

**Statistical analysis of dependencies within
insurance portfolios**

The research presented in this thesis was carried out at the group of Statistics and Probability, Department of Applied Mathematics, Faculty of Electrical Engineering, Mathematics and Computer Science, Twente University, Enschede, The Netherlands.

This research was financially supported by the Dutch Technology Foundation STW (TWI.6314)



© V. Lukočius, 2008.

No part of this work may be reproduced by print, photocopy or any other means without the permission in writing from the author.

Printed by Wöhrmann Printing Service, The Netherlands

ISBN: 978-90-365-2739-2

**STATISTICAL ANALYSIS OF
DEPENDENCIES WITHIN
INSURANCE PORTFOLIOS**

DISSERTATION

to obtain
the degree of doctor at the University of Twente,
on the authority of the rector magnificus,
prof. dr. W.H.M. Zijm,
on account of the decision of the graduation committee,
to be publicly defended
on Thursday, 27 November 2008 at 15:00 hrs

by

Viktor Lukočius
born on 9 March 1979
in Vilnius, Lithuania

Dit proefschrift is goedgekeurd door de promotor

Prof. dr. W. Albers

en de assistent-promotor,

dr. W.C.M. Kallenberg

Contents

1	Introduction	9
1.1	Classical insurance modeling theory	10
1.2	Models and assumptions	11
1.3	Development of the models	15
1.4	Risk measures	19
1.5	Why do we need approximations?	21
1.6	Types of influences on the quantities considered	22
1.7	Data for the model implementation	24
1.8	Outline of the thesis	26
2	Illustrative example: an illness contract	29
2.1	Simulation algorithm	31
2.2	Sensitivity analysis	35
2.3	Fitting simulated data	40
2.3.1	Fitting N and H	41
2.3.2	Fitting G	43
2.4	Conclusions	44
3	The Model	47
3.1	Main characteristics of the model	48
3.2	Suggested distributional assumptions	51
3.3	Model parameters	53
3.3.1	Parameters μ_C , ϵ and μ_G	54
3.3.2	Parameter λ	54
3.3.3	Parameters γ_C and γ_G	54
3.4	Values of the model parameters	56
3.4.1	Situations when $L \sim \text{Gamma}(\alpha_L, \beta_L)$	58
3.4.2	Situations when $L \sim \text{IG}(\alpha_L, \beta_L)$	61
3.4.3	Situations when $L \sim \text{Lognormal}(\alpha_L, \beta_L)$	63

4	Quantities of interest and calculation methods	67
4.1	Net Stop-Loss premium	69
4.1.1	Convolutions	69
4.1.2	Direct Monte-Carlo Simulations	79
4.1.3	Hybrid Monte-Carlo Simulations	80
4.1.4	Approximations	84
4.2	Variance of a Stop-Loss contract	86
4.2.1	Convolutions	86
4.2.2	Direct Monte-Carlo Simulations	88
4.2.3	Hybrid Monte-Carlo Simulations	90
4.2.4	Approximations	92
4.3	Value at Risk of the aggregated claim amount	93
4.3.1	Convolutions	94
4.3.2	Direct Monte-Carlo Simulations	94
4.3.3	Approximations	96
5	Accuracy of approximations	97
5.1	Retention level and criterion	98
5.2	Testing	102
6	Dependence Effect	107
6.1	The data we need	108
6.1.1	Workers compensation insurance	111
6.2	Workers compensation data	112
6.3	Two scenarios	116
6.3.1	The first scenario	117
6.3.2	The second scenario	118
6.4	Conclusions	122
7	Estimation Effects	125
7.1	Observations and estimators	126
7.2	Local behavior of the risk measures	129
7.2.1	Local behavior of VaR	129
7.2.2	Local behavior of SLP	135
7.3	Asymptotic behavior of the estimators	136
7.4	Estimation effect on SLP	150
7.4.1	Protection approach	153
7.5	Estimation effect on VaR	158
7.5.1	Protection approach	161
8	Fire data	169
8.1	Description of the data	169
8.2	Number of claims	170
8.2.1	Time grouping	171

8.2.1.1	Simple claims	173
8.2.1.2	Special claims	176
8.2.2	Space grouping	177
8.3	Claim sizes	181
8.4	Examples and discussion	184
8.4.1	Dependence and estimation effects	186
9	Review and discussions	191
A	Convolutions calculation algorithms	199
A.1	Situation when $C \sim \text{Gamma}$ and $L \sim \text{Gamma}$	200
A.2	Situation when $C \sim \text{Gamma}$ and $L \sim \text{IG}$	203
A.3	Situation when $C \sim \text{IG}$ and $L \sim \text{Gamma}$	204
	Summary	207
	Samenvatting	211
	Acknowledgements	215

Chapter 1

Introduction

From the title of the thesis it is clear that the basic aspect underlying all the mathematical derivations and conclusions of the thesis is insurance. Insurance is a service which is provided by some commercial organization, called the insurance company. The service takes the form of an insurance contract (also called a policy or a set of policies), which can be sold to an individual or some organization (policyholder). The main idea of the insurance contract is to transfer the risk of the policyholder (insured risk) to the insurance company. Hence, the contract is an obligation for the insurance company to cover the loss which is produced by the insured risk of the policyholder in case it occurs during the entire life of the contract. For a certain amount of money (the price of the insurance contract, which is also called the premium) the individual (or organization) transfers its personal risk to the insurance company for a fixed period of time, which is specified in the contract. Speaking of risk demonstrates that an insurance contract deals with uncertain events. That is why statistics plays an important role in this area.

It is useful to mention a specific type of an insurance contract, which is concluded between two insurance companies. Such a contract is called a reinsurance contract and has the same idea as the simple insurance contract, but now the role of the individual is played by an insurance company. For instance, one insurance company transfers part of its risk to another (bigger) company in order to protect against huge losses and hence giving itself the possibility to offer higher protection to its own clients than the company could have been giving otherwise.

Insurance nowadays has become a very powerful tool in financial management. Using the insurance contracts, companies can decrease the uncertainty factor in their business. It significantly simplifies the internal control of the different types of risks and reduces the possibility of insolvency. For individuals, insurance, in addition to its main goal, plays a great psychological role. In modern society more and more people become risk averse. They agree to pay a fixed amount of money (which is even larger than the expected value of the insured risk) to reduce the uncertainty factor from their lives. It is not far from reality to say that

insurance is becoming part of our life. More and more countries obligate citizens to insure certain types of risks (for instance in some countries the common obligated insurances are health and third-party liability insurance for drivers), and people like it, it is handy and simple.

As a result a lot of new insurance companies have been born. Millions of people are obligated to insure their risks which gives rise to huge insurance portfolios with complicated structures and complex internal relationships. Due to involved uncertainty, mathematical models with a stochastic element became very popular in the insurance practice. Whole departments in the large insurance companies were organized to measure and control risks for different lines-of-business.

