/\boldsymbol{R} = 0$, $\forall\ \boldsymbol{R} > 0$). In case that $\boldsymbol{Q} > 0$, the regression coefficients are time dependent and the ratio $\boldsymbol{Q}/\boldsymbol{R}$ is not zero and both variances should be estimated.

4.4 Optimization criteria

In order to determine the unknown model's coefficients there are two different options for the optimization criterion (Petris et al., 2010, Tusell, 2011):

(i) **Least - Squares Method**

Using this optimization criterion, the estimations of the Kalman filter model is subject of the minimization of the squared residuals

$$\sum_{i=1}^{n} \boldsymbol{e}_i' \boldsymbol{e}_i.$$

(ii) **Maximum Likelihood Method**

The second optimization method relies on the maximum likelihood function when the errors follow the normal distribution. In that case, the unconditional distribution of $\boldsymbol{y}_1$ and the sequence of the conditional distributions of $\boldsymbol{y}_t|\boldsymbol{y}_{t-1},\ \boldsymbol{y}_{t-2},\ldots$ come from a multivariate normal distribution with density functions

$$f(\boldsymbol{y}_1) = (2\pi)^{-r/2}|\boldsymbol{P}_{1|0}|^{-1/2}\exp\Big[-\frac{(\boldsymbol{y}_1 - \boldsymbol{F}_1'\hat{\boldsymbol{\theta}}_{1|0})'(\boldsymbol{P}_{1|0}^{-1})(\boldsymbol{y}_1 - \boldsymbol{F}_1'\hat{\boldsymbol{\theta}}_{1|0})}{2}\Big]$$

and

$$f(\boldsymbol{y}_t|\boldsymbol{y}_{t-1},\ldots) = (2\pi)^{-r/2}|\Xi_t|^{-1/2}\exp\Big[-\frac{(\boldsymbol{y}_t - \hat{\boldsymbol{y}}_t)'(\Xi_t)^{-1}(\boldsymbol{y}_t - \hat{\boldsymbol{y}}_t)}{2}\Big],$$

where r denotes the dimension of $\boldsymbol{\theta}$. Note that $\boldsymbol{y}_t|\boldsymbol{y}_{t-1},\ \boldsymbol{y}_{t-2},\ldots$ are independent random variables normally distributed with mean $\hat{\boldsymbol{y}}_t$ and variance

$$\Xi_t = Var[\boldsymbol{y}_t|\boldsymbol{y}_{t-1},\ \boldsymbol{y}_{t-2},\ldots] = Var[\boldsymbol{F}_t'\boldsymbol{\theta}_{t|t-1} + \boldsymbol{w}_t] = \boldsymbol{F}_t'\boldsymbol{P}_{t|t-1}\boldsymbol{F}_t + \boldsymbol{R}_t.$$

Therefore, the common distribution function should be maximized, i.e.,

$$f(\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_t) = f(\boldsymbol{y}_1)\prod_{t=2}^{T} f(\boldsymbol{y}_t|\boldsymbol{y}_{t-1},\ldots) \to \max_{\boldsymbol{\theta}},$$

where $\boldsymbol{\theta}$ are all unknown parameters including the initial values.

In cases of mumerical data, it is easier to minimize the negative log likelihood function

$$-\ln(f(\boldsymbol{y}_1)) - \sum_{t=2}^{T}\ln(f(\boldsymbol{y}_t|\boldsymbol{y}_{t-1},\ldots)) = \ln(\boldsymbol{P}_{1|0}) + \sum_{t=2}^{T}\ln(\boldsymbol{F}_t'\boldsymbol{P}_{t|t-1}\boldsymbol{F}_t + \boldsymbol{R}_t) +$$
$$+(\boldsymbol{y}_1 - \boldsymbol{F}_1'\hat{\boldsymbol{\theta}}_{1|0})'(\boldsymbol{P}_{1|0}^{-1})(\boldsymbol{y}_1 - \boldsymbol{F}_1'\hat{\boldsymbol{\theta}}_{1|0}) + \sum_{t=2}^{T}(\boldsymbol{y}_t - \hat{\boldsymbol{y}}_t)'(\Xi_t)^{-1}(\boldsymbol{y}_t - \hat{\boldsymbol{y}}_t).$$

4.5 Identification of Outliers and Robust Kalman Filter in the Claim Reserving State Space Models

While studying a data set, statistical methods should be applied using either the diagnostic or accommodation approaches because some of the observations may not be homogeneous (Maronna, et al. 2006). These types of methods on one hand identify potential outliers and on the other hand discard them from all calculations. All these methods are used mainly because they can be easily applied in an estimator without making changes to the algorithm itself. Nevertheless, discarding observations which do not follow the pattern of the majority of the data may not be a solution because the data may contain a large percentage of outliers. For that reason, outliers should always be examined exhaustively in order to determine whether they follow any pattern or if they could be studied adequately by an alternative method. The regression equivariant estimator's maximum breakdown point, under the assumption of general position, is given by $\epsilon^*_{\max} = [(m-n)/2]/m$, in which we have m data points and n state variables. When using the classical recursive KF model, $m = n+1$ are the total number of data points at each step k of the algorithm, as long as we have n predictions (1 for each variable) and 1 data point which is used at each period. According to Maronna, et al. 2006, the maximum breakdown point is calculated by

$$\epsilon^*_{\max} = [(n+1-n)/2]/m = [1/2]/m = 0/m = 0.$$

In order to have a positive breakdown and flexibility to outliers, some extra observations need to be used. If only one measurement gives $m = n+2$ data points, the maximum breakdown point is

$$\epsilon^*_{\max} = [(n+2-n)/2]/m = [2/2]/m = 1/m,$$

which means that the filter is able to handle more than one outliers! As we can see, the estimator handles an extra outlier when we have 2 observations; so, $m_r = m-1$ redundant observations allows the estimator to manage $m/2$ outliers. For example, in order to be robust to 2 outliers, i.e. $m/2 = 2$, we need $m = 4$ data points, and proportionately, $m_r = 3$ additional measurements. According to all these we take

$$\epsilon^*_{\max} = [(n+4-n)/2]/m = [4/2]/m = 2/m.$$

4.5.1 Robustification of the Kalman Filter

An important assumption of the KF model is that the errors follow the normal distribution. The majority of the models are based on basic conditions, such as the normality of the errors in the response observations. In case that the errors' distribution is not symmetric or prone to outliers then the assumptions are not validated which means that the coefficient estimations, the confidence intervals, and more statistical quantities will not be reliable.

In the Kalman Filter model, in Step 2 (see Section 4.2) the lack of normality of the errors or the presence of outliers may not only distort the update of $\boldsymbol{\theta}_{t|t}$ but also a part of the sequence of $\boldsymbol{\theta}_{t+i|t+i}$, $\forall\, i > 0$. For that reason, a simple robust strategy for eliminating the effects of such observations, is the replacement of the error $\boldsymbol{\epsilon}_t$ with a bounded function $\psi(\boldsymbol{\epsilon}_t)$.

A practical procedure to find an outlier is to identify it using the studentized residuals of the model

$$r_i = \frac{\epsilon_i}{s(\epsilon_i)} = \frac{\epsilon_i}{\sqrt{MSE(1-h_{ii})}},$$

which depends on the errors-residuals, the mean square error (MSE) and the leverage

$$h_{ii} = [\boldsymbol{H}]_{ii}\,,$$

that is the i-th diagonal element of the projection matrix $\mathbf{H} = \boldsymbol{X}\left(\boldsymbol{X}^\top \boldsymbol{X}\right)^{-1}\mathbf{X}^\top$, where $\boldsymbol{X}$ is the design matrix. To make it clear, first it should be noted that $(\boldsymbol{I} - \boldsymbol{H})$ is idempotent and symmetric. This gives,

$$\begin{aligned} Var(\boldsymbol{\epsilon}) &= Var[(\boldsymbol{I} - \boldsymbol{H})\boldsymbol{Y}] \\ &= (\boldsymbol{I} - \boldsymbol{H})Var(\boldsymbol{Y})(\boldsymbol{I} - \boldsymbol{H})^\top \\ &= \sigma^2(\boldsymbol{I} - \boldsymbol{H})^2 \\ &= \sigma^2(\boldsymbol{I} - \boldsymbol{H}). \end{aligned}$$

Thus, $\mathrm{Var}(\epsilon_i) = (1 - h_{ii})\sigma^2$.

When estimating the parameters of a model and the studentized residual of an observation (or more) takes values greater than 3, then it can be characterized as an outlier observation. Note that we take the absolute value of the studentized residuals.

Instead of using ϵ_t, the Huber function is going to be applied (see Huber, 1981)

$$H_b(X) = X \min\left\{1, \frac{b}{\|X\|}\right\}, \tag{4.7}$$

where $\|\cdot\|$ denotes the Euclidean norm and $0 < b < \infty$. In that case, Step (4.2) of the KF algorithm will be transformed as

$$\hat{\boldsymbol{\theta}}_{t|t} = \hat{\boldsymbol{\theta}}_{t|t-1} + \boldsymbol{P}_{t|t-1}\boldsymbol{F}_t(\boldsymbol{F}_t'\boldsymbol{P}_{t|t-1}\boldsymbol{F}_t + \boldsymbol{R}_t)^{-1}H_b(\boldsymbol{\epsilon}_t), \tag{4.8}$$

where the value of the parameter b must be specified. An approach for the selection of b could be the use of the efficiency that someone wants so as to have robustness in the ideal model compared to the Kalman Filter. This efficiency counts the difference between the robust methods and the Kalman Filter methods. When b takes small values then the relative efficiency takes higher values (Moore and Anderson, 1980).

4.6 Forecasting, Smoothing and Interpolation the Claims

One of the crucial aspects of Solvency II is the estimation of the one year reserve risk. Up until now, actuaries have been concerned about the estimation of the reserve and if it is adequate so as to be covered the ultimate loss. The arrival of Solvency II, made the actuarial companies think the one year perspective. According to the new legislation, the solvency capital requirement (SCR) for unpaid loss is the amount that is sufficient to cover risk over a year. One year reserve risk shows how much the mean estimate of the reserve can be changed during the year. According to the regulations of European Insurance and Occupational Pensions Authority (EIOPA), that risk is calculated either using a Standard Formula or with a specifically approved internal model. Then, we discount the unbiased estimates and we compute the risk margin. According to Solvency II, the one year risk estimates the SCR and the development of the reserve.

The Standard Formula makes the assumption of a lognormal distribution for every line of business. Moreover, each line of business has been characterized by a unique coefficient of variation (CV) that can be applied to all unpaid losses and each year and have been a priori selected by EIOPA. Wüthrich et al. (2008) created a model which computes the variation of the reserves over one year time horizon.

Under Solvency II, during the reserve process it is necessary to estimate the uncertainty which comes from the re-estimation of the best estimate in two consecutive calendar years n and $n + 1$. This means that not only the ultimate payments should be estimated but also the Claims Development Result (CDR) which is an important quantity. Generally, CDR quantifies how much the total reserves will be changed in the next year and depends on D_n and D_{n+1}. In one year the prediction will be updated as

$$\Delta_i^n = \widehat{C}_{i,\infty}^{n-i+1} - \widehat{C}_{i,\infty}^{n-i} = CDR_i(n).$$

If the $CDR_i(n) > 0$, the insurer will pay more than what was expected but if $CDR_i(n) < 0$, the insurer will pay less. In Solvency II, $E[\Delta_i^n | C_{i,n-i}] = 0$, but the actuary must estimate it and create a reserve for studying the uncertainty which can be described by the quantity $Var[\Delta_i^n | C_{i,n-i}]$.

Using the Kalman Filter model, forecasting concerns about the estimation of the parameters $\boldsymbol{\theta}_{t+s}$ or $\boldsymbol{y}_{t+s}$, $s > 0$ given the data $\{\boldsymbol{y}_1, \ldots, \boldsymbol{y}_t\}$, till time t. Filtering concerns about the case of $s = 0$, whilst smoothing deals with the case of $s < 0$ which is not interesting. Using the state equation we have

$$\begin{aligned} \boldsymbol{\theta}_{t+1} &= \boldsymbol{G}_{t+1}\boldsymbol{\theta}_t + \boldsymbol{v}_{t+1}, \\ \boldsymbol{y}_t &= \boldsymbol{F}_t\boldsymbol{\theta}_t + \boldsymbol{w}_t. \end{aligned} \tag{4.9}$$

Then by using (4.9) recursively we obtain

$$\begin{aligned} \boldsymbol{\theta}_{t+s} &= \boldsymbol{G}_{t+s}\boldsymbol{\theta}_{t+s-1} + \boldsymbol{v}_{t+s} \\ &= \boldsymbol{G}_{t+s}\Big(\boldsymbol{G}_{t+s-1}\boldsymbol{\theta}_{t+s-2} + \boldsymbol{v}_{t+s-1}\Big) + \boldsymbol{v}_{t+s} \\ &= \boldsymbol{G}_{t+s}\boldsymbol{G}_{t+s-1}\boldsymbol{\theta}_{t+s-2} + \boldsymbol{G}_{t+s}\boldsymbol{v}_{t+s-1} + \boldsymbol{v}_{t+s} \\ &= \boldsymbol{G}_{t+s}\boldsymbol{G}_{t+s-1}\Big(\boldsymbol{G}_{t+s-2}\boldsymbol{\theta}_{t+s-3} + \boldsymbol{v}_{t+s-2}\Big) + \boldsymbol{G}_{t+s}\boldsymbol{v}_{t+s-1} + \boldsymbol{v}_{t+s} \\ &= \boldsymbol{G}_{t+s}\boldsymbol{G}_{t+s-1}\boldsymbol{G}_{t+s-2}\boldsymbol{\theta}_{t+s-3} + \boldsymbol{G}_{t+s}\boldsymbol{G}_{t+s-1}\boldsymbol{v}_{t+s-2} + \boldsymbol{G}_{t+s}\boldsymbol{v}_{t+s-1} + \boldsymbol{v}_{t+s} \\ &\;\vdots \\ &= \left(\prod_{k=1}^{s}\boldsymbol{G}_{t+k}\right)\boldsymbol{\theta}_t + \left(\prod_{k=2}^{s}\boldsymbol{G}_{t+k}\right)\boldsymbol{v}_{t+1} + \left(\prod_{k=3}^{s}\boldsymbol{G}_{t+k}\right)\boldsymbol{v}_{t+2} + \ldots + \boldsymbol{G}_{t+s}\boldsymbol{v}_{t+s-1} + \boldsymbol{v}_{t+s}. \end{aligned}$$

Then, the forecast for $\boldsymbol{\theta}_{t+s}$ is

$$\hat{\boldsymbol{\theta}}_{t+s} = E\Big(\boldsymbol{\theta}_{t+s}\Big) = \left(\prod_{k=1}^{s}\boldsymbol{G}_{t+k}\right)\hat{\boldsymbol{\theta}}_{t|t}, \tag{4.10}$$

because the mean of $\boldsymbol{v}_k$ is equal to 0 for all k.

Then, using the Step 6 of the Kalman Filter algorithm, recursively we obtain

$$\begin{aligned}
\boldsymbol{P}_{t+s|t} &= \boldsymbol{G}_{t+s}\boldsymbol{P}_{t+s-1|t}\boldsymbol{G}'_{t+s} + \boldsymbol{Q}_{t+s} \\
&= \boldsymbol{G}_{t+s}\Big[\boldsymbol{G}_{t+s-1}\boldsymbol{P}_{t+s-2|t}\boldsymbol{G}'_{t+s-1} + \boldsymbol{Q}_{t+s-1}\Big]\boldsymbol{G}'_{t+s} + \boldsymbol{Q}_{t+s} \\
&= \boldsymbol{G}_{t+s}\boldsymbol{G}_{t+s-1}\boldsymbol{P}_{t+s-2|t}\boldsymbol{G}'_{t+s-1}\boldsymbol{G}'_{t+s} + \boldsymbol{G}_{t+s}\boldsymbol{Q}_{t+s-1}\boldsymbol{G}'_{t+s} + \boldsymbol{Q}_{t+s} \\
&\;\vdots \\
&= \left(\prod_{k=1}^{s}\boldsymbol{G}_{t+k}\right)\boldsymbol{P}_{t|t}\left(\prod_{k=1}^{s}\boldsymbol{G}'_{t+k}\right) + \left(\prod_{k=2}^{s}\boldsymbol{G}_{t+k}\right)\boldsymbol{Q}_{t+1}\left(\prod_{k=2}^{s}\boldsymbol{G}'_{t+k}\right) \\
&+ \left(\prod_{k=3}^{s}\boldsymbol{G}_{t+k}\right)\boldsymbol{Q}_{t+2}\left(\prod_{k=3}^{s}\boldsymbol{G}'_{t+k}\right) + \ldots + \boldsymbol{G}_{t+s}\boldsymbol{Q}_{t+s-1}\boldsymbol{G}'_{t+s} + \boldsymbol{Q}_{t+s}.
\end{aligned}$$

The last expression can be written in a recursive way as

$$\boldsymbol{P}_{t+k|t} = \boldsymbol{G}_{t+k}\boldsymbol{P}_{t+k-1|t}\boldsymbol{G}'_{t+k} + \boldsymbol{Q}_{t+k},$$

starting with $k = 1$ and estimate the whole sequence recursively up to $k = s$.

Then, the best prediction of $\boldsymbol{y}_{t+s}$ is

$$\hat{\boldsymbol{y}}_{t+s} = \boldsymbol{F}_{t+s}\hat{\boldsymbol{\theta}}_{t+s|t}$$

and the corresponding variance of the prediction is

$$\begin{aligned}
Var[\boldsymbol{y}_{t+s} - \hat{\boldsymbol{y}}_{t+s}] &= Var[\boldsymbol{F}_{t+s}(\boldsymbol{\theta}_{t+s} - \hat{\boldsymbol{\theta}}_{t+s|t}) + \boldsymbol{w}_{t+s}] \\
&= \boldsymbol{F}_{t+s}\boldsymbol{P}_{t+s|t}\boldsymbol{F}'_{t+s} + \boldsymbol{R}_{t+s}.
\end{aligned}$$

Remark 4.5. There are many cases where the data set has missing observations. In that case, the unobserved data (unknown claims) can be replaced by interpolating them easily. Then, Steps 3 and 5 of the KF algorithm are replaced by the equations (Maronna et al., 2006)

$$\begin{aligned}
\hat{\boldsymbol{\theta}}_{t|t} &= \hat{\boldsymbol{\theta}}_{t|t-1}, \\
\boldsymbol{P}_{t|t} &= \boldsymbol{P}_{t|t-1}.
\end{aligned}$$

because there is no available information to update the state-vector. If there are more than one unobserved values, this procedure can be applied many times. If for example the values $y_t, \ldots, y_{t+m}$ are missing, then we apply the procedure $m+1$ times and the interpolated data are estimated by

$$\boldsymbol{F}_t\hat{\boldsymbol{\theta}}_{t+s|t-1} = \boldsymbol{F}_t\Big(\prod_{m=0}^{s}\boldsymbol{G}_{t+m}\Big)\hat{\boldsymbol{\theta}}_{t-1|t-1},\; s = 0, 1, \ldots, m.$$

4.7 Construction of a Recursive Model on Run-Off Triangles

In order to transform a run-off triangle into a linear model, the logarithm of the triangle data $Y_{i,j}$ are used. Then,

$$\log(Y_{i,j}) = \mu + \alpha_i + \beta_j + e_{i,j}, \tag{4.11}$$

where $\mu > 0$ is the constant of the model, α_i is a parameter which describes the accident year i and β_j is a parameter which describes the development year j of the triangle. Moreover, $e_{i,j}$ are the errors of the model which are normally distributed with mean 0 and variance $\sigma^2 > 0$ (De Jong and Zehnwirth, 1983).

For the construction of the run-off triangle as a state space model, a two way analysis of variance model should be set up in recursive form (see Verrall, 1994). The data are received according to the year of payment

$$Y_{1,1}, \begin{bmatrix} Y_{1,2} \\ Y_{2,1} \end{bmatrix}, \begin{bmatrix} Y_{1,3} \\ Y_{2,2} \\ Y_{3,1} \end{bmatrix}, \ldots$$

and finally in year t the data take the form $\begin{bmatrix} Y_{1,t} & Y_{2,t-1} & \cdots & Y_{t,1} \end{bmatrix}'$.

