THE ACCURACY AND ROBUSTNESS OF THE CORRELATION BOOTSTRAP METHOD AND THE IMPLICIT CORRELATION METHOD FOR LOSS-RESERVING TRIANGLES.

Exploring the impact of data quality and outliers on the correlation estimate and challenging the assumptions of these methods.

Author:
Martijn Remmelink

A thesis submitted for the fulfillment of the requirements for the Financial Engineering and Management track of the MSc program in Industrial Engineering and Management at the University of Twente.

Supervisors:
T.M. van Altena MSc – Senior actuary Achmea
dr. B. Roorda – University of Twente
dr. R.A.M.G. Joosten - University of Twente

University:
Faculty of Behavioural, Management and Social sciences,
University of Twente

April 7, 2021

Preface

It is a pleasure to present my master's thesis to you. This thesis serves as the result of half a year of research at Achmea N.V. and marks the end of my student time at the University of Twente. Although the research was performed at home as a result of the Covid-19 restrictions, I still learned a lot from my colleagues at Achmea and got interesting insights into the work of an actuary.

First of all, I would like to thank Martijn van Altena, my company supervisor, for his time, patience, and the possibility to ask questions any time. I learned a lot during our meetings and often got helpful insights which definitely had a great impact on the quality of the end product. I also want to thank the Actuarial S&I team of Achmea, I enjoyed the Monday-morning sessions and got interesting insights into the work of the department. Furthermore, I would like to thank everyone at Achmea for the opportunity to do the research in these extraordinary times.

Secondly, I want to thank Berend Roorda, my university supervisor, for the guidance during the research. The meetings were helpful and often gave extra insights into how to describe and use the investigated methods. I also want to thank Reinoud Joosten, my second university supervisor, for the critical feedback and the improvement suggestions. This definitely increased the quality of the report.

Finally, I want to thank all people that were involved during my study. My family, study mates, roommates, friends, and all that helped me during my educational journey. I am thankful for the great time and I am excited to start with the first steps of my professional career.

I hope you enjoy reading this thesis.

Martijn Remmelink
Enschede, 27-03-2021

Management summary

This research is performed on behalf of the Actuarial department of Achmea N.V. located in Zeist. Achmea is one of the largest insurance companies in the Netherlands, offering a wide range of insurance products. Since January 2016, Solvency II guides as the leading framework for the prudential supervision of insurance companies. One of the most important aspects of Solvency II is the Solvency Capital Requirement (SCR). The SCR is the capital requirement that makes sure that the insurance company can pay all obligations in the upcoming year with a probability of 99.5%.

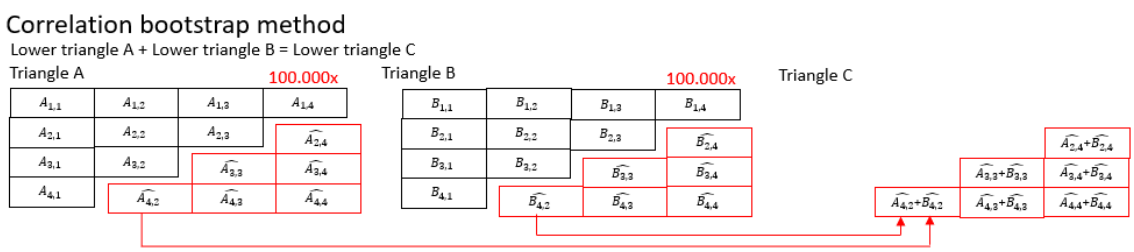
As a result of the size of Achmea and the high variety of insurance products Achmea offers, Achmea uses a partial internal model (PIM) in the aggregation of the reserve risk. In this PIM, the aggregation of the risk capitals starts at the level of homogeneous risk groups (HRGs), a level lower than Solvency II prescribes. As a result, Solvency II does not provide correlation matrices at the level of HRGs and so, Achmea has to come up with these tables themselves. These correlation tables are important, as the correlation parameter decides the amount of diversification benefit that can be obtained, as it is not likely that all HRGs will reach the SCR at the 99.5% level at the same time. The correlation matrices for the HRGs are currently determined by expert panels, using a qualitative method.

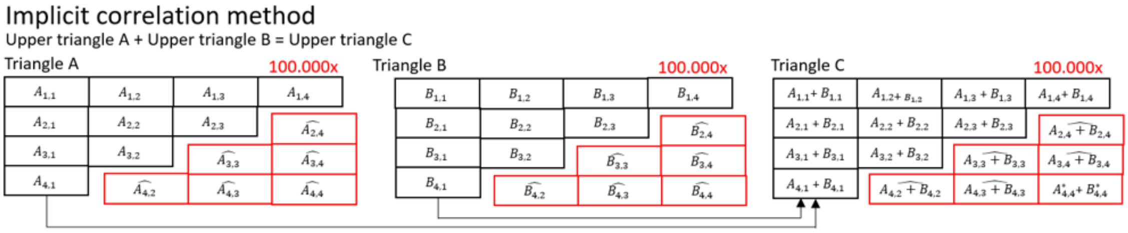
In this research, it is investigated if it is possible to use a quantitative method to determine the correlation parameter between the risk capitals of different HRGs. Specific attention will be paid to the correlation bootstrap method and the implicit correlation method. Furthermore, the performance of these methods is investigated when applied to HRGs with heterogeneous characteristics, as this is currently unknown. The research question which will be answered during this research is:

How accurate are the correlation bootstrap method and the implicit correlation method when estimating the correlation between reserving triangles and how sensitive are the correlation estimates to data quality, outliers, and a violation of the assumptions of these methods?

The correlation bootstrap method and the implicit correlation method both use bootstrapping techniques to create a distribution of the profit and loss at the end of the year. By bootstrapping the reserve triangle of two HRGs and creating a vector with expected settled claims of the combined portfolio of two HRGs, it is possible to derive the implied correlation coefficient by using the standard formula. The difference between the correlation bootstrap method and the implicit correlation method is the way the third vector with the expected settled claims for the upcoming year for the aggregated portfolio is obtained.

In the correlation bootstrap method, the residuals are synchronously bootstrapped, to make sure the dependencies between the residuals of the two HRGs are unchanged during the bootstrapping process. This makes it possible to aggregate the best estimates of both portfolios creating a third value representing the best estimate for the combined portfolio for a specific scenario. By performing this process a sufficient number of times, a distribution of the best estimate for the combined portfolio can be created. In the implicit correlation method, the triangles containing the historical settlement data are added to each other. The two individual triangles and the combined triangle are individually bootstrapped, creating three vectors of best estimates. The difference is visualized in the following figures:





To investigate what characteristics of an HRG have an impact on the correlation estimates, a dataset generator is developed. This dataset generator makes it possible to generate two triangles by setting the initial claim amounts, the settlement period, the MAS parameter denoting the variance within a development period, and drawing residuals with a dependency from a statistical distribution. These parameters are derived from real datasets except for the residuals, these are drawn from the bivariate normal distribution. Before the parameters are derived from a real data set, the data sets need to be corrected for inflation as well as for the portfolio developments. The simulations are performed on incurred triangles as well as paid triangles. For both types of triangles, annually-annually (AA) triangles and annually-quarterly (AQ) triangles are used in the simulations.

By using 100.000 simulation runs per dataset, we could conclude that the correlation estimates are stable with a maximum observed standard deviation of 0.014. In the robustness test, we found that especially the correlation bootstrap method gives robust outcomes if one year of additional data is added to the dataset. This implies, that the correlation estimates will not differ too much from year to year, which is important for a potential implementation in practice. In the simulations in which one parameter of the triangles was tested at a time, it became clear that the MAS parameter creates the most deviations from the pre-set correlation for the correlation bootstrap method. For the implicit correlation method, the MAS parameter and the residual set create the most deviations from the pre-set correlation. The settlement period parameter and the initial claim amounts do not substantially affect the correlation estimates. Furthermore, the deviations are bigger at the higher correlation levels and the deviations are bigger for the AQ-triangle compared to the AA-triangle. We can also conclude that the outcomes for the paid triangles are significantly better than the outcomes for the incurred triangles.

In the second part of the simulation study, we challenged the assumptions belonging to both methods to see what the impact on the correlation estimates is. We can conclude that the correlation bootstrap method deals significantly better with outliers compared to the implicit correlation method. Notable is the impact one outlier has on the other residuals belonging to a development period and the corresponding correlation. As a result, we conclude that outliers need to be excluded from the dataset before the correlation bootstrap method and the implicit correlation method can be used. From the tests with not normally distributed residual sets, we concluded that the correlation estimates became slightly worse for both methods. This is in line with the observation for the outliers, as more extreme residuals will occur in a not normally distributed dataset. The last and probably most promising finding of this research is that it is possible to obtain an accurate correlation estimate between triangles with a different tail length. The SCR is mostly determined by the most recent accident years, as these years contain the most uncertainty. We can conclude that if 80% of the uncertainty in the SCR of both triangles is covered, an accurate correlation estimate can be derived.

We can conclude that the correlation bootstrap method consistently performed better compared to the implicit correlation method. The correlation bootstrap method can be used if there is no autocorrelation available in the triangles and there are no dependencies between the accident years. The process error needs to be excluded when using the correlation bootstrap method, as this reduces the available correlation between the residuals significantly.

Contents

1 Introduction

This study is conducted as the final part of the Master Financial Engineering and Management of the University of Twente. The study is performed at the Actuarial department of Achmea group, which will be introduced in Section 1.1. The focus of this master thesis will be on the aggregation of reserving capitals, which is an important aspect of Solvency II. In Section 1.2, an introduction to Solvency II will be given, followed by an introduction to the research problem in Section 1.3. In Section 1.4, the research questions are introduced and this section is finished with the outline of the remainder of the thesis in Section 1.5.

1.1 Company description

Achmea N.V. is one of the largest insurance companies in the Netherlands and has a rich history that dates to 1811. The company was founded in Friesland as cooperation to insure farmers against the risk of a fire destroying the farm. The cooperation grew fast to are one of the biggest insurance companies in the Netherlands today.

Over the years, many other insurance companies were acquired, of which Centraal Beheer, Interpolis, De Friesland Zorgverzekeraar, FBTO, Zilveren Kruis, and Avéro are the best known brands. Achmea employs approximately 13,800 employees and has an annual turnover of almost € 20 billion. Besides the insurance part of the company, it has a pension division as well.

The master thesis is carried out on behalf of the Actuarial department of Achmea Group. The Actuarial department is responsible for the methodology in determining the best estimates, the Risk Management department is responsible for the methodology of internal models used for determining the capital requirements for market risk and underwriting riks. This study will focus on non-life reserve risk on which this study will focus. In April 2020, a reorganisation took place, in which some responsibilities shifted from the Actuarial department to the Risk Management department and vice versa. This makes it possible to do a study into a risk management subject while performing the study at the Actuarial department. The study took place from 31-08-2020 until 28-02-2021.

1.2 Introduction to Solvency II

Solvency I was introduced in 1973 as a first step to the harmonization of the supervision of insurance companies in Europe. Solvency I set capital requirements for insurance companies, to ensure they are able to settle all claims to a certain degree and create a more competitive landscape within the European Union. However, the rules set in Solvency I were not capable of dealing with the high variety of risk profiles in the insurance sector and were therefore not aligned with the corresponding risks. As a result, it was decided that a new legislative framework was needed which was more widely applicable, and was able to generate capital requirements based on the specific markets a European insurance company finds itself in [Bafin, 2006*a*].

Since January 2016, Solvency II is the leading framework for the prudential supervision of insurance companies. The Solvency II framework serves 4 purposes [Bafin, 2006*b*]:

1. Make sure an insurance company has enough money available to settle all claims;
2. Prevent policyholders, the ones who bought the insurance policy, from the bankruptcy of an insurance company;
3. Having more insights into the financial position of the insurer and so giving the supervisor the possibility to intervene earlier;
4. Improve the trustworthiness of the financial sector, in particular the insurance sector.

The Solvency II framework consists of three pillars: risk quantification, risk management and transparency. These pillars are in line with the legislation of Basel III for the banking sector.

The first pillar, risk quantification, sets out qualitative and quantitative requirements for the calculation of the technical provisions and the Solvency Capital Requirement (SCR) using either a standard formula given by the European Insurance and Occupational Pensions Authority (EIOPA) or a partial internal model developed by the insurance company and approved by the Dutch National Bank (DNB).

The technical provisions exist of two parts, the best estimate, and the risk margin. The technical provisions are meant to quantify the amount another insurance company would have to pay for an immediate transfer of its obligations. The relations between the best estimate, the risk margin, the MCR and the SCR are visualised in Figure 1.

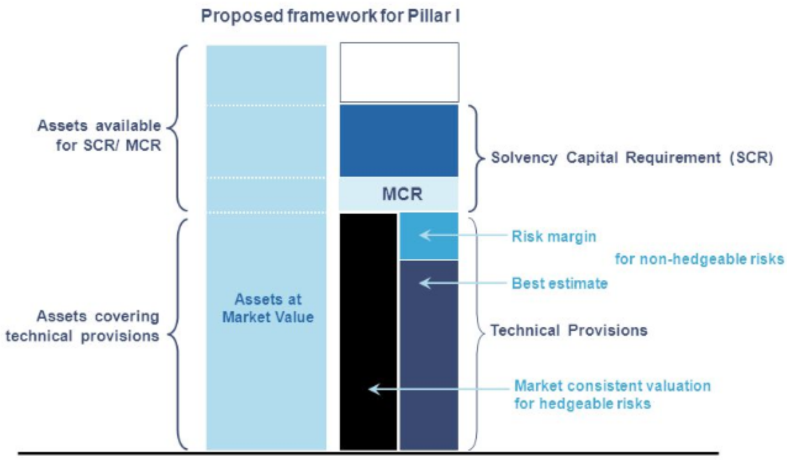


Figure 1: Pillar 1 Solvency Capital Requirements.

Solvency II requires the technical provisions to be a "best estimate" of the current liabilities relating to insurance contracts and a risk margin. The best estimate consists of the best estimate claim provisions which relate to events that have already occurred, and the best estimate premium provision which relates to future claim events.

The SCR is the capital requirement to make sure that the insurance company can pay all obligations in the upcoming year with a probability of 99.5%. The SCR incorporates main risks such as non-life underwriting, life underwriting, health underwriting, market, credit, operational and counterparty default risks, and must be determined and reported to the supervisor every quarter. Each of these main risks consists of one or more sub risks. Non-life underwriting risk consists of non-life catastrophe risk, lapse risk, premium risk and reserve risk. In this study, the focus will be on the non-life reserve risk, which is the risk that the currently available reserves are insufficient to cover their run-off over a one year time horizon [England et al., 2019].

To determine the Non-Life reserve risk, models are needed to have insights into the potential development patterns of the individual components belonging to the Non-Life reserve risk. In Section 1.3 these models are introduced.

1.3 Problem identification

The standard formula is not representative for the risk profile of Achmea due to the size of Achmea and the high variety of products Achmea offers. Therefore, Achmea developed and uses a partial internal model (PIM) to determine the SCR as well as most other large insurance companies in the Netherlands. In a PIM certain risks are quantified by using a company's own quantitative model (an internal model) and the remaining risks are quantified by using the standard formula.

One of the risks for which Achmea developed an internal model is reserve risk. Reserve risk is defined as the uncertainty about the amount and timing of the ultimate claim settlements in relation to existing liabilities, the best estimate. Achmea determines reserve risk at the level of homogeneous risk groups (HRG). Reserve Risk is determined by applying the bootstrap method, which is used to quantify the distribution of the ultimate claims (explained in Section 2.5). The total reserve risk is determined by the aggregation of the reserve risks at the level of HRG.

If it is not likely that two HRGs will yield unexpected liabilities in the same year, a diversification benefit can be reached. This diversification benefit is already reached if the correlation between the risk capitals of two HRGs is lower than 1. When the correlation is lower than 1, the combined required risk capital is lower than the sum of the individual risk capitals. So, the correlation parameter is important in aggregating risk capitals and determining the SCR.

Supervising parties like the DNB prefer a quantitative model in determining the correlation matrix between the reserving triangles of different HRGs over qualitative methods. Especially because quantitative methods give more exact insights into the methods used to come up with a correlation estimate and make it easier to reproduce the obtained numbers. However, to use quantitative models, enough representative data are needed to get reliable correlation estimates. As a result of the use of partial internal models in determining the reserve risk, it is not possible to use the standard two-level approach known in Solvency II [Filipović, 2009], as Solvency II does not prescribe the correlation matrices at the level of the reserve risk of HRGs. As a result, Achmea currently uses expert panels to determine the correlation estimates between reserve risks based on qualitative methods.

In literature, numerous papers have been written about the way the correlation parameter can be estimated between the variability in the reserving triangles of HRGs. However, these papers are often only applicable to specific theoretical circumstances that do not include all deviations seen in practice. Especially determining the correlation between reserving triangles of HRGs with heterogeneous characteristics is a topic that has not been addressed a lot, whereas HRGs are often mutually different in terms of settlement years, available historic data, variance, or distribution channel.

In this study, the correlation bootstrap method and the implicit correlation method will be investigated in depth. These two methods are chosen, as multiple papers are written about these methods and these seem to work properly on HRGs with homogeneous characteristics [Brickman et al., 1993] [Mack, 1993] [England & Verrall, 1999] [Kirschner et al., 2002]. Besides, the two methods have a lot in common and both use bootstrapping techniques to come up with the final correlation.

However, there is not enough knowledge about the quality and the robustness of the correlation estimates when these methods are used to determine the correlation between the reserving triangles of HRGs. Especially when aggregating HRGs with heterogeneous characteristics there are assumptions that make it currently impossible to determine the correlation between these HRGs, as the assumptions do not allow all heterogeneous characteristics.

As there is currently not enough insight into the accuracy, robustness and stability of the correlation estimates between different reserving triangles of HRGs based on the correlation bootstrap method and the implicit correlation method, additional research is required. To get more insights into the performance of both methods, research questions are defined. This will be the topic of Section 1.4.

1.4 Research objective

The objective of the study is to investigate if the correlation bootstrap method and the implicit correlation method can be used to obtain appropriate correlation estimates between different HRGs. To assess whether these methods can be used or not, we need to know how well both methods are capable of estimating the correlation parameter in terms of accuracy, robustness and stability. This is formulated in the following research question:

"How accurate are the correlation bootstrap method and the implicit correlation method when estimating the correlation between reserving triangles and how sensitive are the correlation estimates to data quality, outliers, and a violation of the assumptions of these methods?"

The research question will be answered by 4 sub-questions:

1. What is known in the literature about the correlation bootstrap method and the implicit correlation method in determining the correlation parameter between reserving triangles?
2. How accurately are the methods able to derive the correlation from datasets of which the correlation is known and what is the impact of outliers in the datasets for paid and incurred triangles?
3. How accurately are the correlation estimates if the assumptions of the methods are challenged?
4. How well are the methods capable of determining the correlations between reserving triangles with heterogeneous characteristics?

1.5 Thesis outline

In Section 2, we explain the main concepts needed to understand the process of obtaining the best estimate and the SCR. Furthermore, all different kinds of triangles will be introduced as well. In Section 3, we will investigate what is already known in the literature about deriving the correlation parameter from the datasets of different HRGs, together with the assumptions belonging to these methods and the role of the quality of the data. After we gathered the already known information of the methods, we will explain the setup of the simulation study in Section 4. To perform the simulations, we need clean datasets, which we will create by using a dataset generator. We will discuss the method used to built the dataset generator and how the dataset generator can be used in this study in Section 5. In Section 6, we will verify and validate the simulation models and finally, in Section 7 we will analyse the results of the simulation study. The report finishes with the conclusions, the recommendations, the limitations and the suggestions for future research in Section 8.

2 Best estimate reserve and reserve risk

The first section is dedicated to the best estimate. To obtain the best estimate, we need to acquire more knowledge about the data types and the methods used to derive the best estimate from a dataset. In the second section, we will explain the data types. With the knowledge about the data types, it is possible to explain the Chain Ladder method in the third section. The fourth section gives some insights into the history of bootstrapping techniques followed by an explanation of the bootstrapping technique in Section 2.5. This section is finished with how we can derive the best estimate and the profit and loss from the data sets. With the profit and loss, we are able to determine the SCR which can be used to derive the correlation between different HRGs.

2.1 Best estimate

The best estimate (BE) consists of the best estimate claim provisions which relate to events that have already occurred but have not been settled yet. In other words, it contains the expected required capital to fulfil all outstanding liabilities for an HRG. To obtain the BE, patterns available in historical data of a specific HRG can be used to predict the future. To understand how the BE is estimated, first the way historical data are ordered is addressed in Section 2.2. In Section 2.3, the Chain Ladder method is explained to come from historical data to a projection for the future.

By applying the Chain Ladder method, only a projection can be made based on the historical data available in the triangle. However, it is highly unlikely that the settlement pattern will be exactly the same in the future as it has been in the past. By applying bootstrapping techniques, we are able to obtain a distribution of potential future scenarios. This is the topic of Section 2.4. Based on the outcomes of the bootstrapping process, it is possible to determine the BE.

2.2 Triangles

To model the development of the payment streams related to a HRG and to make projections of the potential developments of these payment streams, run-off triangles are used. These triangles give insights into the timing and amount of claims that are settled at a specific moment in time. To understand the way the used triangles are built up, first some additional knowledge about the claim handling process needs to be acquired.

Not all claims are settled in the same year as they occur, in many cases it takes some time before the claim is received, judged and ultimately settled. Figure 2, visualizes a timeline of the development of a claim. The period between the occurrence and the moment the claim is reported to the insurance company is referred to as incurred but not reported (IBNR), in case a reported claim requires more money than initially estimated, this is denoted as incurred but not enough reported (IBNER). The period between the moment the claim is reported and the moment it is settled is denoted as reported but not settled (RBNS). The time between the occurrence of the claim and the ultimate settlement of the claim differs per HRG. In case of a car accident, the ultimate settlement of the damage to the car will in most cases take less than four years, whereas the ultimate settlement of physical damage may take more than a decade.

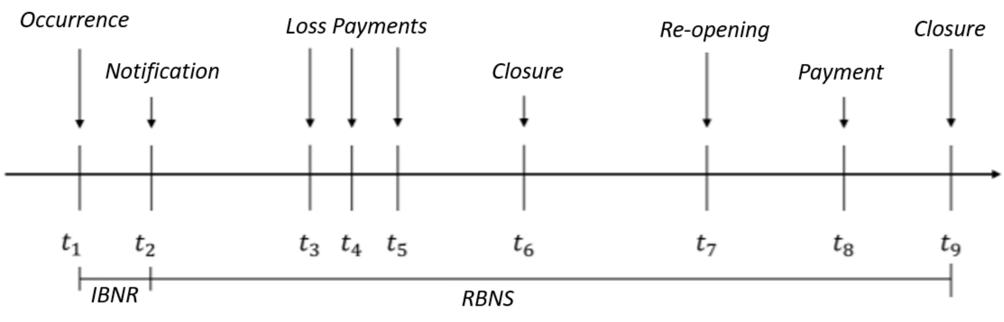


Figure 2: Development of a general insurance claim [Antonio & Plat, 2014].

To get insights into the development of claims relating to an accident year, run-off triangles are used. In a run-off triangle, the development periods are located in the columns and the accident years in the rows. An accident year tracks claims paid and reserves on accidents occurring within a particular year, regardless of when the claim occurred or when the policy was issued. The development period indicates the period in which the payments are settled belonging to a certain accident year. Triangles can have different structures and can contain different kinds of data, the ones relevant for this research are the paid triangle and the incurred triangle. Furthermore, a distinction is made between annually-annually (AA) triangles and annually-quarterly (AQ) triangles. In the next section, we will introduce the different kinds of triangles and explain their implications.

2.2.1 Paid triangles

The paid triangle is mostly used in literature and is as well the easiest triangle to model in a simulation study. When denoting the incremental claims paid by accident year i and development year j by $I_{i,j}$. Then the complete dataset can be described by:

$$\{I_{i,j}|\ i, j \in \mathbb{N} : 1 \le i \le n, 1 \le j \le n - i + 1\} \tag{1}$$

Which can be represented in the following way:

Accident year i	Development year j: 1	2	...	4	5
1	$I_{1,1}$	$I_{1,2}$	...	$I_{1,n-1}$	$I_{1,n}$
2	$I_{2,1}$	$I_{2,2}$	...	$I_{2,n-1}$	
⋮	⋮	⋮	$I_{i,n-i+1}$		
n-1	$I_{n-1,1}$	$I_{n-1,2}$			
n	$I_{n,1}$				

Figure 3: Incremental payments triangle.

In many cases, the incremental paid claims triangle is rewritten as a cumulative paid claims triangle, which is denoted by $C_{i,j}$, again i indicates the accident year and j the development year. The first development year is not different from the first development year of the incremental paid claims triangle. In the remainder of the triangle, the incremental claims belonging to a development period are added to the cumulative payments in the previous development period:

$$C_{i,1} = I_{i,1},\ \ 1 \le i \le n,\ \ i \in N \tag{2}$$

$$C_{i,j} = C_{i,j-1} + I_{i,j},\ \ \ 1 \le\ \ i \le\ \ n - j + 1,\ \ \ 2 \le\ \ j \le\ \ n,\ \ \{i, j \in \mathbb{N}\} \tag{3}$$

This results in a triangle comparable to the incremental paid claims triangle in which all $I_{i,j}$ are replaced by $C_{i,j}$. In numerical examples, this will make it easier to compare accident years to each other.

2.2.2 Incurred triangles

The triangles described in the previous section represented paid triangles, i.e. the data in the triangle are based on paid claims. Another type of run-off triangle is the incurred triangle. An incurred triangle is the sum of two triangles, the paid triangle and the outstanding triangle as visualized in Figure 4. The outstanding triangle can be defined as an estimate to cover the liability over any reported and not settled claim [Norberg, 1993].

Cumulative claims

Accident year i	Development period j: 1	2	···	n
1	$C_{1,1}$	$C_{1,2}$	···	$C_{1,n}$
2	$C_{2,1}$	$C_{2,2}$		
⋮	⋮			
n	$C_{n,1}$			

+

Outstanding claims

Accident year i	Development period j: 1	2	···	n
1	$O_{1,1}$	$O_{1,2}$	···	$O_{1,n}$
2	$O_{2,1}$	$O_{2,2}$		
⋮	⋮			
n	$O_{n,1}$			

=

Incurred claims

Accident year i	Development period j: 1	2	···	n
1	$R_{1,1}$	$R_{1,2}$	···	$R_{1,n}$
2	$R_{2,1}$	$R_{2,2}$		
⋮	⋮			
n	$R_{n,1}$			

Figure 4: Incurred triangle.

The paid triangle is not different from the cumulative triangle described in (2) and (3). In the outstanding $O_{i,j}$ triangle are all liabilities towards already reported but not settled claims. The sum of the Cumulative paid claims triangle and the outstanding claims triangle results in the incurred claims triangle $R_{i,j}$. So in Figure 4, $R_{1,2}$ denotes the cumulative claims up to and including development year 2 for accident year 1 added up with the outstanding claims seen from the year-end of development year 2 for accident year 1.

In case the expectation and timing of all claims would be completely right, the incurred triangle would contain a constant number in every development period. In practice, the expectation is adjusted every period to incorporate IBN(E)R claims.

2.2.3 AA- and AQ-triangles

The example in Figure 3 represents an annually-annually triangle (AA-triangle). In the triangle, the accident periods and the development periods are both determined on a yearly basis. A good characteristic of the AA-triangle is that it is symmetric, which makes it easier to model in a simulation study.

Different from the AA-triangle is the annually-quarterly (AQ-) triangle. In an AQ-triangle, the accident years are still annually as in the AA-triangle, but the development periods are based on quarters. The complete set of incremental claims is then given by:

$$\{I_{i,j} | i, j \in \mathbb{N} : 1 \leq i \leq n \;\;, \; 1 \leq j \leq 4\,(n - i + 1)\} \tag{4}$$

This results in an incremental AQ-triangle:

Accident year i	Development period j						
	1	2	3	4		...	4n
1	$I_{1,1}$	$I_{1,2}$	$I_{1,3}$	$I_{1,4}$	...	...	... $I_{1,4n}$
2	$I_{2,1}$	$I_{2,2}$	$I_{2,3}$	$I_{2,4}$	...	$I_{2,4(n-i+1)}$	
⋮	⋮	⋮	⋮	⋮			
n	$I_{n,1}$	$I_{n,2}$	$I_{n,3}$	$I_{n,4}$			

Figure 5: Incurred triangle.

It is possible to create an AQ-triangle with the cumulative paid claims following the same procedure as in (2) and (3), having the difference that the number of development periods is 4 times as large compared to the AA-triangle. The advantage of an AQ-triangle over an AA-triangle is that it is possible to model more precisely the development periods, as well as the possibility to extract more residuals. This number of residuals will be important in the bootstrapping process later on. In the remainder of the report, all formulas will be written in the form of an AA-triangle. By adjusting the development periods j, it is possible to obtain the formulas for the AQ-triangle.

Now the types of triangles are known, the next step is to generate a development pattern from the triangles which can be used to predict the claims that need to be settled for a HRG. To detect this development pattern, we will use the Chain Ladder method.

2.3 The Chain Ladder method

The most popular method in determining the best estimate is the Chain Ladder method [Mack, 1993]. The advantage of the Chain Ladder method is that it is distribution-free, easy to apply and holds only a limited number of constraints [Mack, 1993]. The Chain Ladder method can be applied on paid as well as on incurred triangles [Liu & Verrall, 2010]. The Chain Ladder method starts with determining the average development factor $\hat{f}_j$ per development period, which can be obtained by:

$$\hat{f}_j = \frac{\sum_{k=1}^{n-j+1} C_{k,j}}{\sum_{k=1}^{n-j+1} C_{k,j-1}} \quad 2 \leq \; j \leq \; n \quad j \in \mathbb{N} \tag{5}$$

It is then possible to estimate the next period by multiplying the average development factor corresponding with the last observation for an accident year:

$$\hat{C}_{i,n-i+2} = C_{i,n-i+1} * \hat{f}_{n-i+2} \quad 2 \leq \; i \leq \; n \quad i \in \mathbb{N} \tag{6}$$

Using the same steps, it is possible to forecast how an incurred triangle will develop in the upcoming periods.