Consequently (says the European Commission), existing solvency requirements (the requirements for the reserve level, so called *Minimum Guarantee Fund*) fail to provide the necessary level of policyholder protection. This is actually not a shocking surprise since the current solvency regime was introduced in the early 1970s when insurance was not so extremely powerful and complicated as nowadays. New solvency regulations (Solvency II) are going to be introduced in order to control the modern insurance (financial) world of the European Union. In contrast with Solvency I, Solvency II envisages to introduce risk management as a basic tool for the capital requirement evaluation.

All this makes risk management very important. Insurance companies need new, more advanced modeling techniques, more complicated risk control strategies, which could provide better approximations of reality.

This thesis investigates a new insurance modeling technique which can be viewed as a generalisation of the classical theory. Before introducing the new model, we present the description of the basic, well known classical insurance modeling theory.

1.1 Classical insurance modeling theory

Let us consider the insurance portfolio of some large insurance company. The whole portfolio can be presented as a final set of policyholders which bought the insurance contracts and transferred their personal risks to the insurance company. Each policyholder is a potential source of loss for the insurance company, and during some reference period (let us say one year) they altogether produce the aggregate sum of their possible claims. That is why it is natural to model the insurance risk by some random variable, and the main interest for the insurance company is to analyse the behavior of the aggregated sum of all the possible claims, which are modeled by the random variables.

There are two major approaches in the classical insurance modeling. The first approach is the individual risk modeling.

In the individual risk model we denote by n the total number of insurance policies. Each policy 'has a chance' to produce a claim during the entire life of the contract. The claim which is made in respect of the policy i we denote by X_i ,

with $i = 1, \dots, n$. Then the total (or aggregate) claim amount at the end of the reference period is

$$S = X_1 + X_2 + \dots + X_n. \quad (1.1)$$

Often X_i is presented as

$$X_i = I_i C_i. \quad (1.2)$$

The random variable I_i is an indicator which defines whether or not the i th policy produced at least one claim. If no claim has occurred, then $I_i = 0$; otherwise, $I_i = 1$. The random variable C_i can have an arbitrary distribution and represents the total amount of the payment in respect of the i th policy. So, one C_i can consist of several claims.

The model is individual in a sense that all the policies are included in the model and each policy is considered individually. The individual model is not efficient from the calculational point of view. Analyzing huge portfolios can take a lot of time even when claims occur very seldom.

The alternative model, which also describes the total claim amount in a fixed period of insurance contracts, is the collective risk model. Let N denote the number of claims arising from policies during the reference period. Let C_1 denote the amount of the first claim, C_2 the amount of the second claim and so on. In the collective risk model, the random sum

$$S = C_1 + C_2 + \dots + C_N \quad (1.3)$$

represents the aggregate claims generated by the portfolio for the considered period.

Now we do not speak about the separate (individual) policies. Instead, we consider the number of claims which occurred during the reference period (this number is not known in advance and modeled by a random variable N). We do not specify which claim arises from which policy. Therefore, each policy can have several claims during the reference period. In a collective model, the individual information of the considered policies is hidden. Instead, we get complete information about the number of claims together with their sizes, which is hidden in an individual model. Therefore, each model (individual or collective) carries different types of information and both of them can find an application, depending on the situation.

The main advantage of a collective risk model is that it is more efficient from the calculational point of view in case S has to be simulated. We do not need anymore to consider each policy separately even if it has not produced any claims. Instead, we consider only real claims, which did occur in reality.

1.2 Models and assumptions

So far we have described the classical insurance modeling in a very general way. We have not introduced any distributional assumptions and have not specified

any relationships between the random variables involved. We are going to start this section by describing the general assumptions which are mostly used in the classical modeling.

All the assumptions we divide into two parts. The first one deals with the distributional assumptions for the random variables involved (the claim sizes and the number of claims (in the collective risk model case)), which are commonly used in practice. As a second part we consider the relationships between the random variables within the models considered.

The most simple random variables which were introduced in the individual risk model are I_i , with $i = 1, \dots, n$. These are indicator random variables (naturally having the Bernoulli distribution), which can take one of two values, 1 or 0. For a single random variable I_i , the value 1 occurs with probability p_i , and 0 with probability $1 - p_i$. Therefore it is possible to specify the occurrence probability for each policy individually, but usually all the I_i 's, $i = 1, \dots, n$ are assumed to be identically distributed as some random variable I , with $P(I = 1) = p$ and $P(I = 0) = 1 - p$. Such random variables are included only in the individual risk models.

The random variables which occur both in the individual and collective models, are the claim sizes C_i , with $i = 1, \dots, n$ in the individual model case, and with $i = 1, \dots, N$ in the collective model case. In general, every distribution with non-negative values only can be considered as a claim size distribution, but there are several additional features which are very desirable in the modeling of the claim size. As an example we can mention closedness under convolution (which significantly simplifies matters when analyzing the aggregate distribution), and positive skewness, which is a quite common feature for insurance data. There are quite a few families of distributions available in literature. In Kaas et al. [2001] various types of distributions are described which are often used in the insurance practice. Reijnen et al. [2005] contains an overview of the classical and some new assumptions for the claim size modeling. The classical, well-known, candidates for the claim size distribution are the Gamma, Log-normal and Pareto distributions. As an alternative can be considered the Weibull, Burr and Inverse Gaussian (*IG*) distributions.

The last distributional assumption of the classical modeling which is going to be discussed in this section, is the assumption for the random variable N , the number of claims in a collective risk model. The most common assumptions for N are Poisson, binomial and negative binomial distributions. In case N is assumed to be Poisson, the distribution of S is called a compound Poisson distribution. In case N is (negative) binomially distributed, S has a compound (negative) binomial distribution.

The second part of the assumptions which are described here, is the relationships between the random variables within the considered models. The random variables in the model can be independent or dependent (correlated) or, it can be a mixture of dependent and independent variables. In the classical models everything is simple, all the variables are assumed to be mutually independent.

In the individual risk model $I_1, \dots, I_n, C_1, \dots, C_n$ are independent, while in the collective risk model N, C_1, \dots are independent.

However, this does not mean that people who apply these models, do not believe in a possible correlation between the random variables. Dependencies do occur in practice. For examples we can refer to Albers et al. [2008] where were mentioned such examples as man and wife which are both insured in the same portfolio, carpoolers using a collective company insurance, catastrophes like hurricanes or floods hitting huge numbers of insured at the same time. People clearly realize that, but often do not introduce any dependence structure in the model. The argument is that dependence does not have much influence on the final result and, moreover, significantly complicates all the calculations. Nevertheless, different types of dependencies have been a subject for study for quite some time. Dhaene et al. [2002a,b] investigate the dependence by using the comonotonicity approach. Hayne [2007] introduces the dependence structure called a shock model, in which claims from different lines of insurance are linked by a common variation (or shock) in the parameters of each of these lines. Copula models have become a popular alternative in the modeling of dependencies between random variables. A copula is a function that represents a joint distribution of random variables with uniform marginal distributions, specifying the underlying dependence structure. The first books in this area were written by Joe [1997] and Nelsen [1999]. More information can be found e.g. in Lindskog [2000] or Embrechts et al. [2003].