According to Verrall (1989), the state space depiction of a log-normal model has the observation equation

$$\underbrace{\begin{bmatrix} \log(Y_{1,t}) \\ \log(Y_{2,t-1}) \\ \vdots \\ \log(Y_{t,1}) \end{bmatrix}}_{\substack{\text{observation} \\ \text{vector}}} = \underbrace{\begin{bmatrix} 1 & 0 & \ldots & & & & & & & & \ldots & 0 & 1 \\ 1 & 1 & 0 & \ldots & & & & & \ldots & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & \ldots & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ \vdots & & & & & \ddots & & & & & & & \vdots \\ \vdots & & & & & \ddots & & & & & & & \vdots \\ 1 & 0 & 1 & 0 & \ldots & & \ldots & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & \ldots & & & & & & & \ldots & 0 & 1 & 0 \end{bmatrix}}_{\text{system matrix}} \underbrace{\begin{bmatrix} \mu \\ \alpha_1 \\ \beta_1 \\ \vdots \\ \alpha_t \\ \beta_t \end{bmatrix}}_{\substack{\text{state} \\ \text{vector} \\ \text{in t}}} + \underbrace{\begin{bmatrix} w_{1,t} \\ w_{2,t-1} \\ \vdots \\ w_{t,1} \end{bmatrix}}_{\substack{\text{measurement} \\ \text{noise} \\ \text{vector}}}, \tag{4.12}$$

for calendar year $t = 0, \cdots, I$ which implies (4.11) for each $Y_{i,j}$ with $t = i + j$, while the state equation is

$$\underbrace{\begin{bmatrix} \mu \\ \alpha_1 \\ \beta_1 \\ \vdots \\ \alpha_{t+1} \\ \beta_{t+1} \end{bmatrix}}_{\substack{\text{state}\\\text{vector}\\\text{in t+1}}} = \underbrace{\begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & \ddots & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \dots & 0 & 1 \\ 0 & \dots & 1 & 0 \\ 0 & \dots & 0 & 1 \end{bmatrix}}_{\text{transision matrix}} \underbrace{\begin{bmatrix} \mu \\ \alpha_1 \\ \beta_1 \\ \vdots \\ \alpha_t \\ \beta_t \end{bmatrix}}_{\substack{\text{state}\\\text{vector}\\\text{in t}}} + \underbrace{\begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ u_t^{(\alpha)} \\ u_t^{(\beta)} \end{bmatrix}}_{\substack{\text{process}\\\text{noise}\\\text{vector}\\\text{in t}}} \tag{4.13}$$

permits the estimation of the accident and development year parameters. According to this notation, the set $\boldsymbol{Y}_1, \boldsymbol{Y}_2, \dots, \boldsymbol{Y}_t$ of vectors, fulfill the run-off triangle. The dimension of each vector $\boldsymbol{Y}_t$ in this time series model is t which means that the analysis of the data set could be based on a multivariate time series model. For example, a linear model with $n = 3$ accident years has the form

$$\begin{bmatrix} Y_{1,1} \\ Y_{1,2} \\ Y_{2,1} \\ Y_{1,3} \\ Y_{2,2} \\ Y_{3,1} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \mu \\ \alpha_2 \\ \beta_2 \\ \alpha_3 \\ \beta_3 \end{bmatrix} + \begin{bmatrix} e_{1,1} \\ e_{1,2} \\ e_{2,1} \\ e_{1,3} \\ e_{2,2} \\ e_{3,1} \end{bmatrix},$$

where $\mu, \alpha_2, \beta_2, \alpha_3, \beta_3$ are the unknown regression coefficients and $e_{i,j}$ are the nuisance parameters.

In state space form, the observation vector is defined by

$$Y_{1,1} = \mu + e_{1,1},$$

$$\begin{bmatrix} Y_{1,2} \\ Y_{2,1} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} \mu \\ \alpha_2 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} e_{1,2} \\ e_{2,1} \end{bmatrix},$$

$$\begin{bmatrix} Y_{1,3} \\ Y_{2,2} \\ Y_{3,1} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \mu \\ \alpha_2 \\ \beta_2 \\ \alpha_3 \\ \beta_3 \end{bmatrix} + \begin{bmatrix} e_{1,3} \\ e_{2,2} \\ e_{3,1} \end{bmatrix}.$$

At time t the state vector is defined by

$$\boldsymbol{\theta}_t = \begin{bmatrix} \mu \\ \alpha_2 \\ \beta_2 \\ \vdots \\ \alpha_t \\ \beta_t \end{bmatrix}. \tag{4.14}$$

But, the state vector $\boldsymbol{\theta}_t$ at time t is related to the state vector $\boldsymbol{\theta}_{t-1}$ at time $t-1$ by the following recursive equation

$$\boldsymbol{\theta}_{t+1} = \begin{bmatrix} 1 & & & & \\ & 1 & & & \\ & & 1 & & \\ & & & \ddots & \\ & & & & 1 \\ 0 & 0 & \dots & \dots & 0 \\ 0 & 0 & \dots & \dots & 0 \end{bmatrix} \boldsymbol{\theta}_t + \begin{bmatrix} 0 & 0 \\ \vdots & \vdots \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \boldsymbol{v}_t,$$

where $\boldsymbol{v}_t$ is a vector of the prior information about $\begin{bmatrix} \alpha_{t+1} \\ \beta_{t+1} \end{bmatrix}$.

Remark 4.6. In cases that the variance of the nuisance parameters, e_{ij}, is known and vague priors for the unknown parameters are being used, then the estimates of the parameters of this method are exactly the same as in the ordinary least-squares estimation method (Verrall, 1989).

The general form which make up the run-off triangle consists of the following equations:

$$\begin{aligned} \boldsymbol{Y}_1 &= \boldsymbol{F}_1\boldsymbol{\theta}_1 + \boldsymbol{e}_1 \\ \boldsymbol{Y}_2 &= \boldsymbol{F}_2\boldsymbol{\theta}_2 + \boldsymbol{e}_2 \\ &\vdots \\ \boldsymbol{Y}_t &= \boldsymbol{F}_t\boldsymbol{\theta}_t + \boldsymbol{e}_t \end{aligned} \tag{4.15}$$

Using the equation (4.15), the Kalman filter method can be applied so as to estimate the unknown parameters. The state space vector $\boldsymbol{\theta}_t$ at time t is linked with the state vector $\boldsymbol{\theta}_{t-1}$ at time $t-1$ and simultaneously with the observation equation constitute the state space model.

4.8 Application of Kalman Filter to Loss Reserving

In the first step, we apply the linear regression Kalman Filter model (KF). Afterwards, we suppose that the parameters are not fixed and we produce the corresponding Random linear regression Kalman Filter model (RKF) to Taylor and Ashe (1983) data set (for more details see Moryson, 1998, Petris et al., 2010). We suppose that $\boldsymbol{Q}_t \neq 0$ and we allow the algorithm to estimate the variance. This means that the parameters of the model are not constant but they have non-zero variance. Figures 4.3 and 4.4 show the development of the first 9 coefficients (excluding the intercept) of KF model and the development of the coefficients of RKF model. As we can observe from these Figures, the estimation of some of the coefficients begins after some data are inputted into the model. This is also the key feature of Kalman Filter's method and explains the concept of its dynamic estimation.

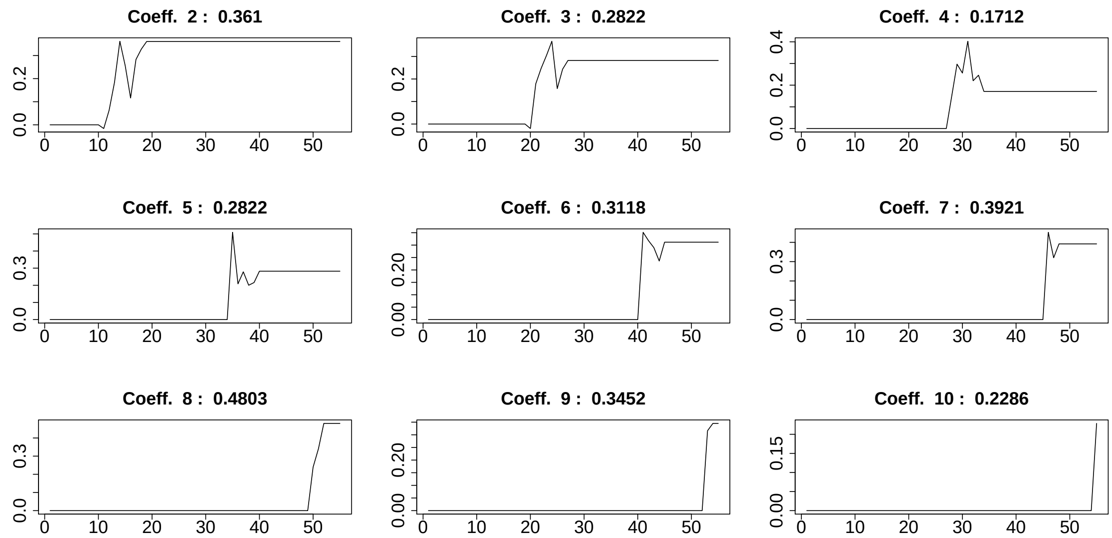

Figure 4.3: Coefficients of the Linear Regression Kalman Filter Model

Comparing the KF and RKF models, it is obvious that a similar pattern is observed. Nevertheless, the estimation of the coefficients is different and the standard errors are also different.

Table 4.1 shows the estimates of the coefficients of the non random and random Kalman Filter model (they are compared with the linear regression model). As we can see, the estimates of the Kalman Filter algorithm have slight differences with the linear regression model.

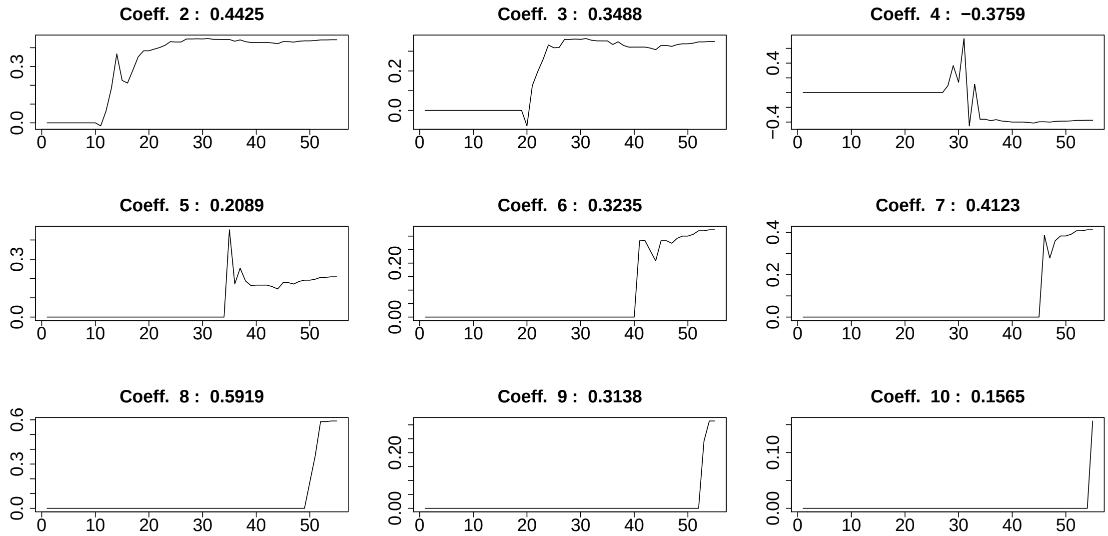

Figure 4.4: Coefficients of the Linear Regression Random Kalman Filter Model

Table 4.1: Coefficients of KF, RKF and LS Models

	Kalman Filter		**LS**
Coefficient	**Non-random**	**Random**	
Intercept	12.519837	12.591888	12.519840
α_2	0.361003	0.442485	0.361002
α_3	0.282241	0.348775	0.282240
α_4	0.171195	-0.375912	0.171194
α_5	0.282223	0.208921	0.282222
α_6	0.311750	0.323464	0.311749
α_7	0.392050	0.412289	0.392049
α_8	0.480272	0.591932	0.480270
α_9	0.345165	0.313837	0.345163
α_{10}	0.228600	0.156549	0.228598
β_2	0.911191	0.856685	0.911190
β_3	0.938721	0.901726	0.938720
β_4	0.964982	0.870378	0.964981
β_5	0.383203	0.466211	0.383202
β_6	-0.004908	0.316003	-0.004909
β_7	-0.118068	0.036046	-0.118069
β_8	-0.439275	-0.460489	-0.439277
β_9	-0.053505	-0.012819	-0.053507
β_{10}	-1.393338	-1.269632	-1.393342

Table 4.2 shows the reserve estimates for each cell of the run-off triangle. The standard error (SE) of these estimates is also provided (for more details of SE, see Section 2.3.1).

Table 4.2: Reserve Estimations and Standard Error from Kalman Filter Method

	KF Linear Model		**KF Random Linear Model**	
Cell	Estimated Reserves	Standard Error	Estimated Reserves	Standard Error
$x_{10,2}$	973602	671916	1004889	755891
$x_{9,3}$	1090101	729299	1168135	834331
$x_{10,3}$	1001990	692345	1053309	793914
$x_{8,4}$	1268870	840863	1471524	1034490
$x_{9,4}$	1120657	750780	1134698	812320
$x_{10,4}$	1030077	712737	1023159	772969
$x_{7,5}$	646825	427014	815581	569733
$x_{8,5}$	710403	471589	985126	694549
$x_{9,5}$	627423	421067	759635	545386
$x_{10,5}$	576709	399731	684964	518966
$x_{6,6}$	404479	266741	641040	447011
$x_{7,6}$	439787	291009	704555	494086
$x_{8,6}$	483014	321387	851020	602329
$x_{9,6}$	426595	286956	656225	472971
$x_{10,6}$	392114	272416	591719	450059
$x_{5,7}$	351068	231764	432840	302365
$x_{6,7}$	362429	239822	487254	341698
$x_{7,7}$	394066	261641	535532	377682
$x_{8,7}$	432800	288953	646860	460424
$x_{9,7}$	382246	257997	498796	361542
$x_{10,7}$	351350	244924	449765	344028
$x_{4,8}$	228731	151577	147724	103851
$x_{5,8}$	256033	169963	265884	187457
$x_{6,8}$	264319	175872	299309	211843
$x_{7,8}$	287392	191873	328965	234152
$x_{8,8}$	315640	211902	397351	285449
$x_{9,8}$	278771	189201	306399	224146
$x_{10,8}$	256238	179614	276281	213287
$x_{3,9}$	379508	253898	484712	346202
$x_{4,9}$	340091	227843	235374	168502
$x_{5,9}$	380684	255479	423643	304158
$x_{6,9}$	393005	264361	476901	343724
$x_{7,9}$	427311	288414	524154	379922
$x_{8,9}$	469312	318520	633116	463155
$x_{9,9}$	414493	284397	488198	363686
$x_{10,9}$	380990	269986	440209	346068
$x_{2,10}$	110928	76555	159533	120003
$x_{3,10}$	102650	70928	145556	109710
$x_{4,10}$	91988	63649	70681	53398
$x_{5,10}$	102968	71370	127217	96387
$x_{6,10}$	106301	73851	143210	108925
$x_{7,10}$	115580	80570	157400	120396
$x_{8,10}$	126940	88981	190120	146772
$x_{9,10}$	112113	79448	146602	115251
$x_{10,10}$	103051	75422	132192	109668
Total	**19511642**	**13194628**	**23597356**	**17142859**
% SE of Reserves	**32.11%**		**36.70%**	

It can be observed from Table 4.2 that the RKF model produces greater total reserves (23597356) than the KF model (19511642). Similar results we obtain for the standard errors (SE) as displayed in Table 4.2. For the KF model the total SE is 13194628 (32.11% of the total reserves), while for the RKF model the total SE is 17142859 (36.70% of the total reserves). The smaller the coefficient of variation is, the better model is, which means that KF model is better than the RKF model.

The run-off triangle for the KF model with the estimations of the future claims and the total reserve is shown at Table 4.3. The corresponding run-off triangle for the RKF model is shown at Table 4.4.

Table 4.3: Run-off Triangle of KF Model and Reserve Estimation

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**Ult. Claims**	**Reserve**
1	357848	1124788	1735330	2218270	2745596	3319994	3466336	3606286	3833515	3901463	3901463	0
2	352118	1236139	2170033	3353322	3799067	4120063	4647867	4914039	5339085	**5450013**	5450013	110928
3	290507	1292306	2218525	3235179	3985995	4132918	4628910	4909315	**5288823**	**5391473**	5391473	482158
4	310608	1418858	2195047	3757447	4029929	4381982	4588268	**4816999**	**5157090**	**5249079**	5249079	660811
5	443160	1136350	2128333	2897821	3402672	3873311	**4224379**	**4480411**	**4861096**	**4964064**	4964064	1090753
6	396132	1333217	2180715	2985752	3691712	**4096191**	**4458621**	**4722940**	**5115945**	**5222245**	5222245	1530533
7	440832	1288463	2419861	3483130	**4129955**	**4569742**	**4963809**	**5251200**	**5678511**	**5794091**	5794091	2310961
8	359480	1421128	2864498	**4133368**	**4843771**	**5326785**	**5759584**	**6075224**	**6544536**	**6671476**	6671476	3806978
9	376686	1363294	**2453395**	**3574052**	**4201475**	**4628070**	**5010315**	**5289086**	**5703579**	**5815692**	5815692	4452398
10	344014	**1317616**	**2319606**	**3349683**	**3926393**	**4318507**	**4669856**	**4926095**	**5307085**	**5410136**	5410136	5066122
Total											**53869732**	**19511642**

Table 4.4: Run-off Triangle of RKF Model and Reserve Estimation

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**Ult. Claims**	**Reserve**
1	357848	1124788	1735330	2218270	2745596	3319994	3466336	3606286	3833515	3901463	3901463	0
2	352118	1236139	2170033	3353322	3799067	4120063	4647867	4914039	5339085	**5498618**	5498618	159533
3	290507	1292306	2218525	3235179	3985995	4132918	4628910	4909315	**5394027**	**5539583**	5539583	630268
4	310608	1418858	2195047	3757447	4029929	4381982	4588268	**4735992**	**4971366**	**5042047**	5042047	453779
5	443160	1136350	2128333	2897821	3402672	3873311	**4306151**	**4572034**	**4995677**	**5122894**	5122894	1249583
6	396132	1333217	2180715	2985752	3691712	**4332752**	**4820006**	**5119315**	**5596216**	**5739426**	5739426	2047714
7	440832	1288463	2419861	3483130	**4298711**	**5003267**	**5538798**	**5867764**	**6391917**	**6549317**	6549317	3066187
8	359480	1421128	2864498	**4336022**	**5321148**	**6172168**	**6819028**	**7216380**	**7849496**	**8039616**	8039616	5175118
9	376686	1363294	**2531429**	**3666126**	**4425761**	**5081987**	**5580783**	**5887182**	**6375380**	**6521983**	6521983	5158689
10	344014	**1348903**	**2402212**	**3425371**	**4110334**	**4702054**	**5151819**	**5428099**	**5868309**	**6000500**	6000500	5656486
Total											**57955446**	**23597356**

Figure 4.5 shows the total reserves using the KF method and the RKF method for each accident year. According to this figure, it is observed that the RKF algorithm gives bigger values for the reserve especially at the last accident years. The reason for this difference could be the fact that there is no much data information at the last accident years which may lead to greater variability. On the other hand, for the first 5 accident years, the reserves for both methods are equivalent. Note that, from Figure 4.5 we also observe a peak at the 8^{th} accident year. The RKF method identifies this sharp increase of the reserve this accident year while line the KF models seems more stable.