Cumulative triangle

	1	2	3	4
1	100	210	290	350
2	120	200	300	
3	135	220		
4	115			
Development factors	(1)1,77	1,44	1,21	

Projected lower triangle

	1	2	3	4
1	100	210	290	350
2	120	200	300	362
3	135	220	317	382
4	115	(2)204	294	354

(1) $\hat{f}_j = \dfrac{\sum_{k=1}^{n-j+1} C_{k,j}}{\sum_{k=1}^{n-j+1} C_{k,j-1}} \rightarrow \hat{f}_2 = \dfrac{\sum_{k=1}^{4-2+1} C_{k,2}}{\sum_{k=1}^{4-2+1} C_{k,2-1}} = \dfrac{210+200+220}{100+120+135} = 1{,}77$

(2) $C_{i,n-i+2} = C_{i,n-i+1} * \hat{f}_{n-i+2} \rightarrow C_{4,4-4+2} = 115 * 1{,}77 = 204$

Figure 6: The Chain Ladder method.

In Figure 6, an example of the Chain Ladder method is given including the calculations. A more extensive example can be found in Appendix A.

2.3.1 Model assumptions Chain Ladder method

The Chain Ladder method holds 4 constraints that are needed to let the method work properly [Mack, 1993]:

1. The combination $\{C_{i,j}|1 \leq j \leq N\}$ and $\{C_{k,j}|1 \leq j \leq N\}$ needs to be independent when $i \neq \; k$ with $i, k \in \{1, \ldots, N\}$, i.e. the accident years are independent.

2. The development factors $\widehat{f_2}, \ldots, \widehat{f_N} > 0$, so that for every $i \in \{1, \ldots, N\}$ and $j \in \{2, \ldots, N\}$: $E\left[C_{i,j+1}|\; C_{i,2}, \ldots, C_{i,j}\right] = E\left[C_{i,j}|C_{i,j-1}\right] = \widehat{f_j} C_{i,j-1}$.

3. There exist a variance parameter $\sigma_2^2, \ldots, \sigma_{J-1}^2 > 0$, so that for every $i \in \{1, \ldots, N\}$ and $j \in \{2, \ldots, N\}$, $\mathrm{Var}(C_{i,j}|\; C_{i,j-1}) = \sigma_j^2 C_{i,j-1}$.

4. The development factors $\widehat{f_2}, \ldots, \widehat{f_N}$ are uncorrelated, i.e. $E\left[\widehat{f_2} * \cdots * \widehat{f_N}\right] = E\left[\widehat{f_2}\right] * \cdots * E\left[\widehat{f_N}\right]$.

Constraint 1 and 4 need the most attention before the Chain Ladder technique is used in the bootstrap method. Constraint 2 could only be a problem in the incremental paid triangle, as it would imply that no new claims were settled during a development period. However, if there are no claims settled, the variance is automatically zero. If the variance is zero, the residuals are automatically zero, which are excluded in the bootstrapping process. So, Constraints 2 and 3 do not form a problem in this study. Besides, the incremental data are transformed into cumulative data, which prevents the development factor from being zero.

With the Chain Ladder method, it is possible to detect the development pattern in a dataset and make a projection of the claims which still need to be settled. However, one projection will not directly give a complete view of the expectation of the claims which still need to be settled. It is for instance possible that more people reopen a case and get additional compensation as a result of law changes or new evidence. This may have happened once in history but could happen more often in the future. To incorporate these potential patterns based on historical data, we will use bootstrapping techniques.

2.4 History of bootstrapping techniques

Bootstrapping methods became popular in modeling reserve risk during the nineties when several papers about the application of bootstrapping methods on insurance portfolios were published. Most notable are the papers of Brickman et al. [1993], Mack [1993], England & Verrall [1999] and Kirschner et al. [2002].

The bootstrap method has three main advantages according to Shapland & Leong [2010]. The first advantage is the possibility to obtain a distribution of the possible claim amounts without knowing the statistical distribution beforehand. Secondly, the bootstrap method uses all features available in the data, the data is not modified or generalized to make it usable. Thirdly, the bootstrap method deals perfectly with the skewness available in the data and as insurance loss distributions are often skewed to the right, not having to correct for this makes modeling easier.

2.5 The bootstrap method

To get an impression of the distribution of the reserve risk, we use a bootstrapping technique which is a Monte Carlo simulation approach [Robinson, 2014]. When using the bootstrapping technique, the deviations from the expected development factor are determined and called residuals. These residuals are randomly replaced in the triangle creating a new history, slightly changing the development pattern and so changing the required capital. This process can be repeated many times to create a distribution of potential scenarios.

To obtain the residuals, we need to undertake a couple of steps. In the remainder of this section, the process is equal for the cumulative paid triangle as for the incurred triangle. In every formula where $C_{i,j}$ is mentioned for the cumulative paid triangle, $R_{i,j}$ can be written for the incurred triangle.

Besides the average development factors (5) the individual development factors ($f_{i,j}$) need to be calculated:

$$f_{i,j} = \frac{C_{i,j}}{C_{i,j-1}} \quad 1 \leq \ i \leq \ n-1 \quad 2 \leq \ j \leq \ n-i+1 \quad \{i, j \in \mathbb{N}\} \tag{7}$$

Based on the average development factor, the individual development factors and the cumulative or incurred triangle, it is possible to determine the unscaled Pearson residuals ($U_{i,j}$)[Braun, 2004]:

$$U_{i,j} = \sqrt{C_{i,j-1}}\left(f_{i,j} - \widehat{f}_j\right) \quad 1 \leq \ i \leq \ n-1 \quad 2 \leq \ j \leq \ n-i+1 \quad \{i, j \in \mathbb{N}\} \tag{8}$$

Now, it is possible to calculate Mack's alpha squared (MAS) k_j. The MAS parameter denotes the variance between the individual development factors and the average development factor by summing up the squared unscaled residuals of a column and dividing it by the number of residuals in the column. The unbiased estimator of the MAS parameter is given by [Braun, 2004]:

$$k_j = \frac{\sum_{i=1}^{n-j+1} U_{i,j}^2}{n-j} \quad 2 \leq \ j \leq \ n \quad \{j \in \mathbb{N}\} \tag{9}$$

Finally, it is possible to derive the unscaled Pearson residuals $r_{i,j}$ and perform the bias adjustment b_j to allow for over-dispersion in the residuals in the sampling process [England & Verrall, 2002]:

$$r_{i,j} = b_j \frac{U_{i,j}}{\sqrt{k_j}} \quad with \quad b_j = \sqrt{\frac{n-j+1}{n-j}} \quad 1 \leq \ i \leq \ n-1 \quad 2 \leq \ j \leq \ n-i+1 \quad \{i, j \in \mathbb{N}\} \tag{10}$$

After the residuals are determined, the residuals are shifted to make sure the average of the residuals equals zero [Huergo et al., 2010]. This is done by determining the average of all the residuals of a triangle and increasing or decreasing every residual with the same number to make sure the average equals 0. The average of the residuals can be determined by:

$$\bar{r} = \frac{\sum_{i=1}^{n-1}\sum_{j=2}^{n} r_{i,j}}{\sum_{i=1}^{n-1} i} \tag{11}$$

With the average known, it is possible to correct all residuals [Huergo et al., 2010]:

$$r^*_{i,j} = r_{i,j} - \bar{r} \quad 1 \leq \ i \leq \ n-1 \quad 2 \leq \ j \leq \ n-i+1 \quad \{i,j \in \mathbb{N}\} \tag{12}$$

In Figures 7 and 8, the steps of a standard bootstrap are visualized for the cumulative paid triangle. In Figure 7, the steps are visualized to come from the initial triangle to the scaled residuals based on the formulas explained.

Cumulative triangle

	1	2	3	4
1	100	180	290	340
2	140	290	400	
3	125	240		
4	120			
Development factors		1,95	1,47	1,17

Individual development factors

	1	2	3	4
1		1,8	1,61	1,17
2		2,07	1,38	
3		1,92		
4				

Unscaled residuals

	1	2	3	4
1		-1,5	1,92	0
2		1,49	-1,5	
3		-0,3		
4				
Variance		2,21	5,97	

Scaled residuals

	1	2	3	4
1		-1,20	1,11	0
2		1,23	-0,88	
3		-0,23		
4				
Bias correction		1,22	1,41	

Figure 7: Deriving residuals from a cumulative paid triangle.

In the triangle furthest to the right, the final residuals can be seen. In the following step, the residuals are resampled. In the resampling process, the zero-residuals are excluded as these do not reflect a deviation from the average development factor. To allow for a wide distribution of potential situations, the resampling process is performed with replacement. This implies that it is possible that a residual is placed multiple times in the newly created triangle and so reflects the reality where it is possible that circumstances from the past could have happened more often.

To come from the resampled residuals to a new best estimate, some steps need to be undertaken. Based on the resampled residuals $r^*_{i,j}$ the new individual development factors $f^*_{i,j}$ can be determined [England, 2003]:

$$f^*_{i,j} = \frac{r^*_{i,j}\sqrt{k_j}}{\sqrt{C_{i,j-1}}} + \widehat{f}_j \quad 1 \leq \ i \leq \ n-1 \quad 2 \leq \ j \leq \ n-i+1 \quad \{j \in \mathbb{N}\} \tag{13}$$

Now we know the new individual development factors, it is possible to calculate the new upper triangle. This can be done by multiplying the cumulative paid claims from the original triangle with the newly created individual development factors:

$$C^*_{i,1} = C_{i,1} \quad 1 \leq \ i \leq \ n \quad \{i \in \mathbb{N}\} \tag{14}$$

$$C^*_{i,j} = C_{i,j-1} * \ f^*_{i,j} \quad 1 \leq \ i \leq \ n-1 \quad 2 \leq \ j \leq \ n-i+1 \quad \{i,j \in \mathbb{N}\} \tag{15}$$

Based on the individual development factors and the cumulative triangle, it is possible to determine the new average development factors:

$$\widehat{f^*_j} = \frac{\sum_{i=1}^{n-j+1} C^*_{i,j}}{\sum_{i=1}^{n-j+1} C_{i,j-1}} \quad 2 \leq \ j \leq \ n-i+1 \quad \{j \in \mathbb{N}\} \tag{16}$$

In the same way, the lower triangle is calculated in (6) it is possible to calculate the required capital to settle all claims in this scenario. By repeating this process many times, we can obtain a distribution of potential settlement scenarios.

In Figure 8, the residuals from Figure 7 are resampled and based on these newly distributed residuals a new history is created. In the figure furthest to the right, the lower triangle is created based on the newly determined average development factors. The difference between the already settled claims (in black) and the ultimate claims in the fourth development year (in red) indicates the reserves that are needed to settle all claims for an accident year. The best estimate in this situation is indicated in green.

Resampled residuals

	1	2	3	4
1		-0,88	1,23	-1,2
2		1,11	-0,88	
3		1,23		
4				

New development factors

	1	2	3	4
1		1,81	1,69	1
2		2,08	1,34	
3		2,11		
4				

New history

	1	2	3	4
1	100	181	305	290
2	140	292	389	
3	125	264		
4	120			
Dev. factors		2,02	1,476	1

Projected lower triangle

	1	2	3	4	
1				340	
2			389	389	0
3		264	389	389	125
4	120	242	358	358	238
Outstanding claims:					363

Figure 8: From residuals to required capital - Cumulative paid triangle.

In the case of an incurred triangle, the process of estimating the best estimate is slightly different. In practice, the values estimated after the first development year are quite accurate and differ only slightly. Especially further in the tail, the adjustments in the incurred claims for an accident year are extremely stable. Remarkably, the residuals follow the same distribution as in the paid triangle, making it possible to apply the same methodology as for the paid triangle.

The difference between bootstrapping the paid triangle and the incurred triangle lies in the interpretation of the resulting lower triangle. In Figure 8, the expected outstanding claims can easily be derived from the triangle. In case of the incurred triangle, the ultimate outcomes need to be corrected for the already paid claims to know how much is expected to be outstanding. A step by step example of the bootstrapping process is worked out in Appendix B.

To incorporate also situations that did not happen in the past but might happen in the future, a process variance is added to the model in practice. This process variance follows a gamma distribution and has a significant impact on the residuals. However, in obtaining the correlation parameter, this additional (uncorrelated) error would lower the overall correlation between the residuals as these would not reflect the real deviations which happened in the past. This is why it is decided to let the process variance out of the bootstrapping process.

With the knowledge of the bootstrap method, we are now capable of getting a distribution of the expected claims that need to be settled for a HRG. With these data, we can obtain the BE and the profit and loss (P&L).

2.6 Determining the BE and P&L

The best estimate is differently determined for the cumulative paid triangle compared to the incurred triangle. In the cumulative paid triangle, the best estimate can be determined by filling the complete lower-triangle based on the average development factors and then subtracting the already paid claims from the ultimately expected claims for every accident year.

$$BE_{Paid} = \sum_{i=2}^{n}((C_{i,n-i+1}\prod_{j=n-i+1}^{n}\widehat{f}_j) - C_{i,n-i+1}) \tag{17}$$

To get the best estimate from the incurred triangle, the lower triangle of the incurred triangle needs to be constructed. The settled claims need then to be subtracted from the ultimate claims to obtain the best estimate.

$$BE_{Incurred} = \sum_{i=2}^{n}((R_{i,n-i+1}\prod_{j=n-i+1}^{n}\widehat{f}_j) - C_{i,n-i+1}) \tag{18}$$

In both cases, the best estimate reflects the capital that needs to be present to cover the expected claim settlements related to an HRG in a specific scenario. By creating 100.000 different scenarios and so 100.000 best estimates, it is possible to determine the best estimate for an HRG as the average best estimate of the 100.000 situations. To cover the risk of ending up with more claims than the expectation, additional capital needs to be in place. The additional capital which needs to be available to cover 99.5% of the scenarios in the upcoming year will be determined based on the SCR.

Before we can determine the SCR, first the expected P&L for the upcoming year needs to be determined for all the 100.000 outcomes of the BE. The process to obtain the P&L is slightly different for the cumulative paid triangle compared to the incurred triangle. The process to obtain the P&L for the cumulative paid triangle is prescribed first, subsequently the differences for the incurred triangle will be illustrated.

To obtain the P&L, the following formula needs to be used:

$$P\&L_{paid} = BE_t - \bar{Settlements}_t - BE_{t+1} \quad 1 \leq \ t \leq \ N-1 \quad t \in \mathbb{N} \tag{19}$$

In this formula BE_t denotes the best estimate at time t, BE_{t+1} denotes the best estimate at time t+1 and $\bar{Settlements}_t$ denotes the expected settlements between time t and t+1. In a world without uncertainty, the P&L would always be zero as the expected settlements at time t minus the settlements during year t would result in the best estimate at t+1. The best estimate at time t can be obtained by using (17) and does not differ from the process described in Figure 7 and Figure 8. To obtain the expected settlements during year t, the cumulative paid triangle can be bootstrapped multiple times and every time the expected payments in the newly created diagonal can be determined:

$$Expectedsettlements_t = \sum_{i=2}^{n} ((C_{i,n-i+1} * \widehat{f_{n-i+1}}) - C_{i,n-i+1}) \tag{20}$$

By taking the average of all newly created diagonals, it is possible to determine the expected settlements during year t. The best estimate at time t+1 can be determined by adding one development year to the initial triangle, determining new average development factors and then again calculating the lower triangle with (6). The P&L is calculated over a period of a year, so in the case of the AQ-triangle 4 diagonals need to be projected to determine the expected settled claims in the upcoming year. The same holds for the best estimate at time t+1.

For the incurred triangle, the process of obtaining the P&L is slightly different. As the incurred triangle on its own gives an indication of the expectation of the ultimately settled claims for an accident year, the expectation becomes more accurate as more claims are settled. The difference between the ultimately expected claims at time t = 0 and the ultimately expected claims at time t = 1 is the adjustment in the expectation of the ultimately expected claims.

$$P\&L_{Incurred} = \sum_{i=1}^{N} (R_{i,N,t=0}) \quad - \sum_{i=1}^{N} (R_{i,N,t=1}) \tag{21}$$

For both the cumulative paid triangle and the incurred triangle holds that if the bootstrap technique is applied several times for a triangle, a vector of P&L is created. The values in this vector can be ordered in descending order and the SCR at 99.5% can be found.

Figure 9 visualizes the distribution of the P&L. Most P&L center around the middle, however some simulation results will have a much higher projected loss. The SCR is the P&L at the 99.5% percentile.

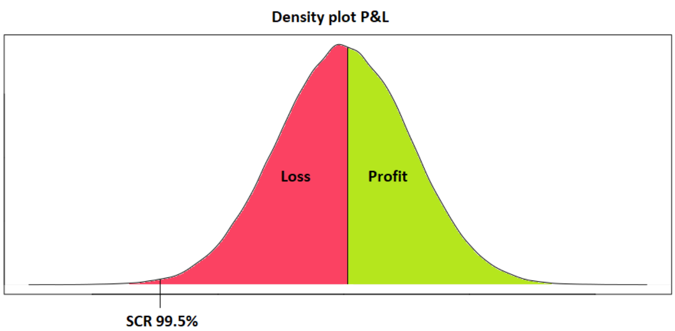


Figure 9: Density plot of the P&L.

2.7 Aggregating reserve capitals

In the previous section, we explained how the reserve capitals can be calculated for a single HRG. In reality, an insurer has many HRGs that all represent a specific product group and even within a specific product group there are often subgroups. To come to the total required capital for all HRGs, it is not as simple as adding up all the individual capitals.

In most situations, it is not likely that HRGs will reach the 99.5% level in the same year. The degree to which these events likely happen in the same year is captured in the correlation parameter, which reaches from -1 indicating a reverse dependency to 1 indicating a strong dependency. When the correlation between two HRGs is lower than 1, a diversification benefit can be reached.

The standard formula proposed by Solvency II is:

$$SCR_{x,y} = \sqrt{SCR_x^2 + SCR_y^2 + 2\rho_{x,y} SCR_x SCR_y} \tag{22}$$

The $\rho_{x,y}$ denotes the correlation between the reserve capitals of HRG x and HRG y at the 99.5% percentile. By rewriting (22), it is possible to derive the correlation between two HRGs by [Devineau & Loisel, 2009]:

$$\rho_{x,y} = \frac{SCR_{x,y}^2 - SCR_x^2 - SCR_y^2}{2 * SCR_x * SCR_y} \tag{23}$$

In this section, we described how SCR_x and SCR_y are determined. However, there are multiple ways to determine $SCR_{x,y}$. This will be investigated in Section 3.

2.8 Summary

We will investigate two types of triangles, triangles containing paid data and triangles containing incurred data. For both types, AA-triangles as well as AQ-triangles, are used. We will use the Chain Ladder method to determine the development factors, which play a major role in the bootstrapping process. Furthermore, we explained the bootstrapping process for a single HRG, and an introduction to the aggregation of reserve capitals is given. In Section 3, we will investigate how $SCR_{x,y}$ can be determined.

3 Estimating the correlation parameter from historical datasets

Here, we present what is already known in the literature about estimating the correlation between different HRGs. We start with an introduction to the concept of correlation. In the second section, we will investigate what methods are already known in the literature to capture the correlation between insurance triangles. Two methods will be chosen which will be investigated further. These methods will be described and the implications of specific features found in the data will be discussed.

3.1 Correlation

Correlation is defined as 'a relation existing between phenomena or things or between mathematical or statistical variables which tend to vary, be associated, or occur together in a way not expected by chance alone' [Akoglu, 2018]. To test the degree of correlation, the bestknown methods are Pearson's product-moment correlation method, Spearman's rank correlation method and Kendall's Tau correlation method, where Pearson's method is widely used for normally distributed datasets and Spearman's and Kendall's method are used for non-normally distributed datasets [Artusi et al., 2002].

Pearson's r method uses the statistical features found in the two datasets x and y, assuming these are normally distributed. The Pearson correlation can then be determined by using [Havlicek & Peterson, 1976]:

$$\rho_{Pearson} = \frac{Cov\left(x, y\right)}{\sigma_x \sigma_y} \tag{24}$$

In determining the correlation between the residuals of reserving triangles, rank methods like Spearman's and Kendall's method perform considerably worse compared to Pearson's r method, especially for relatively small triangles [Huergo et al., 2010].

3.2 Determining the correlation between run-off triangles

In literature, different methods are known to derive the correlation between two lines of business based on bootstrapping techniques [Taylor & McGuire, 2005]. In Brehm [2002] the basis of the correlation bootstrap method is explained and compared to the outcomes of a rank-correlation method. In 2002, Solvency I was still the leading framework, which did not pay attention to diversification benefits [Chandra Shekhar et al., 2008]. This makes that the research conducted back then was not conducted with the purpose of being used in the capital aggregation process. An important remark is that bootstrapping techniques only contain the correlation that is based on situations that happened in the past, where especially those situations that did not happen in the past but could happen in the future may have a lot of impact on the correlation estimate.

Braun [2004] describes how the correlation can be determined between two HRGs based on the residuals in a development period. This results in a correlation parameter for every development period. However, it is questionable how reliable the correlation estimates are for the development periods with a limited number of accident years, in Section 3.5 this will be discussed further.

Kirschner et al. [2002] extended the work of Brehm. A rank-correlation approach was again compared to the outcomes of a synchronous bootstrap methodology. The paper gives a more in-depth explanation of both methods and provides several examples. In line with Brehm [2002], the conclusion is that it is possible to obtain a correlation estimate with both methods which may not differ too much from each other. The conclusion is that correlation estimates obtained by both methods may support decision-makers and help to get a sense of direction. However, it is as well stressed that a lot of knowledge is required to be able to judge the correlation estimates obtained by bootstrapping techniques.

The increased capital requirements in combination with the increased attention for diversification benefits in Solvency II made the correlation parameter become more important recently. The papers of Brehm [2002] and Kirschner et al. [2002] give a good starting point to investigate further. As Brehm [2002] already denoted, the correlation bootstrap method might give a better insight into the correlation estimates at a higher percentile. This method will be further investigated combined with the implicit correlation method. The implicit correlation method is a combination of the rank correlation method and the application of (23). The implicit correlation method is not explicitly described in the literature but follows the same principles as the rank-correlation method.

3.3 The correlation bootstrap method

The correlation bootstrap method or simultaneous bootstrap approach is the method in which the bootstrapping process is simultaneously performed, to make sure the correlation between the residuals of multiple triangles is not lost. If there is a correlation between two HRGs, the residuals likely show to some degree a comparable pattern. If there are for instance more boat incidents in a period than expected, it is also likely that there are more than expected boat incident-related surgeries in that period. This will yield two positive residuals that are linked in the correlation bootstrap method.

After all corresponding residuals are linked to each other, the residuals which are linked to a 0-residual are removed as well as all residuals which don't have a matching residual. If the residual table has the same length and all 0-residuals are removed, it is possible to perform the bootstrapping process. In this process, the linked residuals are synchronously replaced in the reserving triangles to create a new history of possible events.

Define S_k as the complete set of residuals $(r_{i,j}^k)$ belonging to HRG k, where $r_{i,j}^k$ will denote the same moment in time for every k. Let's introduce permutation matrix M, which represents the changed position of the residuals during the bootstrapping process. So, the new order of the residual set can be denoted as S_k^*:

$$S_k^* = MS_k \quad \{k \in \mathbb{N}\} \tag{25}$$

As we take the correlation into account during the bootstrapping process, it is now allowed to add up the P&L of the two triangles, to come to a P&L for the combination of the two lines of businesses. Denote the lower triangle of triangle A as $\hat{A}$, the lower triangle of triangle B as $\hat{B}$ and the combined lower triangle as $\widehat{AB}$:

$$\widehat{AB_{i,j}} = \widehat{A_{i,j}} + \widehat{B_{i,j}} \quad 2 \leq \; i \leq \; n \quad\; n-i+2 \leq \; j \leq \; n \quad \{i,j \in \mathbb{N}\} \tag{26}$$

By repeating this process multiple times, three vectors with P&L are created: one for the first triangle, one for the second triangle and a last one for the combined portfolio. These vectors can be rearranged in descending order to find the scenario at the 99.5% percentile, indicating the expected loss at the 99.5% percentile. As all parameters from (23) are now known, it is possible to calculate the correlation.

In Figure 10, the correlation bootstrap method is visualized. Triangle A and triangle B represent the run-off triangles of HRG A and HRG B. In red are the projected outcomes of the bootstrapping technique. By simultaneously bootstrapping the residuals of triangle A and triangle B, it is allowed to add up both lower triangles, generating the combined expectation in triangle C.

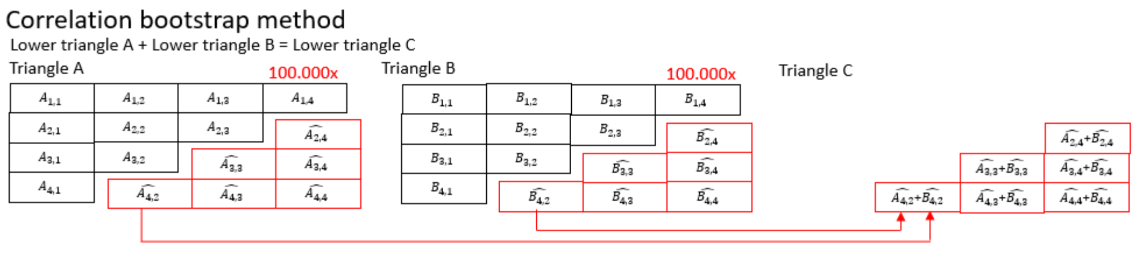


Figure 10: The correlation bootstrap method

3.4 The implicit correlation method

In the implicit correlation method, the combined triangle of the two HRGs is constructed by adding up the two triangles. Denote the upper triangle of triangle A as A, the upper triangle of triangle B as B and the combination of triangle A and triangle B as AB.

$$AB_{i,j} = A_{i,j} + B_{i,j} \quad 1 \leq \ i \leq \ n, \qquad 1 \leq \ j \leq \ n-i+1, \quad \{i,j \in \mathbb{N}\} \tag{27}$$

All three triangles can be independently bootstrapped as described in Chapter 2. This results in an SCR for every bootstrapped triangle. We can then use these SCR estimates to derive the correlation using (23).

As the bootstrapping process is not performed simultaneously, it is not needed to link the residuals to each other and so it is only required to delete the 0-residuals from the individual triangles. This reduces the number of residuals that are deleted, which would in practice mean that more potential developments are taken into account in the bootstrapping process compared to the correlation bootstrap method.

In Figure 11, the implicit correlation method is visualized. Triangle A and triangle B represent the same triangles as in Figure 10. The difference between the two methods is the method used to create the lower triangle of triangle C. In Figure 11, the correlation between the residuals is captured in the simultaneous bootstrapping process, which makes it unnecessary to create the upper triangle. In the implicit correlation method the residuals are not simultaneously bootstrapped and to capture the correlation the upper triangles are added to each other. This is the upper triangle of triangle C, by bootstrapping this upper triangle 100.000 times it is possible to construct a vector with 100.000 P&L estimates. These can then be handled in the same way as the P&L vector of triangle A and triangle B and makes it possible to determine the SCR for all three triangles. By applying (23), we can then determine the correlation parameter.

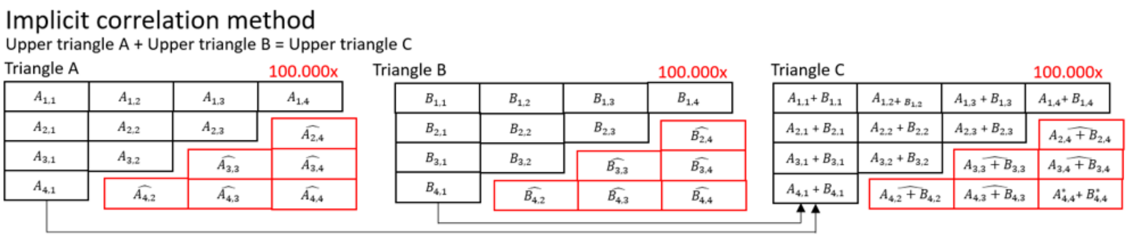


Figure 11: The implicit correlation method

3.5 Model assumptions

To use the Chain Ladder method as well as the correlation bootstrap method and the implicit correlation method, several assumptions need to be fulfilled. The assumptions for the Chain Ladder method have already been discussed in Section 2.3.1. In this section, the assumptions belonging to the correlation bootstrap method and the implicit correlation method will be discussed.

3.5.1 Model assumptions correlation bootstrap method and implicit correlation method

As in the correlation bootstrap method and the implicit method the residuals are extracted from the triangles, the model assumptions regarding the residuals hold for both methods [Huergo et al., 2010]:

1. There exist constants $\hat{f}_j, \hat{\sigma}_j > 0$ and random variables $\epsilon_{i,j}$ such that for all $i \in 1, \ldots, N$ and $j \in \{2, \ldots, N\}$ we have: $C_{i,j} = \hat{f}_j C_{i,j-1} + \sigma_j \sqrt{C_{i,j-1}} \epsilon_{i,j}$.
2. The residuals of triangle A and B, $\epsilon_{i,j}^{(a)}$ and $\epsilon_{k,l}^{(a)}$ are independent if $i \neq \ k$ or $j \neq \ l$ and it holds $E\left[\epsilon_{i,j}^{(a)} \middle| \ B_0^A\right] = 0$, $Var\left(\epsilon_{i,j}^{(a)} \middle| \ B_0^A\right) = 1$ and $P\left(C_{i,j}^{(a)} > 0 \middle| \ B_0^A\right) = 1$ with $B_0^A = \{C_{1,1}^{(a)}, \ldots, C_{N,1}^{(a)} | \ a = 1, \ldots, A\}$ for all $i \in 1, \ldots, N, j \in \{2, \ldots, N\}$ and $a = 1, \ldots, A$. I.e. no autocorrelation is allowed.