All these studies clearly demonstrate that the effects of dependence on some insurance (financial) products can be really huge. However, some arguments can still be suggested against the introduction of dependence. In most cases the dependency structure significantly complicates all the calculations. The distribution of the aggregate sum S usually takes a very complicated form and often does not have an explicit formula. Moreover, it can be stated that such strong dependencies rarely occur in practice. In most cases dependence will be ignored anyway, and as long as dependence effects are 'small in average', the effects on relevant risks (financial or insurance products) are small as well.

Precisely this optimism was crushed in Albers [1999], using local dependence models. Assuming that on average 1% – 5% (which can be considered a small fraction) of the total portfolio claims arise due to dependence, the author showed that Stop-Loss premiums increase by a factor (not a percentage!) 2–6, compared to the independent case. This was a strong signal to take small dependencies seriously and to continue investigation in same direction. This thesis is part of that investigation. The same approach to dependence is analyzed here in detail and extended to more general and flexible models.

Our approach to dependence inside the insurance portfolio differs a bit from the models which were mentioned above. The main goal of our research is to concentrate on weak forms of dependence, in between the independence model and the comonotonic case. In fact, on the scale independent-comonotone our models are still close to the independent end-point. The basic assumption of our modeling technique is the possibility of claims occurring in groups. To be more

precise, we assume that the claims can be of two types. The first type are the individual (ordinary or 'simple') claims. These claims arise one by one and all time points (times when the claims arise) and claim sizes are mutually independent. The sum of such claims in itself can be the individual or collective model. The second type of claims are group claims ('special' claims). We assume that during a single time point several claims can come at once. This part can be seen as a comonotonic part: a whole group has damage. By considering such possibility we certainly introduce positive correlation between the random time points (times at which the claims arrive).

The idea is that claims can arise as a group due to some special cause. As an example of such a special cause we can mention a car crash when more than one car from the same insurer is involved in the accident. Couples or family members, usually having the same insurance company, have a similar life style with similar risks (like travelling together, living in the same house). Bad weather can be considered as a special cause in car insurance as well. Earthquakes and floods also fall in the category of the special cause and can be covered by the model. But these are really extreme cases, which are not of main interest in our research. The notion of groups is not necessarily linked up with time. Groups may also be formed on the basis of spatial arguments or otherwise. In this sense our models are a quite flexible way to describe forms of dependence, see also Albers [1999, p. 175].

So far we explained the influence of a special cause on the total number of claims, but it is certainly possible to introduce an influence on the claim sizes as well. It is noted by some insurance companies that bad weather conditions (heavy rain or fog) not only increase the number of claims, but also decrease the average claim sizes. This means that people indeed produce more car accidents during bad weather, but the damages are smaller on average. On the other hand, the opposite also occurs. In some cases the special cause can significantly increase the claim sizes of the involved policyholders. An example could be a workers compensation insurance, where the ordinary claims can be small, compared to the 'special' claims, which can occur (for instance) due to some explosion in the laboratory.

In general, the group size can be any positive integer, but we mostly concentrate on relatively small groups of average size 5-20. Moreover, we assume that the expected number of special claims is small as well (1% – 5% of the expected total number of claims). Such a setup first time was introduced in Albers [1999]. The research was continued with Reijnen's [2003] master thesis, which was partially summarized in Reijnen et al. [2005]. The present thesis can be considered as a continuation of that work. Therefore, to make it self-contained, we briefly describe the work which has been done before. We start with the description of the technical development of the models.

1.3 Development of the models

In this section we present a somewhat more technical development of the dependence model, which first time was introduced in Albers [1999]. More details will be given in Chapter 3. The same model, in a slightly modified form and with different distributional assumptions appeared later in Reijnen et al. [2005]. During our research the model was generalized several times. Each generalization, in fact, can be considered as a separate model with its own specific features. Therefore we point out three different forms, which can find an application in different fields of insurance modeling.

First briefly consider the independence model (later we will call it Model 1). Consider a large insurance portfolio with n insured members (policyholders). Each of these members has a small probability p to submit a claim C , which is modeled by a random variable. The aggregate sum of all the claims during some reference period (let say one year) can be represented as

$$S = \sum_{i=1}^N C_i. \quad (1.4)$$

All random variables which are involved in the model (i.e. N, C_1, C_2, \dots), are independent, N represents the total number of claims (usually following a binomial or Poisson distribution) and all the C_i 's are distributed as C for all $i = 1, \dots$. This is a general form of the collective risk model, which will be used as basis for our dependence model. We do not introduce any specific distribution for the claim sizes C_i since it is not an essential point at this stage.

To introduce dependence, Albers [1999] assumed that the whole portfolio can be divided into several groups of a fixed size g with similar (dependent) risks. Since n is assumed to be large, the total number of groups $h = n/g$, without loss of generality, is assumed to be an integer. Different group sizes represent different situations of dependent risk, which can occur in practice. For example, $g = 2$ can be assumed to model the dependent risk of couples, $g = 4$ can represent car crashes, $g = 200$ can be used in a modeling of a plain crash, etc. The extreme case $g = n$ can occur in the form of a catastrophe, when all the members in the insurance portfolio fall in the same dependence group. Now, the aggregate sum of the claims takes a form

$$S = \sum_{k=1}^h \sum_{i=1}^g Y_{ik}, \quad (1.5)$$

where

$$Y_{ik} = (1 - V_k)X_{ik}C_{ik} + V_kZ_{ik}D_{ik},$$

for $i = 1, \dots, g$ and $k = 1, \dots, h$. All $V_1, \dots, V_h, X_{11}, \dots, X_{gh}, Z_{11}, \dots, Z_{gh}, C_{11}, \dots, C_{gh}, D_{11}, \dots, D_{gh}$ are assumed here to be independent. Basically Y_{ik} represents a single claim ('simple' or 'special') in case one has occurred during the reference period, or no claim otherwise. Here the author utilizes three types

of indicator random variables. The first one, V_k , represents occurrence of a special cause (for the discussion about the possible special causes see the previous section), affecting the k -th group with probability

$$P(V_k = 1) = 1 - P(V_k = 0) = \theta_k,$$

where θ_k are supposed to be small and usually assumed to be the same, i.e. equal to some θ for all $k = 1, \dots, h$. Both the X_{ik} 's and Z_{ik} 's (with $P(X_{ik} = 1) = 1 - P(X_{ik} = 0) = q$ and $P(Z_{ik} = 1) = 1 - P(Z_{ik} = 0) = r$ for all $i = 1, \dots, g$ and $k = 1, \dots, h$) are indicator random variables, indicating the death (in case of a life insurance or submission of a claim in case of a non-life insurance) of the i -th policyholder in the k -th group during the reference period. The random variable which indicates occurrence of any claim ('simple' or 'special') is denoted by

$$I_{ik} = (1 - V_k)X_{ik} + V_kZ_{ik},$$

with $P(I_{ik} = 1) = 1 - P(I_{ik} = 0) = (1 - \theta)q + \theta r = p$. In case the distributions of all the C_{ik} 's and D_{ik} 's are assumed to be the same, the model can be written as

$$S = \sum_{k=1}^h \sum_{i=1}^g I_{ik} C_{ik},$$

which form is very similar to the independent risk model.