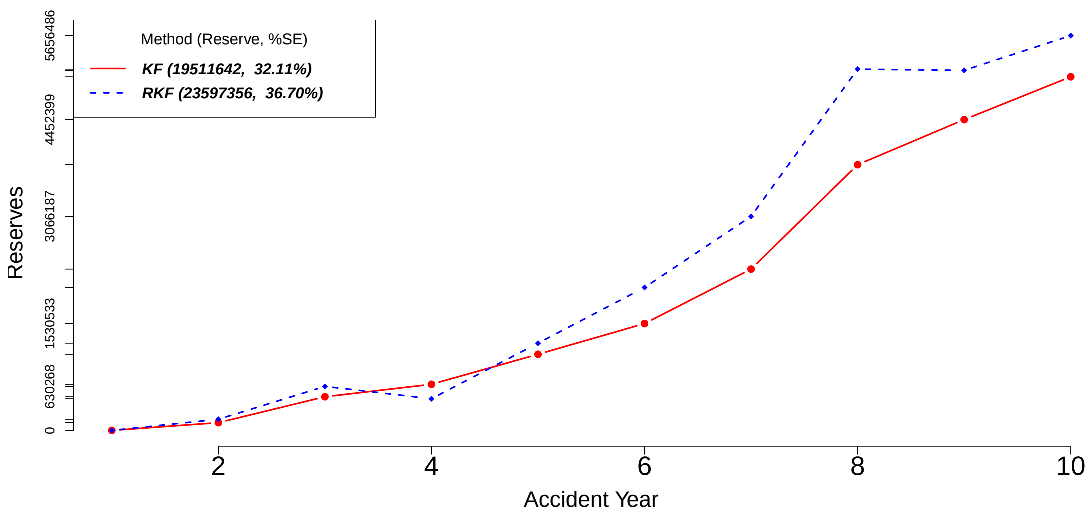

Figure 4.5: Reserves by Accident Year with KF and RKF Methods

Finally, we apply the robust function Kalman Filter model to the Taylor and Ashe triangle. The coefficients of the model for different values of b are shown at Table 4.5. It is observed that as the parameter b increases, the coefficients are stabilized. The coefficients for big values of b are identical to the values of the parameters of the non robust Kalman Filter. For values $b \geq 12.8$, the coefficients are equal which means that the total reserve does not change for $b \geq 12.8$.

Table 4.5: Robust KF Coefficients for Different Values of b

b	**10**	**10.2**	**10.4**	**10.6**	**10.8**	**11**	**11.2**	**11.4**	**11.6**	**11.8**	**12**	**12.2**	**12.4**	**12.6**	**12.8**	**13**
Intercept	11.870	11.916	11.963	12.010	12.056	12.103	12.149	12.196	12.243	12.289	12.336	12.383	12.429	12.476	12.520	12.520
α_2	0.671	0.649	0.626	0.604	0.582	0.560	0.537	0.515	0.493	0.471	0.449	0.426	0.404	0.382	0.361	0.361
α_3	0.611	0.588	0.564	0.541	0.517	0.493	0.470	0.446	0.422	0.399	0.375	0.352	0.328	0.304	0.282	0.282
α_4	0.517	0.492	0.467	0.443	0.418	0.393	0.368	0.343	0.318	0.294	0.269	0.244	0.219	0.194	0.171	0.171
α_5	0.645	0.619	0.593	0.567	0.541	0.515	0.489	0.463	0.437	0.411	0.385	0.359	0.333	0.307	0.282	0.282
α_6	0.693	0.665	0.638	0.611	0.583	0.556	0.529	0.501	0.474	0.447	0.419	0.392	0.365	0.337	0.312	0.312
α_7	0.796	0.767	0.738	0.709	0.680	0.651	0.622	0.593	0.564	0.535	0.506	0.477	0.448	0.419	0.392	0.392
α_8	0.918	0.886	0.855	0.823	0.792	0.761	0.729	0.698	0.667	0.635	0.604	0.572	0.541	0.510	0.480	0.480
α_9	0.841	0.805	0.769	0.734	0.698	0.663	0.627	0.592	0.556	0.521	0.485	0.450	0.414	0.379	0.345	0.345
α_{10}	0.879	0.832	0.786	0.739	0.692	0.646	0.599	0.552	0.506	0.459	0.412	0.366	0.319	0.272	0.229	0.229
β_2	1.221	1.199	1.177	1.154	1.132	1.110	1.088	1.065	1.043	1.021	0.999	0.977	0.954	0.932	0.911	0.911
β_3	1.268	1.244	1.221	1.197	1.173	1.150	1.126	1.103	1.079	1.055	1.032	1.008	0.985	0.961	0.939	0.939
β_4	1.311	1.286	1.261	1.236	1.211	1.187	1.162	1.137	1.112	1.087	1.063	1.038	1.013	0.988	0.965	0.965
β_5	0.746	0.720	0.694	0.668	0.642	0.616	0.590	0.564	0.538	0.512	0.486	0.460	0.434	0.408	0.383	0.383
β_6	0.376	0.349	0.321	0.294	0.267	0.239	0.212	0.185	0.157	0.130	0.103	0.075	0.048	0.021	-0.005	-0.005
β_7	0.286	0.257	0.228	0.199	0.170	0.141	0.112	0.083	0.054	0.025	-0.004	-0.033	-0.062	-0.091	-0.118	-0.118
β_8	-0.002	-0.033	-0.065	-0.096	-0.127	-0.159	-0.190	-0.222	-0.253	-0.284	-0.316	-0.347	-0.378	-0.410	-0.439	-0.439
β_9	0.442	0.406	0.371	0.335	0.300	0.264	0.229	0.193	0.158	0.122	0.086	0.051	0.015	-0.020	-0.054	-0.054
β_{10}	-0.743	-0.790	-0.836	-0.883	-0.930	-0.976	-1.023	-1.070	-1.116	-1.163	-1.210	-1.256	-1.303	-1.350	-1.393	-1.393

The triangle with the the reserves, when the robust Kalman Filter algorithm with $b = 10$ is applied, are presented at Table 4.6. The corresponding triangle for $b = 12.8$ is also presented (see Table 4.7).

Table 4.6: Run-off Triangle of Robust KF Model ($b = 10$) and Reserve Estimation

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**Ult. Claims**	**Reserve**
1	357848	1124788	1735330	2218270	2745596	3319994	3466336	3606286	3833515	3901463	3901463	0
2	352118	1236139	2170033	3353322	3799067	4120063	4647867	4914039	5339085	**5498992**	5498992	159907
3	290507	1292306	2218525	3235179	3985995	4132918	4628910	4909315	**5380667**	**5531613**	5531613	622298
4	310608	1418858	2195047	3757447	4029929	4381982	4588268	**4859689**	**5289410**	**5427025**	5427025	838757
5	443160	1136350	2128333	2897821	3402672	3873311	**4282370**	**4591503**	**5080931**	**5237666**	5237666	1364355
6	396132	1333217	2180715	2985752	3691712	**4160604**	**4591257**	**4916709**	**5431974**	**5596983**	5596983	1905271
7	440832	1288463	2419861	3483130	**4236801**	**4759373**	**5239328**	**5602039**	**6176293**	**6360193**	6360193	2877063
8	359480	1421128	2864498	**4370185**	**5227922**	**5822651**	**6368878**	**6781671**	**7435219**	**7644511**	7644511	4780013
9	376686	1363294	**2717209**	**4133213**	**4939862**	**5499166**	**6012859**	**6401066**	**7015686**	**7212512**	7212512	5849218
10	344014	**1747503**	**3220923**	**4761913**	**5639762**	**6248435**	**6807469**	**7229941**	**7898812**	**8113012**	8113012	7768998
Total											**60523970**	**26165880**

Table 4.7: Run-off Triangle of Robust KF Model ($b = 12.8$) and Reserve Estimation

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**Ult. Claims**	**Reserve**
1	357848	1124788	1735330	2218270	2745596	3319994	3466336	3606286	3833515	3901463	3901463	0
2	352118	1236139	2170033	3353322	3799067	4120063	4647867	4914039	5339085	**5450013**	5450013	110928
3	290507	1292306	2218525	3235179	3985995	4132918	4628910	4909315	**5288823**	**5391473**	5391473	482158
4	310608	1418858	2195047	3757447	4029929	4381982	4588268	**4816999**	**5157090**	**5249079**	5249079	660811
5	443160	1136350	2128333	2897821	3402672	3873311	**4224379**	**4480411**	**4861096**	**4964064**	4964064	1090753
6	396132	1333217	2180715	2985752	3691712	**4096191**	**4458621**	**4722940**	**5115945**	**5222245**	5222245	1530533
7	440832	1288463	2419861	3483130	**4129955**	**4569742**	**4963809**	**5251200**	**5678511**	**5794091**	5794091	2310961
8	359480	1421128	2864498	**4133368**	**4843771**	**5326785**	**5759584**	**6075224**	**6544536**	**6671476**	6671476	3806978
9	376686	1363294	**2453395**	**3574052**	**4201475**	**4628070**	**5010315**	**5289086**	**5703579**	**5815692**	5815692	4452398
10	344014	**1317616**	**2319606**	**3349683**	**3926393**	**4318507**	**4669856**	**4926095**	**5307085**	**5410136**	5410136	5066122
Total											**53869732**	**19511642**

Table 4.8: Robust KF for Different Values of b

b	Ult. Claims	Total Reserves	σ^2	Efficiency	%SE
10.00	60523969	26165879	0.17	1.43	0.36
10.20	59885962	25527872	0.16	1.37	0.36
10.40	59277788	24919698	0.15	1.32	0.35
10.60	58697973	24339883	0.15	1.27	0.35
10.80	58145135	23787045	0.14	1.22	0.35
11.00	57617971	23259881	0.14	1.18	0.34
11.20	57115261	22757171	0.13	1.14	0.34
11.40	56635855	22277765	0.13	1.11	0.33
11.60	56178672	21820582	0.13	1.08	0.33
11.80	55742696	21384606	0.12	1.05	0.33
12.00	55326970	20968880	0.12	1.03	0.33
12.20	54930596	20572506	0.12	1.02	0.32
12.40	54552725	20194635	0.12	1.01	0.32
12.60	54192561	19834471	0.12	1.00	0.32
12.80	53869732	19511642	0.12	1.00	0.32
13.00	53869732	19511642	0.12	1.00	0.32

Table 4.8, shows for different values of the parameter b, the ultimate claims, the total reserve, the estimation of σ^2, the efficiency (compared to the non robust KF model) and the %SE. It is observed that when the parameter b increases, the total reserve decreases while the efficiency tends to 1.

Figure 4.6 shows the evolution of the reserve when robust Kalman Filter is applied to the data (see Table 4.8). This Figure shows also the %SE of the total reserve. We observe that, when the parameter b increases, the ultimate reserve and the %SE decrease.

Remark 4.7. For values of $b \geq 12.80$, with $\sigma^2 = 0.12$ and efficiency$= 1.00$, the values of total reverses are all equal to 19511642. This is due to the fact that the increase of the parameter b creates more tolerance in the error values of the model and thus the errors are not considered as outliers. We also observe that this value is not very far from the value of the total reserves under the log-linear regression model as was presented in Table 2.2 in Section 2.

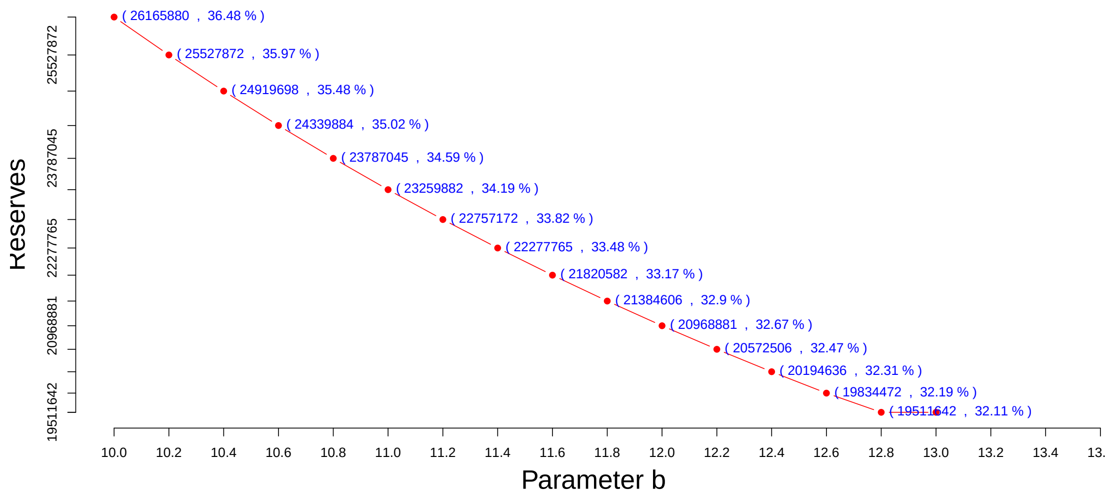

Figure 4.6: Ultimate Reserves and %SE for different robust values of the parameter b

4.9 Concluding Remarks

Even if the state space models are complex and take long time to develop, they can be used as a framework for automating the reserving procedure. They can incorporate subjective expert judgement and data from any relevant source and they also allow the parameters to be dynamic. States and predictions have probabilistic representation which makes the state space models useful of quantifying the uncertainty of the reserves. In contrast with other loss reserving techniques, the state space models provide future estimates for more than one calendar years and this is very important in the reserving process. This loss reserving method uses the Kalman Filter, an algorithm applied to systems that receive external physical disturbances (noises) and aims to create a new estimate of the state of the system without disturbances. In addition, the algorithm is extended by making it robust to extreme values, which, if ignored, will overestimate the final reserve required for the company's liability.

CHAPTER 5

Loss Reserving in a Quantile Regression Model

The protection of the policyholders and the financial stability of the insurance market industry is a crucial aspect and the regulatory authorities intervene to ensure it. Based on Solvency II and IFRS Phase II regulations, each insurance or reinsurance company is obliged to evaluate their insurance liabilities on a risk-adjusted basis to allow for uncertainty in cash flows that arise from liability of insurance contracts. In addition to the calculation of the reserve risk margin, the calculation of the confidence level of the risk margin is also required. Australian Prudential Regulation Authority (APRA) requires to estimate a 75^{th} percentile of the distribution of outstanding claims for recording in profit and loss statements and the risk margin should be established on a basis that is intended to secure the insurance liabilities of the insurance company at a given level of sufficiency (75%).

In recent years quantile regression has become a very popular methodology that satisfies several of the new reforms in insurance and finance. The least squares estimators investigate only changes in the mean when the entire shape of the claims distribution may change dramatically. Quantile regression characterizes a particular point of a distribution and thus provides a more complete description of the distribution in comparison to linear regression. Quantile regression technique can differentiate risk factors that lead to high-level claims than those lead to low-level claims. Quantile regression estimation may be more efficient from the ordinary least squares when the distribution is not normal. Furthermore, quantile regression is more robust against outliers and does not require specifying any error distribution. Therefore, quantile regression may be more appropriate than least squares estimation in the context of the insurance industry (see Buchinsky, 1998 and Koenker, 2005).

In the actuarial literature, few papers involved with quantiles Pitt (2006) used censored regression quantiles to analyze claim termination rates at different quantiles of the distribution of claim duration for income protection insurance. Chan (2015) proposed a quantile regression loss reserving model as the model offers potentially different solutions at distinct quantiles so that the effects of risk factors are differentiated at different points of the conditional loss distribution. Chan et al. (2007) proposed a robust Bayesian analysis of loss reserves data using the generalized -t distribution. Dong et al. (2015) presented in detail the use of parametric and nonparametric quantile regression in non-life applications. One of their contributions is the use of quantile regression for loss reserving. They have shown how one can provide an accurate estimation of risk margin and hence provision, instead of estimating the mean then applying a risk margin. Their method is more robust when the data is heavy-tailed. Nevertheless, the above approached which have been used are for univariate quantile regression models and are suitable for a simple line of business (one run-off triangle). As pointed out by Ajne (1994) and many other researches, when dealing with a portfolio of several lines of business (LOB), the chain ladder predictors for the whole portfolio differ from the sums of the chain ladder predictors for the different individual LOB, because the dependence structure between the sub-portfolios of a portfolio is not taking in to consideration.

In this chapter, we consider a quantile regression application in a multivariate context alternative to a multivariate chain-ladder model for a portfolio of correlated run-off triangles. We propose a reserving problem for a non-life insurance portfolio consisting of several run-off sub-portfolios corresponding to a different line of business that can be embedded within the quantile regression model for longitudinal data. One implication of our model is the diversification effect of a portfolio of reserve risks and can be used as a risk measure with applications in finance and actuarial science. Antonio et al. (2006) by using the theory of linear mixed model built a flexible loss reserving in the framework longitudinal data.

During the last years, many attempts were actualized to extend these methods but that all was done for the situation where one has only one run-off-triangle. Zehnwirth et al., (2001) was considered a more complicated situation where n correlated loss triangles were observed. Our proposal is to apply quantile regression procedures to the loss reserving estimation when multiple lines of business are available incorporating the correlation which exists between them.

5.1 Quantile Regression embedded in a Loss Reserving Model

5.1.1 Quantile Function

For a random variable Y with cumulative distribution function $F_Y(y) = P(Y \leq y)$, the θ^{th} quantile of Y is defined as the inverse function

$$Q_Y(\theta) = F_Y^{-1}(\theta) = \inf\{y : F(y) \geq \theta\}, \tag{5.1}$$

where $0 \leq \theta \leq 1$. In case that $F(\cdot)$ is a strictly increasing and continuous probability distribution function, then $F_Y^{-1}(\theta)$ is the unique real number t such that $F(t) = \theta$ (Gilchrist, 2000).

Quantiles are connected with operations of ordering the sample observations that are used to define them. For a random sample $\{y_1, \ldots, y_n\}$ of Y, the general θ^{th} sample quantile $\xi(\theta)$ may be formulated as the solution of the optimization problem

$$\min_{\xi \in \Re} \sum_{i=1}^{n} \rho_\theta(y_i - \xi), \quad \text{where} \quad \rho_\theta(z) = z(\theta - I(z < 0)) \tag{5.2}$$

and $I(\cdot)$ denotes the indicator function. This loss function is an asymmetric absolute loss function because it is a weighted sum of absolute deviations, where the weight $(1 - \theta)$ is assigned to the negative deviations while the weight θ is assigned to the positive deviations.

When Y is discrete with probability distribution function $f(y) = P(Y = y)$, the minimization problem takes the form:

$$\begin{aligned} Q_Y(\theta) &= \underset{c}{\operatorname{argmin}} E[\rho_\theta(Y - c)] \\ &= \underset{c}{\operatorname{argmin}} \left\{ (1 - \theta) \sum_{y \leq c} |y - c| f(y) + \theta \sum_{y > c} |y - c| f(y) \right\}. \end{aligned} \tag{5.3}$$

In the case of a continuous random variable Y, we substitute the summation with integrals at 5.3:

$$\begin{aligned} Q_Y(\theta) &= \underset{c}{\operatorname{argmin}} E[\rho_\theta(Y - c)] \\ &= \underset{c}{\operatorname{argmin}} \left\{ (1 - \theta) \int_{-\infty}^{c} |y - c| f(y) dy + \theta \int_{c}^{\infty} |y - c| f(y) dy \right\}, \end{aligned} \tag{5.4}$$

where $f(y)$ is the probability density function of the random variable Y.

5.1.2 Quantile Regression Estimation

Just as the sample mean, which minimizes the sum of squared residuals

$$\hat{\mu} = \underset{\mu \in \Re}{\operatorname{argmin}} \sum_{i=1}^{n} (y_i - \mu)^2,$$

can be extended to the linear conditional mean function $E(Y|\boldsymbol{X} = \boldsymbol{x}) = \boldsymbol{x}'\beta$ by solving

$$\hat{\beta} = \underset{\beta \in \Re^p}{\operatorname{argmin}} \sum_{i=1}^{n} (y_i - \boldsymbol{x}_i'\beta)^2.$$

Similarly, the linear conditional quantile function, $Q(\theta|\boldsymbol{X} = \boldsymbol{x}) = \boldsymbol{x}'\beta(\theta)$, can be estimated by solving

$$\hat{\beta}(\theta) = \underset{\beta \in \Re^p}{\operatorname{argmin}} \sum_{i=1}^{n} \rho_\theta(y_i - \boldsymbol{x}_i'\beta), \tag{5.5}$$

for any quantile $\theta \in (0, 1)$. The case $\theta = 1/2$, which minimizes the sum of absolute residuals, corresponds to median regression, which is also known as L_1 regression. The minimization of (5.5) produced by Koenker and D'Orey (1994).

There are advantages of using the Quantile Regression method: Quantile regression allows us to study the impact of predictors on different quantiles of the response distribution, and thus provides a complete picture of the relationship between $\boldsymbol{Y}$ and $\boldsymbol{X}$. Quantile regression is robust to outliers in $\boldsymbol{Y}$ observations.