3. The residuals are normally distributed, i.e. $\epsilon_{i,j}^{(a)} \sim N(0,1)$.

4. The N-dimensional random variables $\epsilon_{i,j} = \left(\epsilon_{i,j}^{(1)}, \ldots, \epsilon_{i,j}^{(A)}\right)^T$ have the correlation-matrices:

   $$\sum\nolimits_j = Corr(\epsilon_{i,j} \,|B_0^A) = \begin{pmatrix} 1 & \rho_j^{(1,2)} & \cdots & \rho_j^{(1,A)} \\ \rho_j^{(2,1)} & 1 & \cdots & \rho_j^{(2,A)} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_j^{(A,1)} & \rho_j^{(A,2)} & \cdots & 1 \end{pmatrix}, \text{ where } \rho_j^{(a,b)} \in (-1,1) \text{ for}$$

   $a, b \in (1, \ldots, A)$ and $a \neq b$. As $Var\left(\epsilon_{i,j}^{(a)} \middle| B_0^A\right) = 1$ holds, the correlation matrices are also co-variance matrices of $\epsilon_{i,j}$ [Quarg & Mack, 2004].

Furthermore, to be able to use the found correlation in the standard formula, it is required that the dependency between triangles can be fully captured by using a linear correlation coefficient approach. This implies that there may not be tail-dependencies available between the residuals of the used triangles [EIOPA, 2014].

Strictly speaking, resampling the residuals over the whole triangle is not in line with the Chain Ladder method. The Chain Ladder method does not necessarily assume the development years to be independent, this could imply that an extreme positive residual is followed by some negative residuals or vice versa. Incorporating these possible dependencies in the model would only make it possible to simultaneously bootstrap within the columns and only determine correlations between the separate development years of the triangles as proposed by Braun [2004].

The method proposed by Braun [2004] has several disadvantages which makes it less suitable for aggregating risk capitals. Pearson correlation estimates need at least 8 observations to generate valid correlation estimates, this implies that the correlation estimates in the last 7 development years are not reliable [Hulley et al., 2013]. As most triangles only have a limited number of development years, this would mean that a significant part of the triangles does not generate reliable correlation estimates.

When aggregating risk capitals, typically one correlation parameter is used to describe the correlation between two triangles, making it possible to use the standard formula (23), so using the prescribed method of Braun [2004] will lead to a lot of extra uncertainty in the correlation estimate. For every disadvantage of the method proposed by Braun [2004], there are statistical procedures known in literature to overcome the limitations such as extrapolating the residuals when there is a low sample size. However, these techniques make the simultaneous bootstrapping process more complicated and add extra uncertainty to the model.

Huergo et al. [2010] investigated the impact of the assumption that the correlation matrices are the same for all development years. This implies that all residuals can be written to one vector, without the need for an extrapolation method to compensate for low sample sizes. The outcome of the research is that the differences in outcomes are neglectable, which justifies the simplifying assumptions that all correlation matrices are the same for all development years j. To make sure the finding of Huergo et al. [2010] holds for other datasets, the independence of the development periods is added as a constraint in this research.

3.5.2 Individual model assumptions correlation bootstrap method

As residual pairs are simultaniously drawn between multiple triangles in the correlation bootstrap method, both triangles must contain the same dimensions. Furthermore, it is important that the distribution of the data in both triangles complies with the assumptions in the previous section.

3.6 Impact of the data quality on the correlation estimates

The data available in the triangles are hardly ever perfectly fit for its purpose, in many cases the data is incomplete, there are booking errors, there are strange patterns in the tail or datasets reflect different periods [Busse et al., 2010]. It is not always possible to fix these issues, making it necessary to find methods to deal with them. Thereby it is good to keep in mind that insurance policies change over the years, making the loss pattern seen in one year not necessarily reflect the loss pattern in another year. These changes may have an impact on the residuals found in the dataset and so are good to keep in mind. Especially because outliers may have a significant impact on the outcomes generated when bootstrapping [Peremans et al., 2017].

3.6.1 Correlation between datasets with outliers

Outliers in a dataset tend to have a lot of impact on the correlation estimate when parameterized methods are used, such as Pearson's method [Kim et al., 2015]. However, it depends on the nature of the triangle what can be seen as an outlier. A triangle with a high MAS-parameter, like an insurance policy insuring the policyholder against the risk of a storm destroying the house, will have specific outliers due to the nature of the policy which cannot be left out of the model without making the model less valid.

In literature, several researchers published articles about detecting and dealing with outliers. Busse et al. [2010] propose to apply an algorithm that detects outliers caused by administrative reasons based on the contribution of a specific period to the total (expected) claims of an accident year. In Appendix D the method of Busse et al. [2010] is explained in depth.

After an outlier is detected, it still cannot automatically be deleted from the dataset. Only after the nature of the outlier is completely known, it is possible to exclude an outlier if it is sure the situation will not happen again in the future. Basically, as few as possible data should be deleted to make the model robust and more reliable.

Another method to detect outliers is to critically evaluate the residuals. However, extreme residuals may be invisible due to the masking effect [Peremans et al., 2017]. This effect occurs if an extreme residual causes to shift the average development factor of a column, resulting in a high deviation of the average for the whole column and so relatively high residuals in the whole development period. On the other hand, a swamping effect may occur if all observations deviate only a little bit and one regular observation deviates a bit more. As the impact of outliers may have a significant impact when using bootstrapping techniques, Verdonck & Debruyne [2011] proposed to weight the residuals and allocate lower weights to potential outliers. The method proposed by Verdonck & Debruyne [2011] is explained in Appendix D.

In practice, software is used to detect outliers. The detected residuals are evaluated by an expert panel to get insight into the nature of the outlier. If it is likely an outlier belonging to an event will happen again in the future, this outlier will be included in the bootstrapping process. If it is not likely that the specific situation will happen again, for instance, because there is an administrative cause of the outlier, it is tried to only delete the part in the residual which is caused by the administrative failure. If it is unlikely that an event will happen again within 200 years or it is impossible to resolve the administrative failure to a correct estimate, a residual might be excluded from the bootstrapping process.

3.7 Correlation between datasets with missing values

In case there are values missing or values excluded from the dataset, this does not cause a problem. The divisor of the average development factor (2.5) needs to be adjusted to the number of resulting values in the column, giving an accurate representation of the actual average development factor. The residuals at the places in the triangle where there are missing values are zero and zero residuals are left out when applying bootstrapping techniques [Huergo et al., 2010].

In the case of the correlation bootstrap method, the corresponding residuals in the other triangle(s) need to be deleted as well. This will lower the number of residuals but makes sure the correlation between the residuals is unchanged.

The good thing about both methods is that it is possible to fill the places where there are missing values with other residuals in the triangle. In this way, it is still possible to generate reliable claim estimates at the places where there is missing data on the basis of the development factor, MAS parameter and a residual.

It is good to understand why data points are missing in a dataset, even though it does not directly influence the method. The missing or excluded data points may have contained extreme residuals which could have had a major impact on the claim reserves that should be held to cover the risks of an HRG.

3.8 The tail of the reserving triangle

Different HRGs may have different settlement periods, a policy containing material claims may settle most of the claims within 4 development years and only contain little volume for special cases in the development years afterward. For immaterial claims, it may take decades before the last payments are settled as many policies guarantee a payout till the retirement age. The HRGs with a relatively short settlement period are referred to as short tail policies and the policies with a long settlement period are referred to as long tail policies. A distinction can be made between the correlation bootstrap method and the implicit method.

Before going into the implications of determining the correlation between long-tail and short-tail policies, it is important to determine the length of the tail of a policy. The tail of a triangle is often set after the first time the development factor is lower than a preset marker or after a certain percentage of the expected ultimate claims are settled. The reason for not taking the complete tail, is that the last development periods only contain little deviations. A little deviation in one accident year without a development in the other accident years might result in extreme residuals that do not reflect an extreme event. The other residuals for this development period will be constant and deviate a little bit from 0, which will let them count in the bootstrapping process while they do not represent a development. To deal with these tail developments, curve fitting techniques are often used [England & Verrall, 2002]. However, smoothing techniques will not be used in this study, as smoothing techniques could potentially change the correlation available between the residuals.

3.8.1 Implicit correlation method

In the implicit correlation method, two HRGs with different development patterns will not directly cause a problem as the residuals in the three triangles are bootstrapped independently. However, it is questionable to what extent the triangle containing the combination of both HRGs hold the specific characteristics of the two individual triangles.

The problem that may occur when two triangles with a different tail are added to each other in the implicit correlation method, is that the residuals belonging to the triangle with the long tail will be unchanged. These unchanged residuals will also end up in the parts of the triangle in which there is a potential correlation between the two triangles resulting in a less accurate SCR of the combined triangle and so, a less accurate correlation estimate.

3.8.2 Correlation bootstrap method

Taylor & McGuire [2005] propose to simultaneously bootstrap the matching residuals and separately bootstrap the residuals of the HRG with the long tail for the part of the triangle which does not have a corresponding part in the short tail triangle when using the correlation bootstrap method. This creates the possibility to obtain corresponding risk capitals, in which the correlation between the triangles is available. However, this method cannot be used when deriving the correlation between two policies with different tails.

Enlarging the tail of a short tail policy to the same dimensions as the long tail policy does not work for the correlation bootstrap method, as zero residuals and the corresponding residuals are deleted from the model. This would still imply that the observations in the tail of the long tail policy are neglected and possibly creates a strange pattern in the not existing tail of the short tail policy.

Shortening the tail of the policy with a long tail seems like the only feasible solution. It is then possible to aggregate all claims from the long tail policy which fall outside the model and place them in the last development period. Not aggregating the values in the tail creates a loss of potential extreme residuals but makes the corresponding residuals hold their original correlation. Aggregating the values in the tail makes the settlements in the tail count but creates a column with residuals that are potentially uncorrelated, affecting the overall correlation.

In the simulation study, the values in the tail are not aggregated, as it is likely that the correlation estimate get worse when the residuals belonging to the aggregated development period are added. We will investigate how much impact it has on the correlation estimate when a part of the tail is neglected.

3.9 Summary

In this section, we explained the methods known in the literature to capture the correlation between different HRGs. Furthermore, we discussed the implications of the specific features found in the datasets on the performance of the implicit correlation method and the correlation bootstrap method. We will investigate the implicit correlation method and correlation bootstrap method further in the simulation study. The setup of the simulation study is the topic of Section 4.

4 Simulation setup

Performing a simulation study requires a robust structure, clear guidelines and a precise process description to be able to verify and reproduce the simulation results [Robinson, 2014]. In this section, the simulation study outline will be discussed, including the relevance of the different parts of the simulation study. This will give insights into the different phases of the simulation study.

The simulation study involves 5 phases, which are depicted in Figure 12 and form the backbone of this section.

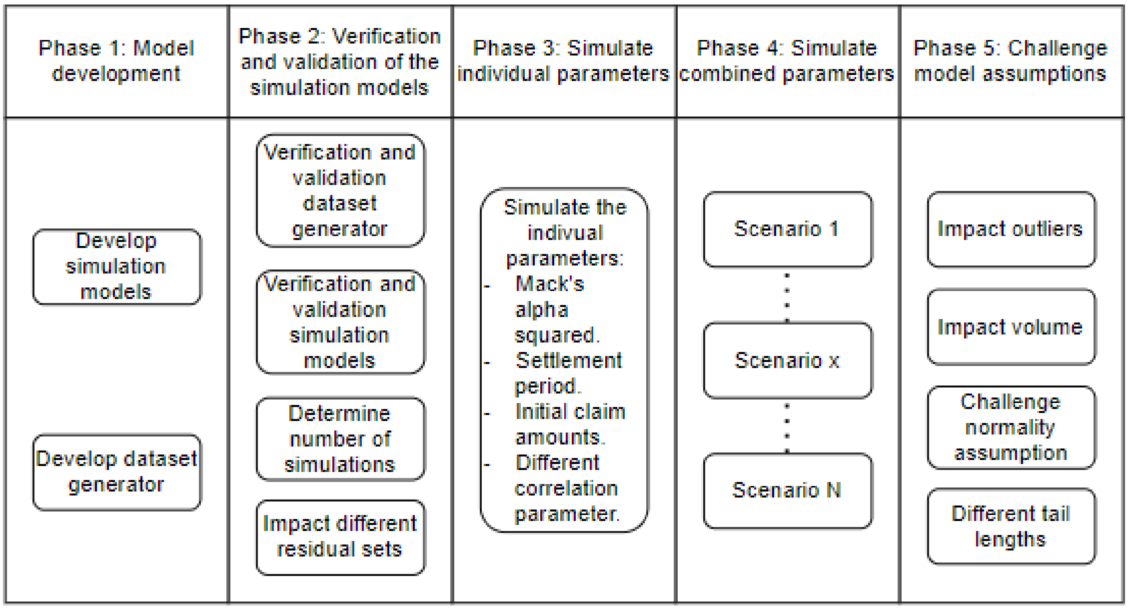


Figure 12: The phases of the simulation study.

In the first phase, we develop the required models. Before it is possible to simulate, the developed models first need to be verified, validated and calibrated which we will do in Phase 2. In the third phase, we will investigate the impact of a deviation of the individual parameters on the correlation estimate. In the fourth phase, the individual parameters are combined, to see whether certain combinations increase or decrease the individual effects. In the fifth phase, we will challenge the assumptions of both models. We are going to use this process for paid as well as for the incurred triangles of which both AA- and AQ-triangles will be tested.

4.1 Phase 1: Model development

Phase 1 contains two aspects, the development of the simulation model used to perform the simulations with the correlation bootstrap method and the implicit correlation method, and the development of the dataset generator.

4.1.1 Develop simulation model

To perform the simulations for the correlation bootstrap method and the implicit correlation method, we will develop a simulation model in R. We chose this open-source programming platform as it is the most used modeling software in the insurance sector. Many preprogrammed packages are already available and a lot of support can be found on the internet. We will develop the models based on the prescribed procedure in Sections 2.3 - 2.7, Section 3.3 and Section 3.4.

4.1.2 Develop dataset generator

Historical data can be used in the simulation study, however, these data are often polluted and therefore do not fulfill all requirements discussed in Section 3.5.1. Besides, understanding what the data are representing in a specific portfolio requires a lot of expert knowledge which makes it more likely mistakes are made when this data is interpreted by a non-expert. Data of a specific portfolio may represent different information over the years due to policy changes, acquisitions and model changes. Besides, the detection of outliers and the decision to include or exclude them is also performed by expert panels. The creation of a clean dataset is a highly complex job, however, the patterns available in the data can be used as a starting point for the creation of a valid dataset.

To be able to create datasets with specific characteristics, we will use a dataset generator to obtain clean datasets that hold the preset characteristics. The creation of the dataset generator will be discussed in Section 5. With this generator, it will be possible to create triangles based on the variables: The MAS parameter, the settlement periods, the initial claim amounts and the residuals. Besides, it will be possible to use this generator for triangles with a varying number of development- and accident years as well as it is possible to use it for AA-, AQ-, paid and incurred triangles. The dataset generator can be extended to create two triangles with a set correlation. This makes it possible to use the datasets with a known dependency in the correlation bootstrap method and the implicit correlation method and see how much the outcomes deviate from the preset correlation. The produced datasets will be tested for their validity, this will be explained in Section 6.2.

The relationship between the dataset generator and the simulation model is visualized in Figure 13.

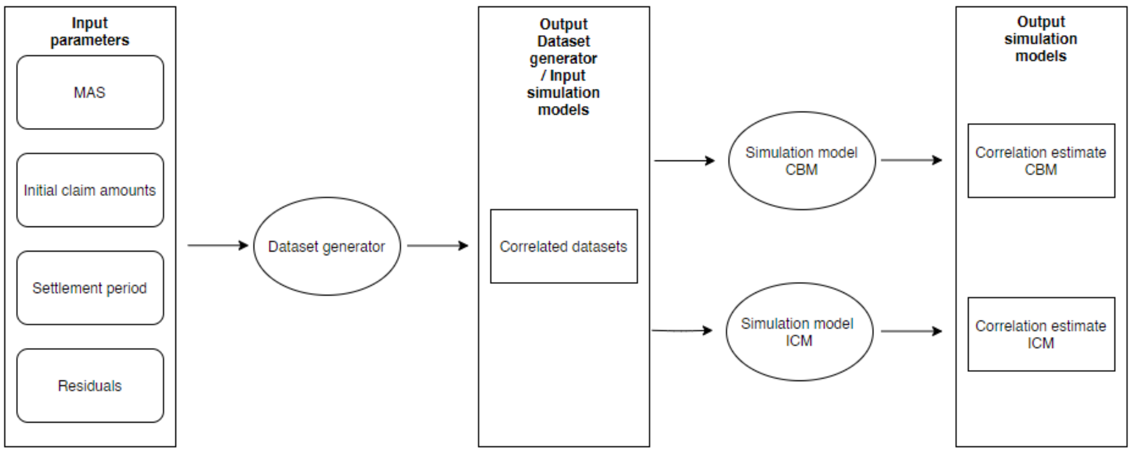


Figure 13: Relationship between the models.

In Figure 13, the input parameters which are required to generate a triangle and the process that needs to be fulfilled to obtain the correlation estimates is schematically visualized. In 4.3.1, we will introduce the different input parameters.

4.2 Phase 2: Verification and validation of the simulation models

In the verification and validation phase, we will investigate whether the simulation models create the expected output. Furthermore, we will investigate whether the created datasets comply with the assumptions of Section 3.5.1, to make sure we use valid triangles in the simulations. We will also investigate how many simulation runs are required per simulation and lastly, we will investigate what the impact of different residual sets is on the correlation estimates.

The verification and validation of the simulation models we developed is one of the most important aspects of the simulation study [Robinson, 2014]. Programming mistakes are easily made and preprogrammed packages may work slightly differently compared to the prescribed methods. This makes it necessary to verify and validate the simulation models before we can perform the simulations.

4.2.1 Verification and validation of the dataset generator

For the verification and validation of the dataset generator, we will investigate how well the created triangles contain the pre-set correlation. This can slightly differ from the pre-set correlation as a result of the limited sample size and the adjustment of the residuals in the triangle generation process. Furthermore, we will investigate whether the created triangles comply with the assumptions of both methods, to make sure the simulations are performed with clean triangles holding the assumptions described in Section 3.5.1.

4.2.2 Verification and validation of the simulation model

For the verification and validation of the simulation model, the output of the different functions is compared to the Excel sheet of England [2003]. This guarantees that we programmed all functions correctly in R. Furthermore, we extended the model of England [2003] with the correlation bootstrap method and the implicit correlation method. By comparing the correlation estimates of the model in R and Excel, it is possible to mitigate the occurrence of programming errors.

4.2.3 Determine number of simulations

The required number of simulations is important, as a too low number of simulations can result in inaccurate correlation estimates and a too high number of simulations will significantly increase the run time, which decreases the number of situations that can be tested. To determine the number of simulations required for stable outcomes, two datasets will be used. One dataset containing an AQ-triangle and one containing an AA-triangle. For both types of triangles, the correlation is determined based on 10.000, 20.000, 50.000 and 100.000 bootstrapped sets of triangles. This process is repeated 100 times for every number of simulation runs. From the outcomes, the standard deviation is determined which we can use to determine which standard deviation is acceptable.

4.2.4 Impact of different residual sets

The last part of Phase 2 is testing the impact of different residual sets. To be able to draw conclusions based on the results of the simulations performed in Phase 3, it is important to understand what impact different residual sets have on the correlation estimates. For both the AA- and the AQ-triangle, 10 sets containing 2 triangles will be generated using the dataset generator. In these datasets, the only parameter which is different for every dataset is the residual set. To these 10 datasets, we apply the correlation bootstrap method and the implicit method, resulting in 10 correlation estimates for every method. If the impact of the residuals is known, we can start with Phase 3.

4.3 Phase 3: Simulate individual parameters

After we tested what impact the different potential errors have on the correlation estimates, we can start with testing the individual parameters. The three parameters used to create the datasets are the settlement period, Mack's alpha squared and the initial claim amounts. To see the impact of these parameters, only one parameter is changed at a time. This makes it possible to derive the impact one parameter has on the correlation estimates.

4.3.1 The parameters

The first parameter we will be testing is Mack's alpha squared (MAS), the parameter which denotes the variance between the individual development factors belonging to a development period (9). The second parameter we are going to test is the settlement periods parameter, denoting the vector with the average development factors (5). The third parameter we will test is the initial claim amounts parameter, denoting the settled claims in the first development year.

Table 1: Simulation study outline paid datasets

		MAS						Settlement periods				First development year			
		Long tail&Long tail		Short tail&Short tail		Long tail&Short tail		Short tail&Short tail		Long tail&Short tail		Short tail&Short tail		Long tail&Short tail	
Correlation	Real	CBM	ICM	CBM	ICM	CBM	ICM	CBM	ICM	CBM	ICM	CBM	ICM	CBM	ICM
-0,25															
0															
0,25															
0,5															
0,75															

In Table 1, the outline for the simulations which we will perform in Phase 3 of the simulation study is visualized. This matrix is used for all 4 types of triangles. In the most left column are the correlation levels at which the triangles will be tested. The set of $\{-0.25, 0, 0.25, 0.5, 0.75\}$ is chosen as this is the range of correlations that are most often observed between the different HRGs. Correlations lower than -0.25 hardly ever occur between HRGs, so this is chosen as the lowest correlation that will be tested. As the correlation estimates currently are rounded to quarters, measuring the correlation at these quarters gives insights into the performance of the methods at the different quarters. It is not likely that the correlation between HRGs is higher than 0.75, so this is taken as the upper-bound. The second column is the correlation that is available between the randomly drawn residuals at a certain correlation level. Especially for the AA-triangles, the real correlation between the residuals may be different from the preset correlation, making it important to take this into account when analyzing the results. More about the deviations is explained in Section 6.

In the remainder of Table 1, are the three situations which will be tested for every parameter. For instance for the MAS, the situation "Long tail&Long tail" will be tested which denotes two triangles with the same relatively high number of settlement periods. The "Long tail" MAS vector is derived from a HRG that has a lot of settlement periods, a relatively stable initial claims and so a relatively low MAS vector. The "Short tail" MAS vector is derived from a HRG that has a relatively low number of settlement periods, relatively unstable initial claims and so a relatively high MAS vector. The "Long tail&Long tail" simulation is only performed once, as the parameters would be exactly the same when simulated for the settlement periods and the initial claims vector. In Table 1, CBM denotes the correlation bootstrap method and ICM denotes the implicit correlation method.

The simulations for both the incurred AA-triangle and the incurred AQ-triangle are performed in the same manner as for the paid triangles. Different is that for the incurred triangles, the parameters are derived from two incurred triangles instead of from the two paid triangles. For the paid as well as for the incurred triangles, the same underlying HRGs are used. The parameter for the "Long tail" MAS from the paid triangle will reflect the same claims as for the "Long tail" MAS parameter of the incurred triangle. This makes sure that it is possible to compare the outcomes.

4.4 Phase 4: Simulate combined parameters

With the outcomes of the first part of the simulation study, it is possible to construct specific situations in which parameter settings can be combined and look whether the deviations in the correlation estimates are getting stronger or weaker. It is also possible to investigate whether certain parameters are more decisive than others. This gives a complete overview of the performance of both methods within the assumptions belonging to the methods.

4.5 Phase 5: Challenge model assumptions

Phase 5 of the simulation study consists of two parts, firstly we will investigate what the impact of outliers and different SCR-volumes is on the correlation estimates. Secondly, we will challenge the assumptions belonging to both methods discussed in Section 3.5.1.

4.5.1 Impact outliers

Outliers may have a significant impact on the SCR and so potentially on the correlation estimates as well [Verdonck et al., 2009]. To investigate how well both methods are able to come up with an accurate correlation estimate, we will add outliers to the dataset. Besides the correlation estimates, we will also investigate what the impact of an outlier is on the other residuals. To have a good insight into the performance of both methods, we will test a dataset containing one outlier and a dataset containing several outliers.

4.5.2 Impact volume

The SCR of HRGs can differ significantly in volume. We will investigate if this causes a problem for the correlation estimates obtained by both methods. Furthermore, we will investigate whether the correlation estimates improve if complete triangles are scaled, to make sure the SCRs of two triangles are approximately equal. By scaling the complete triangle, the residuals do not change and so the available correlation will not be lost during the scaling process.

4.5.3 Challenge normality assumption

The residual sets of most HRGs are assumed to be normally distributed, however, this assumption does not hold for all HRGs. It is interesting to know how well both methods are capable of estimating the correlation in those cases. To resemble this situation, the residuals will be drawn from a skewed distribution. With these outcomes, it is possible to assess how important the normality assumption of the residuals is for both methods.

4.5.4 Different tail lengths

As already discussed in Section 3.8, two HRGs having a different tail length can cause problems in both methods. We will investigate this by performing simulations in which one development year is added at a time. This gives insights into the number of development periods it takes before the correlation estimates become stable. If we can reach a stable correlation estimate at a certain number of development periods which is lower than the total number of relevant development periods, we can conclude that we can use one or both methods to derive an accurate correlation estimate for HRGs with a different tail length.

4.5.5 Dependencies between accident years

If there are dependencies between accident years, they may be lost during the bootstrapping process. This might have a significant impact on the SCR as certain dependencies may cause the correlation between the risk capitals to be much higher or lower than the bootstrapping process would suggest. As the bootstrap yields an unrealistic SCR, it is not possible to capture the correlation using the implicit correlation method or the correlation bootstrap method.

Both methods could be applied and would yield comparable results as the ones in Phase 3 and Phase 4. However, this would mean that only the dependencies between the residuals of the two triangles are taken into account and not the dependencies within a triangle. So, we will not challenge this assumption, as it would definitely not result in an accurate correlation estimate.

4.5.6 Autocorrelation

Autocorrelation can occur in case there are, for instance, seasonal effects, which have an impact on the development pattern. If the bootstrapping process is applied on a triangle containing autocorrelation, it is possible that the residuals belonging to a certain season end up in another season while this is not likely in practice. This may have a significant impact on the SCR and thus it is not allowed to bootstrap triangles that contain autocorrelation. Autocorrelation is most likely to occur in AQ-triangles, as the residuals in an AQ-triangle are determined on a quarterly basis.

Deriving the correlation parameter between two triangles containing autocorrelation will create problems in both methods. As both triangles do not take the dependencies within the triangles into account, it will be impossible to derive an accurate SCR and so an accurate correlation estimate. So, the assumption that no autocorrelation may be available in the datasets will not be challenged.

All assumptions discussed in Section 3.5.1 have now been treated. We will investigate what effect outliers have on the correlation estimates, what impact not normally distributed residuals sets has on the correlation estimates, and how accurate both methods are capable of estimating the correlation parameter between triangles with a significantly different number of settlement periods. The assumptions related to independent accident years and autocorrelation will not be challenged, as the SCRs obtained during the bootstrapping process will not represent the real risk belonging to the datasets if there are dependencies available.

4.6 Summary

In this section, an outline of the simulation study is provided. Figure 12 gives a schematic overview of the phases of the simulation study. Furthermore, specific attention is given to the outline of Phase 3, the testing process of the individual parameters. To perform these simulations, firstly we need to develop and verify the dataset generator and the simulation model. These will be the topics of Section 5 and Section 6.

5 The dataset generator

To be able to simulate many scenarios and to draw general conclusions in the end, we developed a dataset generator. With this generator, it will be possible to vary the settlement periods, the MAS parameter, the initial claim amounts, and the correlation between the residuals of two datasets. Based on these parameters it will be possible to construct two triangles containing the pre-set characteristics which then can be used in the simulation study.

This section is structured in the same order as the steps that need to be taken to construct a new triangle based on the input parameters. This starts with the creation of the correlated normally distributed residuals, followed by adding the MAS parameter to create the unscaled residuals, in the last step the development pattern and the first development year are added and used to create the two triangles.

5.1 Creating correlated residuals

The dataset generator starts with the creation of the correlated residuals. To make the triangles suitable to be used in the bootstrapping process, the residuals need to be normally distributed with mean 0 and variance 1. To make sure the residuals of the two triangles share a preset correlation, we draw the residuals from the bivariate normal distribution. By first creating a vector with randomly drawn observations from the normal distribution (A), it is possible to create a second vector (B) with the desired correlation [Kaas et al., 2008]:

$$B|A = a \sim N\left(\mu_B + \rho\frac{\sigma_B}{\sigma_A}\left(a - \mu_A\right), \sigma_B^2\left(1 - \rho^2\right)\right) \tag{28}$$

In case more than two portfolios are compared at the same time, it is possible to use the multivariate normal distribution. By applying the Cholesky decomposition on the covariance matrix, it is possible to get multiple vectors with the desired correlation structure [Hull, 2018].

5.2 From correlated residuals to correlated triangles

The correlated residuals need to be simultaneously placed in the two upper triangles. This makes sure the correlation between the residuals of the two triangles is maintained in the triangle generation process. In equations (7)-(12), we explained all steps to come from a triangle to the residuals. Recursively solving these equations yields the following equation for deriving new triangles based on the residuals:

$$C_{i,j} = C_{i,j-1}\left(\frac{\frac{r_{i,j}}{b_j} * \sqrt{k_j}}{\sqrt{C_{i,j-1}}} + \widehat{f}_j\right) \quad 1 \leq\ i \leq\ n,\ 2 \leq\ j \leq\ n, \{i, j \in \mathbb{N}\} \tag{29}$$

In Equation (29), there are three unknown parameters, the cumulative claims in the previous development period belonging to an accident year ($C_{i,j-1}$), the MAS parameter (k_j) and the development factor between two consecutive periods ($\widehat{f}_j$). These three parameters seem strongly related in practice, in many cases, a relatively high development factor corresponds with a relatively high MAS parameter. Furthermore, the MAS parameter is related to the initial claim sizes. To get an impression of these dependencies, it is possible to derive the three parameters from historical datasets. More about these three parameters will be explained in Section 5.3.

An important constraint of Equation (29) is that $C_{i,j-1}$ needs to be positive. As soon as $C_{i,j-1}$ becomes negative for an accident year, it will not be possible to calculate $\sqrt{C_{i,j-1}}$ when determining $C_{i,j}$. This problem can be solved by taking ($\sqrt{|C_{i,j-1}|}$) instead of ($\sqrt{C_{i,j-1}}$). However, if $C_{i,j-1}$ would be negative, it would be impossible for $C_{i,j}$ to become positive. It is also not likely in reality that the cumulative claims would become negative, as this would mean that the insurer received money for the claims instead of settling claims.