To introduce a more intuitive way of defining the volume of dependence, a new parameter

$$\epsilon = P(V_k Z_{ik} = 1 | I_{ik} = 1)$$

was introduced by the author. Typically, ϵ will be small, at most a few percent, and $\epsilon = 0$ will represent the boundary case of independence. Note that $\theta = \epsilon p / r$, implying that θ runs from the extremely small ϵp to ϵ itself.

Using these indicators, the whole portfolio is divided into an individual (independent) and a common (dependent) part and dependence is introduced through the V_k 's. The structure of the model allows to introduce different assumptions for the distributions of the claim sizes of the dependent and the independent parts. However, in most cases the claim sizes are assumed to be identically distributed and the normal distribution was used to illustrate the impact of dependence on the different kinds of *SLP*'s.

In the paper the author also shows how the model is related to different, well known types of dependence models (some of them were mentioned in the previous section). We refer to Albers [1999] for the description of these relationships.

A similar (with the same approach to dependence) model, in a slightly different form and with different distributional assumptions, appeared later in Reijnen et al. [2005]. The total aggregate sum of all the claims was presented there as

$$S = \sum_{i=1}^{gW} D_i + \sum_{j=1}^{n-gW} X_j C_j, \quad (1.6)$$

with $P(X_j = 1) = 1 - P(X_j = 0) = q$, $h = n/g$ and $W \sim \text{binomial}(h, \theta)$. Note that parameters q , g , and θ have similar but not the same values when compared with the parameters from model (1.5). Now the model is divided into two parts. In the first part the total claim sizes D_i 's include both the 'simple' (independent) and the 'special' (dependent) claims. Therefore, it is natural to take the D_i 's stochastically larger than the C_j 's, which are the 'simple' claims of the independent part: those who have no 'special' claims. That is why the D_i 's and C_j 's are assumed to come from the same family of distributions (in the paper they assumed to be lognormal), have the same skewness, but different expectations. All the D_i 's are identically distributed, and the same holds for both the C_j 's and the X_j 's as well. Moreover the D_i 's, C_j 's and X_j 's are independent. The random variables X_j indicate the occurrence of (at least one) individual claim in the second part of the model. The W defines the random number of groups of a common risk and is independent of the D_i 's, C_j 's and X_j 's. Note that the model assumes that all the group members submit at least one claim due to the special cause. This is one of the features which distinguishes models (1.5) and (1.6). Speaking about differences, it is clear that model (1.6) is a collective model. In model (1.5) the potential groups have to be defined beforehand, which is often very hard (or even impossible) to do. Model (1.6) uses the collective approach. Therefore, it deals directly with claims which occurred in reality. In that sense model (1.6) can be applied more easily in practice.

The model which is considered in the present thesis has a slightly different structure than models (1.5) and (1.6). Generally, we would like to point out three different forms of the same model, which in a sense can be considered as separate models with their own specific features. The first form of the model (we call it Model 2) is written as

$$S = \sum_{i=1}^N C_i + \sum_{k=1}^H \sum_{j=1}^g D_{kj}. \quad (1.7)$$

It is a collective model, just like (1.6), and it is clearly divided into two parts, but now the two parts refer to the independent and dependent claims and the dependent part does not contain independent claims. Therefore we will often assume the same distribution for the claim sizes C_i and D_{kj} . This feature significantly simplifies all the calculations. Therefore, it will be a typical assumption in the entire thesis. The random variables $N, C_1, \dots, H, D_{11}, \dots$ are independent. In the previous models n was assumed to be large and p typically was small. It allows us to replace the binomial distribution, which appeared in model (1.6), by the Poisson distribution. Hence, the total number of 'simple' claims is written now as $N \sim P(np(1 - \epsilon))$, and the number of groups $H \sim P(\epsilon \frac{np}{g})$. The parameter ϵ appears here in a very natural way, representing the expected percentage of the 'special' claims, compared to the total expected number of claims. Parameters n and p everywhere appear together, as a product. Therefore it is natural to replace np (the total expected number of claims) by a single parameter λ . The special feature of Model 2 is a fixed group size g . Note that g here is similar to the g

from model (1.6), but completely different from the one in model (1.5), which was a potential group size. In fact, gr from model (1.5) corresponds to g from model (1.6) and (1.7). In that sense Model 2 is closer to model (1.6) than to model (1.5).

Certainly, in some situations it is possible (or even necessary) to define fixed groups of a common risk, but in real life more often we deal with uncertain group sizes. Often a fixed group size is not a reasonable assumption. The example could be car crashes due to bad weather (special cause). It is clear that it is impossible to predict how many cars will crash during a rainy day. Therefore, the second form of the model (we call it Model 3) assumes that the group sizes are not fixed but stochastic. We replace the fixed parameter g by a random variable G_k for each $k = 1, \dots, H$. In that sense Model 3 is closer to the model (1.5) where the author fixed only the potential groups. The real groups (with were formed using the indicator random variables) were stochastic as well. Model 3 is written as

$$S = \sum_{i=1}^N C_i + \sum_{k=1}^H \sum_{j=1}^{G_k} D_{kj}, \quad (1.8)$$

assuming that all G_k follow the Poisson distribution with a fixed parameter μ_G . The random variables $N, C_1, \dots, H, G_1, \dots, D_{11}, \dots$ are independent.

The final model (Model 4), has has the same form as Model 3, i.e.

$$S = \sum_{i=1}^N C_i + \sum_{k=1}^H \sum_{j=1}^{G_k} D_{kj}. \quad (1.9)$$

It will be the main object of our investigation, therefore it will be motivated and explained in a more precise way in Chapter 3. Here we present only the general form and basic motivation. The main difference between models (1.8) and (1.9) is the distributional assumption for the group sizes G_k . In Model 3 all the G_k 's were assumed to be random, following the Poisson distribution with a fixed parameter μ_G . Model 4 introduces the overdispersion aspect for the group sizes by changing the fixed parameter μ_G into the random variable L with expectation $E[L] = \mu_G$. The reason for this step is the heterogeneity of the data considered. While considering huge insurance portfolios we have to keep in mind that they certainly include members from different risk levels. An example of this fact could be two collective contracts which appear in the same portfolio: a university collective contract and a contract of some construction company. These two contracts, having completely different risk levels, certainly require different assumptions for the grouping procedure. In order to reflect this fact in the model, we introduced a random expectation of the Poisson group sizes. Basically, for each G_k we have a different L_k , but in the present thesis all the L_k 's are assumed to be independent and identically distributed (iid). Therefore we will use L as a representative of all the L_k 's. This step will be motivated in more detail in Chapters 2 and 3.