In the regression case we assume a sample $(Y_i, \boldsymbol{x}_i)$, $i = 1, ..., n$, where Y_i is the dependent variable and $\boldsymbol{x}_i$ is a $k \times 1$ vector of explanatory variables and $\boldsymbol{\beta}$ is a $k \times 1$ vector of coefficients. The general linear model has the form

$$Y_i = \boldsymbol{x}_i^{'}\boldsymbol{\beta} + u_i, \quad \text{and} \quad E(Y_i|\boldsymbol{x}_i) = \boldsymbol{x}_i^{'}\boldsymbol{\beta}, \tag{5.6}$$

while the $\theta - th$ conditional quantile of Y_i given $\boldsymbol{x}_i$ can be written as (see Bassett and Koenker, 1982)

$$Q_{Y_i}(\theta|\boldsymbol{x}_i) = \boldsymbol{x}_i^{'}\boldsymbol{\beta}_\theta. \tag{5.7}$$

We consider the θth sample quantile $\widehat{q}_i(\theta)$. Mosteller (1946) proved the limiting normality of $\widehat{\xi}^{\theta}_{Y_i|x_i}$ that provides a realization of the least estimation of the form

$$\widehat{q}_i^{\theta} = \boldsymbol{x}_i^{'}\boldsymbol{\beta}_\theta + u_i, \tag{5.8}$$

where $\boldsymbol{\beta}(\theta)$ is a vector to be estimated and u_i is the error term. For independent observations, under certain conditions, the asymptotic variances for u_i, can be obtained as (see Koenker, 2005, p. 132)

$$w_{ii}^{\theta} = \frac{\theta(1-\theta)}{n_i f_i^2(F_i^{-1}(\theta))}. \tag{5.9}$$

Instead of reweighting the observations by $\sqrt{n_i}$ to correct for the heterogeneity in sample size effect, it would be preferable to weight by $1/\sqrt{w_{ii}^{\theta}}$, to obtain the weighted least-squares estimator

$$\widehat{\beta}_{\theta} = (\boldsymbol{X}'\boldsymbol{\Omega}_{\theta}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Omega}_{\theta}^{-1}\widehat{q}, \tag{5.10}$$

that achieves an asymptotic efficiency bound and has limiting covariance matrix

$$\widehat{\boldsymbol{\Sigma}}_{\theta} = (\boldsymbol{X}'\boldsymbol{\Omega}_{\theta}^{-1}\boldsymbol{X})^{-1}. \tag{5.11}$$

With quantile regression we can show how various financial characteristics are different at different quantiles. Thus, the quantile regression method allows the marginal effects to change for claims at different points in the conditional distribution by estimating $\boldsymbol{\beta}_{\theta}$ using several different values of θ, $\theta \in (0,1)$. This means that the quantile regression allows for parameter heterogeneity across different types of claims.

In a similar way can be defined the $\theta - th$ quantile for the linear model. Let $\{\boldsymbol{x}_i, i = 1, ..., n\}$ denote a sequence of (row) k-vectors of a known design matrix and suppose $\{Y_i, i = 1, ..., n\}$ is a random sample on the regression process $u_i = Y_i - \boldsymbol{x}_i'\boldsymbol{\beta}$ having distribution function F. Then the $\theta - th$ regression quantile $0 < \theta < 1$, can be defined as any solution to the minimization problem [see Koenker and Basset (1978), Buchinsky (1998)]

$$\min_{\boldsymbol{\beta}} \frac{1}{n}\sum_{i=1}^{n} \rho_{\theta}(u_i) = \min_{(\boldsymbol{\beta})}\left(\sum_{i:y_i \geq \boldsymbol{x}_i'\boldsymbol{\beta}} \theta|Y_i - \boldsymbol{x}_i'\boldsymbol{\beta}| + \sum_{i:Y_i < \boldsymbol{x}_i'\boldsymbol{\beta}} (1-\theta)|Y_i - \boldsymbol{x}_i'\boldsymbol{\beta}|\right), \tag{5.12}$$

where $\rho_{\theta}(t) = (\theta - I(t < 0))t$ is a check function, and $I(\cdot)$ is the indicator function.

5.2 Correlated Run-off Triangles in a Quantile Longitudinal Model

The reserving procedure for multiple run-off triangles is an important issue of an insurance company because the connections among the triangles may show correlations which are initially unknown. The correlations of different lines of business may produce more efficient estimations for the total reserve. If for example the two run-off triangles are positively correlated, then the variability of the total reserves exceeds the sum of total reserve variability from each triangle. Ajne (1994) notes that commonly used approach in actuarial practice division of the portfolio into several subportfolios and then make calculations using each single line of business. But, this method ignores the or can consist of several lines with homogeneous the dependencies among the subportfolios.

When the run-off triangles are linked with a known structure, such as the paid and incurred triangles, then the Munich Chain Ladder (MuCL) model by Quarg and Mack (2004) is a good method of estimation. Moreover, instead of studying the structural correlations, the correlations between the triangles is an important issue and several papers have been produced such as Braun (2004), Kremer (2005), Schmidt (2006) and Merz and Wuthrich (2008a, 2008b). According to Holmberg (1994), correlations in a run-off triangle may arise among losses as they develop over time or in different year of accident. Other authors have studied correlations over calendar year incorporating the trends of inflation which appear.

Here, we are not going to use a triangulation form to model the data. Let y_{ik} be the k^{th} measurement for the i^{th} subject (triangle), which describes the total claims amount or the number of claims at the i run-off triangle for $i = 1, \ldots, N$, $k = 1, \ldots, n_i$ where n_i is the number of the observed data of the triangle i. We consider the case where, longitudinal data analyses are based on a linear regression model such as

$$y_{i,k} = \boldsymbol{x}'_{i,k}\boldsymbol{\beta} + \epsilon_{i,k} = \beta_1 x_{ik1} + \beta_2 x_{ik2} + \ldots + \beta_p x_{ikp} + \epsilon_{i,k}, \tag{5.13}$$

where $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)$ is a p-vector of unknown regression coefficients while $\epsilon_{i,k}$ is a random variable with mean zero and represents the deviation of the response from the model prediction $\boldsymbol{x}'_{i,k}\boldsymbol{\beta}$. Usually, $x_{ik1} = 1$ for all $i = 1, \ldots, N$ and all $k = 1, \ldots, n_i$, and then the coefficient β_1 is the intercept term of the regression model. In the classical linear model, the $\epsilon_{i,k}$ would be mutually independent $N(0, \sigma^2)$ random variables and represent the error term of the model. Although the most interesting scientific interest is about the mean response which is a function of some covariates, it is also very important the within correlation of observations. Mathematically, the $Cov(y_{i,j}, y_{i,k})$ of two different observations of the same subject, is not equal to zero. In the longitudinal structure it is expected the errors $\epsilon_{i,k}$ to be correlated within subjects (see Diggle, et al., 2002).

Harrison and Hulin (1989) used generalized estimating equations (GEE) as a promised analytic tool that takes into consideration the correlation of responses within a specific subject for response variables. A more interesting characteristic of these equations is the flexibility they have so as to analyze not normally distributed response variables.

Using matrices, the regression equation for the i^{th} subject has the following form:

$$\boldsymbol{Y}_i = \boldsymbol{X}'_i\boldsymbol{\beta} + \boldsymbol{\epsilon}_i, \tag{5.14}$$

where $\boldsymbol{X}'_i$ is a $n_i \times p$ matrix and $\boldsymbol{\epsilon}_i = (\epsilon_{i,1}, \ldots, \epsilon_{i,n_i})$. We should mention that in longitudinal studies, the experimental unit is not the individual measurement $y_{i,k}$, but the whole

Table 5.1: N Run-off Triangle in a Longitudinal Form

Subject	Observation	Response	Covariates		
1	1	$y_{1,1}$	$x_{1,1,1}$	$\dots$	$x_{1,1,p}$
1	2	$y_{1,2}$	$x_{1,2,1}$	$\dots$	$x_{1,2,p}$
$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
1	n_1	y_{1,n_1}	$x_{1,n_1,1}$	$\dots$	$x_{1,n_1,p}$
$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
N	1	$y_{N,1}$	$x_{N,1,1}$	$\dots$	$x_{N,1,p}$
N	2	$y_{N,2}$	$x_{N,2,1}$	$\dots$	$x_{N,2,p}$
$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
N	n_N	y_{N,n_N}	$x_{N,n_N,1}$	$\dots$	$x_{N,n,p}$

sequence, $\boldsymbol{Y}_i$, of measurements on a subject. Let $\boldsymbol{X}$ be an $\sum_{i=1}^{N} n_i \times p$ matrix of explanatory variables and $\sigma^2\boldsymbol{V}$ be a block-diagonal matrix with non-zero $n_i \times n_i$ blocks $\sigma^2\boldsymbol{V}_i$, each representing the variance-covariance matrix for the vector of measurements on the ith subject. Then, $\boldsymbol{y} = (\boldsymbol{y}_1, \dots, \boldsymbol{y}_N)$, is a realization of a multivariate Gaussian random vector, $\boldsymbol{Y}$, with

$$\boldsymbol{Y} \sim N_p(\boldsymbol{X\beta}, \sigma^2\boldsymbol{V}). \tag{5.15}$$

In case we want to analyze data generated by the model 5.15, then the block-diagonal structure of $\sigma^2\boldsymbol{V}$ is very important, because we will use each subject in order to estimate $\sigma^2\boldsymbol{V}$ without making any parametric assumptions about this form. The replication across the subjects is a very crucial characteristic because it affects the structure of the matrix $\sigma^2\boldsymbol{V}$ (Diggle, et al., 2002).

In order to estimate the regression parameter $\boldsymbol{\beta}$, it is usually used the weighted least-squares estimator of $\boldsymbol{\beta}$ using a symmetric weighted matrix, $\boldsymbol{W}$, which can be produced by minimizing the quadratic form

$$(\boldsymbol{y} - \boldsymbol{X\beta})'\boldsymbol{W}(\boldsymbol{y} - \boldsymbol{X\beta}). \tag{5.16}$$

After some matrix manipulations by taking the derivative with respect to $\boldsymbol{\beta}$, the explicit result is

$$\tilde{\boldsymbol{\beta}}_W = \boldsymbol{\beta} + (\boldsymbol{X'WX})^{-1}\boldsymbol{X'W\epsilon} \tag{5.17}$$

which is unbiased estimator for $\boldsymbol{\beta}$ as far as $E[\boldsymbol{\epsilon}|\boldsymbol{x}] = 0$, whatever the choice of the weight matrix $\boldsymbol{W}$. If $\boldsymbol{W} = \boldsymbol{I}$, the identity matrix, then we take the classical OLS estimator

$$\tilde{\boldsymbol{\beta}}_I = (\boldsymbol{X'X})^{-1}\boldsymbol{X'y}, \tag{5.18}$$

with

$$Var[\tilde{\boldsymbol{\beta}}_I] = \sigma^2\{(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{V}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\}. \tag{5.19}$$

If $\boldsymbol{W} = \boldsymbol{V}^{-1}$, the estimator becomes

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{y}, \tag{5.20}$$

with

$$Var[\widehat{\boldsymbol{\beta}}] = \sigma^2(\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}. \tag{5.21}$$

The "hat" notation states that the estimator is the maximum likelihood estimator for $\boldsymbol{\beta}$ under the multivariate Gaussian assumption which is the most efficient weighted least-squares estimator for $\boldsymbol{\beta}$. However, in order to identify the optimal weighting matrix it is necessary to know the complete correlation structure of the data.

5.2.1 Quantile Regression with Longitudinal Data

Fu and Wang (2012) considering the linear quantile regression model by Chen et al. (2004), proposed a combination of the between and within subject estimating functions for parameter estimation, which take into account the correlations and variation of the repeated measurements for subjects. Their model is an extension of the univariate quantile regression proposed by (Wang et al., 2009; Pang et al., 2010). Let y_{ik} be the k^{th} measurement for the i^{th} subject, where $k = 1, \ldots, n_i$ and $i = 1, \ldots, N$. We also suppose that $\boldsymbol{x}_{ik}$ is the corresponding covariate vector and measurements from the same subject are dependent while those from different subjects are independent. We assume that the 100θth quantile of y_{ik} is $\boldsymbol{x}_{ik}^T\boldsymbol{\beta}$, where $\boldsymbol{\beta}$ is a $p \times 1$ unknown parameter vector. Using this notation we consider the following model for the conditional quantile functions

$$Q_\theta(\boldsymbol{y}_{ik}|\boldsymbol{x}_{ik}) = \boldsymbol{x}_{ik}^T\boldsymbol{\beta}_0, \tag{5.22}$$

where $\boldsymbol{\beta}_0$ is the true value of the vector $\boldsymbol{\beta}$. Let the error term $\boldsymbol{\epsilon}_{ik} = \boldsymbol{y}_{ik} - \boldsymbol{x}_{ik}^T\boldsymbol{\beta}_0$ which satisfies the condition $P(\boldsymbol{\epsilon}_{ik} \leq 0) = \theta$. Then, we can find an efficient estimate for the unknown vector $\boldsymbol{\beta}$ for a particular value of θ. According to Chen et al. (2004), under the independence working model assumption, the estimates $\hat{\boldsymbol{\beta}}_I$ are obtained by minimizing the function

$$L_\theta(\boldsymbol{\beta}) = \sum_{i=1}^{N}\sum_{k=1}^{n_i}\rho_\theta(y_{ik} - \boldsymbol{x}_{ik}^T\boldsymbol{\beta}). \tag{5.23}$$

We differentiate (5.23) with respect to $\boldsymbol{\beta}$ and take the following estimating functions to make inferences about the unknown vector $\boldsymbol{\beta}$:

$$W_\theta(\boldsymbol{\beta}) = \sum_{i=1}^{N}\sum_{k=1}^{n_i} \boldsymbol{x}_{ik} S_{ik}$$

where $S_{ik} = \theta - I(y_{ik} - \boldsymbol{x}_{ik}^T\boldsymbol{\beta} \leq 0)$ is a discontinuous function which takes the value $\theta - 1$ when $y_{ik} - \boldsymbol{x}_{ik}^T\boldsymbol{\beta} \leq 0$ and the value θ otherwise.

5.2.2 The Uniform Correlation Model

In the uniform correlation model (also known as exchangeable or compound symmetry correlation model), it is assumed that there is correlation, ρ, between any two measurements on the same subject. In matrix notation, this corresponds to

$$\boldsymbol{V}_i = (1-\rho)\boldsymbol{I}_{n_i} + \rho\boldsymbol{J}_{n_i}, \tag{5.24}$$

where I_{n_i} denotes the $n_i \times n_i$ identity matrix and J_{n_i} the $n_i \times n_i$ matrix all of whose elements are 1 (Searle, et al., 1992). In order to justify the uniform correlation model we should think that the observed measurements, y_{ik}, are realizations of random variables, Y_{ik}. But,

$$Y_{ik} = \mu_{ik} + U_i + Z_{ik},\ i = 1, \ldots, N,\ k = 1, \ldots, n_i, \tag{5.25}$$

where $\mu_{ik} = E[Y_{ik}]$, U_i are mutually independent $N(0, v^2)$ random variables, Z_{ik} are mutually independent $N(0, t^2)$ random variables and the U_i and Z_{ik} are independent of one another. We should mention that 5.25 gives a simple interpretation of the uniform correlation model as one in which a linear regression model for the mean response incorporates a random intercept term which has variance t^2 between the subjects.

Theorem 5.1. *In the case of modeling the correlation between the same subject we assume that $P(\epsilon_{ik} \leq 0, \epsilon_{il} \leq 0) = \delta$ for any $k \neq l$ and the covariance matrix of $\boldsymbol{S}_i = (S_{i1}, \ldots, S_{in_i})^T$ is given by*

$$\boldsymbol{V}_i = (\theta - \theta^2)[(1-\rho)\boldsymbol{I}_{n_i} + \rho\boldsymbol{J}_{n_i}], \tag{5.26}$$

where ρ is the correlation coefficient of S_{ik} and S_{il} and equals $(\delta - \theta^2)/(\theta - \theta^2)$, $\boldsymbol{I}_{n_i}$ is the $n_i \times n_i$ identity matrix and $\boldsymbol{J}_{n_i}$ is the $n_i \times n_i$ matrix of 1s.

Proof. The form of the covariance matrix of $\boldsymbol{S}_i$ is

$$\boldsymbol{V}_i = \sigma^2 \begin{pmatrix} 1 & \rho & \rho & ... & \rho \\ \rho & 1 & \rho & ... & \rho \\ ... & ... & ... & ... & ... \\ \rho & \rho & \rho & ... & 1 \end{pmatrix}, \tag{5.27}$$

because there is correlation between S_{ij} and $S_{ij'}$ with $j \neq j'$ $j, j^{'} = 1, .., n_i$. We have

$$\rho = Corr(S_{ij}, S_{ij'}) = \frac{Cov(S_{ij}, S_{ij'})}{\sqrt{Var(S_{ij})}\sqrt{Var(S_{ij'})}} = \frac{Cov(S_{ij}, S_{ij'})}{\sigma^2}. \tag{5.28}$$

Moreover,

$$Cov(S_{ij}, S_{ij'}) = E[S_{ij}S_{ij'}] - E[S_{ij}]E[S_{ij'}] = \delta - \theta^2. \tag{5.29}$$

Using the fact that $P(\epsilon_{ik} \leq 0) = \theta$, we have

$$\begin{aligned} E[S_{ij}S_{ij'}] &= E\Big\{[\theta - I(\epsilon_{ij} \leq 0)][\theta - I(\epsilon_{ij'} \leq 0)]\Big\} \\ &= \theta^2 - \theta E\Big\{I(\epsilon_{ij} \leq 0)\Big\} - \theta E\Big\{I(\epsilon_{ij'} \leq 0)\Big\} + E\Big\{I(\epsilon_{ij} \leq 0)I(\epsilon_{ij'} \leq 0)\Big\} \\ &= \delta - \theta^2, \end{aligned}$$

and

$$E[S_{ik}] = E\Big\{\theta - I(\epsilon_{ij} \leq 0)\Big\} = \theta - E\Big\{I(\epsilon_{ij} \leq 0)\Big\} = 0, \ \forall k.$$

We used the fact that $I(\epsilon_{ij} \leq 0)$ is a binary variable which takes the value 1 when $\epsilon_{ij} \leq 0$ and the value 0 otherwise with mean θ and variance $\theta(1-\theta)$. Similarly, the variable $I(\epsilon_{ij} \leq 0)I(\epsilon_{ij'} \leq 0)$ is a binary variable with mean δ and variance $\delta(1-\delta)$. Then, we have that

$$Var(S_{ik}) = Var[\theta - I(y_{ik} - \boldsymbol{x}_{ik}^T\boldsymbol{\beta} \leq 0)] = Var[\theta - I(\epsilon_{ik} \leq 0)] = \theta(1-\theta) \tag{5.30}$$

From (5.28) we take that the correlation coefficient is equal to $\rho = \frac{\delta - \theta^2}{\theta - \theta^2}$. Moreover, by (5.27) and (5.30), the covariance matrix $\boldsymbol{V}_i$ is

$$\begin{aligned} \boldsymbol{V}_i &= (\theta - \theta^2)\left[(1-\rho)\begin{pmatrix} 1 & 0 & 0 & .. & 0 \\ 0 & 1 & 0 & ... & 0 \\ ... & ... & ... & ... & ... \\ 0 & 0 & 0 & . & 1 \end{pmatrix} + \rho \begin{pmatrix} 1 & 1 & 1 & ... & 1 \\ 1 & 1 & 1 & ... & 1 \\ ... & ... & ... & ... & ... \\ 1 & 1 & 1 & ... & 1 \end{pmatrix}\right] \\ &= (\theta - \theta^2)\Big[(1-\rho)\boldsymbol{I}_{n_i} + \rho \boldsymbol{J}_{n_i}\Big]. \end{aligned}$$

■

Let now $\boldsymbol{X}_i = \{\boldsymbol{X}_{i1}, \ldots, \boldsymbol{X}_{in_i}\}^T$. In order to obtain efficient estimators we should incorporate an appropriate weighted function that takes into account the correlation for each subject. According to Jung (1996) based on the exchangeable correlation structure assumption

$$Corr(S_{ij}, S_{ik}) = \begin{cases} 1, & j = k \\ 0, & j \neq k \end{cases}, \tag{5.31}$$

the generalized least squares estimate of $\boldsymbol{\beta}$ obtained by minimizing

$$\boldsymbol{S}_i \boldsymbol{V}_i^{-1} \boldsymbol{S}_i \tag{5.32}$$

and differentiating with respect to $\boldsymbol{\beta}$, the following weighted functions are used:

$$\boldsymbol{U}_\theta(\boldsymbol{\beta}) = \sum_{i=1}^{N} \boldsymbol{X}_i^T \boldsymbol{V}_i^{-1} \boldsymbol{S}_i, \tag{5.33}$$

where $\boldsymbol{V}_i^{-1}$ is the inverse matrix of $\boldsymbol{V}_i$.