In the dataset generator, it could occur that negative cumulative claims are obtained in case there is relatively high MAS parameter, combined with a relatively low development factor, relatively low initial claim amounts and an extreme negative residual. In the remainder of this section, we will explain the methods used to derive the parameters of (29) from real data. We will pay specific attention to how it is possible to avoid $C_{i,j-1}$ to become negative.

5.3 Estimating the parameters

In this section, we will discuss how the parameters can be derived from a dataset containing historical data. From Section 3.5.1 we know that it is not allowed to have dependencies between the residuals of a triangle when applying bootstrapping techniques. To make sure these dependencies are not available, the datasets first will be corrected for inflation and portfolio developments. From the adjusted datasets, it is possible to derive the parameters, which will guide as a starting point in the triangle generation process.

Before going into the used methods to correct the datasets, we need to determine which data is representative for the current situation. Analysing the historical data, it becomes apparent that the datasets containing incurred data are quite inaccurate before 2000, creating many deviations and so significantly increasing the MAS parameter. As a result, we choose to only include the data representing the period between 2000 and 2020 for all types of triangles. Furthermore, the AA-triangles are generated from the available AQ-triangles, by taking the cumulative claims belonging to the end of every year.

To make sure no dependencies exist between the residuals as a result of inflation, all claims are represented in current money. By slightly changing the formula for continuously compounding interest of Hull [2015], it is possible to set the historical claims in current money under the assumptions that the claims are settled equally over the development period. In the following formula, $Cq_{i,j,z}$ indicates the inflation-adjusted cumulative payments where q_z denotes the inflation corresponding to the diagonal z where $Cq_{i,j,z}$ finds itself in.

$$Cq_{i,j,z} = C_{i,j} * e^{\frac{1}{2}x_z} * \prod_{k=z+1}^{I} (1 + q_k) \quad 1 \leq i \leq n, \ 1 \leq j \leq n, \ 1 \leq z \leq n \quad \{i, j \in \mathbb{N}\} \tag{30}$$

With,

$$x_z = \ln(1 + q_z) \tag{31}$$

For the values of the inflation parameter, we used the consumer price index (CPI) [CBS, 2020]. It is as well possible to take a more specific inflation parameter that better fits with the nature of the HRG compared to the CPI. However, to assess whether another inflation parameter would better fit with the HRG expert knowledge is required. So, the CPI is chosen as it gives a good indication of the average inflation development across all HRGs.

Besides inflation, there can be other developments which create dependencies between the accident years of a triangle. This may involve portfolio growth, policy changes and/or different claim behaviour, all potentially having a significant impact of the ultimate claims. To make sure that these developments do not cause the accident years to be dependent on each other, the trend is taken out of the triangles. This can be done by fitting a trendline through the data points which minimizes the aggregated deviations between the trendline and the data points. Both a linear as well as an exponential trendline can be constructed and the one with the highest explanatory power (R^2) is the one that we will use. Based on the trendline, it is possible to come up with index numbers, which can be used to scale the historical data, to make them comply with the current situation. For this correction, we use the cumulative claims at the end of the first development year for both the AA- and the AQ-triangles. This makes sure, enough data are in the sample to have insights into the portfolio developments.

From the triangles corrected for the inflation and the portfolio development, it is possible to derive the initial claim amounts by taking the first development period. The MAS parameter can be derived by using Equations (7) - (9). The settlement periods parameter can be derived by using (5).

5.4 Using the dataset generator in the simulation study

In the simulation study, we will create the triangles based on the input parameters discussed in this section. However, to test the performance of the correlation bootstrap method and the implicit correlation method on a wide range of triangles, it will not be possible to manually check every created triangle. If a created triangle contains a negative cumulative payment, it will automatically stop and continue with the next scenario.

To prevent the triangles from containing negative cumulative values, the combination of the settlement periods parameter and the MAS parameter is important. It is most likely that negative cumulative values occur in case of a low settlement periods parameter and a high MAS parameter. In the case of the AQ-triangle, often a high MAS parameter is seen between the individual development factors belonging to the first development period. This is not strange, if for instance only in a couple of years there was a specific event in the first quarter, these years which contained a specific event will have a significant other development pattern compared to the years in which this specific event did not occur. It might as well occur that a specific event took place during the last weeks of a quarter and most claims were not settled at the end of the quarter.

To lower the MAS parameter in the simulation model, the AQ-triangles will be generated based on the cumulative claims at the end of the first development year. The same will happen for the incurred triangle, as the MAS parameter in the first development periods is for the AQ-triangle as well very high as a result of the same reason. The implication of starting the bootstrapping process at the start of the second development year is that approximately 10% of the most relevant residuals are deleted from the model. For the datasets created by the dataset generator, this is not a problem as all residuals are normally distributed. In practice, if there is a high MAS parameter in the first development periods, it is likely that there is autocorrelation available in the model. In those cases, it might as well be necessary to start the bootstrapping process after the first year.

5.5 Output of the dataset generator

When creating a triangle based on randomly drawn residuals, it will be impossible to create a triangle exactly containing the pre-set parameters. This can be compared to the result of a bootstrap triangle as described in 13 - 16. The randomly placed residuals will create a slightly different development pattern and slightly differently scale the MAS parameter. However, as long as the real dataset does contain normally distributed residuals, the resulting parameters will not differ too much from the preset parameter.

5.6 Summary

In this section, we explained the way the dataset generator is developed. Furthermore, we discussed the way the data can be derived from a real dataset. Finally, we paid specific attention to the way it can be made sure that the triangles do not contain negative values. To know whether the outcomes of the dataset generator are in line with the constraints of the Chain Ladder method, the implicit correlation method and the correlation bootstrap method, we will validate and verify the generated datasets in Section 6.

6 Model verification and validation

We describe the verification and validation process of the dataset generator and the simulation model in this section. In the verification process, we will check if both models are rightly implemented, i.e., do they perform the simulations as prescribed in the previous section. In the validation process, we will investigate whether the output of the models is in line with the expected output as prescribed in Robinson [2014].

The verification process is an important aspect of the simulation study, as it reduces the probability that coding errors are made and forces the researcher to overthink the methodology again. The validation process gives insights into the accuracy and correctness of the simulation model, which can be used to come up with an acceptable number of simulations. We will dedicate the first part of this section to the verification of the data set generator. The second part will be dedicated to the validation of the output of the dataset generator. In the third part, we will verify the simulation model. In the fourth part, we will investigate how many times the triangles need to be bootstrapped to have enough observations to derive a constant correlation estimate. In the last section, we will analyse the impact of different residual sets.

6.1 Reproducibility

To make sure all simulations can be reproduced in exactly the same manner, it is possible to set a fixed random seed in R and Excel. This makes sure that all randomly drawn values are drawn in the same order. In the simulations in R and Excel, we use seed 1701. The type of seed that is used in R is the default type, the Mersenne-Twister [Hechenleitner & Entacher, 2002]. The default type of seed used in the Excel model is a linear congruential generator. Furthermore, the use of a seed makes it easier to verify whether there are differences between the different simulation models and trace back potential mistakes. The seeds of Excel and R pick random numbers in a different order, this difference will become clear in the next section.

6.2 Validation of the dataset generator

We explained the formulas used to come from residuals to a triangle in Section 5 and these are all derived from the formulas in Section 2. The goal of the dataset generator is to generate two datasets with a set correlation between the residuals. To investigate whether the preset correlation is also the correlation resulting from the residuals of the created datasets, we test for 10.000 datasets whether the preset correlation corresponds to the correlation obtained from the created dataset.

In the process from residuals to a dataset, there are two occasions where the correlation between the preset correlation and the final correlation may slightly differ. The first point is the moment the residuals are drawn from the bivariate normal distribution. Due to the relatively low sample size, the correlation in the sample may not exactly reflect the preset correlation. The second point is the combination of residuals and the previous development period in (29). The formula would work perfectly if the value in the previous development period is constant and the average of the residuals for a development period is exactly 0, this is hardly ever the case. The combination of previous development periods and randomly drawn residuals may slightly change the settlement pattern, the MAS parameter and ultimately the residuals.

To know how much impact these deviations have on the preset correlation, 10.000 combinations of triangles were constructed with a preset correlation of -0.5, 0, and 0.5. In this process, the correlation is measured between the randomly drawn residuals from the bivariate distribution and the residuals derived after the triangle generation process. In Table 2 are the outcomes of these simulations.

Table 2: Accuracy of the dataset generator.

	Correlation from bivariate distribution				Correlation from final triangles				Final correlation - correlation bivariate distribution			
Correlation	Mean AA	Stdev AA	Mean AQ	Stdev AQ	Mean AA	Stdev AA	Mean AQ	Stdev AQ	Mean AA	Stdev AA	Mean AQ	Stdev AQ
-0,5	-0,499	0,054	-0,500	0,026	-0,479	0,030	-0,481	0,030	0,020	0,030	0,019	0,014
0	0,000	0,073	0,000	0,035	-0,001	0,077	0,000	0,037	0,000	0,034	0,000	0,016
0,5	0,500	0,055	0,500	0,026	0,480	0,062	0,481	0,029	-0,021	0,030	-0,019	0,014

The results of the accuracy-test are as expected, the AQ-triangles better reproduce the preset correlation as these contain 4 times as many residuals as the AA-triangles. The standard deviation in the results of the AA-triangles is quite high, especially when realizing that more than 30% of the created AA-triangles will deviate even more than 1 standard deviation.

The standard deviations in Table 2 are also in line with the standard error for Pearson's correlation coefficient, which can be determined by Holland [2019]:

$$Stdev_{\rho} = \sqrt{\frac{1-\rho^2}{n-2}} \tag{32}$$

The AA-triangles contain 190 residuals, the AQ-triangles contain 820 residuals. By applying (32), the results in Table 3 can be obtained.

Table 3: Expected standard error for Pearson's correlation coefficient.

Correlation	AA-triangle	AQ-triangle
-0.5	0.063	0.030
0	0.073	0.035
0.5	0.063	0.030

The results in Table 3 come close to the outcomes of in Table 2, so it can be concluded that the standard deviation is within an acceptable range.

6.3 Validity of the created triangles

In Section 3, we explained the assumptions underlying the correlation bootstrap method and the implicit correlation method. The triangles generated by the dataset generator need to fulfill these requirements before we can use them in the simulations. To be able to derive the residuals, the triangles need to fulfill the requirements for the Chain Ladder method. Constraints 2 and 3 of Section 2.3.1 are easily met, as the MAS parameter and development factors need to be inserted manually and are only adjusted slightly in the dataset generation process.

To test whether the accident years of a triangle are independent of each other, we performed the Chi-Square test on the residuals as proposed by Bowerman [2016]. The Chi-Square test is performed with P=0.05 on 10 incremental paid triangles with each 20 accident years. Of these triangles, only the oldest 15 accident years are tested, as the last 5 accident years do not contain enough observations. The 15 accident years are tested pairwise. We test H_0 'Accident year i is independent of accident year j with P=0.05' against H_1 'Accident year i is dependent of accident year j with P=0.05'. In all cases, we could not reject the null hypothesis, which indicates that the accident years are independent of each other. In practice, two accident years might be dependent while being not, as these are randomly generated. In those cases it is best to generate another random triangle, however, simulating with a triangle that has accidentally some dependencies will hardly influence the results.

The same 10 triangles we used for the Chi-Square test are used to test if the development factors are uncorrelated. For all triangles, the equation $E\left[\widehat{f_2} * \cdots * \widehat{f_N}\right] = E\left[\widehat{f_2}\right] * \cdots * E\left[\widehat{f_N}\right]$ is met, indicating that there is no correlation between the development factors. So, the triangles generated by the dataset generator fulfill the constraints of the Chain Ladder method.

In Section 3.5.1, the constraints for the correlation bootstrap method and the implicit method are described. We will test these assumptions one by one.

The first assumption is easily met, as the MAS-parameter and the settlement pattern are set manually. Nevertheless, it is important to keep in mind that the variance and development factors need to be unequal to zero. This will make sure the complete triangle can be used in the bootstrapping process.

To test the second constraint, we performed the Durbin-Watson test to test for autocorrelation [Bowerman, 2016]. For all 10 triangles, the first 15 accident years are used to perform the Durbin-Watson test to make sure that there are enough residuals available. These 150 tests all yield a test statistic between 1.5 and 2.5, which indicates that there is no autocorrelation in the triangles. This is in line with expectations, as the residuals are randomly drawn from the normal distribution.

For the third constraint, we derived the residuals from the 10 generated datasets. The average of the residuals is within +/- 0.05 from 0 and the variance is within +/- 0.05 from 1. As with the correlation between the residuals, a higher number of residuals in the triangle improves the deviations from the desired mean and variance.

To test whether the residuals are normally distributed, we used the Kolmogorov-Smirnov normality test and the Jarque-Bera test [Thadewald & Büning, 2007] [Bowerman, 2016]. In both tests, we tested H_0 'The data follows a normal distribution with P=0.05' against H_1 'The data does not follow the normal distribution with P=0.05'. It is possible to reject H_0 if the decision variable resulting from the test is lower than 0.05.

For the Kolmongorov-Smirnov test, all the P-value are higher than 0.05 for all 10 triangles, indicating that the residuals are normally distributed. By applying the Jarque-Bera test on the same 10 triangles, the P-value is not in all cases higher than 0.05. This difference can be explained by the nature of the tests, the Kolmonogorov-Smirnov test compares the residuals to a set of randomly drawn numbers from the normal distribution. As the residuals derived from the triangle are based on the same seed as the random draws from the normal distribution, this lowers the difference between the two distributions, resulting in a higher test-statistic. Furthermore, the Jarque-Bera test focuses on the skewness and kurtosis of the distribution of the residuals, both can be present in the triangles if relative extreme residuals end up in the tail of the triangles. By only including the development periods with at least 4 residuals, the tail behaviour is less present and the Jarque-Bera test also yields P-values higher than 0.05 for all 10 triangles.

6.4 Verification of the simulation model

The verification process of the simulation model contains multiple steps. To start the verification process, the code in R was first made more efficient, generalizing the functions and avoiding unnecessary loops. To make sure all functions used were implemented correctly, we checked the description of the functions again to avoid the risk of wrong interpretation of the functions. We also compared the outcomes of the functions to the outcomes of the Excel sheet of England [2003] in which Mack's bootstrapping technique was already worked out.

As a final check, we rebuilt the simulation model developed in R by extending the Excel sheet of England [2003] with the correlation bootstrap method and the implicit correlation method. 10 different datasets are created in the dataset generator and simulated in R and Excel. The bootstrapping process will be performed in a slightly different order in both programs as a result of the different seed that is used in R and Excel, but as the same triangles are used, this may not have too much impact on the outcomes of the simulations. The bootstrapping process is repeated 100.000 times in both models. In Table 4 are the resulting correlation estimates of the 10 simulation runs.

Table 4: Correlation estimates Excel model vs outcomes R model.

	1	2	3	4	5	6	7	8	9	10
CBM Excel	-0,19	0,06	0,22	0,14	0,41	0,04	0,51	-0,2	0,03	0,26
CBM R	-0,19	0,03	0,23	0,17	0,41	0,03	0,48	-0,19	0	0,21
ICM Excel	-0,17	0	0,22	0,01	0,28	-0,02	0,57	-0,08	-0,08	0,36
ICM R	-0,16	-0,03	0,23	0,08	0,26	-0,03	0,54	-0,07	-0,09	0,29

For the correlation bootstrap method (CBM), it can be concluded that the difference between the row containing the results from Excel and the rows containing the results from R are within an acceptable range. Only in the case of the 10th dataset, a relatively high deviation can be seen. However, this difference is likely the result of the different used seeds. For the implicit correlation method, the 4th and the 10th dataset generated a relatively high difference. To make sure that this difference is not created by a programming error, we bootstrapped both datasets as well using seed 1702 instead of 1701, this results for both datasets in a difference of 0.03. So, it can be concluded that there are no significantly different outcomes between the model developed in R and the model developed in Excel.

6.5 Calibration of the simulation model

To determine the number of simulations, we measured the standard deviation in the outcomes by performing 100 times 10.000, 20.000, 50.000, and 100.000 simulation runs for the same dataset. This process is performed for an AA-triangle and an AQ-triangle. In Table 5, the results of these simulations can be found.

Table 5: Standard deviation at different number of simulations.

Number of simulations	CBM AA	ICM AA	CBM AQ	ICM AQ
100.000 x100	0,012	0,014	0,012	0,009
50.000 x100	0,016	0,017	0,019	0,016
20.000 x100	0,027	0,030	0,027	0,021
10.000 x100	0,015	0,116	0,037	0,029

From Table 5, it becomes clear that the standard deviations decreases as the number of simulation runs increases. Based on these results, we decide to use 100.000 simulation runs for every simulation. A higher number of simulation runs is possible, but this increases the run-time of the simulation model significantly, reducing the number of simulations that can be performed during the available simulation time. It can be expected that 99.7% of the outcomes will be within 3 times the standard deviation of the mean as it is assumed the correlation estimates are normally distributed, which would mean that the maximum deviation for the AA-triangle in the case of the implicit correlation method would be 0.042. This will be taken into account when analysing the results in Section 7.

6.6 The impact of different residual sets

The first parameter we will test is the impact different residual sets have on the correlation estimates when keeping the other parameters equal. The parameters used for this test are of the dataset described as "Long tail" in Section 4 at the correlation level of 0, as the deviation might be the highest around this correlation level according to (32). In Tables 6 and 7 the outcomes of 10 pairs of triangles containing the "Long tail&Long tail" parameters are depicted to get an idea of the deviations per residual set. In the first row of the tables are the correlation derived from the residual set which was drawn from the bivariate distribution as explained in Section 5.1. So, the first row of the tables contains the input correlation, the second and third line the output correlation estimates after the methods are used.

Table 6: Correlation estimates AA-paid triangles with different residuals.

AA-Paid

	1	2	3	4	5	6	7	8	9	10
Bivariate	0,01	-0,12	0,07	0,06	0,10	-0,11	-0,11	-0,03	0,01	-0,05
CBM	-0,03	-0,11	0,08	0,14	0,10	-0,13	-0,10	0,00	0,02	-0,05
ICM	0,04	-0,30	0,23	0,02	0,17	-0,01	-0,30	-0,09	-0,07	0,00

Table 7: Correlation estimates AA-incurred triangles with different residuals.

AA-Incurred

	1	2	3	4	5	6	7	8	9	10
Bivariate	0,01	-0,12	0,07	0,06	0,10	-0,11	-0,11	-0,03	0,01	-0,05
CBM	-0,02	-0,11	0,08	0,15	0,08	-0,13	-0,11	0,03	0,05	-0,08
ICM	-0,22	-0,32	0,15	-0,01	0,33	0,03	-0,23	-0,22	-0,20	-0,25

As a result of the same used seed and the equal dimensions of the AA-triangles used in Tables 6 and 7, the correlations obtained from the residuals in the bivariate-row are equal. It also becomes apparent that the differences between the bivariate row and the correlation bootstrap method (CBM) row are relatively small in both tables. Only at the fourth simulation, there is a significant difference between the correlation estimate in the bivariate-row and the CBM row.

The difference between the correlation estimates in the bivariate-row and the correlation obtained by the implicit correlation method (ICM) are significantly bigger compared to the differences in the case of the correlation bootstrap method. In almost all cases, the difference between the correlation estimates in the bivariate row and the ICM row are higher for the incurred triangles compared to the paid triangles. This difference can most likely be explained by the fact that the MAS parameter is higher for the incurred triangles. A higher MAS parameter results in a higher SCR and so potentially gives a less smooth tail, which could cause the relative high differences to occur.

However, when analyzing the correlation estimate at different percentiles, it becomes clear that this estimate is quite smooth for all tested percentiles as can be seen in Table 8. The values in Table 8 are of the same run as the first run of the AA-incurred simulation run of Table 7. A comparable smooth pattern can be seen at all the other simulation runs and this emphasizes that the number of simulations is enough to get an accurate estimation at the 99.5% percentile.

Table 8: The correlation estimate at different percentiles.

Percentiles	80%	90%	95%	99%	99,5%
CBM	-0,01	-0,03	-0,04	-0,04	-0,02
ICM	-0,23	-0,24	-0,24	-0,23	-0,22

The difference between the two methods may be explained by a different MAS parameter for the combined triangle in the implicit correlation method compared to the correlation bootstrap method, resulting in different P&L estimates and so in a significantly different correlation estimate. We will investigate this further in Section 7.

Table 9: Correlation estimates AQ-paid triangles with different residuals.

AQ-paid

	1	2	3	4	5	6	7	8	9	10
Bivariate	-0,03	0,02	0,04	0,00	0,00	0,02	0,02	-0,02	0,00	0,03
CBM	-0,02	0,03	0,04	0,03	-0,01	0,01	0,01	-0,03	0,02	0,02
ICM	0,06	-0,01	-0,01	0,00	0,07	0,00	-0,03	0,09	-0,04	0,00

Table 10: Correlation estimates AQ-incurred triangles with different residuals.

AQ-Incurred

	1	2	3	4	5	6	7	8	9	10
Bivariate	-0,03	0,02	0,04	0,00	0,00	0,02	0,02	-0,02	0,00	0,03
CBM	-0,03	0,03	0,00	0,01	0,01	-0,01	0,01	0,01	0,02	0,00
ICM	-0,33	0,00	-0,14	-0,19	-0,17	0,00	-0,13	-0,06	0,10	-0,18

In general, the observations for the AA-triangles hold also for the AQ-triangles as can be seen in Table 9 and Table 10. Notable is that the differences are generally smaller, which can be explained by the increased number of residuals used in these triangles. The impact of individual extreme residuals on the P&L is smaller in these triangles, which creates a smoother tail pattern.

From the analysis of the residual sets, we can conclude that the residual sets may have a significant impact on the correlation estimates obtained by the correlation bootstrap method and the implicit correlation method. Especially for the AA-triangles, this observation needs to be taken into account when analyzing the results in Section 7. To investigate the impact of the individual parameters on the correlation estimates in Section 7, every table will start with the correlation estimates obtained from the residuals of the datasets used, indicating the correlation estimate at the base scenario. If the correlation estimate changes significantly, this will be the result of the parameter change. If there is almost no deviation, we can conclude that a parameter has only a little or no impact on the correlation estimates.

6.7 Summary

In this section, we performed the verification of both the dataset generator and the simulation model. We can conclude that there are no significant differences between the outcomes of the simulation model in Excel and the simulation model in R. Secondly, we investigated what the accuracy of the dataset generator is. We can conclude that the obtained differences are in line with the expected differences as a result of the limited sample size. To make sure the differences which occur as a result of this limited sample size do not influence the analysis in Section 7, the correlation estimates belonging to the residual sets will be available in the result tables. Thirdly, we investigated whether the dataset generator creates datasets that fulfil the assumptions. This is the case, so we can use the dataset generator to create datasets for the simulations. Fourthly, we determined that the triangles will be bootstrapped 100.000 times before the SCR is calculated, to make sure that there are enough observations in the tail. Lastly, we investigated the impact of different residual sets. We concluded that the deviations from the intended correlation can be quite high, especially for the implicit correlation method. To make sure the other parameters can be analysed properly, every result table in Section 7 contains a base scenario. This makes sure that it is possible to compare the differences in the simulation outcomes.

7 Simulation results

In this section, we discuss the simulation results. We will follow the same order in this section as the order in which the simulation study is performed. Firstly, we test the parameters one by one. Secondly, we will investigate whether certain combinations of parameters create a higher deviation from the preset correlation compared to their individual deviation. Thirdly, we challenge the assumptions of the correlation bootstrap method and the implicit correlation. Finally, we translate the outcomes to the reality and we discuss their implications.

7.1 The impact of the MAS parameter, the settlement periods and the initial claim amounts

With the results of the impact of different residual sets known, it is possible to individually check the impact of the other parameters which can be changed in the dataset generator. To get more insights in the outcomes of the correlation bootstrap method and the implicit correlation method, a grouping-method will be used. We will group the simulation results based on the difference between the simulation result and the correlation derived of the residuals after they were drawn from the bivariate normal distribution. Three groups are made, deviations smaller than 0.05, deviations between 0.05 and 0.1, and deviations bigger than 0.1. In Table 11 an example table of the output of the simulations is given. In Table 12, the deviations between the simulation output and the correlation measured after the residuals are displayed, making it easier to analyse the simulation outcomes. These two tables guide as a reading example to make sure the simulation outcomes are interpreted correctly.

Table 11: Example correlation table.

		Example absolute					
		Long tail&Long tail		Short tail&Short tail		Long tail&Short tail	
Correlation	Bivariate	CBM	ICM	CBM	ICM	CBM	ICM
-0,25	-0,30	-0,26	-0,31	-0,26	-0,33	-0,27	-0,25
0	-0,12	-0,10	0,01	-0,09	-0,01	-0,09	0,03
0,25	0,26	0,26	-0,02	0,26	-0,01	0,22	-0,07
0,5	0,50	0,49	0,56	0,49	0,58	0,47	0,56
0,75	0,72	0,76	0,67	0,74	0,65	0,74	0,64

Table 12: Example result table.

		Example deviation					
		Long tail&Long tail		Short tail&Short tail		Long tail&Short tail	
Correlation	Bivariate	CBM	ICM	CBM	ICM	CBM	ICM
-0,25	-0,30	0,04	-0,01	0,03	-0,04	0,03	0,05
0	-0,12	0,01	0,13	0,03	0,11	0,03	0,15
0,25	0,26	0,01	-0,27	0,00	-0,27	-0,04	-0,33
0,5	0,50	-0,01	0,06	-0,02	0,08	-0,03	0,06
0,75	0,72	0,04	-0,05	0,02	-0,07	0,01	-0,09

Legend

	Deviation within 0,05 of the bivariate correlation.
	Deviation between 0,05 and 0,1 of the bivarate correlation.
	Deviation of more than 0,1 of the bivariate correlation.

In the most left column of Table 11 are the correlation levels, as described in Section 4. On the right side of the correlation column, is the column containing the correlation estimates derived from the residuals after they were drawn from the bivariate normal distribution with the correlation from the most left column. From Section 6.4 we know that the deviation between these first two columns can be quite high, so it is important to known this deviation when analyzing the result. In the remainder of Table 11 are the correlation estimates for the correlation bootstrap method (CBM) and the implicit correlation method (ICM) per tested situation. For example, if Table 11 would represent the correlation estimates from varying the MAS parameter in the dataset generator, the results below Long&Long would represent the outcomes of a simulation in which a low MAS parameter was used for both triangles. So, the correlation estimate of -0.26 below CBM of Long&Long represents the measured correlation from the simulation in which the residuals contained a correlation of -0.3.

In Table 12, the difference between the correlation obtained from the triangle before it was bootstrapped and the correlation obtained after the triangle was bootstrapped is displayed. Taking the same example of Table 11, the difference between -0.3 and -0.26 is 0.04, so this is the number below CBM at the -0.25 correlation level in Table 12. If the difference is within -0.05 and 0.05 of the correlation derived from the residuals, we assume the correlation estimate is accurate enough and does not get a color. If the difference is between -0.1 and -0.05 or 0.05 and 0.1, the deviation gets a light orange color indicating a moderate difference. If the deviation is less than -0.1 or more than 0.1, the correlation estimate gets a dark orange color indicating a strong difference from the correlation derived from the residuals.

In the remainder of this section, the results will be presented in the form of Table 12, to make it easier to interpret the results. We introduced the data used in the simulations briefly in Section 4, but this needs extra attention before the results can be explained correctly.

In Table 13, the parameters are compared. So in case of the MAS parameter of a paid triangle, the triangle denoted as the “Long tail” triangle has a relatively low MAS parameter compared to the triangle with a relatively “Short tail”. In case of the settlement pattern in the paid triangles, the settlement pattern in the triangle with a “Long tail” has a relative stable settlement pattern whereas the triangle with a “Short tail” has a more unstable settlement pattern. For the initial claim amounts, the variance is measured between the accident years, in case of the triangle with a “Long tail”, the variance is relatively low compared to the variance in the initial claim amounts of the triangle with a “Short tail”.

Table 13: Data used in the simulation runs.

	MAS		Settlement periods		Initial claim amounts	
	Long	Short	Long	Short	Long	Short
Paid	Low	High	Low	High	Low	High
Incurred	High	Low	High	Low	Low	High

In the simulation performed with an AQ-paid triangle, the first three development periods are excluded from the model, taking the cumulative claims at the end of the fourth quarter as the starting point for the simulation. This is done to exclude the first 3 observations from the MAS vector as these are relatively high and resemble the real world in a better way, in which the first development year often needs to be excluded as a result of the possibility of seasonal effects, creating autocorrelation. In Appendix E, the differences between the outcomes of the AQ-paid triangle with and without the first development year can be analyzed. Furthermore, the outcomes of Sections 7.1.1, 7.1.2 and 7.1.3 are visualized in a different order, making it easier to see the development in a particular triangle. Lastly, Appendix E contains the outcomes as given in Table 11.

7.1.1 The MAS parameter

The first parameter we test, is the MAS parameter. In every simulation, the settlement periods parameter and the initial claim amounts are taken from the dataset which corresponds to Long tail&Long tail, as these input parameters have the lowest possibility to yield negative cumulative claims. For every type of triangle, the first tested situation contains the MAS parameter of the dataset containing a low MAS parameter, in the second simulation run the MAS parameter of the dataset containing a high MAS parameter is taken and in the third simulation one triangle got the low MAS parameter and the other the high MAS parameter. In Tables 14 – 17, the outcomes for the simulations in which only the MAS parameter is changed are shown.

Table 14: AA paid – MAS parameter.

		AA paid - MAS					
		Long tail&Long tail		Short tail&Short tail		Long tail&Short tail	
Correlation	**Bivariate**	CBM	ICM	CBM	ICM	CBM	ICM
-0,25	-0,30	0,01	-0,01	0,01	-0,06	0,01	-0,02
0	-0,12	0,00	-0,19	0,00	-0,20	-0,01	-0,14
0,25	0,26	-0,03	-0,11	-0,05	0,01	-0,05	-0,04
0,5	0,50	-0,06	-0,03	-0,05	0,02	-0,11	-0,06
0,75	0,72	0,00	-0,05	0,01	-0,08	-0,08	-0,13

Table 15: AQ paid - MAS parameter.