1.4 Risk measures

In the previous sections we described models which can be used to model the aggregate sum of claims S of some insurance portfolio. This sum certainly is a subject of interest to insurance companies, but very often not the sum itself, but a function of it attracts more attention. As an example we can consider (re)insurance products (contracts), the prices of which (premiums) are based on S . The important feature of insurance business is that premiums are paid by policyholders in advance, before the insurer pays claims. An insurance company may get into problems in case the total number of claims during some reference period exceeds the total premium (the sum of all the premiums which were paid by policyholders in advance). Therefore, insurance companies have to keep a certain amount of money in reserve. The amount of reserve is controlled by regulatory authorities and depends on the volume and type of risk which the insurer has collected in the portfolio. At the beginning of this chapter we mentioned that in order to improve policyholders protection, the new solvency regulation is going to be introduced soon. Solvency II is going to utilize risk management as a basic tool for the capital requirement evaluation. Consequently, it becomes very important for insurance companies to analyze their current situation by measuring the total portfolio risk.

Mathematically, a risk measure is a mapping from a space of random variables to the real line. It is a tool which interprets a certain risk (modeled by random variables) by an economically motivated number. Usually risk measures should satisfy a number of properties, which are considered to be helpful in the decision making procedure. Following Laeven [2005, p. 2], such properties can generally be divided into three parts. The first part of properties are the so called basic rationality properties, such as monotonicity, which intuitively is very important and is satisfied by most risk measures studied today. The second part consists of technical properties, such as continuity. These are mostly necessary for the theoretical analysis (mathematical proofs) and rarely have any economical interpretation. The third part is the 'intuitive' part, containing such important properties as additivity.

In general it is not necessary that a considered risk measure satisfies all the properties. Usually the properties are chosen according to the specific features of the risk which is going to be measured by the risk measure constructed. However, when the risk is defined and all the goals are clear, it is usually not easy to define a 'best' risk measure. Different risk measures can have different positive and negative features, which could be handy in different situations.

One of the risk measures which is quite popular in financial management is the standard deviation. It has a clear economical interpretation, illustrating the spread of the risk (values of the random variable) around its expectation. It certainly gives a lot of information about the investigation target, but has several important drawbacks. One of them is that it treats the negative and the positive deviations from the expectation in the same way. This could be a big problem

when the considered risk is not symmetric. In that case alternative risk measures like semi-deviation can be used to get a better picture of the considered situation. Another drawback of the standard deviation can be noticed when applying it with heavy-tailed distributions, which are widely used in actuarial practice. For such distributions the standard deviation may not even exist and, in case it does exist, it certainly does not provide adequate probabilistic information.

In the present research we mostly concentrate on distributions which are far from the normal, i.e. not symmetric or having heavy tails. This is because the basic target of our consideration is the random sum of random variables, the distribution of which is usually far from normal. Moreover, in the previous research, Albers [1999] and Reijnen et al. [2005] showed that the tail of the distribution is very sensitive to the dependence effect. It certainly is present everywhere, but in the tail it simply blows up, which can lead to huge risk underestimation. Therefore the measures which concern the tail behavior of distributions are of main interest in the present thesis. The tool which was used by Albers [1999] and Reijnen et al. [2005] to illustrate the effect of dependence, was the net Stop-Loss premium (*SLP*). From one side, the *SLP* can be simply considered as a price (or part of the price) of the reinsurance contract, but it certainly can also serve as a risk measure, illustrating the tail behavior of the considered quantity. In the present thesis we continue to use *SLP* as a basic tool of the dependence analysis.

The idea of a Stop-Loss contract is simple. If S denotes the total risk amount, the Stop-Loss contract covers the part of this risk which is above some retention level a . Hence, the total risk S is divided into two parts: retained risk $S - (S - a)^+$ and the Stop-Loss risk $(S - a)^+$ (here $(S - a)^+ = \max(S - a, 0)$). If we consider reinsurance as an application example of the Stop-Loss contract, the reinsurer gets the risk $(S - a)^+$ and the remaining part $S - (S - a)^+$ is left for the insurance company which buys the contract. The *SLP* is defined as $E[(S - a)^+]$. It is clear that in reality the *SLP* cannot be the total price of a Stop-Loss contract because the re-insurer's position becomes very risky with expected profit 0. Some positive loading definitely has to be added to the *SLP*. The value of such loading depends on many factors and it has to be acceptable for both sides: the insurance company and the reinsurer. It can be some fixed value or it can depend on some risk characteristic like mean or standard deviation. Therefore, variance (or standard deviation) of the Stop-Loss contract is analyzed here as well.

Yet another risk measure which will be used as a tool for the dependence analysis is the Value at Risk (*VaR*). This quantity, which is sometimes also called the Quantile (or Percentile) function, is defined as the inverse of the distribution function of the considered risk (in our case it is the aggregate sum S). It presents an upper limit for the aggregated sum of claims, given some probabilistic risk value. Writing $F_S(s) = P(S \leq s)$, for given p , it is defined by $VaR = F_S^{-1}(1 - p)$.

VaR is a popular risk measure in the insurance world. It deals with the tail behavior of the underlying distribution, which is an important feature in our research. Moreover, Laeven [2005, p. 117-118] in his thesis illustrated that *VaR* arises as the optimal solvency capital requirement in the considered minimization

problem. Furthermore, as stated in Bouma [2006], the formula used to calculate the Solvency Capital Requirements is likely to be based on *VaR* (or Tail Value at Risk). All this makes *VaR* an important economic quantity. Thus, it is certainly important to analyse its behavior taking dependence into account.

1.5 Why do we need approximations?

In the previous section several risk measures were introduced, which in the present thesis are going to be used as dependence analysis tools. Once the risk measures have been selected, the next step is to calculate them in practice. This is a nontrivial exercise, in particular when there is stochastic dependence between the risks under consideration. All the risk measures which will be considered in the thesis are basically functions of S , the aggregate sum of all the claims in the considered insurance portfolio. Therefore, to calculate the chosen risk measure we have to know the distribution of S . It is well-known that distribution functions of (random) sums of (dependent) random variables are typically of a complex form. Albers [1999] showed that even in a very simple case (when the distribution of the claim sizes is normal and the distribution of the total number of claims is binomial) the resulting distribution of S can be calculated only numerically. In most cases, when more complicated (and also more realistic) underlying distributions are assumed, it is not possible to get an analytical expression for the distribution of S at all. Simulations might be a solution, but in most situations (since the target of interest is the tail of distribution) this will be too time consuming. Moreover, to get a clear picture of how the risk measures actually depend on the underlying parameters, it does not help much to obtain numerical outcomes for separate sets of parameter values. With such ‘snapshots’ we will never be able to grasp the broad underlying picture.