Proposition 5.1. *The inverse matrix of $\boldsymbol{V}_i$ can be written as*

$$\boldsymbol{V}^{-1} = \frac{1}{\theta - \theta^2}\Big(\boldsymbol{W}_i^{bet} + \boldsymbol{W}_i^{wit}\Big), \tag{5.34}$$

where $\boldsymbol{W}_i^{bet}$ and $\boldsymbol{W}_i^{wit}$ are quantities related to information from different subjects and from the same subject, respectively

$$\boldsymbol{W}_i^{bet} = \frac{\boldsymbol{J}_{n_i}}{n_i[1 + (n_i - 1)\rho]}, \quad and \quad \boldsymbol{W}_i^{wit} = \frac{1}{1-\rho}\Big(\boldsymbol{I}_{n_i} - \frac{1}{n_i}\boldsymbol{J}_{n_i}\Big) \tag{5.35}$$

Proof. Suppose $\boldsymbol{A}$ is an invertible square matrix and $\boldsymbol{u}$, $\boldsymbol{w}$ are column vectors. Suppose furthermore that $1 + \boldsymbol{w}^T\boldsymbol{A}^{-1}\boldsymbol{u} \neq 0$. Then, the Sherman Morrison formula (Bartlett, 1951) states that

$$(\boldsymbol{A} + \boldsymbol{u}\boldsymbol{w}^T)^{-1} = \boldsymbol{A}^{-1} - \frac{\boldsymbol{A}^{-1}\boldsymbol{u}\boldsymbol{w}^T\boldsymbol{A}^{-1}}{1 + \boldsymbol{w}^T\boldsymbol{A}^{-1}\boldsymbol{u}}. \tag{5.36}$$

Starting from

$$\boldsymbol{V}_i = \sigma^2\Big[(1-\rho)\boldsymbol{I}_{n_i} + \rho\boldsymbol{J}_{n_i}\Big]$$

and supposing that $\rho \boldsymbol{J}_{n_i} = \boldsymbol{u}\boldsymbol{w}^T$ where $\boldsymbol{u} = \boldsymbol{w} = \{\rho, \rho, \ldots, \rho\}^T$ is a $n_i \times 1$ vector, by (5.36) we take

$$
\begin{aligned}
\boldsymbol{V}_i^{-1} &= \frac{1}{\sigma^2}\Bigg[\frac{1}{1-\rho}\boldsymbol{I}_{n_i} - \frac{\left(\frac{1}{1-\rho}\boldsymbol{I}_{n_i}\right)\rho \boldsymbol{J}_{n_i}\frac{1}{1-\rho}\boldsymbol{I}_{n_i}}{1+\frac{n_i\rho}{1-\rho}}\Bigg] \\
&= \frac{1}{\sigma^2}\Bigg[\frac{1}{1-\rho}\boldsymbol{I}_{n_i} - \frac{1}{1-\rho}\Big(\frac{\rho}{1+(n_i-1)\rho}\Big)\boldsymbol{J}_{n_i}\Bigg] \\
&= \frac{1}{\sigma^2}\Bigg[\frac{1}{1-\rho}\boldsymbol{I}_{n_i} + \frac{1}{1-\rho}\Big(\frac{n_i(1-\rho) - n_i - n_i(n_i-1)\rho}{(1+(n_i-1)\rho)n_i^2}\Big)\boldsymbol{J}_{n_i}\Bigg]. \\
&= \frac{1}{\sigma^2}\Bigg[\frac{1}{1-\rho}\boldsymbol{I}_{n_i} + \frac{1}{1-\rho}\Big(\boldsymbol{I}_{n_i}\frac{1-\rho}{[1+(n_i-1)\rho]n_i} - \frac{1}{n_i}\boldsymbol{I}_{n_i}\Big)\boldsymbol{J}_{n_i}\Bigg] \\
&= \frac{1}{\sigma^2}\Bigg[\frac{1}{1-\rho}\boldsymbol{I}_{n_i} + \frac{\boldsymbol{J}_{n_i}}{[1+(n_i-1)\rho]n_i} - \frac{1}{n_i(1-\rho)}\boldsymbol{I}_{n_i}\Bigg] \\
&= \frac{1}{\sigma^2}\Bigg[\frac{\boldsymbol{J}_{n_i}}{[1+(n_i-1)\rho]n_i} + \frac{1}{1-\rho}\Big(\boldsymbol{I}_{n_i} - \frac{1}{n_i}\boldsymbol{J}_{n_i}\Big)\Bigg]
\end{aligned}
$$

that provides (5.34). ■

If there is no correlation between the same subject then the correlation coefficient ρ is zero and the inverse matrix of $\boldsymbol{V}_i$ is equal to

$$
\boldsymbol{V}^{-1} = \frac{1}{\sigma^2}\boldsymbol{I}_{n_i},
$$

and $\boldsymbol{U}_\theta(\boldsymbol{\beta})$ is equivalent to the estimating functions $\boldsymbol{W}_\theta(\boldsymbol{\beta})$. Furthermore, from (5.33), using the result of (5.34) we take

$$
\begin{aligned}
\boldsymbol{U}_\theta(\boldsymbol{\beta}) &= \sum_{i=1}^{N}\boldsymbol{X}_i^T\boldsymbol{V}_i^{-1}\boldsymbol{S}_i \\
&= \sum_{i=1}^{N}\boldsymbol{X}_i^T\frac{1}{\sigma^2}\left[\frac{\boldsymbol{J}_{n_i}}{n_i[1+(n_i-1)\rho]} + \frac{1}{1-\rho}\Big(\boldsymbol{I}_{n_i} - \frac{1}{n_i}\boldsymbol{J}_{n_i}\Big)\right]\boldsymbol{S}_i \\
&= \frac{1}{\sigma^2}\sum_{i=1}^{n}\boldsymbol{X}_i^T\left[\frac{\boldsymbol{J}_{n_i}}{n_i[1+(n_i-1)\rho]}\right]\boldsymbol{S}_i + \frac{1}{\sigma^2}\sum_{i=1}^{N}\boldsymbol{X}_i^T\left[\frac{1}{1-\rho}\Big(\boldsymbol{I}_{n_i} - \frac{1}{n_i}\boldsymbol{J}_{n_i}\Big)\right]\boldsymbol{S}_i \\
&= \frac{1}{\sigma^2}\sum_{i=1}^{N}\boldsymbol{X}_i^T\left[\frac{1}{1+(n_i-1)\rho}\right]\boldsymbol{J}_{n_i}\sum_{k=1}^{n_i}\boldsymbol{S}_i/n_i + \frac{1}{(1-\rho)\sigma^2}\sum_{i=1}^{N}\boldsymbol{X}_i^T\Big(\boldsymbol{S}_i - \boldsymbol{1}_{n_i}\sum_{k=1}^{n_i}\boldsymbol{S}_i/n_i\Big),
\end{aligned}
\tag{5.37}
$$

where $\mathbf{1}_{n_i}$ is a $n_i \times 1$ vector of 1s. Then, by (5.37) and (5.35) we can extract the following two estimating functions:

$$\boldsymbol{U}^{bet}(\beta) = \sum_{i=1}^{N} \frac{1}{1+(n_i-1)\rho} \boldsymbol{X}_i^T \mathbf{1}_{n_i} \sum_{k=1}^{n_i} \boldsymbol{S}_i / n_i = \sum_{i=1}^{N} \boldsymbol{X}_i^T \boldsymbol{W}_i^{between} \boldsymbol{S}_i,$$
$$\boldsymbol{U}^{wit}(\beta) = \frac{1}{1-\rho} \sum_{i=1}^{N} \boldsymbol{X}_i^T \Big(\boldsymbol{S}_i - \mathbf{1}_{n_i} \sum_{k=1}^{n_i} \boldsymbol{S}_i / n_i \Big) = \sum_{i=1}^{N} \boldsymbol{X}_i^T \boldsymbol{W}_i^{within} \boldsymbol{S}_i \qquad (5.38)$$

Remark 5.1. *Note that the estimating functions $\boldsymbol{U}^{wit}(\boldsymbol{\beta})$ indicate the differences within a subject while $\boldsymbol{U}^{bet}(\boldsymbol{\beta})$ indicate the information which comes from different subjects.*

5.2.3 Estimation of the Parameters

Generally, the most difficult issue when using quantile regression is the estimation of the covariance matrix of the parameter estimators because it involves the unknown density functions of the errors. Resampling methods have been proposed to estimate the covariance matrix (Parzen et al., 1994), but these methods cause some potentials especially to real data analysis. These methods are useful because the parameter estimates can be easily obtained but the variance is difficult to be estimated. Moreover, there is no analytical proof for the validity of the traditional bootstrap technique for the quantile regression model (Yin and Cai, 2005). Fu and Wang, (2012) extended the smoothing method of quantile regression with independent data proposed by Wang et al. (2009) and proposed a method for longitudinal data.

Suppose, that $\hat{\boldsymbol{\beta}}_u$ is the estimator which results from $U_\theta(\boldsymbol{\beta})$. Then, under some regularity conditions, $\hat{\boldsymbol{\beta}}_u$, is a consistent estimator of $\boldsymbol{\beta}_0$ and

$$\sqrt{N}(\hat{\boldsymbol{\beta}}_u - \boldsymbol{\beta}_0) \rightarrow N(0, \boldsymbol{\Lambda}). \qquad (5.1)$$

So, the resulting estimator $\hat{\boldsymbol{\beta}}_u$ from 5.37 can be approximated by $\boldsymbol{\beta} + \boldsymbol{\Lambda}^{1/2}\boldsymbol{Z}$ where $\boldsymbol{Z}$ is the standard normal distribution $N(\mathbf{0}, \boldsymbol{I}_p)$ and $\boldsymbol{\Lambda}^{1/2}\boldsymbol{Z}$ is a disturbance quantity to $\boldsymbol{\beta}$. Moreover, according to 5.23, the estimating functions $\boldsymbol{U}_\theta(\beta)$ can be defined as $\tilde{\boldsymbol{U}}_\theta(\beta) = E_{\boldsymbol{Z}}\{\boldsymbol{U}_\theta(\boldsymbol{\beta} + \boldsymbol{\Lambda}^{1/2}\boldsymbol{Z})\}$ where expectation is over $\boldsymbol{Z}$. Nevertheless, the variance-covariance matrix $\boldsymbol{\Lambda}$ is unknown which means that the expectation cannot be computed. For that reason, Brown and Wang (2005) suggested the use of a known matrix $\boldsymbol{\Gamma}$ instead of $\boldsymbol{\Lambda}$ and using appropriate iterative algorithms in order to estimate the matrix $\boldsymbol{\Lambda}$. So, the objective function is $\tilde{\boldsymbol{U}}_\theta(\beta) = E_{\boldsymbol{Z}}\{\boldsymbol{U}_\theta(\boldsymbol{\beta} + \boldsymbol{\Gamma}^{1/2}\boldsymbol{Z})\}$.

Note that

$$E\{L_\theta(\boldsymbol{\beta} + \boldsymbol{\Gamma}^{1/2}\boldsymbol{Z})\} = \theta - P\{\boldsymbol{x}_{ik}^T \boldsymbol{\Gamma}^{1/2}\boldsymbol{Z} \geq b_{ik}\} = \theta - 1 + \Phi\Big[\frac{b_{ik}}{\sigma_{ik}}\Big], \qquad (5.2)$$

where $b_{ik} = \boldsymbol{y}_{ik} - \boldsymbol{x}_{ik}^T\boldsymbol{\beta}$ and $\sigma_{ik}^2 = \boldsymbol{x}_{ik}^T\boldsymbol{\Gamma}\boldsymbol{x}_{ik}$. Then,

$$\tilde{\boldsymbol{U}}_\theta(\beta) = \sum_{i=1}^{N} \boldsymbol{X}_i^T \boldsymbol{V}_i^{-1}\tilde{\boldsymbol{S}}_i, \tag{5.3}$$

where $\tilde{\boldsymbol{S}}_i = (\tilde{\boldsymbol{S}}_{i1}, \ldots, \tilde{\boldsymbol{S}}_{in_i})$ with $\tilde{\boldsymbol{S}}_{ik} = \theta - 1 + \Phi\left[\frac{b_{ik}}{\sigma_{ik}}\right]$. Differentiating (5.3) with respect to $\boldsymbol{\beta}$, we take

$$\tilde{\boldsymbol{D}}_\theta(\beta) = -\sum_{i=1}^{N} \boldsymbol{X}_i^T \boldsymbol{V}_i^{-1}\tilde{\boldsymbol{\Lambda}}_i\boldsymbol{X}_i, \tag{5.4}$$

where $\tilde{\boldsymbol{\Lambda}}_i$ is a diagonal $n_i \times n_i$ matrix with diagonal element $\sigma_{ik}^{-1}\phi[\frac{b_{ik}}{\sigma_{ik}}]$.

In order to produce the estimators and the corresponding covariance matrix, we need iterative methods. We adopt the algorithm of Fu and Wang (2012) who extended the induced smoothing method of Wang et al. (2009) and Pang et al. (2010). A similar algorithm was applied by Stoner and Leroux (2002) for the analysis of clustered data. The steps of the algorithm are the following:

ALGORITHM:

Step 1. Produce some initial values $\tilde{\boldsymbol{\beta}}^0 = \hat{\boldsymbol{\beta}}_I$, which have been obtained by the independence working model and $\boldsymbol{\Gamma}^0 = n^{-1}\boldsymbol{I}_p$.

Step 2. Given $\tilde{\boldsymbol{\beta}}^{k-1}$ and $\boldsymbol{\Gamma}^{k-1}$ from the $k-1$ step, update $\hat{\delta}^{k-1}$, using the following equation

$$\hat{\delta}^{k-1} = \frac{\sum_{i=1}^{N}\sum_{k=1}^{n_i}\sum_{l\neq k}^{n_i} I[\hat{\epsilon}_{ik} \leq 0, \hat{\epsilon}_{il} \leq 0]}{\sum_{i=1}^{N} n_i(n_i - 1)}.$$

Step 3. Update the estimation parameters $\tilde{\boldsymbol{\beta}}^k$ and the matrix $\boldsymbol{\Gamma}^k$ using the equations

$$\tilde{\boldsymbol{\beta}}^k = \tilde{\boldsymbol{\beta}}^{k-1} + \{\tilde{\boldsymbol{D}}_\theta(\tilde{\boldsymbol{\beta}}^{k-1}, \boldsymbol{\Gamma}^{k-1})\}^{-1}\tilde{\boldsymbol{U}}_\theta(\tilde{\boldsymbol{\beta}}^{k-1}, \boldsymbol{\Gamma}^{k-1}, \hat{\delta}^{k-1}),$$
$$\boldsymbol{\Gamma}^k = \tilde{\boldsymbol{D}}_\theta^{-1}(\tilde{\boldsymbol{\beta}}^{k-1}, \boldsymbol{\Gamma}^{k-1})\boldsymbol{V}(\tilde{\boldsymbol{\beta}}^{k-1}, \hat{\delta}^{k-1})\tilde{\boldsymbol{D}}_\theta^{-1}(\tilde{\boldsymbol{\beta}}^{k-1}, \boldsymbol{\Gamma}^{k-1}).$$

Step 4. Repeat Steps 2 and 3 until convergence.

Remark 5.2. *Under some regularity conditions,* $n^{-1/2}\{\tilde{\boldsymbol{U}}_\theta - \boldsymbol{U}_\theta\} = o_p(1)$.

5.3 Numerical Illustrations

In this section, some numerical examples using quantile regression models will be implemented. We suppose that we have two blocks of business for which we are trying to calculate reserve indications. Both companies operate in Greece. Company A mainly focuses on Motor Business and underwrites all vehicle categories apart from taxis and trucks while company B underwrites all vehicle categories for Motor Business. Tables 5.2 and 5.3 show the triangles of the incremental incurred claims (paid and outstanding claims) for both companies.

Table 5.2: Motor Triangle and Premiums for Company A

Accident	**Development Year**										
Year	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**Premium**
2007	58134	162688	101105	100964	61591	71009	34024	2746	646	10190	1051637
2008	51437	197139	120641	74807	76771	77276	39070	4396	13809		1190965
2009	57906	116191	143953	103883	70760	177194	35341	6088			1327568
2010	40352	121837	88389	320429	75127	70190	63723				1418348
2011	82227	279591	151260	230293	82378	47315					1504056
2012	196417	119755	228499	99894	44266						1580233
2013	67161	107098	198252	75172							1619382
2014	78293	141865	106150								1727540
2015	74472	118886									1820104
2016	43281										1883017

Table 5.3: Motor Triangle and Premiums for Company B

Accident	**Development Year**										
Year	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**Premium**
2007	63078	143002	144235	75007	60775	70804	27508	4757	3172	6385	1633833
2008	65567	177292	107870	137305	72741	68708	102864	4335	6107		1675707
2009	87394	146346	158876	199846	53161	72764	42915	10898			1636855
2010	70017	153893	119028	93771	49600	185689	28331				1689715
2011	104638	186326	335477	136857	87941	69248					1649386
2012	76390	190629	192606	121704	66297						1712587
2013	58620	184557	135174	118180							2105361
2014	87845	166511	145385								2265432
2015	53616	152751									1976188
2016	62904										1351719

It is obvious that for Company A for accident year 1, a big claim has been paid 10 years after the accident date (the amount of this claim is embedded at the total incremental amount of 10190) and could be represented as an outlier claim. This claim will dramatically change the development pattern of the payments in case of using the Chain Ladder method because the Loss Development Factor for this year increases from 1.01 to 1.02 (see Table 5.4). This is commonly observed in Motor Business because there are many cases of claims which are settled many years later especially when accidents with large compensations (such as partial or total disability, deaths, etc.) are observed. This can also be observed in accident year 2 where a large amount is observed at the last known development year (13809). On the other hand, for company B, the data of the triangle seem to be more stable without big fluctuations. Moreover, the premiums of the companies by accident year are presented at the run-off triangles. For the implementation of the data set we divided the payments by exposures for each year of business before the analysis is carried out. we will divide the losses by the premium to put the accident years on a more nearly equal basis.

The number of exposures (counts of incurred losses) for each accident year is also provided (see Tables 5.5 and 5.6) for each lines of business.

Table 5.4: Loss Development Factors

Company	0-1	1-2	2-3	3-4	4-5	5-6	6-7	7-8	8-9
A	2.93	1.61	1.37	1.13	1.15	1.07	1.01	1.01	1.02
B	3.25	1.68	1.30	1.12	1.15	1.08	1.01	1.01	1.01

Table 5.5: Counts of Incurred Claims for Company A

Accident	**Development Year**									
Year	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
2007	118	272	241	169	103	106	71	4	3	5
2008	134	287	235	192	121	117	84	14	13	
2009	129	254	267	193	110	136	80	16		
2010	87	255	218	152	111	86	52			
2011	148	285	277	212	127	94				
2012	108	255	227	185	103					
2013	129	215	220	150						
2014	128	277	234							
2015	122	236								
2016	94									

Table 5.6: Counts of Incurred Claims for Company B

Accident	**Development Year**									
Year	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
2007	139	286	276	170	137	140	74	15	8	6
2008	143	337	258	224	158	158	90	20	13	
2009	151	273	310	239	145	135	81	11		
2010	138	285	273	182	122	127	70			
2011	161	372	349	282	185	129				
2012	131	327	297	237	150					
2013	144	345	284	222						
2014	146	337	295							
2015	130	301								
2016	155									

If we were trying to calculate the expected value of the reserve run-off, we could simply calculate the expected value for each line of business separately and add all the expectations together. However, when we quantify a value other than the mean, such as a quantile, we cannot simply sum across the lines of business. In such a case, we will overstate the aggregate reserve need.

Figures 5.1 and 5.2 show claims development charts of Company A and Company B, respectively, with individual panels for each origin period. Chain Ladder loss development factors for each company are also presented in Table 5.4. According to the claims development chart we observe that the pattern of of company A and company B looks similar (see Figure 5.3).