		AQ paid - MAS					
		Long tail&Long tail		Short tail&Short tail		Long tail&Short tail	
Correlation	**Bivariate**	CBM	ICM	CBM	ICM	CBM	ICM
-0,25	-0,21	0,03	0,04	0,04	0,15	0,07	0,16
0	0,02	0,00	-0,06	0,00	-0,15	0,00	-0,09
0,25	0,25	-0,01	-0,05	-0,01	-0,04	-0,04	-0,17
0,5	0,53	0,01	0,03	-0,02	-0,01	-0,14	-0,15
0,75	0,79	-0,02	0,00	-0,02	-0,06	-0,18	-0,20

Table 16: AA incurred – MAS parameter.

		AA Incurred - MAS					
		Long tail&Long tail		Short tail&Short tail		Long tail&Short tail	
Correlation	**Bivariate**	CBM	ICM	CBM	ICM	CBM	ICM
-0,25	-0,30	-0,01	0,28	0,02	0,04	0,06	0,18
0	-0,12	0,00	-0,21	0,01	-0,02	0,02	-0,08
0,25	0,26	-0,04	0,00	-0,03	0,06	-0,12	-0,10
0,5	0,50	-0,07	-0,06	-0,08	0,00	-0,20	-0,23
0,75	0,72	0,02	0,07	0,01	0,03	-0,15	-0,15

Table 17: AQ incurred - MAS parameter.

		AQ Incurred - MAS					
		Long tail&Long tail		Short tail&Short tail		Long tail&Short tail	
Correlation	**Bivariate**	CBM	ICM	CBM	ICM	CBM	ICM
-0,25	-0,21	0,03	-0,07	0,04	0,03	0,14	0,15
0	0,02	0,01	-0,02	0,01	-0,02	-0,01	0,02
0,25	0,25	-0,01	-0,27	-0,02	-0,31	-0,18	-0,26
0,5	0,53	0,00	0,21	0,01	0,04	-0,36	-0,33
0,75	0,79	-0,01	-0,03	-0,01	-0,04	-0,53	-0,54

Overall, we can conclude that the largest differences occur in the simulations in which a triangle with a low MAS parameter and a triangle with a high MAS parameter are simulated based on the number of (dark) orange coloured instances. In the simulations with AQ-triangles there are less deviations compared to the AA-triangles, but if there is a deviation these seem to deviate more compared to the deviations in the AA-triangles. Furthermore, the correlation bootstrap method is performing substantially better compared to the implicit correlation method. Only in a limited number of cases, the implicit correlation method performs slightly better than the correlation bootstrap method.

It is remarkable that the correlation estimates in the Long&Long column are not seen in the other columns. This means that the impact of the residuals set is wiped out by the impact of the MAS parameter.

Another way to look to the outcomes of the simulation is by looking to the difference between the expected SCR and the obtained SCR of the simulation study. The expected SCR can be determined based on the correlation coefficient in the bivariate column and the SCR of the two individual triangles. By rewriting (23), we find the following equation.

$$SCR_{x,y} = \sqrt{(2\rho SCR_x SCR_y + {SCR_x}^2 + {SCR_y}^2)} \tag{33}$$

By applying (33) on the results of Table 15, we can determine the expected SCR for every simulation. To be able to compare the differences between the expected SCR and the SCR of the simulation study with each other, these are translated to percentages.

Table 18: AQ paid - MAS parameter.

		AQ paid - MAS					
		Long tail&Long tail		Short tail&Short tail		Long tail&Short tail	
Correlation	Bivariate	CBM	ICM	CBM	ICM	CBM	ICM
-0,25	-0,21	0,03	0,04	0,04	0,15	0,07	0,16
0	0,02	0,00	-0,06	0,00	-0,15	0,00	-0,09
0,25	0,25	-0,01	-0,05	-0,01	-0,04	-0,04	-0,17
0,5	0,53	0,01	0,03	-0,02	-0,01	-0,14	-0,15
0,75	0,79	-0,02	0,00	-0,02	-0,06	-0,18	-0,20

Table 19: Difference in the obtained SCR.

	AQ paid - MAS difference in SCR					
	Long tail&Long tail		Short tail&Short tail		Long tail&Short tail	
Correlation	CBM	ICM	CBM	ICM	CBM	ICM
-0,25	2,0%	2,7%	2,6%	9,2%	2,5%	5,4%
0	-0,2%	-2,8%	0,2%	-7,8%	-0,1%	-3,0%
0,25	-0,2%	-2,1%	-0,2%	-1,6%	-1,1%	-4,3%
0,5	0,3%	1,1%	-0,5%	-0,4%	-3,2%	-3,4%
0,75	-0,5%	-0,1%	-0,6%	-1,7%	-3,8%	-4,3%

In Table 19 are the percentage differences between the expected SCR and the obtained SCRs of the simulation study belonging to the results in Table 18. This gives some intuition into the impact a wrong estimate has on the SCR. For all result tables in Section 7.1, the percentage difference table can be found in Appendix E.

7.1.2 The settlement periods parameter

The same method as we used for the MAS parameter, is applied for the settlement periods parameter. So, all parameters are kept constant at the Long tail&Long tail level, whereas the settlement pattern is varied per simulation run.

Table 20: AA paid – Settlement periods.

		AA paid - Settlement periods					
		Long tail&Long tail		Short tail&Short tail		Long tail&Short tail	
Correlation	Bivariate	CBM	ICM	CBM	ICM	CBM	ICM
-0,25	-0,30	0,01	-0,01	0,01	-0,01	0,01	0,00
0	-0,12	0,00	-0,19	0,01	-0,19	0,00	-0,19
0,25	0,26	-0,03	-0,11	-0,03	-0,08	-0,02	-0,08
0,5	0,50	-0,06	-0,03	-0,06	-0,03	-0,06	-0,03
0,75	0,72	0,00	-0,05	0,00	-0,03	0,00	-0,05

Table 21: AQ paid – Settlement periods.

		AQ paid - Settlement periods					
		Long tail&Long tail		Short tail&Short tail		Long tail&Short tail	
Correlation	Bivariate	CBM	ICM	CBM	ICM	CBM	ICM
-0,25	-0,21	0,03	0,04	0,03	0,05	0,03	0,06
0	0,02	0,00	-0,06	0,00	-0,06	-0,01	-0,06
0,25	0,25	-0,01	-0,05	-0,01	-0,05	0,00	-0,04
0,5	0,53	0,01	0,03	0,01	0,03	0,01	0,01
0,75	0,79	-0,02	0,00	-0,02	0,00	-0,02	-0,01

Table 22: AA incurred – Settlement periods.

		AA Incurred - Settlement periods					
		Long tail&Long tail		Short tail&Short tail		Long tail&Short tail	
Correlation	Bivariate	CBM	ICM	CBM	ICM	CBM	ICM
-0,25	-0,30	-0,01	0,28	-0,01	0,29	-0,01	0,28
0	-0,12	0,00	-0,21	0,01	-0,21	0,01	-0,21
0,25	0,26	-0,04	0,00	-0,04	0,00	-0,04	0,00
0,5	0,50	-0,07	-0,06	-0,07	-0,06	-0,07	-0,07
0,75	0,72	0,02	0,07	0,02	0,07	0,01	0,08

Table 23: AQ incurred – Settlement periods.

		AQ Incurred - Settlement periods					
		Long tail&Long tail		Short tail&Short tail		Long tail&Short tail	
Correlation	Bivariate	CBM	ICM	CBM	ICM	CBM	ICM
-0,25	-0,21	0,03	-0,07	0,03	-0,07	0,03	-0,07
0	0,02	0,01	-0,02	0,01	-0,02	0,01	-0,02
0,25	0,25	-0,01	-0,27	-0,01	-0,27	-0,01	-0,26
0,5	0,53	0,00	0,21	0,00	0,22	0,00	0,21
0,75	0,79	-0,01	-0,03	-0,01	-0,03	0,00	-0,03

Compared to the results of Section 7.1.1 the deviations for the development factors are quite stable. Just looking to the orange color it becomes clear that the difference in settlement pattern hardly changes the correlation estimates. The difference observed between the results in the CBM columns and the ICM columns can be explained by the impact of the residual sets, which we already investigated in Section 6.6.

7.1.3 The initial claim amounts parameter

The initial claim amounts have a lot of impact on the SCR, especially when the initial claim amounts are characterized with a lot of variance. In Tables 24-27 are the outcomes of the simulation runs in which we simulated the impact of the the initial claim amounts parameter. These simulations are performed in the same way as for the MAS parameter and the settlement periods parameter.

Table 24: AA-paid – Initial claim amounts.

		AA paid - Initial claim amounts					
		Long tail&Long tail		Short tail&Short tail		Long tail&Short tail	
Correlation	Bivariate	CBM	ICM	CBM	ICM	CBM	ICM
-0,25	-0,30	0,01	-0,01	0,01	0,00	0,01	-0,03
0	-0,12	0,00	-0,19	0,00	-0,18	0,02	-0,22
0,25	0,26	-0,03	-0,11	-0,03	-0,11	-0,03	-0,08
0,5	0,50	-0,06	-0,03	-0,05	-0,02	-0,06	-0,05
0,75	0,72	0,00	-0,05	0,02	-0,05	0,00	-0,05

Table 25: AQ-paid – Initial claim amounts.

		AQ paid - Initial claim amounts					
		Long tail&Long tail		Short tail&Short tail		Long tail&Short tail	
Correlation	Bivariate	CBM	ICM	CBM	ICM	CBM	ICM
-0,25	-0,21	0,03	0,04	0,03	0,03	0,03	0,03
0	0,02	0,00	-0,06	0,00	-0,06	0,00	-0,05
0,25	0,25	-0,01	-0,05	0,00	-0,05	-0,01	-0,07
0,5	0,53	0,01	0,03	-0,01	0,02	-0,01	0,04
0,75	0,79	-0,02	0,00	-0,01	0,02	-0,03	-0,03

Table 26: AA incurred – Initial claim amounts.

		AA Incurred - Initial claim amounts					
		Long tail&Long tail		Short tail&Short tail		Long tail&Short tail	
Correlation	Bivariate	CBM	ICM	CBM	ICM	CBM	ICM
-0,25	-0,30	-0,01	0,28	0,00	0,29	0,01	0,28
0	-0,12	0,00	-0,21	0,00	-0,22	0,00	-0,24
0,25	0,26	-0,04	0,00	-0,03	0,02	-0,04	0,01
0,5	0,50	-0,07	-0,06	-0,09	-0,07	-0,09	-0,06
0,75	0,72	0,02	0,07	0,03	0,09	0,01	0,05

Table 27: AQ-incurred – Initial claim amounts.

		AQ Incurred - Initial claim amounts					
		Long tail&Long tail		Short tail&Short tail		Long tail&Short tail	
Correlation	Bivariate	CBM	ICM	CBM	ICM	CBM	ICM
-0,25	-0,21	0,03	-0,07	0,04	-0,05	0,05	-0,26
0	0,02	0,01	-0,02	0,02	-0,01	-0,02	-0,18
0,25	0,25	-0,01	-0,27	-0,01	-0,26	-0,04	0,03
0,5	0,53	0,00	0,21	0,02	0,22	-0,05	0,15
0,75	0,79	-0,01	-0,03	-0,01	-0,03	-0,11	-0,23

The results in Tables 24 - 27 show a similar pattern as the results in Section 7.1.2. This indicates that the initial claim amounts have almost no impact on the correlation estimates and the differences can be attributed to the residual set as described in Section 6.6.

More remarkable are the outcomes in Table 27, where the Long tail&Short tail columns show a deviation from the Long tail&Long tail and Short tail&Short tail columns. Especially for the implicit correlation method, the outcomes are completely different compared to the other outcomes. To get more insights in this specific outcome, the parameters for the log-normal distribution of the initial claim amounts are derived in R from both datasets and three additional simulation runs are performed at every correlation level. In Table 28 the outcomes of this simulation can be seen.

Table 28: Varying the initial claim amounts.

		AQ Incurred - LogNormal first dev. Period					
		Long tail&Long tail		Short tail&Short tail		Long tail&Short tail	
Correlation	Bivariate	CBM	ICM	CBM	ICM	CBM	ICM
-0,25	-0,20	0,02	-0,38	0,01	-0,05	0,14	0,13
0	0,02	0,00	0,06	-0,01	0,26	-0,01	0,06
0,25	0,22	-0,03	-0,12	0,00	0,05	-0,16	-0,17
0,5	0,56	-0,09	-0,02	-0,08	0,07	-0,45	-0,42
0,75	0,77	-0,02	-0,03	-0,02	0,04	-0,57	-0,53

The results in Table 28 are worse compared to the results in Table 17. This can be attributed to the MAS parameter which is not adjusted to the randomly drawn initial claim amounts. If there is a relative low number of extreme observations in the initial claim amounts, the MAS parameter which was meant for the relative bigger initial claim amounts will scale the residuals much more than intended, creating bigger differences from the average development factor.

The outcomes from the simulations in which one parameter was changed in every simulation run shows that the MAS parameter and the residual set have the most impact on the correlation estimates. Furthermore, we can conclude that if there is a higher MAS parameter, the more the outcomes differ, especially when combined with a triangle with another distribution for the initial claim amounts. The settlement pattern hardly has any impact on the outcomes of the simulations. Lastly, we can conclude that the correlation bootstrap method performs substantially better than the implicit correlation method in all circumstances.

7.2 Combined parameters

With the obtained knowledge of the individual impact of the different parameters, it is also interesting to see whether certain combinations of parameters create more or less differences as in the individual examples in Section 7.1.

The situations which we are going to simulate follow on the simulations performed in Section 7.1, now the parameters are changed in pairs to see whether some combinations create stronger or weaker deviations. The situations are displayed in Table 29 and will be simulated for every type of triangle.

Table 29: Other situations simulated.

MAS parameter	Short tail & Short tail	Short tail & Short tail	Short tail & Short tail	Short tail & Short tail	Long tail & Short tail
Settlement periods	Short tail & Short tail	Long tail & Short tail	Short tail & Short tail	Long tail & Long tail	Long tail & Short tail
Initial claim amounts	Long tail & Long tail	Long tail & Long tail	Short tail & Short tail	Long tail & Short tail	Long tail & Short tail

The exact results of these simulations can be found in Appendix F. In general, it can be concluded that the results for the individual parameters also hold for the combination of parameters. In almost all cases, the results resemble the correlation estimates which were already visible in Section 7.1. This underpins the results of Section 7.1 that most deviations are created when the two triangles contain a different MAS parameter.

7.3 Impact volume of SCR

A complication of different MAS parameters is that it is likely that the SCR estimates are quite different in terms of volume. For example, if the portfolio size of portfolio A is 10 times as big as the size of portfolio B, it is likely that only the extreme outliers of portfolio B may have an impact on the residuals of the combined portfolio. If the MAS parameter of the combined model increases or decreases slightly compared to the MAS parameter in portfolio A, it is possible that the SCR changes slightly having an extreme impact on the correlation estimate.

Before the simulations are performed, it is good to have a feeling by the impact a 1% difference in the SCR of portfolio A, the SCR of portfolio B and the SCR of portfolio A+B have on the correlation outcomes. In Table 30 the impact is showed in case there is a lot of difference between SCR A and SCR B, in Table 31 the impact of a 1% difference is showed in case SCR A and SCR B are of a comparable volume. The correlation in these tables is determined with (23).

Table 30: Impact of a 1% difference in SCR if SCR A is 10x bigger than to SCR B.

Scenario	1	2	3	4	5	6	7
SCR portfolio A	100.000	99.000	101.000	100.000	100.000	100.000	100.000
SCR portfolio B	10.000	10.000	10.000	9.900	10.100	10.000	10.000
SCR portfolio A+B	107.000	107.000	107.000	107.000	107.000	105.930	108.070
Correlation	0,67	0,78	0,57	0,68	0,67	0,56	0,79

Table 31: Impact of a 1% difference in SCR if SCR A and SCR b are equal.

Scenario	8	9	10	12	13
SCR portfolio A	100.000	99.000	101.000	100.000	100.000
SCR portfolio B	100.000	100.000	100.000	100.000	100.000
SCR portfolio A+B	183.000	183.000	183.000	181.170	184.830
Correlation	0,67	0,69	0,66	0,64	0,71

As can be seen in Tables 30 and 31, the impact of an 1% change in the SCR has much more impact on the correlation estimate in case there is a lot of difference in the volume of SCR A and SCR B. Especially, when the SCR of portfolio A+B is changed, the impact on the correlation estimate is significant. This same effect may occur when two portfolios are added and the MAS parameter is slightly changed.

To test whether the correlation estimate becomes better if the SCRs are of a comparable size, the simulations in which the MAS parameter was tested with the Short tail&Long tail MAS parameter are tested again. But now, the triangle containing the lowest SCR is scaled by the factor:

$$Scaling\ factor = \frac{SCR_{highest}}{SCR_{lowest}} \tag{34}$$

This scaling is allowed, as it does not influence the residuals and so should still contain the preset correlation.

Table 32: AA-paid correlation estimates.

		AA paid - MAS			
		Unequal SCR		Equal SCR	
Correlation	Bivariate	CBM	ICM	CBM	ICM
-0,25	-0,30	0,18	0,05	0,19	0,00
0	-0,12	0,00	-0,13	0,01	-0,18
0,25	0,26	-0,02	-0,05	-0,06	-0,08
0,5	0,50	-0,06	-0,05	-0,07	-0,04
0,75	0,72	-0,08	-0,13	0,28	0,28

Table 33: AA-Incurred correlation estimates.

		AA Incurred - MAS			
		Unequal SCR		Equal SCR	
Correlation	Bivariate	CBM	ICM	CBM	ICM
-0,25	-0,30	0,08	0,17	0,06	0,03
0	-0,12	0,05	-0,07	0,04	-0,05
0,25	0,26	-0,12	-0,05	-0,10	-0,05
0,5	0,50	-0,19	-0,17	-0,19	-0,19
0,75	0,72	-0,16	-0,12	-0,16	-0,10

Table 34: AQ-paid correlation estimates.

		AQ paid - MAS			
		Unequal SCR		Equal SCR	
Correlation	Bivariate	CBM	ICM	CBM	ICM
-0,25	-0,21	0,07	0,16	0,06	0,13
0	0,02	0,00	-0,09	0,01	-0,09
0,25	0,25	-0,04	-0,17	-0,06	-0,14
0,5	0,53	-0,14	-0,15	-0,11	-0,11
0,75	0,79	-0,18	-0,20	-0,21	-0,20

Table 35: AQ-Incurred correlation estimates.

		AQ Incurred - MAS			
		Unequal SCR		Equal SCR	
Correlation	Bivariate	CBM	ICM	CBM	ICM
-0,25	-0,21	0,14	0,15	0,18	0,17
0	0,02	-0,01	0,02	-0,01	0,00
0,25	0,25	-0,18	-0,26	-0,19	-0,25
0,5	0,53	-0,36	-0,33	-0,39	-0,34
0,75	0,79	-0,53	-0,54	-0,56	-0,56

In Tables 32-35 the correlation estimates before and after the scaling took place can be observed. In case of the AQ-Paid triangle the correlation estimate were in most cases slightly better after the scaling took place. In case of the AQ-incurred triangle, the scaling made the outcomes in most cases worse than they were before the scaling. In case of the AA-triangles, the scaling improved the outcomes of most of the incurred triangles, however, in case of the AA-paid triangles the outcomes became worse.

7.4 Robustness of the models

To check how much the correlation estimates differ if another last development period is added, we added four different diagonals to the triangles. The four diagonals are determined by bootstrapping the triangle belonging to the 0-correlation level for the Long tailLong tail triangle from 7.1.1, and creating a lower triangle. From this lower triangle, the first diagonal is added to the original dataset for the AA-triangles, in case of the AQ-triangles four additional triangles are added. This will give insights in how robust the correlation estimates are when new data is added to the model. The robustness of the models can be important in practice, as it is not desired to have a highly fluctuating correlation parameter over time.

In Tables 36 - 39 the outcomes of the robustness tests can be seen.

Table 36: AA-paid robustness check.

	1	2	3	4
CBM	-0,08	-0,07	-0,10	-0,09
ICM	-0,32	-0,32	-0,34	-0,33

Table 37: AA-incurred robustness check.

	1	2	3	4
CBM	-0,08	-0,09	-0,09	-0,08
ICM	-0,34	-0,33	-0,33	-0,33

Table 38: AQ-paid robustness check.

	1	2	3	4
CBM	0,02	0,02	0,02	0,02
ICM	0,05	0,04	0,05	0,05

Table 39: AQ-Incurred robustness check.

	1	2	3	4
CBM	-0,03	-0,03	-0,03	-0,01
ICM	0,01	-0,02	0,06	-0,06

Looking to the results in Tables 36 - 39, a difference can be noted between the results of the AA-triangles and the AQ-triangles. Based on these observations, it can be concluded that the correlation bootstrap method creates relatively stable results for all four types of triangles. For the implicit correlation method, the results are relatively stable for all triangles except for the AQ-incurred triangle.

7.5 Outliers

To test the impact of outliers, we simulated three scenarios on all four types of triangles. In the first scenario, a triangle is picked which was used in Section 7.1 in the Long&Long scenario at the 0.75 correlation level. This scenario will serve as the base scenario. In the second scenario, one outlier is added which represents an increase of 3 times the initial value in one of the two triangles. This will create two extreme residuals, namely the residual which lets the claims triple and the up following residual which lets the claims decrease to the initial value. In the third scenario a complete accident year of one dataset will be adjusted to create 10 extreme residuals. In Table 40 are the outcomes of the three scenarios for an AQ-paid triangle in which a specification to the capitals is made.

Table 40: Impact of outliers on the correlation estimate.

	CBM	ICM	SCR1	SCR2	SCR3 CBM	SCR3 ICM
Scenario 1	0,77	0,79	100%	100%	100%	100%
Scenario 2	0,23	1,45	922%	100%	514%	569%
Scenario 3	0,19	2,37	1684%	100%	924%	1024%

To interpret the results of Table 40 correctly, it is good to understand the numbers in the first scenario correctly. The values in the CBM and ICM column represent the correlation estimate and not the deviations as in previous tables. The values in the SCR1, SCR2, SCR3 CBM and SCR3 ICM give a percentual representation of the development of the SCR, taking scenario 1 as the base scenario. In this base scenario, SCR1 and SCR2 are quite comparable in volume, SCR3 is approximately 1.8 times SCR1.

When we add an outlier in one triangle and not in the other, it is expected that the correlation will go down between the two triangles. Taking this into account, the outcomes of the correlation estimate for the correlation bootstrap method in scenario 2 and scenario 3 are not surprising. As more extreme residuals are placed in scenario 3 than in scenario 2, it is expected that the correlation is lower in scenario 3.

In case of the implicit correlation method, the outcomes are more counterintuitive. Where in the case of the correlation bootstrap method it did still hold that $SCR_3 \leq SCR_1 + SCR_2$, in scenario 3 of the implicit correlation method $SCR_3 > SCR_1 + SCR_2$ is the case, creating an unrealistic correlation estimate which is not between -1 and 1. The reason SCR_3 is greater than 1 is in the way SCR_3 is determined. Where a relative extreme SCR_3 can be seen as a conditional probability of $P(SCR_2 = extreme|SCR_1 = extreme)$ in the correlation bootstrap method, this conditional probability is lost in the implicit correlation method as SCR_3 is individually bootstrapped.

Table 41: Results of adding outliers to the dataset.

	AA-paid		AA-Incurred		AQ-Paid		AQ-Incurred	
	CBM	ICM	CBM	ICM	CBM	ICM	CBM	ICM
Scenario 1	0,72	0,67	0,74	0,80	0,77	0,79	0,78	0,77
Scenario 2	0,38	2,08	0,32	0,79	0,23	1,45	0,73	0,77
Scenario 3	0,14	4,00	0,15	3,96	0,19	2,37	0,23	0,64

Based on the outcomes of the outlier test in Table 41, it can be concluded that the correlation bootstrap method performs significantly better if outliers are present in the dataset compared to the implicit correlation method. Furthermore, it becomes clear that the implicit method performs better on the incurred triangles than the paid triangles. To get a better understanding of the impact one outlier has on the rest of the residuals in a development year, a step by step description is added in Appendix G. The impact on the correlation estimate is the highest if it occurs in the first development years, as the most residuals are then affected by the outlier.

7.6 Challenging the assumptions

Now we know the performance of both methods within the assumptions underlying the bootstrapping techniques, it is also interesting to see the impact on the correlation estimates when the assumptions are violated. Firstly, we will challenge the normality assumption of the residuals. Secondly, we will investigated how well it is possible to estimate the correlation if triangles with a different tail length are used.

7.6.1 Normality

The outcomes for the outlier test can also be used as a violation of the normality assumption. By adding extreme outliers to the dataset, the residuals may become skewed and the normality assumption in which the kurtosis needs to be close to 0 is lost [Kaas et al., 2008]. To simulate the impact of a violation of the normality assumption, we draw the residuals from the gamma distribution. The residuals cannot directly be drawn from the gamma distribution as the gamma distribution only yield positive outcomes. It is as well possible to use the lognormal distribution and the student-t distribution to generate the residuals which contain a skewed distribution. The gamma distribution is chosen as this distribution is used to model the simulation error in the general bootstrapping process.

To generate the residuals, we first need a vector with values of the gamma distribution. To create these, a scale parameter of 1 and a shape parameter of 0.5 are chosen, as this generates a sample which is clearly skewed. Secondly, a vector with 0 and 1 values is drawn from the binomial distribution. Lastly, the residual set is determined based on (37), which makes approximately half of the residuals negative.

$$X \sim Gamma(1, 0.5) \tag{35}$$

$$Y \sim Binomial(N, 0.5) \tag{36}$$

$$Z \sim X * Y + (-1 + X) * Y \tag{37}$$

The correlation is measured at the same correlation levels as the situations in Section 7.1. To create the correlated residuals from the gamma distribution, a copula approach is used [Hull, 2018].

The simulations in which we draw the residuals from the gamma distribution, contain the same parameters as in the simulations in which the MAS parameter was investigated in section 7.1.1. The result of these simulations can be found in Tables 42-45.

Table 42: AA paid using gamma distribution.

		AA paid Gamma					
		Long tail&Long tail		Short tail&Short tail		Long tail&Short tail	
Correlation	Bivariate	CBM	ICM	CBM	ICM	CBM	ICM
-0,25	-0,21	0,00	0,00	na	na	na	na
0	-0,10	0,05	0,00	na	na	na	na
0,25	0,36	-0,04	0,08	-0,20	-0,15	-0,08	-0,08
0,5	0,55	-0,02	0,03	-0,12	-0,18	-0,10	-0,16
0,75	0,72	-0,08	-0,05	0,02	0,19	-0,20	-0,12

Table 43: AA incurred using gamma distribution.

		AA incurred Gamma					
		Long tail&Long tail		Short tail&Short tail		Long tail&Short tail	
Correlation	Bivariate	CBM	ICM	CBM	ICM	CBM	ICM
-0,25	-0,21	0,03	0,41	0,01	0,05	0,05	0,12
0	-0,10	0,01	-0,20	0,04	-0,03	0,03	-0,17
0,25	0,36	-0,05	-0,08	-0,05	-0,09	-0,12	-0,23
0,5	0,55	0,02	-0,15	-0,01	-0,12	-0,10	-0,17
0,75	0,72	-0,11	0,08	-0,07	0,02	-0,29	-0,28

Table 44: AQ paid using gamma distribution.

		AQ paid Gamma					
		Long tail&Long tail		Short tail&Short tail		Long tail&Short tail	
Correlation	Bivariate	CBM	ICM	CBM	ICM	CBM	ICM
-0,25	-0,24	0,01	-0,14	0,02	-0,17	0,07	-0,05
0	0,05	0,01	0,02	0,04	-0,02	0,02	-0,18
0,25	0,28	0,02	0,04	0,00	0,12	-0,03	0,25
0,5	0,50	-0,04	0,11	-0,05	0,15	-0,14	0,14
0,75	0,75	-0,03	-0,02	-0,02	-0,09	-0,18	-0,29

Table 45: AQ incurred gamma distribution.

		AQ incurred Gamma					
		Long tail&Long tail		Short tail&Short tail		Long tail&Short tail	
Correlation	Bivariate	CBM	ICM	CBM	ICM	CBM	ICM
-0,25	-0,24	0,03	-0,03	0,00	0,20	0,19	0,15
0	0,05	0,03	0,00	0,00	0,13	-0,02	0,02
0,25	0,28	-0,09	-0,06	-0,03	0,07	-0,22	-0,25
0,5	0,50	0,02	-0,32	-0,02	0,08	-0,40	-0,42
0,75	0,75	-0,01	0,08	-0,03	-0,05	-0,51	-0,50

Overall, the deviations in Table 42-45 became a slightly worse compared to the outcomes of Table 14-17. However, in most cases the correlation estimate hardly differs. This can be explained by the same reason as already discussed in the previous section, as outliers will come out much less extreme than they went into the triangle. This can be seen in Figures 14 and 15.

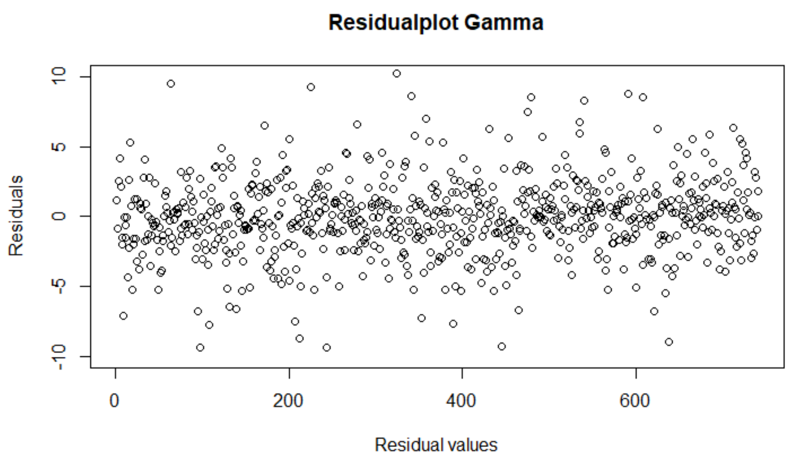


Figure 14: Residuals Gamma ∼ (1,0.5).