Hence what we need to do is to find approximations to the distribution of S which possess the following properties. First of all, they are sufficiently accurate to allow using them instead of the exact results. Moreover, they should be much easier to evaluate. Finally, they should be transparent, in the sense that they reveal how the various underlying parameters influence the final outcome. The possibility to perform such sensitivity analyses is of course very important, possibly even more than the ability to directly and quickly obtain numerical answers for given numerical input combinations. The basic idea in obtaining the approximations in question will run as follows. Some moments of S can be obtained relatively easily. This will allow us to consider approximations based on these moments as an alternative calculation method. Several possible approximation methods (normal power, Edgeworth expansions, gamma, inverse Gaussian and gamma-inverse Gaussian) were already investigated in Reijnen et al. [2005], where model (1.6) was used to construct the aggregate sum S . The main goal of that paper was to find the best approximation (among the ones considered) of the *SLP* and to formulate a relatively simple rule of thumb to choose this optimum

under the distributional assumptions at hand. As a result, two approximations (inverse Gaussian and gamma-inverse Gaussian) were selected by the authors. According to the rule of thumb presented in the paper, these two approximations cover a wide range of distributional and parametric assumptions. Therefore, these two approximation techniques are going to be considered in the present thesis as well as candidates for approximating the distribution function of S when Model 4 is assumed. To judge the accuracy of approximations (and also explain what is actually a reasonable approximation in our case), we are going to introduce a fixed criterion, which will have a clear economic interpretation. Approximations which fall in the range of the criterion, will be accepted as reasonable. Using similar arguments as in Reijnen et al. [2005] we are going to create a rule of thumb for Model 4, which will serve as a 'decision-making' tool for choosing the appropriate approximation.

1.6 Types of influences on the quantities considered

In this section we briefly describe different types of influences on the risk measures considered. Each of the types of influences defined here will be very carefully investigated in the present thesis. Considering a fixed risk measure (or an insurance product), we point out three major types of influences. The first one is model influence.

The calculation procedure of the price of an insurance contract (or some risk measure) requires a lot of information about the future, which is not available in the present. By selling a contract, an insurance company is obligated to cover all future claims (of the types mentioned in the contract) of the buyer, during some time period (which is also mentioned in the contract). The price of the contract certainly depends on the number and amount of claims which are expected to be paid. The precise amount of all the claims is obviously not known (otherwise the contract price would simply be equal to that amount). Therefore, possible claims which are going to arise in future are modeled by random variables. The parameters necessary to characterize these random variables are usually estimated based on available data about similar claims which have already occurred and were covered in the past. In this way, using the present (available) information and introducing a number of assumptions and simplifications, the model defines a general development rule of the portfolio. When the model is defined, it can be used to predict the future, validate insurance products, calculate risk measures, etc. Nevertheless, the contract seller has to keep in mind that the price of the contract is based on the model, which only is a rough simplification of the real situation. Models can go from simple to very complicated. The latter describe reality in more detail, but they still are wrong and being more detailed they may lead to statistical procedures which are less robust against changes in the situation, or suffer more from less fortunate data sets used in the estimation

problem. Several characteristics of models, like soft sensitivity to the underlying parameters, simple application in practice, are considered to be an advantage, and of course it is better to have a rough approximation to the right answer than a mathematically elegant solution for the wrong one. The message here is that models should neither be too simple nor too complicated. In the first case modeling loses its goal of describing reality. On the other hand, models should not be too complicated, as this reduces the possibility of their application in practice and may lead to a lack of robustness. In our case, the more complicated models are generalizations of the simpler ones. Therefore they contain more parameters to be estimated and hence the larger estimation error involved should be compensated by the better fit to reality. Therefore, speaking about the model influence, we basically speak about different choices of the model. It is clear that under such different models we get different contract prices, different risk measures and different pictures of the portfolio development. Later in the thesis instead of 'model influence' a 'dependence effect' statement will be used. This is because the main differences between the models considered will be connected with the dependence structure, which was introduced earlier. Thus, the introduced dependence will be the main source of the model influence on the quantities calculated. We refer to Albers [1999] and Reijnen et al. [2005] for illustrative examples of the dependence effect. Models (1.5) and (1.6) were compared there to the independence model and the dependence effects were presented using the *SLP*. Assuming that on average 1% – 5% (which can be considered a small fraction) of the total portfolio claims arise due to dependence, the authors showed that *SLP*'s increase by a factor (not a percentage!) 2 – 6 when dealing with normally distributed claim size distributions, or even up to factor 500 for more realistic skewed claim size distributions.

The second type of influence which is considered in the present thesis is approximation influence. In the previous section we explained the importance of approximations in our research. By changing the original distribution of S to the approximate one, we introduce an additional source of error for the resulting quantity. This may lead to unexpected results. Therefore, it is very important to analyze this source of error in detail. Note that when we speak about the dependence effect, we expect that risk measures like *SLP* will increase after the introduction of dependence. This is logical since dependence naturally implies that the claims stick together and hence the possibility of a much higher total number of claims increases, which leads to a higher risk in particular with respect to the tail of the distribution of S . We face a completely different situation while dealing with the approximation influence. We certainly do not know in advance whether the final quantity will be increased or decreased after the approximation is applied. Therefore, as was already stated in the previous section, it is very important to make the approximation error as small as possible, especially in the tail of the approximated distribution.

The final type of influence which is going to be investigated here, is estimation influence (or estimation effect). It is clear that before the model can be applied

in practice, the parameters of the underlying distributions have to be estimated from the data. Replacing the true parameters (which are naturally unknown in practice) by estimated values, we introduce one more source of error. Just as with ignoring the dependence effect, it is too optimistic to act as if the estimation errors are negligible (unless we have a very large number of observations). Therefore, we investigate this source of error as well. As usual in statistics we can give confidence upper (and lower) bounds for the error induced by estimation.

1.7 Data for the model implementation

In this section we are going to look a bit closer at the application of our models in practice. More precisely, we are going to discuss possible difficulties which can arise while applying the models. In general, all the arguments which will be given here are applicable to all the models with the dependence structure defined in the thesis. Nevertheless, we shall use Model 4 (which was briefly defined in section 1.3) as representative for all the models in the discussions and illustrative examples that follow. Speaking about application difficulties, we are mainly concerned with the calibration procedure of the model parameters. It is clear that before a model can be used as a prediction (pricing or validation) tool, all its parameters have to be calibrated with historical data. At this stage we may face difficulties like lack of data, when the number of observations is not sufficient for reasonable estimation, or even worse, complete absence of the needed data. To be consistent, we start with a general explanation of the parameters involved and a brief description of the data which are needed for the model implementation.