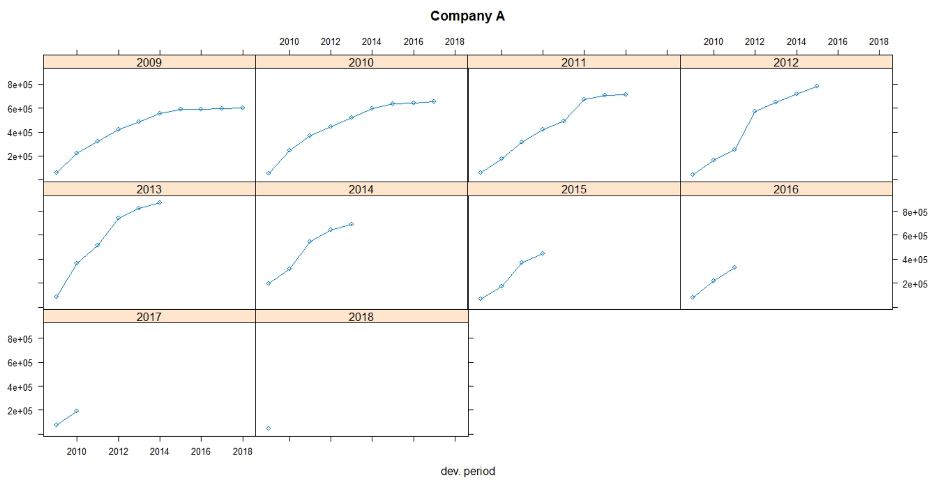

Figure 5.1: Claims development chart of Company A with individual panels for each origin period

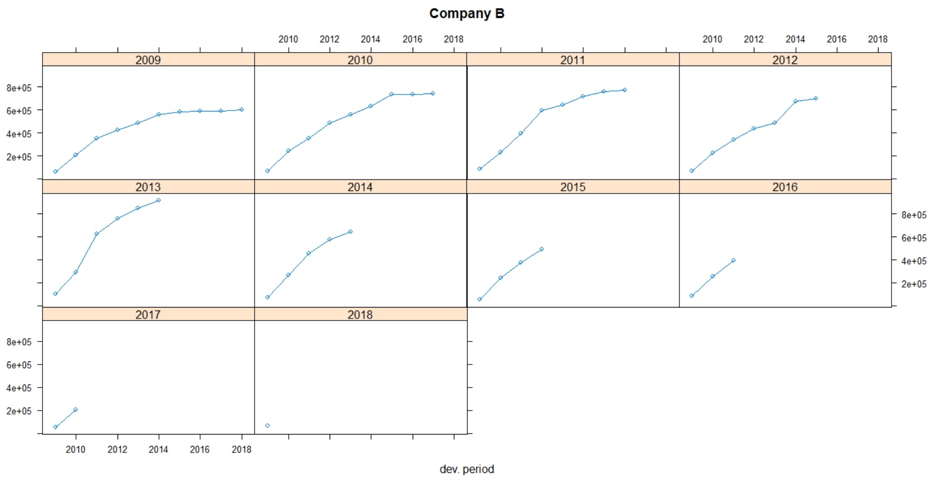

Figure 5.2: Claims development chart of Company B with individual panels for each origin period

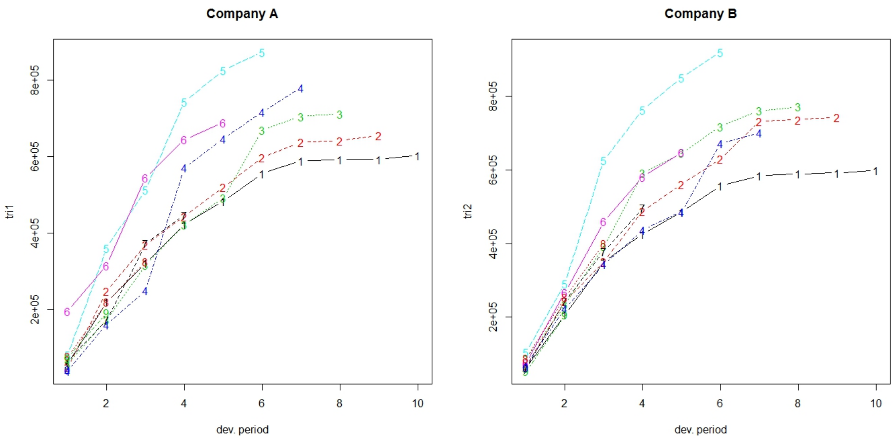

Figure 5.3: Claims development chart of the triangles with one line per origin period.

Table 5.7 and Table 5.8 display the values of reserves and ultimate paid claims based on individual quantile regression method, for company A and company B, respectively, for different quantiles. The loss ratios for each quantile are also provided at the end of each of the Tables 6 and 7. Table 5.9 and Table 5.10 display the values of reserves and ultimate paid claims based on longitudinal quantile regression method, for company A and company B, respectively, for different quantiles. The loss ratios are also included in the end of the tables. Loss ratios for motor car insurance typically range from 40% to 60%. At this case, insurance companies are collecting more premiums than the amount paid in claims. Loss Ratio is considered as one of the tools with which explains a company's suitability for coverage. A high Loss Ratio means is considered bad which leads to bad financial health because the insurance company may not be collect enough premium to pay claims, expenses and make a reasonable profit.

Table 5.7: Reserves and Ultimate Claims of Company A based on Individual QR

Accident	**Quantile 50%**		**Quantile 60%**		**Quantile 75%**		**Quantile 90%**		**Quantile 95%**		**Quantile 99.5%**	
Year	**Reserves**	**Ultimate**	**Reserves**	**Ultimate**	**Reserves**	**Ultimate**	**Reserves**	**Ultimate**	**Reserves**	**Ultimate**	**Reserves**	**Ultimate**
2007	0	603097	0	603097	0	603097	0	603097	0	603097	0	603097
2008	11702	667049	11496	666843	12348	667695	12348	667695	12348	667695	12348	667695
2009	25522	736838	25770	737085	30734	742050	30734	742050	30734	742050	30734	742050
2010	30123	810171	32114	812161	48436	828484	48436	828484	61841	841888	61841	841888
2011	84962	958027	86226	959291	102917	975981	102917	975981	102917	975981	102917	975981
2012	144793	833624	150033	838863	295824	984655	485528	1174359	485528	1174359	485528	1174359
2013	218293	665976	223617	671300	222374	670057	485688	933370	485688	933370	485688	933370
2014	376255	702563	396658	722966	439704	766012	439675	765983	400741	727049	400741	727049
2015	433878	627236	514291	707648	547762	741119	519185	712542	482151	675508	482151	675508
2016	364632	407912	377312	420592	439462	482742	396155	439435	374632	417912	374632	417912
Total	**1690161**	**7012491**	**1817516**	**7139846**	**2139562**	**7461893**	**2520666**	**7842996**	**2436579**	**7758909**	**2436579**	**7758909**
LR	**46.37%**		**47.21%**		**49.34%**		**51.86%**		**51.31%**		**51.31%**	

Table 5.8: Reserves and Ultimate Claims of Company B based on Individual QR

Accident	**Quantile 50%**		**Quantile 60%**		**Quantile 75%**		**Quantile 90%**		**Quantile 95%**		**Quantile 99.5%**	
Year	**Reserves**	**Ultimate**	**Reserves**	**Ultimate**	**Reserves**	**Ultimate**	**Reserves**	**Ultimate**	**Reserves**	**Ultimate**	**Reserves**	**Ultimate**
2007	0	598722	0	598722	0	598722	0	598722	0	598722	0	598722
2008	7915	750705	8760	751550	9338	752128	7642	750431	7642	750431	7642	750431
2009	15633	787833	15012	787213	14631	786831	20011	792211	20011	792211	20011	792211
2010	16638	716969	16780	717111	20452	720783	19694	720025	19694	720025	19694	720025
2011	64164	984651	77249	997735	156940	1077426	232413	1152900	232413	1152900	232413	1152900
2012	124174	771801	131694	779320	212699	860326	348719	996345	348719	996345	348719	996345
2013	184956	681488	196298	692830	270111	766642	413334	909865	413334	909865	413334	909865
2014	326036	725777	308932	708673	422205	821946	621763	1021504	621763	1021504	621763	1021504
2015	398134	604501	382772	589139	477907	684274	614736	821104	614736	821104	614736	821104
2016	514302	577206	499960	562864	588777	651681	742201	805105	742201	805105	742201	805105
Total	**1651953**	**7199653**	**1637457**	**7185156**	**2173060**	**7720759**	**3020513**	**8568213**	**3020513**	**8568213**	**3020513**	**8568213**
LR	**40.68%**		**40.60%**		**43.63%**		**48.42%**		**48.42%**		**48.42%**	

Table 5.9: Reserves and Ultimate Claims of Company A based on Longitudinal QR

Accident	**Quantile 50%**		**Quantile 60%**		**Quantile 75%**		**Quantile 90%**		**Quantile 95%**		**Quantile 99.5%**	
Year	**Reserves**	**Ultimate**	**Reserves**	**Ultimate**	**Reserves**	**Ultimate**	**Reserves**	**Ultimate**	**Reserves**	**Ultimate**	**Reserves**	**Ultimate**
2007	0	603097	0	603097	0	603097	0	603097	0	603097	0	603097
2008	10465	665812	13163	668509	14614	669961	14998	670344	16354	671701	14284	669631
2009	15332	726647	19519	730834	24491	735807	25753	737069	27823	739139	24671	735987
2010	18773	798820	24548	804595	30472	810519	33114	813161	36000	816047	31594	811641
2011	109292	982356	146947	1020012	162394	1035459	280876	1153941	270861	1143926	271168	1144233
2012	142717	831548	195409	884240	227356	916187	388819	1077650	534912	1223743	535184	1224015
2013	121615	569298	164628	612311	205183	652866	313448	761131	303832	751515	304086	751769
2014	143793	470101	216128	542436	274426	600734	369152	695460	398023	724331	624747	951055
2015	156935	350292	245965	439323	255528	448885	319659	513016	354969	548327	355450	548807
2016	108851	152131	205266	248546	205286	248567	211902	255182	211842	255123	211977	255257
Total	**827771**	**6150102**	**1231572**	**6553903**	**1399751**	**6722081**	**1957720**	**7280050**	**2154617**	**7476947**	**2373161**	**7695491**
LR	**40.67%**		**43.34%**		**44.45%**		**48.14%**		**49.44%**		**50.89%**	

Table 5.10: Reserves and Ultimate Claims of Company B based on Longitudinal QR

Accident	Quantile 50%		Quantile 60%		Quantile 75%		Quantile 90%		Quantile 95%		Quantile 99.5%	
Year	Reserves	Ultimate	Reserves	Ultimate	Reserves	Ultimate	Reserves	Ultimate	Reserves	Ultimate	Reserves	Ultimate
2007	0	598722	0	598722	0	598722	0	598722	0	598722	0	598722
2008	14812	757602	19390	762179	21292	764082	21818	764608	23573	766363	20865	763655
2009	23811	796012	31465	803665	38726	810927	40529	812730	43425	815625	38967	811168
2010	31032	731363	42019	742349	51128	751459	54931	755262	59120	759451	52655	752985
2011	146138	1066624	203353	1123839	221691	1142178	370226	1290712	358282	1278769	358674	1279161
2012	220915	868541	313593	961219	357766	1005392	589509	1237135	806340	1453966	806747	1454373
2013	197552	694083	279346	775877	337958	834489	497873	994404	484759	981290	485153	981684
2014	217454	617195	342324	742065	418633	818374	541538	941279	583882	983622	899021	1298762
2015	243640	450007	397190	603557	410815	617182	490441	696808	541561	747928	542253	748621
2016	189118	252022	356630	419534	356665	419569	368159	431063	368056	430960	368290	431194
Total	**1284472**	**6832172**	**1985308**	**7533008**	**2214675**	**7762375**	**2975023**	**8522722**	**3268999**	**8816698**	**3572625**	**9120325**
LR	**38.61%**		**42.57%**		**43.86%**		**48.16%**		**49.82%**		**51.54%**	

To examine the role of dependence is important to calculate the reserves for each individual LOB, and then use the sum to compare it with the sum of the run-off triangles resulting from the RCR cross effects (see Table 5.11).

Table 5.11: Estimated reserves using Individual Quantile Regression (IQR) and the Longitudinal Algorithm (LALG)

	Quantile 50%	**Quantile 60%**	**Quantile 75%**	**Quantile 90%**	**Quantile 95%**	**Quantile 99.5%**
Company A IQR	1690161	1817516	2139562	2520666	2436579	2436579
Company B IQR	1651953	1637457	2173060	3020513	3020513	3020513
Company A LALG	827771	1231572	1399751	1957720	2154617	2373161
Company B LALG	1284472	1985308	2214675	2975023	3268999	3572625
IQR - LALG	**1229871**	**238092**	**698196**	**608436**	**33477**	**-488694**

Applying individual quantile regression, a higher quantile leads to larger total reserve. Nevertheless, for company A quantiles over 95% provide equal values of reserves, while for company B quantiles over 90% provide equal values of reserves. The longitudinal algorithm gives different estimations for each quantile. Applying longitudinal quantile regression, the estimated ultimate reserves for both companies A and B are smaller than the sum of individual estimated reserves for each company A and B based on individual quantile regression.

For model comparison, two criteria, namely the Root Mean Squared Error (RMSE) and Percentage Total (PT) are proposed. The Root Mean Square Error (RMSE) is a well-known and frequently used measure of the differences between the predicted values of a model and the actually observed values. Specifically, the RMSE represents the sample standard deviation of the differences between the predicted and observed values. These differences are usually known as residuals when the calculations are made over the data sample and are called prediction errors when computed out-of-sample. The RMSE is a measure of how to spread out these residuals are. The RMSE aggregates the magnitudes of the errors in predictions for all data into a single measure. RMSE is a measure of accuracy and is useful for comparing different models for a particular data set (Hyndman et al., 2006). RMSE is the square root of the average of squared errors. It should be mentioned that the effect of each error on RMSE is proportional to the size of the squared error. So, larger errors have a disproportionately large effect on RMSE. Consequently, RMSE is sensitive to outliers (Pontius et al., 2008, Willmott et al., 2008).

For the RMSE for one triangle data we have:

$$RMSE = \left[\frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n-i+1} (y_{ij} - \hat{y}_{ij})^2 \right]^{1/2}, \tag{5.1}$$

while when using many run-off triangles the RMSE is defined as

$$RMSE = \left[\frac{1}{nN} \sum_{k=1}^{N} \sum_{i=1}^{n} \sum_{j=1}^{n-i+1} (y_{ij} - \hat{y}_{ij})^2 \right]^{1/2}, \tag{5.2}$$

where k is the counter for each triangle. Moreover, the percentage total (PT) is also a comparison criteria which is defined as

$$PT = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n-i+1} \hat{y}_{ij}}{\sum_{i=1}^{n} \sum_{j=1}^{n-i+1} y_{ij}}, \tag{5.3}$$

while for many triangles becomes

$$PT = \frac{\sum_{k=1}^{N} \sum_{i=1}^{n} \sum_{j=1}^{n-i+1} \hat{y}_{ij}}{\sum_{k=1}^{N} \sum_{i=1}^{n} \sum_{j=1}^{n-i+1} y_{ij}}. \tag{5.4}$$

The RMSE and PT measure the model-fit with respect to observations where PT closest to 100 is accepted while RMSE is preferred to be the smallest.

Table 5.12: RMSE and PT of Individual QR and Longitudinal QR

	Root Mean Square Error			**Percentage Total**		
Quantile	**Company A**	**Company B**	**Londitudinal**	**Company A**	**Company B**	**Londitudinal**
50%	457.52	248.60	398.07	82.78	90.37	84.02
60%	455.12	244.40	389.59	90.26	93.66	93.32
75%	511.97	263.43	388.36	123.37	106.04	106.02
90%	730.79	466.38	536.24	158.87	139.59	151.32
95%	693.67	466.38	762.02	158.87	139.59	185.72
99.5%	730.79	466.38	789.86	158.87	139.59	186.01

According to the comparison criteria, the longitudinal algorithm provides the smaller RMSE when using the 75% quantile resulting to better fit of the data, better estimations.

So, a combination of different companies or lines of business provides a better estimation of the total reserve. In case of using the PT criterion, we take exactly the same results and the 75% quantile produces the best fit. If we make estimations separately, the suggested models for both triangles use quantiles below 75% which means weak prudence.

Backtesting is also important for comparisons. The real IBNR for Company A is 1695553 while for Company B is 2483699. Table 5.11 shows the total reserves for all methods. As we can observe from Table 5.11, mainly the combined algorithm provides total reserve which is smaller than the Individual Quantile Regression. Note that the estimated reserve tends to be similar with the real one when using a quantile between 60% and 75% which leads to stable reserve estimations. This is not happening when using the Individual Quantile Regression where these estimations use smaller quantiles.

In Figure 5.4 we have incorporated the reserve estimation according to the Chain Ladder method. We can observe that the Chain Ladder method underestimates the IBNR in both companies. Specifically, the Chain Ladder estimation for Company A is 1624721 which is smaller than the real IBNR (1695553). The Chain Ladder estimation for company B is 1901883 which is significantly smaller than the real IBNR (2483699). Moreover, for company A, the individual quantile regression method gives the real IBNR at quantile 50% while the longitudinal algorithm gives the real IBNR at almost 80% quantile. For company B both methods give the real IBNR at almost 80% quantile.

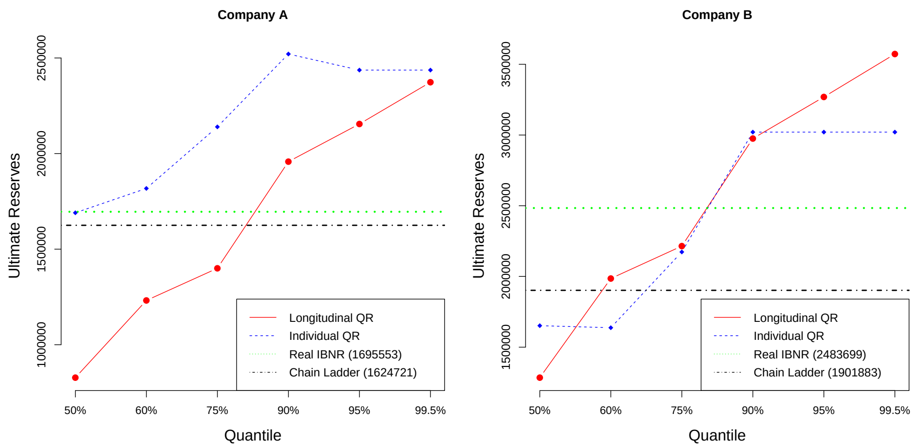

Figure 5.4: Ultimate Reserves for individual QR and londitudinal QR

Figures 5.5 and 5.6 display the reserve estimation for each accident year using the individual quantile regression and the longitudinal quantile regression. Each plot gives the reserve for different quantiles. For 75% quantile, the reserve estimations seem to be more stable in comparison with other quantiles. This is a controversial matter because we have seen that the 75% quantile is the best choice for the algorithm to be used.