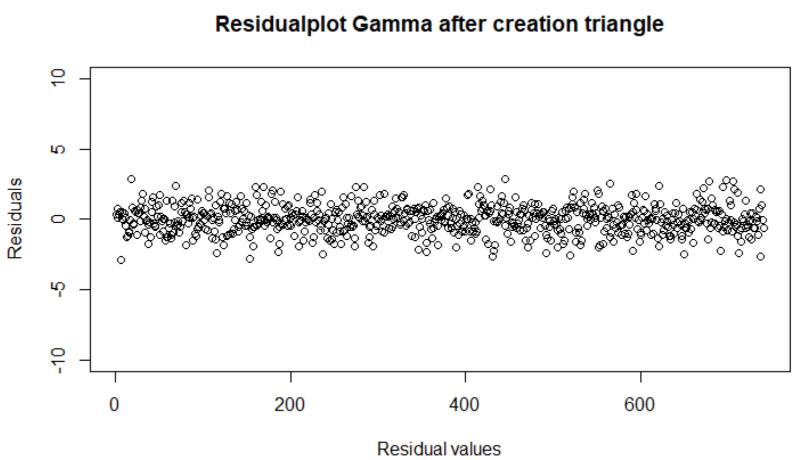


Figure 15: Residuals from triangle with Gamma residuals.

The residuals in Figure 15 are less extreme compared to Figure 14. However, by performing the Kolmongorov-Smirnov normality test on the residuals in Figure 15, it becomes apparent that the residuals are far from normally distributed. This can be explained by the same phenomena that occurred in the previous section about outliers. In Appendix G, this process is step by step explained.

In Table 42, four triangles contained a negative cumulative payment. As explained in Section 5, it is then not possible to complete the triangle and so use it in the simulation model. This is why there is a 'na' in four situations. In Appendix H, the results of a comparable simulation in which 5 scenarios are performed at the 0-correlation level can be found. These outcomes give more insights in the impact of the different sets of residuals. Also when the residuals are chosen from the gamma distribution, comparable deviations can be seen as in Section 6.6 where different residuals sets containing the same correlation parameter are used.

7.6.2 Different tail lengths

In practice the tail lengths are hardly ever the same. With the tail length we mean the number of development periods it takes before 99.5% of the claims are settled. Extending the tail of the shortest triangle results in zero residuals, which will not work in the correlation bootstrap method. To be able to compare the outcomes, we will investigate how important it is to take the complete tail in the correlation estimating process into account. To mimic the situation, we will use a dataset containing the parameters of the long triangle and a dataset containing the parameters of the short triangle, the correlation between the datasets is estimated based on four till eighteen years of development periods. We will perform this test on three AQ-paid triangles, one containing similar parameters, one containing opposite parameters and one containing real data.

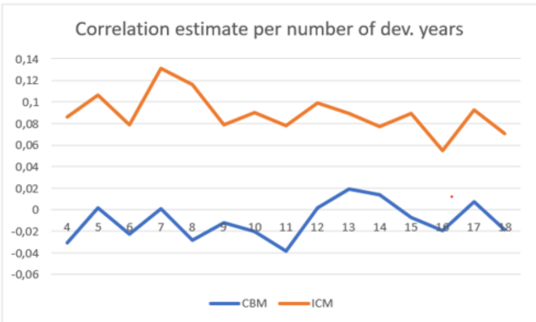


Figure 16: Similar parameters.

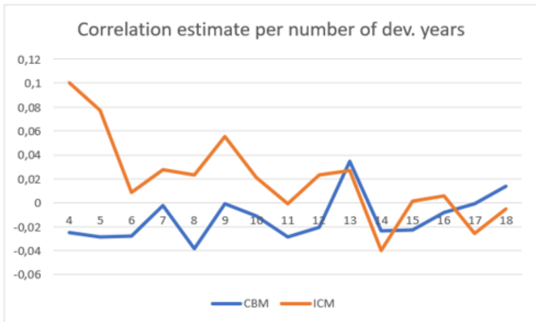


Figure 17: Opposite parameters.

In Figure 16, in which similar parameters are used in the created triangles, the correlation estimate is quite constant over the different number of development periods included. In Figure 17, the situation in which opposite parameters are used, it becomes clear that some time is needed for the implicit correlation method to come to a steady state.

The fact that the first development periods have the most impact on the correlation estimates is not surprising. The most uncertainty is in the development periods with a high MAS parameter and these are mostly located in the first development periods. Looking to the SCR, approximately 70% of the uncertainty is located in the first four development years (i.e. 16 development periods in the AQ-triangle), thereafter a relative high MAS parameter can mostly be dedicated to one event. In case of the two investigated triangles, the portfolio with a long tail has 75% of the uncertainty in the SCR located in the first four development years, the portfolio with a short tail has 65% of the uncertainty located in the first four development years.

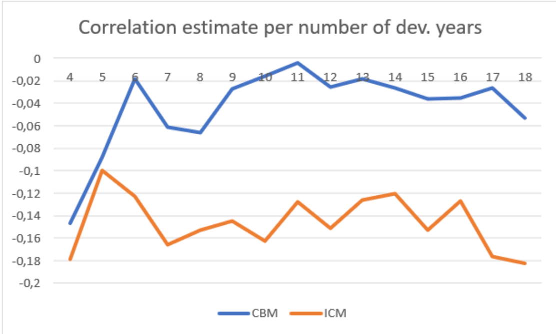


Figure 18: Two real datasets.

By performing the simulation with two real datasets, it becomes apparent that the correlation estimate is quite stable after the first seven years. At this point, approximately 80% of the uncertainty is covered in both datasets. In line with what was already observed in Figure 16 and Figure 17, it is not necessary to include the complete tail to get an accurate correlation estimate. The remarkable thing about this observation is, that it is possible to determine the correlation between triangles with a long and a short tail. As it is expected that both triangles are uncorrelated, it can again be concluded that the correlation bootstrap method comes closer to this expectation than the implicit correlation method.

7.7 Implications implementation in practice

Except for the previous section, all simulations are performed with datasets generated by the dataset generator. In practice, the datasets will hardly ever be perfectly normally distributed, especially further in the tail. This is probably as well a reason for the correlation in Figure 18 to decline after the 16th year. Taking the complete dataset containing 35 years of data and deriving the residuals, the plots in Figure 19 and Figure 20 can be made.

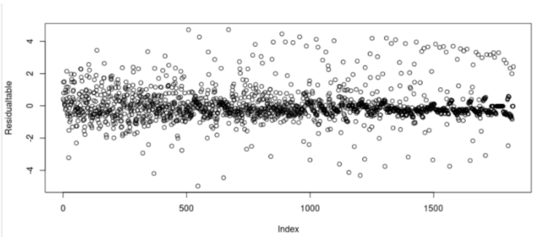


Figure 19: Residuals long tail.

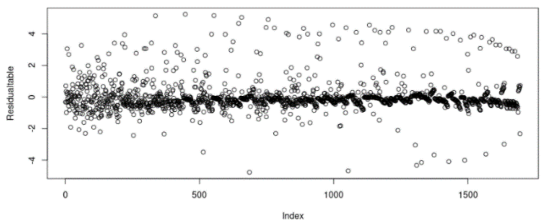


Figure 20: Residuals short tail.

In Figure 19 and 20 are on the y-axis the values of the residuals, on the x-axis the residual number. This number can be interpreted in the following way, the residuals belonging to the first development factor are numbered 1 till 35, the second set of residuals belonging to the second development factor have got the number 36 till 70 etcetera. As can be seen in Figure 19 and Figure 20, the higher the residual number, the lesser the residuals seem normally distributed.

So to apply the correlation bootstrap method and the implicit correlation method, it is important to take the number of development periods as short as possible to have as few as possible non-normally distributed residuals in the dataset but at the same time have enough residuals to determine an accurate correlation estimate.

Furthermore, in practice it will often not be possible to use the residuals of the first development year in the AQ-triangles, as these may contain autocorrelation as a result of seasonal effects. As already discussed earlier, this also improved the correlation estimates for the AQ-triangles as a result of the lowered MAS parameter.

7.8 Summary

The section started with an investigation into the parameters which have the most impact on the correlation estimates when using both methods. For the correlation bootstrap method, the MAS parameter has clearly the most impact on the outcomes. In case of the implicit correlation method, the MAS and the residual set both have a substantial impact on the correlation estimates.

Secondly, we investigated whether scaling the triangles to get comparable SCR estimates does improve the correlation estimates. This improved the correlation estimates in case of the AA-incurred and the AQ-paid triangles but did worsen the correlation outcomes for the AA-paid and AQ-incurred triangle, making it hard to say whether it really improves the outcomes.

Thirdly, we investigated how robust the methods are, i.e., how much the correlation changes if an additional year is added. In case of the AA-triangles, both methods performed relatively well. The correlation estimates over the different runs are close to each other. In the case of the AQ-incurred triangle, the correlation bootstrap method performed substantially better than the implicit correlation method. Two out of four observations had a deviation more than 0.05 from the expected correlation estimate. The difference for the AQ-paid triangle are comparable to the outcomes of the AA-triangles.

Fourthly, we investigated what the impact of outliers is on the correlation estimates. In this test, the correlation bootstrap method performed significantly better compared to the implicit correlation method.

Fifthly, we challenged the assumptions belonging to the used methods. The first assumptions we challenged, is the impact not normally distributed residuals have on the correlation estimate. The outcomes are a bit worse compared to the tests with normally distributed residuals, but overall the patterns are quite comparable. Especially when taking into account that the MAS parameter we used, is of a dataset having normally distributed residuals, making it possible that the outcomes would further improve if an accurate MAS parameter were used. Secondly, we investigated what the impact of different tail lengths is on the correlation estimates. We determined that an accurate correlation estimate can be determined if 80% of the uncertainty is covered in both triangles.

Lastly, we discussed what needs to be taken into account when applying both methods in practice including an analyses of two real datasets.

8 Conclusion and recommendations

This section starts with answering the research questions proposed in Section 1. The section will continue with a general conclusion followed by the scientific contribution of this research. Furthermore, we will discuss the limitations of this research. The section finishes with the recommendations and proposals for future research.

8.1 Conclusions

In Section 1, the research question was formulated which would be answered by 4 sub-questions. In this section, the sub-questions will be answered one by one, finishing with the research question.

Sub-question 1: What is known in the literature about the correlation bootstrap method and the implicit correlation method in determining the correlation parameter between reserving triangles?

In Section 2 and Section 3 the findings from the literature are described. There are plenty of articles in which the correlation bootstrap method is described. However, the method is performed slightly differently in every paper. Furthermore, most papers about the correlation bootstrap method are written between 1990 and 2010, before Solvency II was introduced. The capital requirements proposed by Solvency II are stricter compared to the regulations of Solvency I, making the accuracy of the correlation estimate even more important nowadays.

The implicit correlation method is hardly described in the literature. The method uses some aspects of the rank correlation method proposed by Brehm [2002] and was extended by Kirschner et al. [2002], but is significantly different from the method used.

Underlying both methods investigated is the bootstrapping technique, on which numerous papers are written. This literature is used to investigate what assumptions need to be fulfilled to be able to apply the bootstrapping techniques.

Sub-question 2: How accurate are the methods able to derive the correlation from datasets of which the correlation is known and what is the impact of outliers in the datasets for paid and incurred triangles?

To answer this sub-question, we developed a dataset generator. With the dataset generator, it is possible to create a high variety of triangles based on the parameters: correlation between residuals, the average development factors, the initial claim amounts and the MAS parameter. This makes it possible to construct new triangles based on real data, of which it is known that the residuals follow the intended distribution and make sure that it is possible to construct datasets containing different correlation levels.

With the datasets we generated in the dataset generator, it was possible to measure the impact of the individual parameters by changing one parameter at a time. In the case of the correlation bootstrap method, the parameter with the most impact is the MAS parameter. In the case of the implicit correlation method, the MAS parameter and the residual set have the most impact on the correlation estimates.

When comparing two triangles that contain the same MAS parameter, the correlation estimates are in almost all cases quite accurate, especially for the correlation bootstrap method. In case two different MAS parameters are used, the correlation estimates differ substantially more. In the AA-paid triangle, these deviations are the lowest followed by the AA-incurred triangle, the AQ-paid triangle, and lastly the AQ-incurred triangle which performed substantially worse.

To make sure the MAS parameter is as low as possible, the datasets from which the parameters are derived are corrected for inflation and the trend. Furthermore, the simulation error was not included in the calculation of the SCR. This was excluded as the correlation between the residuals might partly be lost when a gamma distribution is used to skew the outcomes. Lastly, the developments during the first development year are excluded for the AQ-triangles, as these contain a high MAS parameter and potentially autocorrelation in practice.

Table 46: Number of simulations in which the correlation estimate differs more than 0.1 from the preset correlation.

	AA-paid		AQ-paid		AA-Incurred		AQ-Incurred	
	CBM	ICM	CBM	ICM	CBM	ICM	CBM	ICM
Residuals	0	4	0	0	0	7	0	7
MAS parameter	1	5	2	6	3	6	4	7
Settlement periods	0	4	0	0	0	6	0	6
Initial claim amounts	0	6	0	0	0	6	1	8

	Better
	Worse
	Equally good

Another way to look to the results, is by counting the number of instances in which the resulting correlation differed more than 0.1 from the preset correlation. In Table 46, these results can be found. The results for the correlation bootstrap method are substantively better compared to results of the implicit correlation method.

Sub-question 3: How accurate are the correlation estimates if the assumptions of the methods are violated?

To test the importance of the assumptions belonging to the bootstrapping technique, tests are performed with outliers in the dataset, gamma-distributed residuals, different tail lengths, and it is discussed what the impact of autocorrelated residuals and dependent accident years would be on the correlation estimates.

In the outlier test, the correlation bootstrap method performed significantly better compared to the implicit correlation method. Whereas the correlation decreased as a result of the added outliers in the correlation bootstrap method, the correlation increased significantly in the implicit correlation method reaching correlation estimates bigger than 1.

In the test with gamma-distributed residual instead of normally distributed residuals, the outcomes of the correlation bootstrap method are comparable to the outcomes of the simulations with normally distributed residuals. In the case of the implicit correlation method, the results are slightly worse compared to the outcomes of the simulations with the normally distributed residuals. Notable is the way extreme residuals are down-scaled as a result of the changing development vector and increased MAS parameter. This makes sure the variance of the residuals stays close to 1 and the average of the residuals stays close to 0.

Table 47: Results of the tests in which the assumptions are violated.

	AA-paid		AQ-paid		AA-Incurred		AQ-Incurred	
	CBM	ICM	CBM	ICM	CBM	ICM	CBM	ICM
Robustness check								
Outliers								
Normality								

	Better
	Worse
	Equally good

In Table 47, an overview is given in which we see how well the methods performed to each other. As in Table 46, the correlation bootstrap method performed equally good or better than the implicit correlation method.

The impact of autocorrelation and dependent accident years are not simulated as it is logical that these dependencies will vanish when a bootstrapping approach is used. This will lead to an inaccurate SCR and so to an inaccurate correlation estimate. We can conclude that autocorrelation and dependent accident years are not allowed in the triangles when performing the simulations.

Sub-question 4: How well are the methods capable of determining the correlations between reserving triangles with heterogeneous characteristics?

As already discussed in Sub-question 2, the MAS parameter has the most impact on the deviations. However, the observations are often still quite close to the pre-set correlation, especially for the correlation bootstrap method. The MAS parameter is also the parameter that will differ the most between triangles with heterogeneous characteristics. The impact of the differences in the size of the SCR is limited, so this does not form a problem for the obtained correlation estimates.

Specific attention was paid to the comparison of two triangles with different lengths of the development pattern. It could be concluded that only 80% of the uncertainty needs to be covered in both triangles to get an accurate correlation estimate. In case more development periods are needed to fulfil this 80% requirement than there are valid development periods available in the other triangle, a lower percentage can be taken but this correlation estimate is more uncertain.

The biggest concern in comparing triangles with heterogeneous characteristics is the number of normally distributed residuals. Dependent on the settlement pattern, at some time the developments resemble one-time events, increasing the MAS parameter and shifting zero-residuals to non-zero residuals while there is only one deviation. These non-zero residuals are not wanted in the bootstrapping process as these do not represent actual developments. Especially for AA-triangles with a short tail, it might be hard to collect enough normally distributed residuals to get an accurate representation of the potential developments in the upcoming year.

"How accurate are the correlation bootstrap method and the implicit correlation method when estimating the correlation between reserving triangles and how sensitive are the correlation estimates to data quality, outliers, and a violation of the assumptions of these methods?"

Based on the answers to the sub-questions, we can conclude that the correlation bootstrap method performs substantially more accurate compared to the implicit correlation method. In all circumstances, the correlation bootstrap method performed equally good or better than the implicit correlation method. The correlation estimates are the most sensitive to differences in the MAS parameter, to which almost all differences obtained in the tests with data quality, outliers and violations of the methods assumptions can be related to.

8.1.1 Scientific contribution

The scientific contribution of this research is firstly that we performed one of the first studies into the correlation estimate after the Solvency II regulations were installed. Other large insurance companies will also encounter problems when trying to use a quantitative method to obtain the correlation estimate between different HRGs. This research can be seen as the first step towards a more quantitative approach capable of determining the correlation between different HRGs.

Secondly, this study contributes to the scientific body of knowledge as the implicit correlation method and the correlation bootstrap method are both not described and tested before as in this study. Especially, the lessons which we learned about the dynamics of the triangles when outliers and non-normally distributed residuals are added to the triangles, yield valuable insights into the impact these have on the SCR and ultimately on the correlation estimates.

Thirdly, the correlation bootstrap method seems to be able to measure the correlation between triangles with a different length of the development pattern, which was not proven before. This is remarkable, as many papers indicated that it is not possible to use the correlation bootstrap method on triangles with a different number in development periods.

8.1.2 Limitations

Despite the promising outcomes of the study, we encountered a number of limitations that need to be mentioned.

The simulations performed in this research are mainly based on two datasets, one containing a relatively long development pattern and relatively low MAS parameter and one containing a short development pattern having a relatively high MAS parameter. With the parameters of these triangles, many possible scenarios are simulated. However, more real datasets would make it easier to generalize the outcomes.

The simulation time of the simulation model is relatively long, making it necessary to make choices in the simulations which are performed. Especially in the simulation runs containing a violation of the assumptions, more simulations would have strengthened the outcomes.

The most recent literature on the aggregation of reserve capitals is published in Germany in the German language. This forms an obstacle both in language barrier as in permissions required to read the articles.

8.2 Recommendations

This study has shown that it is possible to determine the correlation between reserving triangles using a quantitative method. If Achmea wants to continue with the research into methods to quantify the correlation between different reserving triangles, the following is recommended:

1. Focus on the correlation bootstrap method, as this method performed substantially better in the simulations compared to the the implicit correlation method. Besides, the method mimics more realistically the development of a portfolio containing both triangles.
2. Be aware of the impact outliers may have on the other residuals belonging to the development period the outlier finds itself in. All residuals belonging to the development period containing an outlier will change significantly, eliminating the correlation that might have been available in that development period.
3. Only include the development periods in which real developments occurred. Development periods in which the residuals only represent little changes will decrease the overall correlation in the model.
4. Stay up to date with the developments in the research performed in Germany into the aggregation of risk capitals. However, many papers are written in German, it seems that the most promising works are translated over time before they are published in the more renowned actuarial papers.
5. It would be interesting to apply the method used on a larger number of insurance portfolios to see how much the quantitatively obtained correlation estimates differ from the qualitative correlation estimates obtained by the expert panel.

8.3 Future research

In this research, we investigated how accurately the implicit correlation method and the correlation bootstrap method are capable of estimating the correlation parameter between two reserving triangles. However, the research also generates new questions that could form the basis for future research.

The implicit correlation method has shown to be extremely sensitive to the residual set used in a simulation. However, no unambiguous answer could be given to the question why the correlation estimate differs so much per residual set. Understanding what makes the outcomes differ so much would give more insights in the method and potentially guidelines in which this method can be equally good as the correlation bootstrap method.

We observed that the correlation slightly differs if the volume of the triangles is changed. It would be interesting to know whether it is possible to scale the triangles in such a way that the correlation estimates become even more accurate, without changing the residuals available in the triangle.

An interesting question that could potentially make the correlation estimating process much easier, is to investigate what creates the difference between the pre-set correlation and the output of the correlation bootstrap method. In many situations, the difference between these two is extremely low. In case it is possible to derive the correlation parameter by determining the Pearson correlation of the two vectors of residuals obtained from the dataset which is corrected for inflation, portfolio developments and outliers, it would be possible to skip the complete bootstrapping process.

Bibliography

Akoglu, H. [2018], 'User's guide to correlation coefficients', *Turkish Journal of Emergency Medicine* **18**(3), 91–93.

Antonio, K. & Plat, R. [2014], 'Micro-level stochastic loss reserving for general insurance', *Scandinavian Actuarial Journal* **2014**(7), 649–669.

Artusi, R., Verderio, P. & Marubini, E. [2002], 'Bravais-Pearson and Spearman correlation coefficients: meaning, test of hypothesis and confidence interval', *The International Journal of Biological Markers* **17**(2), 148–151.

Bafin [2006*a*], 'SolvencyI', `https://www.bafin.de/EN/Aufsicht/VersichererPensionsfonds/Aufsichtsregime/SolvencyI/solvency_I_node_en.html`. Accessed: 2021-01-11.

Bafin [2006*b*], 'SolvencyII', `https://www.bafin.de/EN/Aufsicht/VersichererPensionsfonds/Aufsichtsregime/SolvencyII/solvency_II_node_en.html`. Accessed: 2021-01-11.

Bowerman, B. [2016], *Business Statistics in Practice: Using Data, Modeling, and Analytics*, McGraw-Hill Higher Education.

Braun, C. [2004], 'The prediction error of the chain ladder method applied to correlated run-off triangles', *ASTIN Bulletin: The Journal of the IAA* **34**(2), 399–423.

Brehm, P. [2002], Correlation and the aggregation of unpaid loss distributions, *in* 'CAS Forum', Vol. 2, Citeseer, pp. 1–23.

Brickman, S., Barlow, C., Boulter, A., English, A., Furber, L., Ibeson, D., Lowe, J., Pater, R. & Tomlinson, D. [1993], Variance in claim reserving, *in* 'General Insurance Convention, Institute of Actuaries'.

Busse, M. M., Müller, U. & Dacorogna, M. [2010], 'Robust estimation of reserve risk', *ASTIN Bulletin-Actuarial Studies in non LifeInsurance* **40**(2), 453.

CBS [2020], 'Consumentenprijzen; prijsindex', `https://opendata.cbs.nl/statline/?dl=3F0E#/CBS/nl/dataset/83131NED/table`. Accessed: 2020-12-22.

Chandra Shekhar, P., Kumar, N. & Warrier, S. [2008], 'Journey of insurer solvency regulations–2007 and beyond', *Joint Colloquim of the IACA, PBSS and IAAHS Sections. Boston* .

Devineau, L. & Loisel, S. [2009], 'Risk aggregation in Solvency II: How to converge the approaches of the internal models and those of the standard formula?', *Bulletin Français d'Actuariat* **9**(18), 107–145.

EIOPA [2014], 'The underlying assumptions in the standard formula for the solvency capital requirement calculation', *The European Insurance and Occupational Pensions Authority* .

England, P. [2003], 'Bootstrapping Mack's model'. Excel file.

England, P. D., Verrall, R. J. & Wüthrich, M. V. [2019], 'On the lifetime and one-year views of reserve risk, with application to IFRS 17 and Solvency II risk margins', *Insurance: Mathematics and Economics* **85**, 74–88.

England, P. & Verrall, R. [1999], 'Analytic and bootstrap estimates of prediction errors in claims reserving', *Insurance: Mathematics and Economics* **25**(3), 281–293.

England, P. & Verrall, R. [2002], 'Stochastic claims reserving in general insurance', *British Actuarial Journal* pp. 443–544.

Filipović, D. [2009], 'Multi-level risk aggregation', *ASTIN Bulletin, Forthcoming* .

Havlicek, L. L. & Peterson, N. L. [1976], 'Robustness of the Pearson correlation against violations of assumptions', *Perceptual and Motor Skills* **43**(3_suppl), 1319–1334.

Hechenleitner, B. & Entacher, K. [2002], On shortcomings of the ns-2 random number generator, *in* 'Proceedings of the Communication Networks and Distributed Systems Modeling and Simulation', pp. 71–77.

Holland, S. [2019], 'Covariance and Correlation', `http://strata.uga.edu/8370/lecturenotes/correlation.html`. Accessed: 2021-01-24.

Huergo, L., Heberle, J. & Merz, M. [2010], 'Bootstrapping the chain-ladder method for several correlated run-off portfolios', *Zeitschrift für die gesamte Versicherungswissenschaft* **98**(5), 541–564.

Hull, J. C. [2015], *Options, Futures, and Other Derivatives, 9th Edition*, Pearson Education.

Hull, J. C. [2018], *Risk Management and Financial Institutions, 5th Edition*, Hoboken: John Wiley & Sons.

Hulley, S., Cummings, S., Browner, W., Grady, D. & Newman, T. [2013], 'Appendix 6c', *Designing clinical research: An epidemiologic approach* **79**.

Kaas, R., Goovaerts, M., Dhaene, J. & Denuit, M. [2008], *Modern actuarial risk theory: using R*, Vol. 128, Springer Science & Business Media.

Kim, Y., Kim, T.-H. & Ergün, T. [2015], 'The instability of the Pearson correlation coefficient in the presence of coincidental outliers', *Finance Research Letters* **13**, 243–257.

Kirschner, G. S., Kerley, C. & Isaacs, B. [2002], Two approaches to calculating correlated reserve indications across multiple lines of business, *in* 'Casualty Actuarial Society Forum', pp. 211–246.

Liu, H. & Verrall, R. J. [2010], 'Bootstrap estimation of the predictive distributions of reserves using paid and incurred claims', *Variance* **4**, 121–135.

Mack, T. [1993], 'Distribution-free calculation of the standard error of chain ladder reserve estimates', *ASTIN Bulletin: The Journal of the IAA* **23**(2), 213–225.

Norberg, R. [1993], 'Prediction of outstanding liabilities in non-life insurance1', *ASTIN Bulletin: The Journal of the IAA* **23**(1), 95–115.

Peremans, K., Segaert, P., Van Aelst, S. & Verdonck, T. [2017], 'Robust bootstrap procedures for the chain-ladder method', *Scandinavian Actuarial Journal* **2017**(10), 870–897.

Quarg, G. & Mack, T. [2004], 'Munich chain ladder', *Blätter der DGVFM* **26**(4), 597–630.

Robinson, S. [2014], *Simulation: the practice of model development and use*, Palgrave Macmillan.

Shapland, M. R. & Leong, J. [2010], Bootstrap modeling: beyond the basics, *in* 'Casualty Actuarial Society Forum, Fall 2010', pp. 4–52.

Taylor, G. & McGuire, G. [2005], 'Synchronous bootstrapping of seemingly unrelated regressions'.

Thadewald, T. & Büning, H. [2007], 'Jarque–bera test and its competitors for testing normality–a power comparison', *Journal of applied statistics* **34**(1), 87–105.

Verdonck, T. & Debruyne, M. [2011], 'The influence of individual claims on the chain-ladder estimates: Analysis and diagnostic tool', *Insurance: Mathematics and Economics* **48**(1), 85–98.

Verdonck, T., Van Wouwe, M. & Dhaene, J. [2009], 'A robustification of the chain-ladder method', *North American Actuarial Journal* **13**(2), 280–298.

A The Chain Ladder Method

The Chain Ladder method works in the same manner for any kind of triangle, so the approach can be used as well as for AQ-triangles as for AA-triangles and as well for paid triangles as for incurred triangles. Let's take a 5x5 cumulative paid triangle, with on the x-axis the development years and on the y-axis the accident years:

	1	2	3	4	5
2000	100	320	420	490	530
2001	125	310	450	510	
2002	140	380	460		
2003	110	350			
2004	135				

Figure 21: Example AA-triangle.

To estimate the second development year for accident year 2004, the development factor between development year 1 and development year 2 is needed. This development factor can be obtained by the formula:

$$\hat{f}_j = \frac{\sum_{k=1}^{n-j+1} C_{k,j}}{\sum_{k=1}^{n-j+1} C_{k,j-1}} \quad 2 \leq j \leq n \quad j \in \mathbb{N} \tag{38}$$

Filling in the formula for the development factor between development year 1 and development year 2 yields:

$$\hat{f}_2 = \frac{\sum_{k=1}^{5-2+1} C_{k,2}}{\sum_{k=1}^{5-2+1} C_{k,1}} = \frac{320 + 310 + 380 + 350}{100 + 125 + 140 + 110} = \frac{1360}{475} \approx 2.86 \tag{39}$$

The expectation for development year 2 of accident year 2004 becomes: $135 * 2.86 \approx 386$. In the same way, it is possible to fill in the rest of the expected values:

	1	2	3	4	5
2000	100	320	420	490	530
2001	125	310	450	510	551
2002	140	380	460	529	571
2003	110	350	462	531	558
2004	135	386	510	587	634
$\hat{f}_j$		2.86	1.32	1.15	1.08

Figure 22: Calculating the lower triangle.

B The bootstrapping process

In this appendix the steps from Figures 7 and 8 will be explained one by one. In Figure 23, the process from a cumulative triangle to scaled residuals is visualized. In Figure 24, the process from resampled residuals to the outstanding claims is visualized. In the remainder of this appendix, the steps are explained one by one.

Cumulative triangle

	1	2	3	4
1	100	180	290	340
2	140	290	400	
3	125	240		
4	120			
Development factors		1,95	1,47	1,17

Individual development factors

	1	2	3	4
1		1,8	1,61	1,17
2		2,07	1,38	
3		1,92		
4				

Unscaled residuals

	1	2	3	4
1		-1,5	1,92	0
2		1,49	-1,5	
3		-0,3		
4				
Variance		2,21	5,97	

Scaled residuals

	1	2	3	4
1		-1,20	1,11	0
2		1,23	-0,88	
3		-0,23		
4				
Bias correction		1,22	1,41	

Figure 23: From cumulative claims to scaled residuals.