The general form of Model 4 (see (1.9)) includes two types of random variables. The first type, which includes N , H and G_k , deals with the number of claims and the second one, which includes C_i and D_{kj} , deals with the claim amounts. Therefore, for convenience all the parameters will be divided into these two parts as well. First consider the parameters of the number of claims distribution(s). The distribution of N and H is in Model 4 the same as in Model 3: $N \sim P(\lambda(1 - \epsilon))$ and $H \sim P(\lambda\epsilon/\mu_G)$. Here λ is the total expected number of claims, ϵ is the expected percentage (in decimals) of special claims and μ_G is the expected group size, see also Section 1.3. The precise distribution of G_k has not yet been defined. Such a definition requires additional motivation, which is going to be introduced in later chapters of the thesis. Apart from μ_G the remaining parameters of the distribution of G_k will be denoted by some abstract vector, called θ_G . For the C_i and D_{kj} several claim size distributions will be considered, see also Section 1.2, where classical and alternative candidates are presented. As mentioned in Section 1.3 often the same distribution will be taken for C and D . Therefore, we denote the parameters of their distribution by the abstract vector θ_C .

As was mentioned before, all these parameters have to be estimated from the historical data and the estimation error (or even estimation possibility) mostly depends on the available structure and volume of the data. Therefore, it is hard to

say something precise about the estimation without seeing the data. Nevertheless, some general points and possible estimation difficulties can already be mentioned now. From a first glance it is clear that all parameters can be divided into two parts in yet another way: the 'easy' part and the 'complicated' part. This division is basically inspired by the dependence information, which is often not easily available in the data. Therefore, parameters which are included into the 'easy' part are λ and the θ_C set. These parameters do not require any information about the dependence (groups) structure. Parameter λ is the total expected number of claims which can be easily estimated from any reasonable data set. The same situation occurs with θ_C , which are the set of parameters of the claim size distribution(s). As long as we assume that all the claim sizes (C_i 's and D_{ik} 's) follow the same distribution (otherwise θ_C shifts to the 'complicated' part) it is quite easy to estimate them as well.

The most difficult part is the estimation of ϵ and θ_G . For these parameters, the data has to include information about the dependence groups. If such information is not included directly, it has to be possible to distill it from the available data. Otherwise, estimation is impossible at all. It is very important to have data with individual information. If the data is aggregated, it is not flexible anymore and it is impossible to get any information about the dependence structure. Now we are going to present several examples of possible data sets which more or less satisfy our requirements. The first example we call a 'perfect' data set where all the necessary information is already included (see Table 1.1).

The most important field of the table is the second one, named 'Group code'. Obviously it is clear that zeroes mean individual claims and numbers define different groups. In this particular example we have 43 groups of sizes 3, 2, 4, ..., 2. Having such a data set, we directly get all the information we need for the model implementation. The field 'Claim amount' can be used to estimate the total expected number of claims λ and parameters of the distribution of the claim sizes θ_C . The field 'Group code' contains information which is necessary for the estimation of ϵ and θ_G . The main problem is that usually a field like 'Group code' is not available in historical data of insurance businesses. This is not a surprise since advanced modeling of the type we are considering here, is only very slowly getting introduced in the insurance practice. Data sets which for a long time have been collected by insurance companies mostly are analyzed with independence models and do not contain extended information about the claims, which is necessary for implementation of new models. Therefore we anticipate the fact that information about dependence will not be straightforwardly be available in the insurance databases. In that case it should be possible to derive such information from the underlying basic data. An example of such a possibility, which is based on workers compensation insurance, we illustrate in Table 1.2. Here the 'Group code' field is created from the 'Date', 'Incident' and 'Place' fields. Therefore, if several equal accidents happened at the same date and at the same place, we can assume that they form one dependence group. For instance if in some laboratory there was an explosion on June 21, the people which got compensation for this accident are in

Claim amount	Group code
75	0
40	0
5	1
9	1
92	1
44	0
50	0
27	0
88	2
42	2
68	0
26	3
27	3
27	3
96	3
35	0
...	...
85	0
92	43
12	43

Table 1.1: Example of the 'perfect' data set.

the same dependence group.

Note that data sets from Tables 1.1 and 1.2 are only illustrative examples of the structure which is needed for successful estimation. To get reasonable estimates we often have to consider data from several years. Moreover, as was already stated in the previous sections, the notion of groups is not necessarily linked up with time. Groups may also be formed on the basis of spatial arguments or otherwise. In this sense our models are a quite flexible way to describe forms of dependence.

1.8 Outline of the thesis

The present thesis considers the advanced actuarial modeling technique which introduces dependence as a natural extension for the existing independence models. The structure of dependence considered here is briefly introduced in Chapter 1. A more detailed explanation together with the definition of the main model can be found in Chapter 3. The form of dependence introduced here is quite new and differs considerably from the structures available in literature. Therefore, all the assumptions and generalizations of the model have to be motivated. This aspect is covered in Chapter 2, where the recommendations are based on an

Nr	Date	Name	Claim amount	Group code	Incident	Place
1	d1	n1	c1	0	i7	p1
2	d2	n2	c2	0	i2	p2
3	d3	n3	c3	d3.i3.p3	i3	p3
4	d3	n4	c4	d3.i3.p3	i3	p3
5	d3	n5	c5	d3.i3.p3	i3	p3
6	d4	n6	c6	0	i4	p4
7	d4	n7	c7	0	i1	p1
8	d6	n8	c8	0	i6	p4
9	d7	n9	c9	d7.i7.p2	i7	p2
10	d7	n10	c10	d7.i7.p2	i7	p2
11	d8	n11	c11	0	i2	p1
12	d9	n12	c12	d9.i9.p4	i9	p4
13	d9	n13	c13	d9.i9.p4	i9	p4
14	d9	n14	c14	d9.i9.p4	i9	p4
15	d9	n15	c15	d9.i9.p4	i9	p4
16	d9	n16	c16	0	i3	p3
...
18	d10	n18	c18	0	i2	p1
19	d11	n19	c19	d11.i7.p2	i7	p2
20	d11	n20	c20	d11.i7.p2	i7	p2

Table 1.2: Example of creating the "Group code" field from the available information.