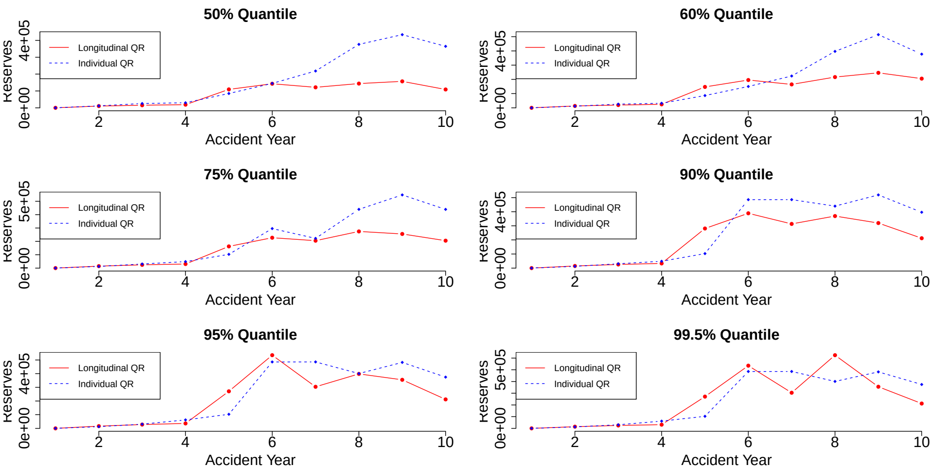

Figure 5.5: Reserves estimation for individual QR and longitudinal QR (Company A)

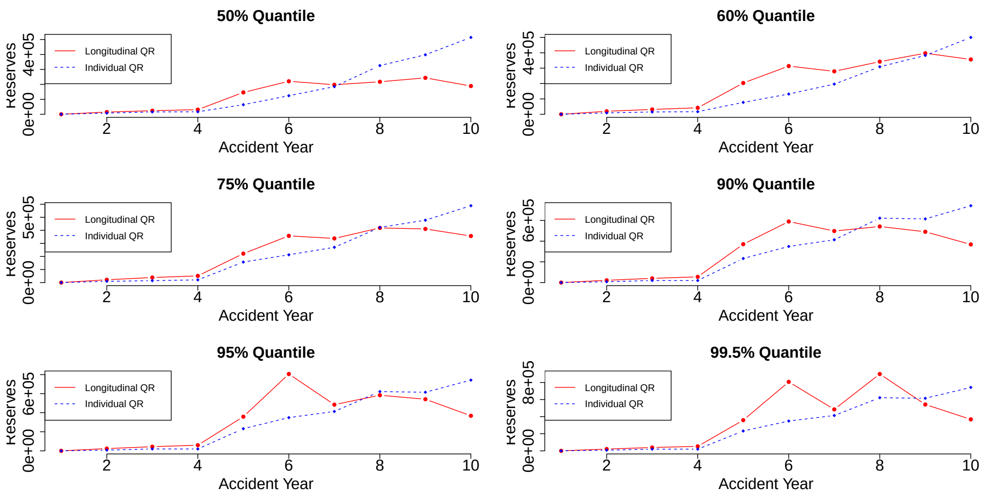

Figure 5.6: Reserves estimation for individual QR and longitudinal QR (Company B)

5.4 Risk Capital Requirement and Risk Margin

Solvency II and IFRS bring some significant changes, particularly in relation to the estimation of insurance liabilities. Generally, the Probability of Sufficiency is a measure of solvency in liability valuation (Dal Moro et al., 2017):

- Probability of Sufficiency below 50% indicates that the technical provisions are set below the central estimate which leads to under-reserved position.
- Probability of Sufficiency with values between 50% and 60% indicates the technical provisions are approximately at the level of central estimate which leads to weak prudence.
- For values of Probability of Sufficiency around to 75% the technical provisions are above the central estimate and this leads to sufficient prudence.
- Finally, if the Probability of Sufficiency is above 75%, the technical provisions are enough so as to lead to strong prudence.

According to Figures 5.1, 5.2 and 5.4 the cumulative evolution of the claims seem to be stable but there are some years where some big claims occurred.

5.4.1 Risk Margin

Dong et al. (2015) have shown how one can provide accurate estimation of risk margin and hence provision, instead of estimating the mean then applying a risk margin. Their method is more robust when the data is heavy tailed, has been used for the univariate quantile regression model and is suitable for a simple line of business (one run-off triangle). Moreover, the Risk Margin (RM) is exclusively based on the Solvency Capital Requirement (SCR) estimation. The overall risk margin (RM) according to the Cost-of-Capital methodology is calculated as follows (CEIOPS, 2009):

$$RM = \sum_{t\geq 0} \frac{SCR(t)}{(1+r_t)^{t+1}} \times CoC = \sum_{t\geq 0} \frac{VaR_{99.5\%}(R_t) - mean(R_t)}{(1+r_t)^{t+1}} \times CoC$$

where R_t is the estimated reserve for the accident year t, r_t is the risk-free rate for maturity, $SCR(t)$ is the Solvency Capital Requirement for the accident year t and CoC is the cost of capital rate.

Along with Best Estimate (BE), Risk Margin makes up the technical provisions and ensures that their value is equivalent to the amount that an (re)insurer would be expected to require in order to take over and meet the insurance obligations. Generally, Risk Margin increases the value of the technical provisions from the BE up to an amount which is equivalent to a theoretical level needed to transfer obligations to another (re)insurer. Risk margin represents what an (re)insurer would have to pay to the market to take on the BE liabilities. When the market takes on your BE liabilities, they will have to set aside capital to cover the SCR. This has a cost as the insurer buying your BE liabilities cannot use the capital backing the SCR for alternative profit generating activities (e.g. writing more new business). Therefore holding the SCR incurs a cost. The Risk Margin represents this cost.

In order to estimate the Risk Margin we are going to estimate the Solvency Capital Requirement as the difference between the 99.5% quantile and the 50% quantile of the reserves. The Cost of Capital is 6% (as Solvency II suggests) and we will suppose that the risk-free rate for maturity is $r_t = 1\%$ for all accident years[1]. We will also estimate the Risk Margin according to the Standard Approach where the SCR will be estimated using the bootstrap procedure. Table 5.13 presents the results of the Longitudinal algorithm while Tables 5.14 and 5.15 present the results of the individual quantile regression model and the bootstrap method.

Table 5.13: Risk Margin based on Longitudinal Quantile Regression (QR)

	Company A			Company B		
Accident Year	SCR	Capital Charge 6%	Discounted Capital Charge (1% discount rate)	SCR	Capital Charge 6%	Discounted Capital Charge (1% discount rate)
2007	0	0	0	0	0	0
2008	3819	229	227	6053	363	360
2009	9339	560	549	15156	909	891
2010	12821	769	747	21623	1297	1259
2011	161876	9713	9334	212537	12752	12255
2012	392467	23548	22405	585832	35150	33444
2013	182471	10948	10314	287601	17256	16256
2014	480954	28857	26916	681567	40894	38143
2015	198515	11911	11000	298613	17917	16546
2016	103126	6188	5658	179172	10750	9829
Total	**1545390**	**92723**	**RM=87148**	**2288153**	**137289**	**RM=128983**

Table 5.14: Risk Margin based on Individual Quantile Regression (QR)

	Company A			Company B		
Accident Year	SCR	Capital Charge 6%	Discounted Capital Charge (1% discount rate)	SCR	Capital Charge 6%	Discounted Capital Charge (1% discount rate)
2007	0	0	0	0	0	0
2008	646	39	38	-274	-16	-16
2009	5212	313	307	4378	263	258
2010	31717	1903	1847	3056	183	178
2011	17954	1077	1035	168249	10095	9701
2012	340735	20444	19452	224544	13473	12819
2013	267395	16044	15114	228378	13703	12909
2014	24486	1469	1370	295727	17744	16550
2015	48272	2896	2675	216602	12996	12002
2016	10000	600	549	227899	13674	12503
Total	**746418**	**44785**	**RM=42387**	**1368560**	**82114**	**RM=76902**

[1] The real risk-free rate for maturity is given by EIOPA

Table 5.15: Risk Margin based on Bootstrap (Poisson)

	Company A			Company B		
Accident Year	**SCR**	**Capital Charge 6%**	**Discounted Capital Charge (1% discount rate)**	**SCR**	**Capital Charge 6%**	**Discounted Capital Charge (1% discount rate)**
2007	0	0	0	0	0	0
2008	81326	4880	4831	112721	6763	6696
2009	80433	4826	4731	142344	8541	8372
2010	79564	4774	4633	127414	7645	7420
2011	152521	9151	8794	231367	13882	13340
2012	203568	12214	11621	280495	16830	16013
2013	187305	11238	10587	286199	17172	16177
2014	286217	17173	16018	500754	30045	28024
2015	462659	27760	25635	690254	41415	38246
2016	1229084	73745	67428	1110921	66655	60946
Total	**2762677**	**165761**	**RM=154279**	**3482470**	**208948**	**RM=195234**

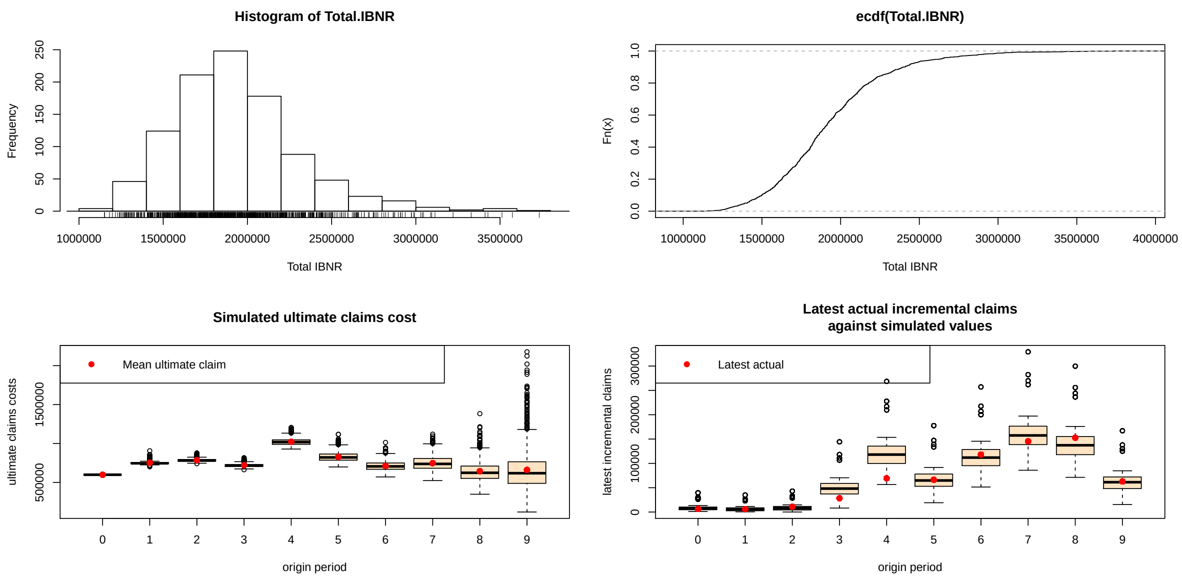

Figure 5.7: Bootstrap graphs for Company A

Remark 5.3. If the distribution of the reserves were known, then the mean of this distribution would be the Best Estimate, i.e. the amount to be paid as compensation to the beneficiaries. Nevertheless, this distribution is not known and for that reason many methodologies are used to estimate the Best Estimate such as the Bootstrap method. In case of Quantile Regression models, a specific quantile, which will provide estimations close to the mean, will be used in order to estimate the Best Estimate. For that reason we use the 50% quantile but it could be used a bigger one especially when the distribution has a long tail.

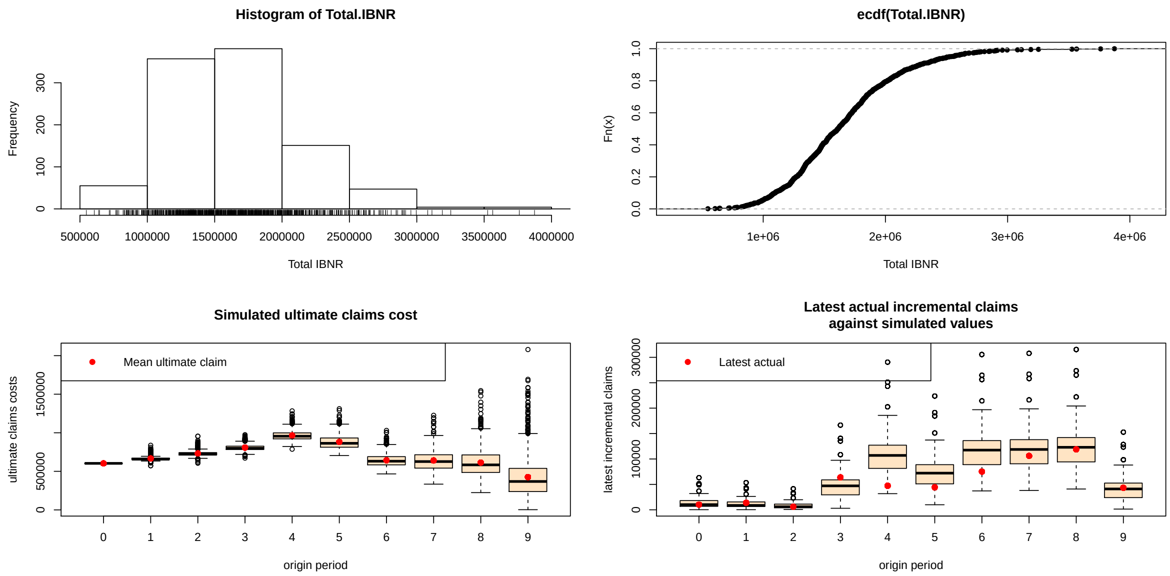

Figure 5.8: Bootstrap graphs for Company B

5.5 Concluding Remarks

We proposed quantile regression for longitudinal data in the framework of a general multivariate loss reserving model. Our model considered a combination of the between and within lines of business, taking into account the correlations and variation of run-off triangles.

We investigated a general insurance portfolio that consists of two correlated sub-portfolios (two auto run-off triangles). The least squares estimators investigated only changes in the mean, while the quantile regression characterized a particular point of a distribution, which provides a more complete description of the entire shape of the claims distribution.

According to Solvency II and IFRS, the solvency capital requirement (SCR) was provided based on the best estimate (BE) and in the sequel, the overall risk margin (RM), based on the Cost-of-Capital methodology, was calculated.

CHAPTER 6

General Conclusions

In this thesis we presented one of the most fundamental aspects of the insurance company, the management of loss reserves. The protection of the policyholders and the financial stability of the insurance market industry is a crucial aspect and the regulatory authorities intervene to ensure it. Reserve risk mainly focuses on the ability of an insurance company to cover the claims payment during the full run-off liabilities. Reliable reserves estimation indicated that the total reserves are sufficient to cover future payments, which arise from the incurred claims.

The key characteristics of the Solvency II were presented in Chapter 2, focusing on non-life risk that is separated into reserve risk and premium risk. Reserve risk deals with the liabilities of insurance policies. Solvency II is the new European Union (EU) legislation risk-sensitive system, which has been designed in order to make policyholders feel more secure and at the same time to create stable financial markets.

In Chapter 3 we have shown how robust estimation techniques can be incorporated in a loss reserving framework, providing a fair value for the estimation of outstanding reserves. Least squares estimators and robust estimators were applied to Ashe and Taylor data. The values of claims reserves were presented and the sensitivity of log-linear loss reserving model was shown by embedding one or two artificial outliers to the data set. The implementation of data showed the superiority of robust M-estimator and Robust ANOVA in comparison with the least squares estimator and the rest of robust estimators for claims reserves estimation.

In Chapter 4 we illustrated how random coefficient regression models can be incorporated in loss reserving techniques for the univariate case (one line of business) and the multivariate cases (several lines of business). These models provided a fair value for the estimation of outstanding reserves in cases we have indications that the run-off patterns are changing. In order to remediate the effect of outliers to the estimation of the total reserves, robust versions of the above two random coefficient models were applied. Implementing the data sets for both models we showed the superiority of robust M-estimator in comparison with the non-robust estimators.

An application of the Kalman Filter to a state-space model recursive algorithm was presented in Chapter 5, in order to estimate the reserves of an insurance company. The Kalman Filter algorithm was applied to systems that receive external physical disturbances (noises), aiming to create a new estimate of the state of the system without disturbances. In addition, the algorithm was extended by making it robust at extreme values, which, if ignored, will overestimate the final reserve required for the company's liability.

Finally in Chapter 6, we considered a quantile regression application in a multivariate context alternative to a multivariate chain-ladder model for a portfolio of correlated run-off triangles. We proposed a reserving problem for a nonlife insurance portfolio consisting of several run-off sub-portfolios corresponding to a different line of business that can be embedded within the quantile regression model for longitudinal data. The basic point of quantile regression is that it constitutes a robust method that helps to eliminate any outliers that appeared in the loss reserving data set. In addition to the estimation of total reserves, the risk margin was estimated based on the SCR as the difference between the 99.5% quantile and the 50% quantile of the total reserves.

From the above analysis, we can conclude that indeed the use of stochastic modelling can lead to the improvement of treatment for the variation problems of the total reserve as well as the correlation issues between different Lines of Business.

Bibliography

[1] Antonio, K. and Beirlant, J., 2006. Actuarial statistics with generalized linear mixed models. Insurance: Mathematics and Economics, 75, 643-676.

[2] Armstrong, R.D. and Frome, E. L., 1979. Least-absolute values estimation for one-way and two-way tables. Naval Research Logistic Quarterly, 26, 79-96.

[3] Bartlett, M.S., 1951. An inverse matrix adjustment arising in discriminant analysis. Annals of Mathematical Statistics. 22(1), 107-111.

[4] Bilodeau, M. and Duchesne, P., 2000. Robust estimation of the SUR model. Canad. J. Statist. 28, 277-288.

[5] Bornhuetter, R. L. and Ferguson, R. E., 1972. The Actuary and IBNR. Proceedings of the Casualty Actuarial Society, 59(112).

[6] Bradu, D. and Mundlak, Y., 1970. Estimation in lognormal linear models. JASA 6, 198-211.

[7] Braun, C., 2004. The prediction error of the chain ladder method applied to correlated run-off triangles. ASTIN Bulletin, 34(2), 399-423.

[8] Brown, B.M. and Wang Y.G., 2005. Standard errors and covariance matrices for smoothed rank estimators. Biometrica 92, 149-158.

[9] Busse, M., Müller, U. and Dacorogna, M., 2010. Robust estimation of reserve risk, *ASTIN Bulletin.* 40(2), 453-489.

[10] Caroni, C., 1987. Residuals and influence in the multivariate linear model. The Statistician, 36, 365-370.

[11] Carrato, A., 2016. A practitioner's introduction to stochastic reserving. PhD Thesis.

[12] Carpenter, G., 2014. http://www.guycarp.com/content/guycarp /en/home/the-company/media-resources/glossary/a.html.

[13] Cazan, I., 2011. Kalman filters, Honor Project.

[14] Chan, J.S.K., 2015. Predicting loss reserves using quantile regression Running title: Quantile regression loss reserve models. Journal of Data Science, 13(1), 127-156.

[15] CEIOPS, 2007. QIS3, technical specifications.

[16] Final CEIOPS' Advice for Level 2 Implementing Measures on Solvency II: Technical Provisions - Article 86 (d), 2009. Calculation of the Risk Margin.

[17] Chan, J.S.K., Choy, S.T.B. and Makov, U.E. 2008. Dynamic and robust models for loss reserves using generalized-t distribution. ASTIN Bulletin, 38, 207-230.

[18] Chan, J.S.K., 2015. Predicting loss reserves using quantile regression. Running title: Quantile regression loss reserve models. Journal of Data Science, 13(1), 127-156.

[19] ChandraShekhar P., Kumar, N. and Warrier, S.R., 2007. Journey of insurer solvency regulations.

[20] ChandraShekar, P. and Warrier, S.R., 2007. Risk-based capital management: A principles based approach to insurer solvency management.

[21] Chen, L., Wei L.J. and Parzen M.I., 2004. Quantile regression for correlated observations. In: Proceedings of the Second Seattle Symposium in Biostatistics. Lecture Notes in Statistics, Vol. 179. New York: Springer, 51-69.

[22] Christofides, S., 1990. Regression models based on log-incremental payments. Claims Reserving Manual 2, Institute of Actuaries, London.

[23] Crow, E.L. and Shimizu, K., 1988. Lognormal distributions, theory and applications. Marcel Dekker, Inc., New York.

[24] Dal Moro, E. and Krvavych, Y., 2017. Probability of Sufficiency of Solvency II Reserve Risk Margins: Practical Approximations. ASTIN Bulletin. Earlier version of the paper was presented at ASTIN Colloquium, Sydney 2015. Available at SSRN: https://ssrn.com/abstract=2652088.

[25] De Jong, P. and Zehnwirth, B., 1983. Claims reserving, state-space models and the Kalman Filter. Journal of the Institute of Actuaries, 110, 157-81.

[26] De Jong, P. and Heller G.Z., 2008. Generalized linear models for insurance data. Cambridge University Press.

[27] De Jong, P., 2012. Modeling dependence between loss triangles. American Actuarial Journal, 16(1), 74-86.

[28] Diggle, P.J., Heagerty P.J., Liang K.Y. and Zeger S.L., 2002. Analysis of longitudinal data. New York: Oxford University Press.

[29] Dong, A.X.D., Chan, J.S.K. and Peters, G.W., 2015. Risk margin quantile function via parametric and non-parametric bayesian approaches. ASTIN Bulletin, 503-550.

[30] Doherty, N.A. and Garven, J.R., 1995. Insurance cycles: Interest rates and the capacity constraint model. The Journal of Business, 68(3), 383.

[31] Efron, B., 1979. Bootstrap methods: Another look at the jackknife. Ann. Stat. 7, 1-26.