Resampled residuals

	1	2	3	4
1		-0,88	1,23	-1,2
2		1,11	-0,88	
3		1,23		
4				

New development factors

	1	2	3	4
1		1,81	1,69	1
2		2,08	1,34	
3		2,11		
4				

New history

	1	2	3	4
1	100	181	305	290
2	140	292	389	
3	125	264		
4	120			
Dev. factors		2,02	1,476	1

Projected lower triangle

	1	2	3	4	
1				340	
2			389	389	0
3		264	389	389	125
4	120	242	358	358	238
Outstanding claims:					363

Figure 24: From resampled residuals to expected outstanding claims.

In the column furthest to the left is the cumulative triangle with the development factors as explained in Appendix A. On the right side of the cumulative claims triangle is the triangle with the individual development factors. The individual development factors can be derived by dividing the cumulative claims in a development period belonging to an accident year by the cumulative claims in the previous development period of that same accident year. This is visualized in Figure 25.

Cumulative triangle

	1	2	3	4
1	100	180	290	340
2	140	290	400	
3	125	240		
4	120			
Development factors		1,95	1,47	1,17

Individual development factors

	1	2	3	4
1		(1) 1,8	1,61	1,17
2		2,07	1,38	
3		1,92		
4				

$$(1)\ f_{i,j} = \frac{C_{i,j}}{C_{i,j-1}} \rightarrow f_{1,2} = \frac{C_{1,1}}{C_{1,2-1}} = \frac{180}{100} = 1,8$$

Figure 25: From cumulative triangle to individual development factors.

In Figure 26, the way the unscaled residuals can be derived is visualized. For the unscaled residuals, the cumulative claims triangle, the average development factors and the individual development factors are needed.

Cumulative triangle

	1	2	3	4
1	100	180	290	340
2	140	290	400	
3	125	240		
4	120			
Development factors		1,95	1,47	1,17

Individual development factors

	1	2	3	4
1		1,8	1,61	1,17
2		2,07	1,38	
3		1,92		
4				

Unscaled residuals

	1	2	3	4
1		(2) -1,45	1,919	0
2		1,49	-1,512	
3		-0,28		
4				
Variance		(3) 2,21	5,97	(4) 5,97

$$(2)\ U_{i,j} = \sqrt{C_{i,j-1}}\left(f_{i,j} - \hat{f}_j\right) \to U_{1,2} = \sqrt{C_{1,2-1}}\left(f_{1,2} - \hat{f}_2\right) = \sqrt{100}*(1,8-1,95) = -1,45$$

Figure 26: Obtaining the unscaled residuals.

To obtain the MAS parameter, the unscaled residuals are squared and all the squared unscaled residuals are added to each other. This number is divided by the number of datapoints in the development period corrected by one. Furthermore, it is not possible to calculate the MAS parameter for the last development period, as the divisor would be 0. The variance in the last development period is determined by taking the maximum MAS parameter in the previous two development periods.

Unscaled residuals

	1	2	3	4
1		(2) -1,45	1,92	0
2		1,49	-1,51	
3		-0,28		
4				
Variance		(3) 2,21	5,97	(4) 5,97

$$(3)\ k_j \frac{\sum_{i=1}^{n-j+1} U_{i,j}^2}{n-j} \to k_2 = \frac{\sum_{i=1}^{4-2+1} U_{i,2}^2}{4-2} = \frac{-1,45^2+1,49^2-0,28^2}{2} = 2,21$$

$$(4)\ k_n = max(k_{n-1}, k_{n-2}) = max(2,21; 5,97) = 5,97$$

Figure 27: Obtaining the variance.

In figure 28, the unscaled residuals are scaled to scaled residuals. Firstly, the residuals get a bias adjustment, to adjust for the limited number of observations in the development period. Secondly, the unscaled residuals (corrected for bias) are divided by the standard deviation of the corresponding development period.

Unscaled residuals

	1	2	3	4
1		-1,45	1,92	0
2		1,49	-1,51	
3		-0,28		
4				
Variance		2,21	5,97	5,97

Scaled residuals

	1	2	3	4
1		(6) -1,19	1,11	0
2		1,23	-0,88	
3		-0,23		
4				
Bias correction		(5) 1,23	1,41	

$$(5)\ b_j = \sqrt{\frac{n-j+1}{n-j}} \to b_2 = \sqrt{\frac{4-2+1}{4-2}} = 1,23$$

$$(6)\ r_{i,j} = b_j \frac{U_{i,j}}{\sqrt{k_j}} \to r_{1,2} = b_2 \frac{U_{1,2}}{\sqrt{k_2}} = 1,23*\frac{-1,45}{\sqrt{2,21}} = -1,19$$

Figure 28: Scaling the residuals.

The next step in the bootstrapping process is to resample the residuals over the triangle. In Figure 29, this bootstrapping process is visualized using colors. The bootstrapping process is performed with replacement, so it is possible that one residual is picked multiple times. Zero-residuals are excluded in the bootstrapping process.

Scaled residuals

	1	2	3	4
1		-1,20	1,11	0
2		1,23	-0,88	
3		-0,23		
4				
Bias correction		1,23	1,41	

Resampled residuals

	1	2	3	4
1		-0,88	1,23	-1,2
2		1,11	-0,88	
3		1,23		
4				

Figure 29: The resampling process.

In Figure 30, the process to obtain new development factors is depicted. For the new development factors, the original cumulative claims, the original average development factor, the original MAS parameter and the resampled residuals are required.

Cumulative triangle

	1	2	3	4
1	100	180	290	340
2	140	290	400	
3	125	240		
4	120			
Development factors		1,95	1,47	1,1724

Unscaled residuals

	1	2	3	4
1		-1,45	1,92	0,00
2		1,49	-1,51	
3		-0,28		
4				
Variance		2,21	5,97	5,97

Resampled residuals

	1	2	3	4
1		-0,88	1,23	-1,2
2		1,11	-0,88	
3		1,23		
4				

New development factors

	1	2	3	4
1		(7) 1,81	1,69	1,00
2		2,08	1,34	
3		2,11		
4				

$$(7)\ f^*_{i,j} = \frac{r^*_{i,j}\sqrt{k_j}}{\sqrt{C_{i,j-1}}} + \widehat{f}_j \rightarrow f^*_{1,2} = \frac{r^*_{1,2}\sqrt{k_2}}{\sqrt{C_{1,2-1}}} + \widehat{f}_2 = \frac{-0{,}88\sqrt{2{,}21}}{\sqrt{100}} + 1{,}95 = 1{,}81$$

Figure 30: Obtaining the new development factors.

In Figure 31, the process of obtaining a new paid claims history is visualized. The first development period is identical to the original first development period. In the remainder of the triangle, the new cumulative paid claims can be determined by multiplying the original cumulative paid claims in the previous development period with the new development factor corresponding to the development period and specific accident year. Based on the newly created cumulative paid claims and the original paid claims in the previous development period, it is possible to obtain new average development factors.

Cumulative triangle

	1	2	3	4
1	100	180	290	340
2	140	290	400	
3	125	240		
4	120			
Development factors		1,95	1,47	1,17

New development factors

	1	2	3	4
1		1,81	1,69	1
2		2,08	1,34	
3		2,11		
4				

New history

	1	2	3	4
1	(8) 100	(9) 181	305	290
2	140	292	389	
3	125	264		
4	120			
Development factors		(10) 2,02	1,48	1

$$(8)\ C^*_{i,1} = C_{i,1} \rightarrow C^*_{1,1} = C_{1,1} = 100$$

$$(9)\ C^*_{i,j} = C_{i,j-1} * f^*_{i,j} \rightarrow C^*_{1,2} = C_{1,1} * f^*_{1,2} = 100 * 1{,}81 = 181$$

$$(10)\ \widehat{f^*_j} = \frac{\sum_i^{n-j+1} C^*_{i,j}}{\sum_i^{n-j+1} C_{i,j-1}} \rightarrow \widehat{f^*_2} = \frac{\sum_{i=1}^{4-2+1} C^*_{i,2}}{\sum_{i=1}^{4-2+1} C^*_{i,2-1}} = \frac{181+292+264}{100+140+125} = 2{,}02$$

Figure 31: Creating a new history.

In Figure 32, the lower triangle is created based on the newly created average development factors. With the lower triangle it is possible to determine the expected outstanding claims by subtracting the already paid claims from the ultimate expected cumulative claims per accident year. By adding up the expected outstanding claims for every accident year, it is possible to determine the total expected outstanding claims for a HRG.

New history

	1	2	3	4
1	100	181	305	290
2	140	292	389	
3	125	264		
4	120			
Dev. factors		2,02	1,48	1

Projected lower triangle

	1	2	3	4		
1				340		0
2			389	389	(12)	0
3		264	389	389		125
4	120	(11) 242	358	358		238
Outstanding claims:						(13) 363

$$(11)\ C^*_{i,j} = C_{i,j-1} * \widehat{f^*_j} \rightarrow C^*_{4,2} = C_{4,1} * \widehat{f^*_2} = 120 * 2{,}02 = 242$$

$$(12)\ O_i = C^*_{i,n} - C^*_{i,n-i+1} \rightarrow O_2 = C^*_{2,4} - C^*_{2,4-2+1} = 389 - 389 = 0$$

$$(13)\ O = \sum_{i=1}^{n} O_i \rightarrow O = \sum_{i=1}^{4} O_i = 0 + 0 + 125 + 238 = 363$$

Figure 32: Deriving the expected outstanding claims.

C Obtaining the P&L

To obtain the expected P&L over a one year period of a triangle, the expected outstanding claims are determined at time t=0 and time t=1 as well as the expected settlements between t=0 and t=1. By first bootstrapping the triangle 100.000 times, it is possible to determine the expected settlements between t=0 and t=1 by taking the average of all expected settlements in the upcoming year. It is then possible to determine the P&L for every bootstrapped triangle.

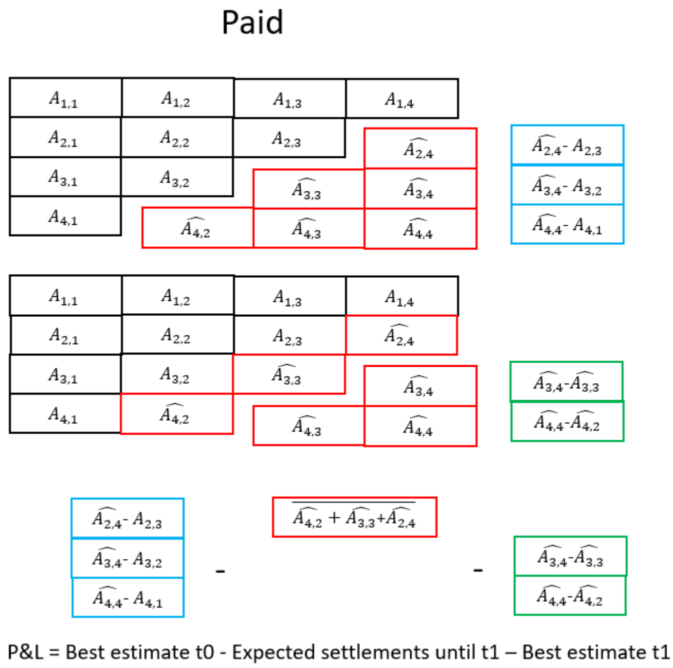


Figure 33: P&L of the cumulative paid claims triangle.

The P&L of the incurred triangle is determined slightly differently. The last development period at t=0 and t=1 denotes the expected ultimate settled claims. The difference between these two expected ultimate settled claims is the adjustment over a one year horizon. If the ultimate expected settled claims at time t=1 is lower than the ultimate expected settled claims at time t=0, there is a profit and otherwise there is a loss. The process is depicted in Figure 34 on the next page.

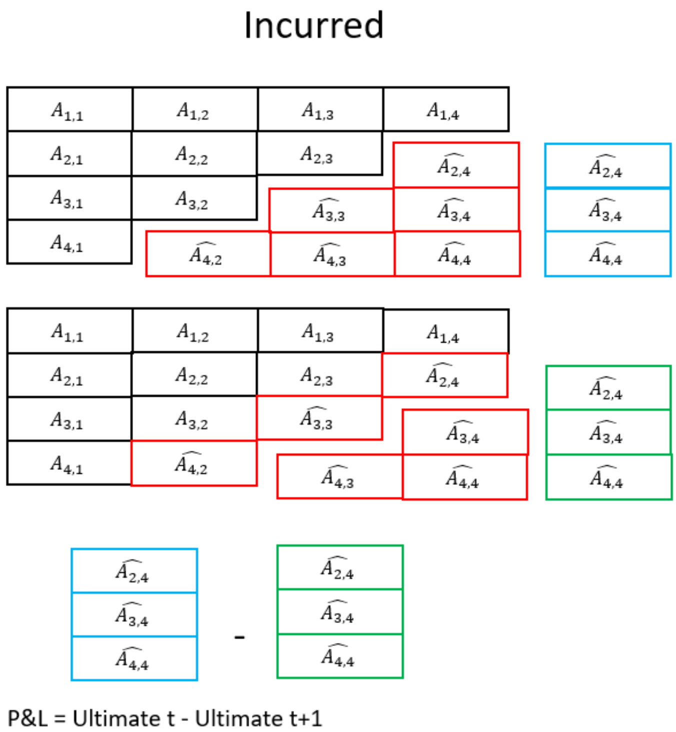


Figure 34: The P&L of the incurred triangle.

D Detecting and dealing with outliers

In this Appendix, the methods of Busse et al. [2010] and Verdonck & Debruyne [2011] are explained.

The algorithm of Busse et al. [2010] only works properly on datasets with a relatively stable development pattern, in which the settlements of the claims follow a stretched pattern. According to the algorithm of Busse et al. [2010], a data point can be denoted as an administrative outlier if it fulfills the following properties:

$$C_{i,j} - C_{i,j-1} \geq \alpha C_{i,N} \tag{40}$$

$$C_{i,j+1} - C_{i,j} \leq -\alpha C_{i,N} \tag{41}$$

Where α is an arbitrary factor between 0.1 and 0.3, depending on the nature of the portfolio. Negative administrative outliers can be detected when the following two conditions hold:

$$C_{i,j} - C_{i,j-1} \leq -\alpha C_{i,N} \tag{42}$$

$$C_{i,j+1} - C_{i,j} \geq \alpha C_{i,N} \tag{43}$$

Outliers caused by other reasons can be detected if there is a strong deviation from the average development factor:

$$f_{i,j} \geq \beta \hat{f}_j \tag{44}$$

Where β is an arbitrary factor between 10 and 100.

The method of Verdonck & Debruyne [2011] proposed to weight the residuals and allocate lower weights to potential outliers. This is done to prevent the outliers from being excluded from the model and makes sure as much as possible data is used in the bootstrapping process. This can be done using the Huberized value:

$$\psi_c\left(r_{i,j}\right) = \begin{cases} (r_{i,j}) & (r_{i,j} \geq c) \\ (c * sign(r_{i,j})) & (r_{i,j} \leq c) \end{cases} \tag{45}$$

And the formula:

$$Weighted\ residuals_{i,j} = \frac{\psi_c\left(r_{i,j}\right)}{r_{i,j}} \tag{46}$$

The tuning constant c typically gets the value 1.345, however in [Verdonck et al., 2009] it is shown that this value is often too low and is it proposed to take c as the 75% percentile of the absolute ranked residuals. This makes sure the c value is adjusted to the specific conditions found in the dataset. Taking a higher percentile will increase efficiency but decrease the breakdown factor.

E Results Section 7.2

In this appendix, the results from Section 7.1 will be extended with additional tables. First, the difference between including and excluding the first development year for the AQ-paid triangle will be depicted. Thereafter, the results of Section 7.1 will be grouped in a different order and the correlation estimates will be available in numbers.

Table 48: AQ-paid including year 1.

		AQ paid - MAS					
		Long tail&Long tail		Short tail&Short tail		Long tail&Short tail	
Correlation	Bivariate	CBM	ICM	CBM	ICM	CBM	ICM
-0,25	-0,21	0,04	0,05	0,03	0,16	0,10	-0,18
0	0,02	0,01	-0,03	0,03	-0,17	0,03	-0,25
0,25	0,25	-0,04	-0,05	-0,05	0,04	-0,11	-0,23
0,5	0,53	-0,01	0,00	0,00	0,02	-0,08	-0,24
0,75	0,72	0,03	0,03	0,07	-0,01	-0,16	-0,43

		AQ paid - Settlement periods					
		Long tail&Long tail		Short tail&Short tail		Long tail&Short tail	
Correlation	Bivariate	CBM	ICM	CBM	ICM	CBM	ICM
-0,25	-0,21	0,04	0,05	0,04	0,04	0,03	0,06
0	0,02	0,01	-0,03	0,02	-0,02	0,03	-0,04
0,25	0,25	-0,04	-0,05	-0,04	-0,04	-0,03	0,03
0,5	0,53	-0,01	0,00	-0,02	0,00	0,01	0,03
0,75	0,72	0,03	0,03	0,04	0,04	0,05	0,05

		AQ paid - Initial claim amounts					
		Long tail&Long tail		Short tail&Short tail		Long tail&Short tail	
Correlation	Bivariate	CBM	ICM	CBM	ICM	CBM	ICM
-0,25	-0,21	0,04	0,05	0,00	0,01	0,02	-0,05
0	0,02	0,01	-0,03	-0,01	-0,04	-0,04	-0,06
0,25	0,25	-0,04	-0,05	0,00	0,00	-0,03	-0,04
0,5	0,53	-0,01	0,00	-0,02	0,03	-0,02	0,04
0,75	0,72	0,03	0,03	0,05	0,05	0,03	0,05

Table 49: AQ-paid excluding year 1.

		AQ paid - MAS					
		Long tail&Long tail		Short tail&Short tail		Long tail&Short tail	
Correlation	Bivariate	CBM	ICM	CBM	ICM	CBM	ICM
-0,25	-0,21	0,03	0,04	0,04	0,15	0,07	0,16
0	0,02	0,00	-0,06	0,00	-0,15	0,00	-0,09
0,25	0,25	-0,01	-0,05	-0,01	-0,04	-0,04	-0,17
0,5	0,53	0,01	0,03	-0,02	-0,01	-0,14	-0,15
0,75	0,79	-0,02	0,00	-0,02	-0,06	-0,18	-0,20

		AQ paid - Settlement periods					
		Long tail&Long tail		Short tail&Short tail		Long tail&Short tail	
Correlation	Bivariate	CBM	ICM	CBM	ICM	CBM	ICM
-0,25	-0,21	0,03	0,04	0,03	0,05	0,03	0,06
0	0,02	0,00	-0,06	0,00	-0,06	-0,01	-0,06
0,25	0,25	-0,01	-0,05	-0,01	-0,05	0,00	-0,04
0,5	0,53	0,01	0,03	0,01	0,03	0,01	0,01
0,75	0,79	-0,02	0,00	-0,02	0,00	-0,02	-0,01

		AQ paid - Initial claim amounts					
		Long tail&Long tail		Short tail&Short tail		Long tail&Short tail	
Correlation	Bivariate	CBM	ICM	CBM	ICM	CBM	ICM
-0,25	-0,21	0,03	0,04	0,03	0,03	0,03	0,03
0	0,02	0,00	-0,06	0,00	-0,06	0,00	-0,05
0,25	0,25	-0,01	-0,05	0,00	-0,05	-0,01	-0,07
0,5	0,53	0,01	0,03	-0,01	0,02	-0,01	0,04
0,75	0,79	-0,02	0,00	-0,01	0,02	-0,03	-0,03

In Table 48, are on the left the outcomes of the simulations with the AQ-paid triangle including the first 3 development periods. In Table 49 on the right are the outcomes when only the values at the end of the first development year are taken as the starting point of the simulation. As can be seen, are the deviations in Table 49 substantially lower compared to the outcomes in Table 48 when looking to the MAS parameter, the other parameters stay close to the same.

As for the AQ-paid triangle, the other tables will also be grouped to each other, to make the differences clearer.

Table 50: AA paid – MAS parameter.

		AA paid - MAS					
		Long tail&Long tail		Short tail&Short tail		Long tail&Short tail	
Correlation	Bivariate	CBM	ICM	CBM	ICM	CBM	ICM
-0,25	-0,30	0,01	-0,01	0,01	-0,06	0,01	-0,02
0	-0,12	0,00	-0,19	0,00	-0,20	-0,01	-0,14
0,25	0,26	-0,03	-0,11	-0,05	0,01	-0,05	-0,04
0,5	0,50	-0,06	-0,03	-0,05	0,02	-0,11	-0,06
0,75	0,72	0,00	-0,05	0,01	-0,08	-0,08	-0,13

Table 51: AA paid – Settlement periods.

		AA paid - Settlement periods					
		Long tail&Long tail		Short tail&Short tail		Long tail&Short tail	
Correlation	Bivariate	CBM	ICM	CBM	ICM	CBM	ICM
-0,25	-0,30	0,01	-0,01	0,01	-0,01	0,01	0,00
0	-0,12	0,00	-0,19	0,01	-0,19	0,00	-0,19
0,25	0,26	-0,03	-0,11	-0,03	-0,08	-0,02	-0,08
0,5	0,50	-0,06	-0,03	-0,06	-0,03	-0,06	-0,03
0,75	0,72	0,00	-0,05	0,00	-0,03	0,00	-0,05

Table 52: AA paid – Initial claim amounts.

		AA paid - Initial claim amounts					
		Long tail&Long tail		Short tail&Short tail		Long tail&Short tail	
Correlation	Bivariate	CBM	ICM	CBM	ICM	CBM	ICM
-0,25	-0,30	0,01	-0,01	0,01	0,00	0,01	-0,03
0	-0,12	0,00	-0,19	0,00	-0,18	0,02	-0,22
0,25	0,26	-0,03	-0,11	-0,03	-0,11	-0,03	-0,08
0,5	0,50	-0,06	-0,03	-0,05	-0,02	-0,06	-0,05
0,75	0,72	0,00	-0,05	0,02	-0,05	0,00	-0,05

Table 53: AQ paid – MAS parameter.

		AQ paid - MAS					
		Long tail&Long tail		Short tail&Short tail		Long tail&Short tail	
Correlation	Bivariate	CBM	ICM	CBM	ICM	CBM	ICM
-0,25	-0,21	0,03	0,04	0,04	0,15	0,07	0,16
0	0,02	0,00	-0,06	0,00	-0,15	0,00	-0,09
0,25	0,25	-0,01	-0,05	-0,01	-0,04	-0,04	-0,17
0,5	0,53	0,01	0,03	-0,02	-0,01	-0,14	-0,15
0,75	0,79	-0,02	0,00	-0,02	-0,06	-0,18	-0,20

Table 54: AQ paid – Settlement periods.

		AQ paid - Settlement periods					
		Long tail&Long tail		Short tail&Short tail		Long tail&Short tail	
Correlation	Bivariate	CBM	ICM	CBM	ICM	CBM	ICM
-0,25	-0,21	0,03	0,04	0,03	0,05	0,03	0,06
0	0,02	0,00	-0,06	0,00	-0,06	-0,01	-0,06
0,25	0,25	-0,01	-0,05	-0,01	-0,05	0,00	-0,04
0,5	0,53	0,01	0,03	0,01	0,03	0,01	0,01
0,75	0,79	-0,02	0,00	-0,02	0,00	-0,02	-0,01

Table 55: AQ paid – Initial claim amounts.

		AQ paid - Initial claim amounts					
		Long tail&Long tail		Short tail&Short tail		Long tail&Short tail	
Correlation	Bivariate	CBM	ICM	CBM	ICM	CBM	ICM
-0,25	-0,21	0,03	0,04	0,03	0,03	0,03	0,03
0	0,02	0,00	-0,06	0,00	-0,06	0,00	-0,05
0,25	0,25	-0,01	-0,05	0,00	-0,05	-0,01	-0,07
0,5	0,53	0,01	0,03	-0,01	0,02	-0,01	0,04
0,75	0,79	-0,02	0,00	-0,01	0,02	-0,03	-0,03

Table 56: AA incurred – MAS parameter.

		AA Incurred - MAS					
		Long tail&Long tail		Short tail&Short tail		Long tail&Short tail	
Correlation	Bivariate	CBM	ICM	CBM	ICM	CBM	ICM
-0,25	-0,30	-0,01	0,28	0,02	0,04	0,06	0,18
0	-0,12	0,00	-0,21	0,01	-0,02	0,02	-0,08
0,25	0,26	-0,04	0,00	-0,03	0,06	-0,12	-0,10
0,5	0,50	-0,07	-0,06	-0,08	0,00	-0,20	-0,23
0,75	0,72	0,02	0,07	0,01	0,03	-0,15	-0,15

Table 57: AA incurred – Settlement periods.

		AA Incurred - Settlement periods					
		Long tail&Long tail		Short tail&Short tail		Long tail&Short tail	
Correlation	Bivariate	CBM	ICM	CBM	ICM	CBM	ICM
-0,25	-0,30	-0,01	0,28	-0,01	0,29	-0,01	0,28
0	-0,12	0,00	-0,21	0,01	-0,21	0,01	-0,21
0,25	0,26	-0,04	0,00	-0,04	0,00	-0,04	0,00
0,5	0,50	-0,07	-0,06	-0,07	-0,06	-0,07	-0,07
0,75	0,72	0,02	0,07	0,02	0,07	0,01	0,08

Table 58: AA incurred – Initial claim amounts.

		AA Incurred - Initial claim amounts					
		Long tail&Long tail		Short tail&Short tail		Long tail&Short tail	
Correlation	Bivariate	CBM	ICM	CBM	ICM	CBM	ICM
-0,25	-0,30	-0,01	0,28	0,00	0,29	0,01	0,28
0	-0,12	0,00	-0,21	0,00	-0,22	0,00	-0,24
0,25	0,26	-0,04	0,00	-0,03	0,02	-0,04	0,01
0,5	0,50	-0,07	-0,06	-0,09	-0,07	-0,09	-0,06
0,75	0,72	0,02	0,07	0,03	0,09	0,01	0,05

Table 59: AQ incurred – MAS parameter.

		AQ Incurred - MAS					
		Long tail&Long tail		Short tail&Short tail		Long tail&Short tail	
Correlation	Bivariate	CBM	ICM	CBM	ICM	CBM	ICM
-0,25	-0,21	0,03	-0,07	0,04	0,03	0,14	0,15
0	0,02	0,01	-0,02	0,01	-0,02	-0,01	0,02
0,25	0,25	-0,01	-0,27	-0,02	-0,31	-0,18	-0,26
0,5	0,53	0,00	0,21	0,01	0,04	-0,36	-0,33
0,75	0,79	-0,01	-0,03	-0,01	-0,04	-0,53	-0,54

Table 60: AQ incurred – Settlement periods.

		AQ Incurred - Settlement periods					
		Long tail&Long tail		Short tail&Short tail		Long tail&Short tail	
Correlation	Bivariate	CBM	ICM	CBM	ICM	CBM	ICM
-0,25	-0,21	0,03	-0,07	0,03	-0,07	0,03	-0,07
0	0,02	0,01	-0,02	0,01	-0,02	0,01	-0,02
0,25	0,25	-0,01	-0,27	-0,01	-0,27	-0,01	-0,26
0,5	0,53	0,00	0,21	0,00	0,22	0,00	0,21
0,75	0,79	-0,01	-0,03	-0,01	-0,03	0,00	-0,03

Table 61: AQ incurred – Initial claim amounts.

		AQ Incurred - Initial claim amounts					
		Long tail&Long tail		Short tail&Short tail		Long tail&Short tail	
Correlation	Bivariate	CBM	ICM	CBM	ICM	CBM	ICM
-0,25	-0,21	0,03	-0,07	0,04	-0,05	0,05	-0,26
0	0,02	0,01	-0,02	0,02	-0,01	-0,02	-0,18
0,25	0,25	-0,01	-0,27	-0,01	-0,26	-0,04	0,03
0,5	0,53	0,00	0,21	0,02	0,22	-0,05	0,15
0,75	0,79	-0,01	-0,03	-0,01	-0,03	-0,11	-0,23

In the following tables, on the left side are the tables in which the correlation estimates are shown instead of the deviations. On the right side are the tables with the %-difference in the SCR. These are shown side by side, as this makes it easier to interpret the results.

Table 62: AA-Paid correlation estimates.

		AA paid - MAS					
		Long tail&Long tail		Short tail&Short tail		Long tail&Short tail	
Correlation	Bivariate	CBM	ICM	CBM	ICM	CBM	ICM
-0,25	-0,30	-0,29	-0,31	-0,28	-0,36	-0,28	-0,32
0	-0,12	-0,11	-0,30	-0,11	-0,32	-0,13	-0,26
0,25	0,26	0,23	0,15	0,21	0,27	0,21	0,22
0,5	0,50	0,44	0,47	0,46	0,53	0,39	0,45
0,75	0,72	0,72	0,67	0,73	0,64	0,64	0,59

		AA paid - Settlement periods					
		Long tail&Long tail		Short tail&Short tail		Long tail&Short tail	
Correlation	Bivariate	CBM	ICM	CBM	ICM	CBM	ICM
-0,25	-0,30	-0,29	-0,31	-0,29	-0,30	-0,28	-0,30
0	-0,12	-0,11	-0,30	-0,11	-0,31	-0,12	-0,31
0,25	0,26	0,23	0,15	0,23	0,18	0,24	0,18
0,5	0,50	0,44	0,47	0,45	0,48	0,44	0,48
0,75	0,72	0,72	0,67	0,72	0,70	0,72	0,68

		AA paid - Initial claim amounts					
		Long tail&Long tail		Short tail&Short tail		Long tail&Short tail	
Correlation	Bivariate	CBM	ICM	CBM	ICM	CBM	ICM
-0,25	-0,30	-0,29	-0,31	-0,28	-0,29	-0,28	-0,32
0	-0,12	-0,11	-0,30	-0,11	-0,30	-0,10	-0,33
0,25	0,26	0,23	0,15	0,23	0,15	0,23	0,18
0,5	0,50	0,44	0,47	0,45	0,48	0,44	0,46
0,75	0,72	0,72	0,67	0,74	0,68	0,72	0,67

Table 63: AA-Paid %-difference in SCR.

	AA paid - MAS					
	Long tail&Long tail		Short tail&Short tail		Long tail&Short tail	
Correlation	CBM	ICM	CBM	ICM	CBM	ICM
-0,25	0,6%	-0,8%	0,8%	-4,4%	0,5%	-0,9%
0	0,2%	-11,1%	0,2%	-12,0%	-0,4%	-4,4%
0,25	-1,2%	-4,3%	-1,8%	0,4%	-1,2%	-0,9%
0,5	-2,0%	-1,1%	-1,6%	0,7%	-2,7%	-1,4%
0,75	0,0%	-1,5%	0,2%	-2,3%	-1,7%	-2,7%

	AA paid - Settlement periods					
	Long tail&Long tail		Short tail&Short tail		Long tail&Short tail	
Correlation	CBM	ICM	CBM	ICM	CBM	ICM
-0,25	0,6%	-0,8%	0,7%	-0,4%	1,0%	-0,2%
0	0,2%	-11,1%	0,4%	-11,5%	0,0%	-11,6%
0,25	-1,2%	-4,3%	-1,1%	-3,0%	-0,7%	-3,1%
0,5	-2,0%	-1,1%	-1,9%	-0,9%	-2,0%	-0,9%
0,75	0,0%	-1,5%	-0,1%	-0,8%	0,0%	-1,4%

	AA paid - Initial claim amounts					
	Long tail&Long tail		Short tail&Short tail		Long tail&Short tail	
Correlation	CBM	ICM	CBM	ICM	CBM	ICM
-0,25	0,6%	-0,8%	1,0%	0,3%	0,8%	-1,8%
0	0,2%	-11,1%	0,3%	-11,0%	1,0%	-12,9%
0,25	-1,2%	-4,3%	-1,1%	-4,2%	-1,1%	-3,3%
0,5	-2,0%	-1,1%	-1,8%	-0,7%	-2,1%	-1,6%
0,75	0,0%	-1,5%	0,5%	-1,4%	0,0%	-1,5%

Table 64: AA-Incurred correlation estimates.

		AA Incurred - MAS					
		Long tail&Long tail		Short tail&Short tail		Long tail&Short tail	
Correlation	Bivariate	CBM	ICM	CBM	ICM	CBM	ICM
-0,25	-0,30	-0,30	-0,01	-0,28	-0,25	-0,23	-0,12
0	-0,12	-0,11	-0,32	-0,10	-0,13	-0,09	-0,20
0,25	0,26	0,22	0,26	0,23	0,32	0,13	0,16
0,5	0,50	0,43	0,44	0,43	0,50	0,30	0,28
0,75	0,72	0,74	0,80	0,73	0,76	0,57	0,57

		AA Incurred - Settlement periods					
		Long tail&Long tail		Short tail&Short tail		Long tail&Short tail	
Correlation	Bivariate	CBM	ICM	CBM	ICM	CBM	ICM
-0,25	-0,30	-0,30	-0,01	-0,30	-0,01	-0,30	-0,02
0	-0,12	-0,11	-0,32	-0,11	-0,32	-0,11	-0,32
0,25	0,26	0,22	0,26	0,22	0,26	0,22	0,26
0,5	0,50	0,43	0,44	0,43	0,44	0,43	0,44
0,75	0,72	0,74	0,80	0,74	0,80	0,74	0,80

		AA Incurred - Initial claim amounts					
		Long tail&Long tail		Short tail&Short tail		Long tail&Short tail	
Correlation	Bivariate	CBM	ICM	CBM	ICM	CBM	ICM
-0,25	-0,30	-0,30	-0,01	-0,30	-0,01	-0,29	-0,02
0	-0,12	-0,11	-0,32	-0,12	-0,33	-0,12	-0,35
0,25	0,26	0,22	0,26	0,23	0,27	0,22	0,27
0,5	0,50	0,43	0,44	0,42	0,43	0,42	0,44
0,75	0,72	0,74	0,80	0,75	0,81	0,74	0,77

Table 65: AA-Incurred %-difference in SCR.

	AA incurred - MAS					
	Long tail&Long tail		Short tail&Short tail		Long tail&Short tail	
Correlation	CBM	ICM	CBM	ICM	CBM	ICM
-0,25	-0,4%	18,1%	1,4%	2,9%	2,8%	7,8%
0	0,3%	-12,5%	0,7%	-1,0%	0,9%	-3,1%
0,25	-1,5%	0,1%	-1,2%	2,5%	-4,0%	-3,2%
0,5	-2,4%	-2,1%	-2,6%	-0,1%	-4,8%	-5,5%
0,75	0,5%	2,1%	0,2%	0,9%	-3,8%	-3,8%

	AA incurred - Settlement periods					
	Long tail&Long tail		Short tail&Short tail		Long tail&Short tail	
Correlation	CBM	ICM	CBM	ICM	CBM	ICM
-0,25	-0,4%	18,1%	-0,5%	18,3%	-0,7%	18,0%
0	0,3%	-12,5%	0,3%	-12,4%	0,4%	-12,5%
0,25	-1,5%	0,1%	-1,6%	0,0%	-1,5%	-0,1%
0,5	-2,4%	-2,1%	-2,4%	-2,1%	-2,4%	-2,3%
0,75	0,5%	2,1%	0,5%	2,1%	0,4%	2,2%

	AA incurred - Initial claim amounts					
	Long tail&Long tail		Short tail&Short tail		Long tail&Short tail	
Correlation	CBM	ICM	CBM	ICM	CBM	ICM
-0,25	-0,4%	18,1%	-0,2%	18,1%	0,4%	16,6%
0	0,3%	-12,5%	-0,3%	-13,2%	-0,2%	-14,1%
0,25	-1,5%	0,1%	-1,0%	0,6%	-1,4%	0,3%
0,5	-2,4%	-2,1%	-3,0%	-2,4%	-2,9%	-2,0%
0,75	0,5%	2,1%	0,8%	2,4%	0,4%	1,4%

Table 66: AQ-Paid correlation estimates.

		AQ paid - MAS					
		Long tail&Long tail		Short tail&Short tail		Long tail&Short tail	
Correlation	**Bivariate**	CBM	ICM	CBM	ICM	CBM	ICM
-0,25	-0,21	-0,19	-0,16	-0,19	-0,06	-0,13	-0,43
0	0,02	0,03	-0,01	-0,01	-0,14	0,01	-0,24
0,25	0,25	0,21	0,20	0,21	0,31	0,14	0,03
0,5	0,53	0,49	0,52	0,51	0,52	0,40	0,25
0,75	0,79	0,77	0,77	0,78	0,71	0,56	0,38

		AQ paid - Settlement periods					
		Long tail&Long tail		Short tail&Short tail		Long tail&Short tail	
Correlation	**Bivariate**	CBM	ICM	CBM	ICM	CBM	ICM
-0,25	-0,21	-0,19	-0,16	-0,19	-0,16	-0,19	-0,12
0	0,02	0,03	-0,01	0,04	-0,01	0,03	-0,01
0,25	0,25	0,21	0,20	0,21	0,21	0,22	0,22
0,5	0,53	0,49	0,52	0,49	0,52	0,51	0,54
0,75	0,79	0,77	0,77	0,78	0,78	0,77	0,80

		AQ paid - Initial claim amounts					
		Long tail&Long tail		Short tail&Short tail		Long tail&Short tail	
Correlation	**Bivariate**	CBM	ICM	CBM	ICM	CBM	ICM
-0,25	-0,21	-0,19	-0,16	-0,19	-0,17	-0,18	-0,19
0	0,02	0,03	-0,01	0,04	0,02	0,03	-0,02
0,25	0,25	0,21	0,20	0,21	0,21	0,20	0,15
0,5	0,53	0,49	0,52	0,50	0,53	0,47	0,50
0,75	0,79	0,77	0,77	0,77	0,78	0,75	0,79

Table 67: AQ-Paid %-difference in SCR.

	AQ paid - MAS					
	Long tail&Long tail		Short tail&Short tail		Long tail&Short tail	
Correlation	CBM	ICM	CBM	ICM	CBM	ICM
-0,25	2,0%	2,7%	2,6%	9,2%	2,5%	5,4%
0	-0,2%	-2,8%	0,2%	-7,8%	-0,1%	-3,0%
0,25	-0,2%	-2,1%	-0,2%	-1,6%	-1,1%	-4,3%
0,5	0,3%	1,1%	-0,5%	-0,4%	-3,2%	-3,4%
0,75	-0,5%	-0,1%	-0,6%	-1,7%	-3,8%	-4,3%

	AQ paid - Settlement periods					
	Long tail&Long tail		Short tail&Short tail		Long tail&Short tail	
Correlation	CBM	ICM	CBM	ICM	CBM	ICM
-0,25	2,0%	2,7%	2,1%	2,9%	1,9%	3,6%
0	-0,2%	-2,8%	-0,2%	-2,8%	-0,3%	-3,2%
0,25	-0,2%	-2,1%	-0,4%	-2,1%	-0,2%	-1,6%
0,5	0,3%	1,1%	0,3%	1,0%	0,2%	0,3%
0,75	-0,5%	-0,1%	-0,5%	-0,1%	-0,5%	-0,1%

	AQ paid - Initial claim amounts					
	Long tail&Long tail		Short tail&Short tail		Long tail&Short tail	
Correlation	CBM	ICM	CBM	ICM	CBM	ICM
-0,25	2,0%	2,7%	1,9%	2,0%	1,6%	1,9%
0	-0,2%	-2,8%	-0,1%	-3,0%	-0,1%	-2,6%
0,25	-0,2%	-2,1%	0,0%	-1,9%	-0,3%	-2,6%
0,5	0,3%	1,1%	-0,2%	0,7%	-0,2%	1,2%
0,75	-0,5%	-0,1%	-0,3%	0,4%	-1,0%	-0,7%

Table 68: AQ-incurred correlation estimates.

		AQ Incurred - MAS					
		Long tail&Long tail		Short tail&Short tail		Long tail&Short tail	
Correlation	**Bivariate**	CBM	ICM	CBM	ICM	CBM	ICM
-0,25	-0,21	-0,18	-0,28	-0,17	-0,18	-0,07	-0,07
0	0,02	0,03	0,00	0,03	0,00	0,01	0,04
0,25	0,25	0,24	-0,02	0,24	-0,05	0,07	-0,01
0,5	0,53	0,53	0,74	0,54	0,56	0,17	0,20
0,75	0,79	0,78	0,76	0,78	0,75	0,26	0,25

		AQ Incurred - Settlement periods					
		Long tail&Long tail		Short tail&Short tail		Long tail&Short tail	
Correlation	**Bivariate**	CBM	ICM	CBM	ICM	CBM	ICM
-0,25	-0,21	-0,18	-0,28	-0,18	-0,28	-0,18	-0,29
0	0,02	0,03	0,00	0,03	0,00	0,03	0,00
0,25	0,25	0,24	-0,02	0,24	-0,01	0,24	-0,01
0,5	0,53	0,53	0,74	0,53	0,74	0,53	0,74
0,75	0,79	0,78	0,76	0,78	0,76	0,79	0,76

		AQ Incurred - Initial claim amounts					
		Long tail&Long tail		Short tail&Short tail		Long tail&Short tail	
Correlation	**Bivariate**	CBM	ICM	CBM	ICM	CBM	ICM
-0,25	-0,21	-0,18	-0,28	-0,17	-0,26	-0,16	-0,47
0	0,02	0,03	0,00	0,04	0,01	0,00	-0,16
0,25	0,25	0,24	-0,02	0,24	-0,01	0,21	0,28
0,5	0,53	0,53	0,74	0,55	0,74	0,48	0,68
0,75	0,79	0,78	0,76	0,78	0,76	0,68	0,56

Table 69: AQ-incurred %-difference in SCR.

	AQ incurred - MAS					
	Long tail&Long tail		Short tail&Short tail		Long tail&Short tail	
Correlation	CBM	ICM	CBM	ICM	CBM	ICM
-0,25	1,9%	-4,3%	2,8%	2,0%	7,5%	7,8%
0	0,6%	-0,8%	0,4%	-1,2%	-0,6%	0,9%
0,25	-0,5%	-11,3%	-0,7%	-13,1%	-6,4%	-9,4%
0,5	-0,1%	6,8%	0,3%	1,1%	-11,6%	-10,4%
0,75	-0,2%	-0,7%	-0,2%	-1,1%	-14,2%	-14,5%

	AQ incurred - Settlement periods					
	Long tail&Long tail		Short tail&Short tail		Long tail&Short tail	
Correlation	CBM	ICM	CBM	ICM	CBM	ICM
-0,25	1,9%	-4,3%	1,8%	-4,2%	1,9%	-4,6%
0	0,6%	-0,8%	0,6%	-0,9%	0,5%	-0,8%
0,25	-0,5%	-11,3%	-0,4%	-11,2%	-0,5%	-11,0%
0,5	-0,1%	6,8%	-0,1%	6,8%	0,1%	6,7%
0,75	-0,2%	-0,7%	-0,2%	-0,8%	-0,1%	-0,8%

	AQ incurred - Initial claim amounts					
	Long tail&Long tail		Short tail&Short tail		Long tail&Short tail	
Correlation	CBM	ICM	CBM	ICM	CBM	ICM
-0,25	1,9%	-4,3%	2,3%	-2,9%	3,8%	-15,4%
0	0,6%	-0,8%	0,7%	-0,6%	-1,0%	-8,2%
0,25	-0,5%	-11,3%	-0,4%	-10,9%	-1,4%	1,3%
0,5	-0,1%	6,8%	0,6%	6,8%	-1,1%	5,2%
0,75	-0,2%	-0,7%	-0,1%	-0,9%	-1,2%	-4,6%

F Outcomes combining parameters

The correlation estimates of the other possible situations which can be made with the parameters can be found in this appendix. The outcomes show a comparable pattern as shown in Section 7.1 in which the individual parameters are discussed.

Table 70: Combining parameters AA-paid.

		AA-Paid									
MAS		Short tail&Short tail		Short tail&Short tail		Short tail&Short tail		Short tail&Short tail		Long tail&Short tail	
Development pattern		Short tail&Short tail		Long tail&Short tail		Short tail&Short tail		Short tail&Short tail		Long tail&Short tail	
First dev. Period		Long tail&Long tail		Long tail&Long tail		Short tail&Short tail		Long tail&Short tail		Long tail&Short tail	
Correlation	**Bivariate**	CBM	ICM	CBM	ICM	CBM	ICM	CBM	ICM	CBM	ICM
-0,25	-0,30	0,01	-0,06	0,01	-0,04	0,01	-0,05	0,01	-0,10	0,05	0,07
0	-0,12	0,00	-0,20	0,01	-0,20	0,00	-0,15	0,00	-0,21	0,03	-0,01
0,25	0,26	-0,04	0,04	-0,04	0,02	-0,03	0,06	-0,04	0,08	-0,07	0,09
0,5	0,50	-0,06	0,02	-0,05	0,02	-0,06	0,01	-0,06	0,01	-0,11	0,13
0,75	0,72	0,00	-0,07	0,00	-0,08	0,01	-0,05	0,00	-0,05	-0,15	-0,10

Table 71: Combining parameters AA-incurred.

		AA-Incurred									
MAS		Short tail&Short tail		Short tail&Short tail		Short tail&Short tail		Short tail&Short tail		Long tail&Short tail	
Development pattern		Short tail&Short tail		Long tail&Short tail		Short tail&Short tail		Long tail&Long tail		Long tail&Short tail	
First dev. Period		Long tail&Long tail		Long tail&Long tail		Short tail&Short tail		Long tail&Short tail		Long tail&Short tail	
Correlation	**Bivariate**	CBM	ICM	CBM	ICM	CBM	ICM	CBM	ICM	CBM	ICM
-0,25	-0,30	0,02	0,04	0,02	0,04	0,02	0,05	0,03	0,02	0,07	0,20
0	-0,12	0,01	-0,02	0,01	-0,01	0,01	0,02	0,00	-0,01	0,03	-0,09
0,25	0,26	-0,03	0,07	-0,02	0,06	-0,03	0,05	-0,04	0,05	-0,13	-0,08
0,5	0,50	-0,07	0,00	-0,08	0,00	-0,07	0,00	-0,06	0,01	-0,20	-0,17
0,75	0,72	0,01	0,03	0,01	0,04	0,01	0,05	0,01	0,04	-0,15	-0,11

Table 72: Combining parameters AQ-paid.

		AQ-paid									
MAS		Short tail&Short tail		Short tail&Short tail		Short tail&Short tail		Short tail&Short tail		Long tail&Short tail	
Development pattern		Short tail&Short tail		Long tail&Short tail		Short tail&Short tail		Short tail&Short tail		Long tail&Short tail	
First dev. Period		Long tail&Long tail		Long tail&Long tail		Short tail&Short tail		Long tail&Short tail		Long tail&Short tail	
Correlation	**Bivariate**	CBM	ICM	CBM	ICM	CBM	ICM	CBM	ICM	CBM	ICM
-0,25	-0,21	0,01	0,14	0,01	0,28	0,02	0,11	0,01	0,09	0,08	0,36
0	0,02	-0,02	-0,18	-0,03	-0,18	-0,02	-0,12	-0,02	-0,18	-0,02	0,04
0,25	0,25	-0,03	0,02	-0,04	0,06	-0,03	0,07	-0,04	-0,03	-0,08	0,14
0,5	0,53	-0,02	0,01	-0,03	-0,02	-0,01	-0,02	-0,02	-0,03	-0,14	0,18
0,75	0,79	0,01	-0,06	0,00	-0,04	-0,01	-0,07	-0,04	-0,12	-0,21	-0,01

Table 73: Combining parameters AQ-incurred.

		AQ-Incurred									
MAS		Short tail&Short tail		Short tail&Short tail		Short tail&Short tail		Short tail&Short tail		Long tail&Short tail	
Development pattern		Short tail&Short tail		Long tail&Short tail		Short tail&Short tail		Long tail&Long tail		Long tail&Short tail	
First dev. Period		Long tail&Long tail		Long tail&Long tail		Short tail&Short tail		Long tail&Short tail		Long tail&Short tail	
Correlation	**Bivariate**	CBM	ICM	CBM	ICM	CBM	ICM	CBM	ICM	CBM	ICM
-0,25	-0,21	0,04	0,03	0,04	0,03	0,04	0,04	0,03	0,02	0,13	0,17
0	0,02	0,01	-0,02	-0,01	-0,03	-0,01	-0,03	0,00	-0,03	-0,03	0,03
0,25	0,25	-0,02	-0,30	-0,02	-0,28	-0,02	-0,28	-0,04	-0,31	-0,18	-0,27
0,5	0,53	0,00	0,02	-0,01	0,03	-0,01	0,04	0,00	0,03	-0,35	-0,31
0,75	0,79	-0,01	-0,04	-0,01	-0,03	-0,01	-0,03	-0,01	-0,04	-0,53	-0,53

G From outliers to residuals

To understand the outcomes of Section 7.5 and Section 7.6.1 it is good to understand the impact one (or several) outliers in a development period have on the rest of the residuals and ultimately on the correlation estimate. The example is made up using the datasets found in the paper of Braun [2004] and the dataset generator described in Section 5. On the left side, an example of a triangle without outliers will be visible, on the right side an example with an outlier will be showed. This makes it easy to follow the difference in the development of the residuals.

	1	2	3	4
1		-0,924	0,945	1,035
2		0,278	0,336	-0,543
3		0,128	-0,791	1,314
4		-0,328	-0,034	-0,607
5		-0,072	0,266	0,250
6		0,173	-1,036	-0,229
7		1,963	-0,728	1,140
8		-1,181	-1,248	0,008
9		-0,669	-1,606	1,142
10		-0,669	0,493	0,650
11		0,427	-1,148	-0,836
12		0,317	-1,030	
13		0,361		
14				

Figure 35: Resdiuals normal distribution.

	1	2	3	4
1		-1,315	-0,007	-0,411
2		1,501	0,343	-1,294
3		-0,465	-0,314	1,049
4		0,041	0,101	-0,997
5		0,034	0,041	-0,131
6		0,779	0,170	1,454
7		0,982	-0,315	1,993
8		-0,386	-1,585	-0,188
9		0,308	-0,440	-0,470
10		20,000	-0,900	0,327
11		-0,118	-1,482	-0,220
12		0,126	-1,471	
13		1,587		
14				

Figure 36: Residuals including an outlier.

In red, the outlier is visualized in Figure 35. This outlier is quite extreme, to show how the residual develops itself during the process.

With the residuals, preset development pattern, preset MAS parameter and preset initial claim amounts it is possible to construct the triangle. By keeping these parameters the same for both triangles, the effect can best be seen.

	1	2	3	4
1	114423	221746	292458	337696
2	152296	350427	449743	497345
3	144325	326371	403173	466660
4	145904	311573	395064	435774
5	170333	375986	481321	542361
6	189643	430053	529198	589899
7	179022	485990	604439	693089
8	205908	401857	490647	549800
9	210951	437175	528814	607360
10	213426	442484	569791	647171
11	249508	577860	712055	785103
12	258425	592195	732236	
13	368762	843915		
14	394997			

Figure 37: Triangle without outlier.

	1	2	3	4
1	114423	207794	263708	291658
2	152296	400728	514060	559289
3	144325	302644	379931	437262
4	145904	326400	415768	454274
5	170333	380612	483791	540378
6	189643	457858	584063	674295
7	179022	442272	556272	649892
8	205908	439875	532504	594138
9	210951	484429	607421	673956
10	213426	1448694	1812344	2038372
11	249508	549161	670065	747487
12	258425	581984	711141	
13	368762	922377		
14	394997			

Figure 38: Triangle with outlier.

In Figure 38, it can be seen that the red block contains a much higher value compared to the value on the same place in Figure 37. This is the result of the outlier. The interesting thing happens when the residuals are again derived from the triangle.

	1	2	3	4
1		-1,164	1,694	1,022
2		0,342	1,043	-1,007
3		0,154	-0,415	1,323
4		-0,417	0,538	-1,066
5		-0,098	0,977	-0,031
6		0,209	-0,638	-0,645
7		2,452	-0,203	1,037
8		-1,487	-0,932	-0,336
9		-0,846	-1,359	1,064
10		-0,846	1,320	0,437
11		0,526	-0,674	-1,458
12		0,388	-0,513	
13		0,442		
14				

Figure 39: Residuals triangle without outlier.

	1	2	3	4
1		-0,497	0,528	-0,490
2		-0,025	1,274	-1,394
3		-0,375	0,167	0,935
4		-0,284	0,829	-1,089
5		-0,309	0,791	-0,243
6		-0,191	1,060	1,310
7		-0,144	0,301	1,845
8		-0,417	-1,649	-0,304
9		-0,295	0,147	-0,591
10		3,285	0,039	0,103
11		-0,403	-1,400	-0,350
12		-0,365	-1,357	
13		-0,175		
14				

Figure 40: Residuals triangle with outlier.

In Figure 40 it becomes clear that the outlier of figure 36 of 20 is reduced to an outlier of 3.29. Furthermore, the rest of the residuals in development period 2 did all became negative. The preset correlation between the two triangles without outliers was 0.6, but the correlation between the residuals in the second development period of figure 39 and 40 is close to 0. So, one outlier can cause the complete correlation between two development periods to vanish.

This is mostly the result of the difference in the MAS parameter and the difference in the development pattern. The preset MAS parameter, which is used to scale the residuals can be seen in Figure 41.

11104,38	607,07	321,80	363,48	156,37	30,81	20,41	4,52	26,45	1,95	10,31	1,86	1,86

Figure 41: The MAS parameter from the initial triangle.

348207,91	599,73	196,33	314,70	274,52	25,09	17,09	0,48	18,64	1,93	7,85	0,46	0,46

Figure 42: The MAS parameter after the outlier is added.

In Figure 41 and 42 it becomes clear that the MAS parameter belonging to the first average development factor increased significantly as a result of the outlier. This is not strange, as a result of the outlier the differences between the individual development factors and the average development factor increases which automatically scales the MAS parameter. A higher MAS parameter means lower residuals as a result of (10). The reduction of the residual from 20 to 3.29 can mainly be explained by this: $\frac{\sqrt{348207.91}}{\sqrt{11104.38}} \approx 5.6$ which is almost in line with $\frac{20}{3.29} \approx 6.1$. The difference between 5.6 and 6.1 can be explained in the difference in development pattern. The initial development pattern can be found in Figure 43 and the development pattern from the created triangle can be found in Figure 44.

2,23	1,27	1,12	1,07	1,04	1,02	1,01	1,00	1,00	1,00	1,00	1,00	1,00

Figure 43: The MAS parameter from the initial triangle.

2,59	1,27	1,12	1,07	1,04	1,02	1,02	1,00	1,00	1,00	1,01	1,00	1,00

Figure 44: The MAS parameter after the outlier is added.

The difference between Figure 43 and Figure 44 is mainly in the first average development factor, the rest of the development factors did hardly change in the triangle generation process. The difference between the two development factors is not strange, the outlier is so extreme that it is impossible to reproduce the same average development factor. The increase of the average development factors causes the other residuals to become negative earlier. As can be seen in Figure 40, the residuals other than the outlier are all negative and relatively close to 0. Especially when these are compared to the residuals in Figure 39.

H Not normally distributed residual sets

In this appendix are the results of the test with not normally distributed datasets. In the first four figures are the outcomes of the gamma distribution with a scale parameter of 1, a shape parameter of 0.5 and a correlation parameter of 0.

Table 74: AA paid using gamma distribution.

		AA paid gamma					
		Long tail&Long tail		Short tail&Short tail		Long tail&Short tail	
Correlation	**Bivariate**	CBM	ICM	CBM	ICM	CBM	ICM
0	0,06	-0,01	0,04	na	na	na	na
0	-0,10	0,05	0,00	na	na	na	na
0	0,12	-0,03	0,08	-0,03	0,02	-0,02	-0,02
0	-0,07	0,06	0,01	0,00	-0,06	0,01	-0,02
0	-0,06	-0,02	-0,08	0,09	0,29	0,13	0,18

Table 75: AA incurred using gamma distribution.

		AA incurred gamma					
		Long tail&Long tail		Short tail&Short tail		Long tail&Short tail	
Correlation	**Bivariate**	CBM	ICM	CBM	ICM	CBM	ICM
0	0,06	0,03	0,43	0,00	0,02	-0,01	0,12
0	-0,10	0,01	-0,20	0,04	-0,03	0,03	-0,17
0	0,12	-0,04	-0,26	-0,03	-0,11	-0,06	-0,25
0	-0,07	0,06	-0,42	0,04	-0,13	0,04	-0,16
0	-0,06	-0,03	0,26	-0,04	0,07	0,00	0,06

Table 76: AQ paid using gamma distribution.

		AQ paid gamma					
		Long tail&Long tail		Short tail&Short tail		Long tail&Short tail	
Correlation	**Bivariate**	CBM	ICM	CBM	ICM	CBM	ICM
0	0,03	0,02	-0,16	-0,01	-0,21	-0,01	-0,14
0	0,05	0,01	0,02	0,04	-0,02	0,02	-0,18
0	0,03	0,03	0,05	0,02	0,21	0,01	0,34
0	-0,02	0,03	0,16	0,05	0,33	0,00	0,32
0	-0,01	0,02	0,02	0,03	0,00	0,03	-0,13

Table 77: AQ incurred using gamma distribution.

		AQ incurred gamma					
		Long tail&Long tail		Short tail&Short tail		Long tail&Short tail	
Correlation	**Bivariate**	CBM	ICM	CBM	ICM	CBM	ICM
0	-0,01	0,03	-0,13	0,03	0,21	0,03	0,03
0	0,03	-0,01	0,01	-0,01	0,16	-0,01	-0,03
0	0,05	0,03	0,00	0,00	0,13	-0,02	0,02
0	0,03	0,00	-0,03	-0,03	0,04	-0,01	-0,09
0	-0,02	0,02	-0,13	0,06	0,08	0,01	0,04

In the following tables are the absolute correlation estimates of the tables which are in Section 7.6.1.

Table 78: AA paid using gamma distribution.

		AA paid Gamma					
		Long tail&Long tail		Short tail&Short tail		Long tail&Short tail	
Correlation	**Bivariate**	CBM	ICM	CBM	ICM	CBM	ICM
-0,25	-0,21	-0,21	-0,21	na	na	na	na
0	-0,10	-0,05	-0,10	na	na	na	na
0,25	0,36	0,32	0,44	0,17	0,22	0,28	0,28
0,5	0,55	0,53	0,57	0,43	0,37	0,45	0,39
0,75	0,72	0,64	0,67	0,73	0,91	0,52	0,60

Table 79: AA incurred using gamma distribution.

		AA incurred Gamma					
		Long tail&Long tail		Short tail&Short tail		Long tail&Short tail	
Correlation	**Bivariate**	CBM	ICM	CBM	ICM	CBM	ICM
-0,25	-0,21	-0,18	0,20	-0,20	-0,16	-0,16	-0,09
0	-0,10	-0,08	-0,30	-0,06	-0,13	-0,06	-0,27
0,25	0,36	0,31	0,28	0,32	0,27	0,24	0,14
0,5	0,55	0,56	0,40	0,53	0,43	0,44	0,38
0,75	0,72	0,61	0,80	0,65	0,74	0,43	0,44

Table 80: AQ paid using gamma distribution.

		AQ paid Gamma					
		Long tail&Long tail		Short tail&Short tail		Long tail&Short tail	
Correlation	**Bivariate**	CBM	ICM	CBM	ICM	CBM	ICM
-0,25	-0,24	-0,23	-0,38	-0,23	-0,42	-0,17	-0,29
0	0,05	0,06	0,07	0,09	0,02	0,07	-0,14
0,25	0,28	0,30	0,31	0,28	0,39	0,24	0,53
0,5	0,50	0,46	0,60	0,45	0,64	0,35	0,63
0,75	0,75	0,71	0,72	0,73	0,66	0,57	0,46

Table 81: AQ incurred using gamma distribution.

		AQ incurred Gamma					
		Long tail&Long tail		Short tail&Short tail		Long tail&Short tail	
Correlation	**Bivariate**	CBM	ICM	CBM	ICM	CBM	ICM
-0,25	-0,24	-0,21	-0,27	-0,24	-0,05	-0,06	-0,09
0	0,05	0,07	0,05	0,04	0,17	0,02	0,07
0,25	0,28	0,18	0,22	0,25	0,34	0,06	0,03
0,5	0,50	0,51	0,18	0,48	0,57	0,10	0,08
0,75	0,75	0,74	0,83	0,72	0,70	0,24	0,25