[32] England, P.D. and Verrall, V.J., 1999. Analytic and bootstrap estimates of prediction errors in claims reserving. Insurance: Mathematics and Economics, 25, 281-93.

[33] England, P.D. and Verrall, V.J., 2002. Stochastic claims reserving in general insurance. Presented to the institute of actuaries, 28 January 2002.

[34] Farcomeni A., 2012. Quantile regression for longitudinal data based on latent markov subject-specific parameters. Statistics and Computing, 22(1), 141-152.

[35] Finney, D.J., 1951. On the distribution of a variate whose logarithm is normally distributed. Supplement to the journal of the royal statistical society, 7, $155-161$.

[36] Fitzenberger, B., Koenker R. and Machado J.F., 2001. Economic applications of quantile regression. Special issue of empirical economics. Springer, Heidelberg.

[37] Fu, L. and Wang, Y.G., 2012. Quantile regression for longitudinal data with a working correlation model. Comput. Stat. Data Anal., 56, 2526-2538.

[38] Gesmann, M. and Zhang, Y., 2010. Chain Ladder: Mack, bootstrap, Munich and multivariate-chain-ladder methods. R package version 0.1.3.

[39] Gervini, D. and Yohai, V.J., 2002. A class of robust and fully efficient regression estimators. Annals of Statistics, 30, 583-616.

[40] Geraci, M. and Bottai M., 2007. Quantile regression for longitudinal data using the asymmetric Laplace distribution. Biostatistics, 8(1), 140-154.

[41] Geraci, M., 2014. lqmm: Linear quantile mixed models. R package version 1.5.1, URL http://CRAN.R-project.org/package=lqmm.

[42] Gilchrist, W.G., 2000. Statistical modelling with quantile functions. Chapman & Hall.

[43] Greene, W., 2003. Econometric Analysis, 5th ed. Prentice Hall, Upper Saddle River, N.J.

[44] Greene, W., 2000. Econometric Analysis. Prentice Hall, Upper Saddle River, N.J.

[45] Griffiths W.E., 1972. Estimation of actual response coefficients in the Hildreth-Houck random coefficient model. Journal of the American Statistical Association, 67(339), 633-635.

[46] Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J. and Stahel, W.A., 1986. Robust Statistics; The Approach Based on Influence Functions. Wiley, New York.

[47] Hardle, W., Klinke, S. and Muller, M., 1999. XploRe Learning Guide. Springer, Berlin.

[48] Harrison, D.A. and Hulin, C.L., 1989. Investigations of absenteeism: Using event-history models to study the absence-taking process. Journal of Applied Psychology, 74, 300-316.

[49] Heien, D.M., 1968. A Note on log-linear regression, Journal of the American Statistical Association, 63.

[50] Hildreth, C. and Houck, J.P., 1968. Some estimators for a linear model with random coefficients. Journal of American Statistical Association.

[51] Hogg, R.V. and Craig, A.T., 1978. Introduction to mathematical statistics. Macmillan Publishing Co., Inc., New York.

[52] Holmberg, R.D., 1994. Correlation and the measurement of loss reserve variability. In CAS Forum, 1-247.

[53] Hössjer, O., 1992. On the optimality of S-estimators, *Statistics and Probability Letters,* 14, 413-419.

[54] Hsiao, C., 1986. Analysis of Panel Data. Econometric Society Monographs.

[55] Hsiao, C., 1974. Statistical Inference for a model with both random cross-sectional and time effects. X●nternational economic review, 15, 12-30.

[56] Hsiao, C., 2003. Analysis of Panel Data, 2nd ed. New York: Cambridge University Press.

[57] Huber, P.J., 1973. Robust Statistics: asymptotics, conjectures and Monte Carlo. Annals of Statistics, 1, 799-821.

[58] Huber, P.J. and Dutter, R., 1974. Numerical solutions of robust regression. COMPSTAT 1974, *Proc. in Computational Statistics*, Ed.: G. Bruckman, Physika Verlag, Vienna.

[59] Huber, P.J., 1981. Robust Statistics. Wiley, New York.

[60] Hubert, M., Verdonck, T. and Yorulmaz, O., 2017. Fast robust SUR with economical and actuarial applications. Statistical Analysis and Data Mining, 10(2), 77-88.

[61] Hyndman, R.J. and Koehler A.B., 2006. Another look at measures of forecast accuracy. International Journal of Forecasting, 22(4), 679-688.

[62] Jung, S., 1996. Quasi-likelihood for median regression models. J. Am. Statist. Ass., 91, 251-257.

[63] Kaas, R., Goovaerts, M., Dhaene, J. and Denuit, M., 2009. Modern Actuarial Risk Theory: Using R. Springer, Berlin.

[64] Kailath, T., Sayed, A.H. and Hassibi, B., 2000. Linear Estimation. Prentice Hall, New Jersey.

[65] Kalman, R. E., 1960. A new approach to linear filtering and prediction problems. Journal of Basic Engineering, 82, 340-345.

[66] Klugman, S.A., Panjer, H.H. and Willmot, G.E., 2008. Loss Models: From Data to Decision, Third edition. John Wiley, Hoboken NJ.

[67] Koenker, R. and Basset G., 1978. Regression quantiles. Econometrica, 46(1), 33-50.

[68] Koenker, R., 2001. Linear hypothesis: regression (quantile). In International Encyclopedia of Social & Behavioral Sciences (Smelser NJ and Baltes PB eds), 8893-8899. Elsevier.

[69] Koenker, R., 2005. Quantile Regression. Cambridge University Press.

[70] Koenker, R., 2011. quantreg: Quantile regression. R package version 4.76. https://cran.r-project.org/web/packages/quantreg/index.html.

[71] Koenker, R. and D'Orey, V., 1987. Computing regression quantiles. Applied statistics, 36, 383-393.

[72] Koenker, R. and D'Orey, V., 1994. A remark on algorithm AS229: Computing dual regression quantiles and regression rank scores. Applied Statistics, 43, 410-414.

[73] Koenker, R. and Portnoy, S., 1990. M-estimation of Multivariate Regressions. JASA, 85, 1060-1068.

[74] Koller D. and Friedman N., 2009. Probabilistic graphical models. Cambridge, MA: MIT Press.

[75] Kremer, E., 1982. IBNR claims and the two-way model of ANOVA. Scandinavian Actuarial Journal, 47-55.

[76] Kremer, E., 1997. Robust lagfactors. ASTIN colloquium 1997.

[77] Kremer, E., 2005. The correlated chain-ladder method for reserving in case of correlated claims developments. Blatter DGVFM, 27, 315-322.

[78] Krishnakumar, J., 1995. A General methodology of robust estimation for panel data models.

[79] Laird, N. and Ware, J., 1982. Random-effects models for longitudinal data. Biometrics, 38(4), 963-974.

[80] Lee, L.F. and Griffiths, W.E., 1979. The prior likelihood and best linear unbiased prediction in stochastic coefficient linear models. Center for economic research, Department of Economics, University of Minnesota, 79-107.

[81] Liang, K.Y. and Zeger S.L., 1986. Longitudinal data analysis using generalized linear models. Biometrika, 73, 13-22.

[82] Li, J., 2006. Comparison of stochastic reserving methods. Australian Actuarial Journal, 12(4), 489-569.

[83] Louis, G.D., 1996. UMVUE of the IBNR reserve in a lognormal linear regression model. Insurance Mathematics and Economics, 18(1), 43-57.

[84] Mack, T., 1991. A simple parametric model for rating automobile insurance or estimating IBNR claims reserves. ASTIN Bulletin, 21(1), 93-109.

[85] Mack, T., 1993. Distribution-free calculation of the standard error of chain-ladder reserve estimates. ASTIN Bulletin, 23, 213-225.

[86] Mack, T., 1994a. Measuring the variability of chain-ladder reserve estimates. Spring Forum.

[87] Mack, T., 1994b. Which stochastic model is underlying the chain-ladder method? Insurance: Mathematics and Economics, 15, 133-138.

[88] Marco A. Ramez Corzo, Enrique de Alba, 2005. Bayesian claims reserving when there are negative values in the runoff triangle. In 40^{th} Actuarial Research Conference, ITAM, Mexico.

[89] Maronna, R.A., Martin, R.D. and Yohai, V.J., 2006. Robust Statistics: Theory and Methods, John Wiley and Sons, New York.

[90] McCullagh, P. and Nelder J.A., 1989. Generalized Linear Models (2nd ed.). London: Chapman and Hall.

[91] McCulloch, C.E., Shayle R.S. and Casella G., 2006. Variance Components (Wiley Series in Probability and Statistics). New York: Wiley-Interscience. ISBN 0-470-00959-4.

[92] Merz, M. and Wuthrich, M., 2008. Stochastic claims reserving methods in Insurance, Wiley Finance.

[93] Merz, M. and Wuthrich, M., 2008*a*. Prediction error of the chain ladder reserving method applied to correlated run off trapezoids. Annals of Actuarial Science, 2(1), 25-50.

[94] Merz, M. and Wuthrich, M., 2008*b*. Prediction error of the multivariate chain ladder reserving method. North American Actuarial Journal, 12(2), 175-197.

[95] Merz, M., Wuthrich, M. and Lysenko, N., 2009. Uncertainty of the claims development result in the chain ladder method. Scandinavian Actuarial Journal, 63-84.

[96] Michael, R.C. and Robert, A.L., 2001. An introduction to bootstrap methods with applications to R. Wiley Publishing.

[97] Moore, J.B. and Anderson, B.D.O., 1980. Coping with singular transition matrices in estimation and control stability theory, Int. J. Contr., 3.

[98] Moryson, M., 1998. Testing for random walk coefficients in regression and state space models. Physica-Verlag, New York.

[99] Mosteller, F., 1946. On some useful inefficient statistics. Ann. Math. Statist., 17(4), 377-408.

[100] Nelder, J.A. and Wedderburn R.W.M., 1972. Generalized linear models. Journal of the Royal Statistical Society, Series A, 135, 370-384.

[101] Ohlsson, E. and Lauzeningks, J., 2008. The one-year non-life insurance risk. Conference paper, Astin Colloquium, Manchester.

[102] Pang, L., Lu, W. and Wang, H.J., 2012. Variance estimation in censored quantile regression via induced smoothing. Computational Statistics and Data Analysis, 56, 785-796.

[103] Parzen, M.I., Wei, L.J. and Ying Z., 1994. A resampling method based on pivotal estimating functions. Biometrika, 81, 341-350.

[104] Pena, D., and Yohai, V., 1999. A fast procedure for outlier diagnostics in large regression problems. Journal of the American Statistical Association, 94, 434-445.

[105] Pentikainen, T., 1982. Solvency of insurers and equalization reserves, Vol. I, General Aspect.

[106] Peremans, K., Van Aelst, S. and Verdonck, T., 2018. A robust general multivariate chain ladder method, Risks.

[107] Petris, G., Petrone, S. and Campagnoli, P., 2010. An R package for dynamic linear models. Journal of Statistical Software, 36(12), 1-16.

[108] Pinheiro, J. and Bates, D.M., 2000. Mixed-Effects Models in S and S-PLUS. New York: Springer.

[109] Pinheiro, J., Andrade e Silva, J. M. and De Lourdes Centeno, M., 2003. Bootstrap methodology in claim reserving. Journal of Risk and Insurance, 70(4), 70-714.

[110] Pitselis, G., 2005. Application of M-estimators to cross-section effect models. Communication in Statistics-Simulation and Computation, 34, 1-16.

[111] Pitselis, G., Grigoriadou, V., Badounas, I., 2015. Robust loss reserving in a log-linear model. Insurance: Mathematics and Economics, 64, 14-27.

[112] Pizzinga, A., Simoes A.R. and Fernandes, C., 2010. A row-wise stacking of the runoff triangle: State space alternatives for IBNR reserve prediction, Astin Bulletin.

[113] Pontius, R., Thontteh O. and Chen H., 2008. Components of information for multiple resolution comparison between maps that share a real variable. Environmental Ecological Statistics, 15, 111-142.

[114] Pröhl, C. and Schmidt, K.D., 2005. Multivariate chain-ladder, Dresdner Schriften zur Versicherungsmathematik.

[115] Quarg, G. and Mack T., 2004. Munich chain ladder. Blatter DGVFM, Band XXVI, 4, 597-630.

[116] Radtke, M., Schmidt, K.D. and Schnaus, A., 2012. Handbook on Loss Reserving, EAA Lecture Notes, European Actuarial Academy, Springer.

[117] Raj, B. and Ullah, A., 1981. Econometrics: A Varying coefficients approach. Econometrics.

[118] Rantala, J., 1982. Solvency of insurers and equalization reserves, Vol. II, Risk Theoretical Model.

[119] Renshaw, A.E., 1989. Chain-ladder and interactive modeling (claims reserving and GLIM). *Journal of the Institute of Actuaries,* 116, 559-587.

[120] Renshaw, A.E. and Verrall, R.J., 1998. A stochastic model underlying the chain-ladder technique. British Actuarial Journal, 4(4), 903-923.

[121] Rousseeuw, P.J., 1984. Least median of squares regression. Journal of the American Statistical Association 79, 871-881.

[122] Rousseeuw, P.J. and Yohai, J.V., 1984. Robust regression by means of S-estimator. In Robust and Nonlinear Time Series (J. Franke, W. Härdle and D. Martin, (Eds.), Lecture Notes in Statistics., Springer, New York, 26, 256-272.

[123] Rousseeuw, P. and Leroy, A., 1987. Robust Regression and Outlier Detection. Wiley, New York.

[124] Schlemmer, D.D. and Tarkowski, T., 2013. Reserving in two Steps: Total IBNR = Pure IBNR + IBNER, in: CAS E-Forum.

[125] Schmidt, K.D., 2006. Methods and models of loss reserving based on run-off triangles: a unifying survey. Casualty Actuarial Society. Forum, 269-317.

[126] Schmidt, K.D., 2006. Optimal and additive loss reserving for dependent lines of business. CAS Forum (fall), 319-351.

[127] Searle, S.R., Casella G. and McCulloch C.E., 1992. Variance Components. New York: John Wiley and Sons.

[128] Serfling, R., 2002. Efficient and robust fitting of log-normal distributions. North American Actuarial Journal, 6, 95-109.

[129] Shimizu, K., 1988. In Lognormal Distributions, Theory and Applications. Marcel Dekker, Inc., New York.

[130] Shi, P. and Frees, E., 2011. Dependent loss reserving using copulas. ASTIN Bulletin, 41(2), 449-486.

[131] Swamy, P.A.V.B., 1970. Efficient inference in a random coefficient regression model. Econometrica, 38, 311-323.

[132] Swamy, P.A.V.B., 1971. Statistical inference in random coefficient models. New York: Springer-Verlag.

[133] Swamy, P.A.V.B., 1973. Criteria, constraints, and multicollinearity in random coefficients regression models. Annals of Economic and Social Measurement, 2, 429-50.

[134] Swamy, P.A.V.B., 1974. Linear models with random coefficients, in Paul Zarembka (ed.), Frontiers of Econometrics. New York Academic Press, 143-68.

[135] Solvency II Implementation Date, 2014. http://www.lloyds.com/themarket/operating-at-lloyds/solvency-ii/legislativedevelopments/solvency-ii-implementation-date.

[136] Taylor, G.C., Ashe, F.R., 1983. Second moments of estimates of outstanding claims. Journal of Econometrics, 23, 37-61.

[137] Taylor, G.C., 1986. Claim reserving in nonlife insurance. North-Holland, Amsterdam.

[138] Taylor, G.C., 2000. Loss reserving - An actuarial perspective. Kluwer Academic Publishers, Norwell, Mass.

[139] Taylor, G.C. and Mcguire, 2005. Synchronous bootstrapping of seemingly unrelated regressions. Paper presented at the 36th International ASTIN Colloquium, September 4-7, ETH, Zurich.

[140] Taylor, G.C. and Mcguire, 2007. A synchronous bootstrap to account for dependencies between lines of business in the estimation of loss reserve prediction error. North American Actuarial Journal, 11(3), 70-88.

[141] Taylor, G. and Jing X., 2016. An empirical investigation of the value of claim closure count information to loss reserving. Variance, 10(1), 75-120.

[142] Tukey, J.W., 1977. Explanatory Data Analysis, Addison-Wesley.

[143] Tusell, F., 2011. Kalman filtering in R. Journal of Statistical Software. 39. 10.18637/jss.v039.i02.

[144] Van Eeghen, J., 1981. Loss reserving methods. Surveys of Actuarial Studies No. 1, Nationale Nederlanden N.V.

[145] Venter, G.G., Tampubolon and Dumaria R., 2010. Robustifying reserving. Variance, 4(2), 136-154.

[146] Verdonck, T., Van Wouwe, M. and Dhaene, J., 2009. A robustification of the chain-ladder method. North American Actuarial Journal, 13(2), 280-298.

[147] Verdonck, T. and Debruyne, M., 2010. The influence of individual claims on the chain-ladder estimates: Analysis and diagnostic Tool. Insurance: Mathematics and Economics, 48, 85-98.

[148] Verrall, R.J., 1989. A state space representation of the chain ladder linear model. Journal of the institute of actuaries, 116: 589-609.

[149] Verrall, R.J., 1991. On the estimation of reserves from log-linear models. Insurance: Mathematics and Economics, 10, 75-80.

[150] Verrall, R.J., 1994. A method for modelling varying run-off evolutions in claims reserving. ASTIN Bulletin, 24, 325-332.

[151] Verrall, R.J., 1994. Statistical methods for the chain-ladder technique. Casualty Actuarial Society Forum, Spring 1994, 393-446.

[152] Verrall, R.J., 2000. An investigation into stochastic claims reserving models and the chain-ladder technique. Insurance: Mathematics and Economics, 26, 91-99.

[153] Verdonck, T., Van Wouwe M. and Dhaene, J., 2009. A robustification of the chain-ladder method. Nourth American Actuarial Journal, 13(2), 280-298.

[154] 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, 85-98.

[155] Wang, Y.G., Lin, X. and Zhu, M., 2005. Robust estimating functions and bias correction for longitudinal data analysis. Biometrics. The International Biometric Society, 61(3), 684-691.

[156] Wang, Y.G., Shao, Q. and Zhu, M., 2009. Quantile regression without the curse of unsmoothness. Computational Statistics and Data Analysis, 53, 3696-3705.

[157] Wiendorfer B., 2012. A practical guide to the use of the chain-ladder method for determining technical provisions for outstanding reported claims in non-life insurance. Working Paper Series 77, Fachhochschule des BFI, Wien.

[158] Willmott, C. and Matsuura, K., 2006. On the use of dimensioned measures of error to evaluate the performance of spatial interpolators. International Journal of Geographic Information Science, 20, 89-102.

[159] Wüthrich, M. and Merz, M., 2008. Stochastic Claims Reserving Methods in Insurance, Wiley Finance.

[160] Yin, G. and Cai, J., 2005. Quantile regression models with multivariate failure time data. Biometrics, 61, 151-161.

[161] Yohai, J.V., Stahel, W.A. and Zamar, R.H., 1991. A procedure for robust estimation and inference in linear regression, in Stahel, W. A. and S. W. Weisberg, (Eds.), Directions in Robust Statistics and Diagnostics,, Part II, Springer-Verlag, New York.

[162] Yohai, J.V. and Zamar, R.H., 1997. Optimal locally robust M-estimates of regression. Journal of Statistical Planning and Inference, 64(2), 309-323.

[163] Yu K.M. and Zhang J., 2005. A three-parameter asymmetric Laplace distribution and its extension. Communications in Statistics: Theory and Methods, 34, 1867-1879.

[164] Zeger, S.L. and Liang, K.Y., 1986. Longitudinal data analysis for discrete and continuous outcomes. Biometrics, 42, 121-130.

[165] Zehnwirth B., 1996. Kalman filters with applications to loss reserving. Research Paper Number 35, Centre for Actuarial Studies, The University of Melbourne.

[166] Zehnwirth, B. and Barnett, G., 2001. Reserving for multiple excess layers. Paper at the ASTIN Colloquium at Washington.

[167] Zellner, A., 1962. An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias. Journal of the American Statistical Association, 57, 348-368.

[168] Zhang, Y., 2010. A general multivariate chain ladder model. Insurance: Mathematics and Economics, 46, 588-599.

[169] Zorn, C.J.W., 2001. Generalized estimating equation models for correlated data: A review with applications. American Journal of Political Science, 45, 470-490.