illustrative example. This concerns a flu epidemic inside some company, introduced as possible realization of our dependence structure. A general scheme of mutual infection is developed and simulated in order to illustrate and motivate the distributional assumptions of our model. Calculation aspects are covered in Chapters 4. Various ways to calculate the risk measures are presented: convolution methods, simulation and approximations. As already explained in Section 1.5 transparent approximations are a basic tool in this thesis. To apply these approximations, they should be sufficiently accurate. Chapter 5 is devoted to this important issue, leading to a simple rule of thumb showing the conditions under which the approximations work well within the assumed parameter range. Next, Chapter 6 uses these methods to illustrate and analyze possible consequences of the dependence and different kinds of impacts to the risk measures. It turns out that the dependence effect can indeed be very large and that even small dependencies cannot be ignored. For illustrative purposes an example of workers compensation data is given. In Chapter 7 the estimation problem is treated. The estimation error is dominated by the part of the parameters related to the special claims, because by their nature we do not have many observations of them (see also Section 1.7). Although the estimation error in this way is restricted to a few

parameters, ϵ , μ_G and γ_G , it is seen that it may be quite substantial. Upper and lower confidence bounds are given for the risk measures, thus protecting against the estimation effect. The real data which are supplied by the insurance company Nationale Nederlanden are analyzed in Chapter 8. Different ways of application of the model are considered there. The main results and conclusions are presented in Chapter 9.

Chapter 2

Illustrative example: an illness contract

In Chapter 1 we briefly introduced the model which will be the main target of investigation in the present thesis. The structure of the model is based on the idea of dependence which appeared for the first time in Albers [1999]. That paper presented strong evidence that even a small amount of dependence can lead to huge risk underestimation. This finding was a strong signal to take small dependencies seriously and to continue investigation in the same direction. The structure of the model, together with its distributional assumptions, were changed and generalized several times in order to increase its practical relevance. The original form of the model (see (1.5)) for convenience uses the normal distribution to model the claim sizes. It is clear that this normality assumption is not the most realistic one for claim sizes. Nevertheless, the author clearly illustrated that even under such a 'soft' assumption, dependence has a strong influence on the resulting *SLP*'s. A more realistic assumption for the claim sizes (lognormal distribution) was used in Reijnen et al. [2005] and the dependence effect in that case was much larger.

In our research, the model has experienced several changes which are mostly related to the dependence part. It is natural to motivate each of such generalizations from real data. This is obvious, as the assumptions and structure of the model first of all should be close to reality and real data are representative of that reality. The main problem is that usually such data (containing information about the dependence structure) are unavailable. This is not a surprise since advanced modeling of the type we are considering here, is only very slowly getting introduced into the insurance practice. Data sets which for a long time have been collected by insurance companies mostly are analyzed with independence models. They typically do not (automatically) contain the additional information about the claims which is necessary for implementation of new models. Therefore, we

had to consider alternative ways in order to get an idea about certain aspects of the dependence assumptions.

This chapter is devoted to an illustrative example of a possible dependence structure. The idea was to create an algorithm which would simulate some concrete dependence pattern, illustrating the needed features which are hidden in the available data sets. Such features are mostly connected to the distributional assumptions of the random variables which appear in the model. In Model 4 the distributional forms of N and H are already settled: $N \sim P(\lambda(1 - \epsilon))$ and $H \sim P(\lambda\epsilon/\mu_G)$. Candidates for the distribution of C (and D) are presented in Section 1.2 of Chapter 1. Finally, G is assumed to be $P(L)$ with L still undetermined. All these distributional assumptions should be motivated, and, as far as they are not yet defined, some proposals should be given. Here we concentrate on the distributions of N, H and G , since these concern the new aspect of our model.

In order to get some idea about the behavior of all these random variables, we are going to simulate a flu epidemic inside a large company, considering several departments as potential places of the mutual infection. When some worker in the company is ill, the company is obliged to pay his (her) salary during the whole period of his illness. This can lead to large losses if the company is hit by an epidemic, during which a significant part of workers do not work but still get their salaries. To make a connection with insurance, we consider a Stop-Loss contract which covers that part of the risk involved which is above some retention level. The payment which people receive during their illness period can be considered as claims and the sum of all these claims then is an aggregated claim amount.

Such a structure perfectly reflects the type of dependence which is considered in the thesis. The groups of a mutual infection (people which got infection from each other) are considered as groups of a common risk and claims from people which got the infection independently or suffer from other types of illness, fall in the category of independent claims. During the simulation process, which is described in Section 2.1, we can easily collect the necessary information about the quantities of interest. Later this information is used to get an idea about the dependence pattern and possible distributional assumptions for N, H and G . The main statistical analysis is presented in Section 2.3, where several suggested distributions are fitted to the simulated data. However, before the statistical analysis can be applied, the method should pass through several sensitivity tests. It is important to show that the simulation algorithm is not extremely sensitive to the input parameters such as the illness probabilities, infection probabilities, number of days people spend at home during illness, etc. The sensitivity analysis is performed in Section 2.2. The last section summarizes the main conclusions of the chapter, which are presented as a motivation for the subsequent research.

2.1 Simulation algorithm

This section describes the algorithm which will be used later in the data simulation process. As was already mentioned, the algorithm simulates a local flu epidemic inside a company. We consider a large company with a fixed number of workers which are divided into several departments. Each department is considered as a potential place of mutual infection. Therefore, each working day of each department will be simulated separately. As the algorithm is analogous for all the departments, we consider only one of them as a representative. The remaining departments will be simulated in the same way. Each simulated (working) day of the department consists of two simulated events: ‘personal infection’ and ‘mutual infection’, as each person in the department can (with some given probability) get flu independently (personal infection) or (in case he/she is already infected) can infect someone else inside the department (mutual infection). For brevity we will from now on simply write ‘he’ where ‘he/she’ is meant. People who were infected independently fall in the category of the ‘simple’ claims and people who were infected by infected staff members form dependence groups. If a person got the infection (no matter in which way), he can infect someone else during a random number of days which he still spends at work. We assume this number to be between 0 (if he got flu at home and did not go to work at all) and 3 (if he stays infected at work for three days). After illness at work follows illness at home, which is also modeled by a random number. Yet another assumption is that after illness a person does not fall ill again for two months.

In order to complete the algorithm, values for the input parameters have to be selected. These parameters are the probability to get flu, the probability to infect someone and those characterizing illness at work (random number of days) and illness at home (random number of days). The probability to get flu independently is assumed to be different for each month. Hence, 12 probabilities are needed in total (see Table 2.1). The probabilities of mutual infection are assumed to be the

Month	Jan-Feb	Mar	Apr-Sep	Oct	Nov	Dec
Probability	0.005	0.007	0.001	0.008	0.02	0.012

Table 2.1: Monthly probabilities to get flu.

same as the probabilities to get flu independently. Values of illness at work are random integers between 0 and 3. The corresponding distribution is presented in Table 2.2. Illness at home is assumed to be a normal random variable with

Nr. of days	0	1	2	3
Probability	0.2	0.5	0.25	0.05

Table 2.2: Probabilities to stay at work for infected person.

mean 5 and variance 3. Of course, it will be rounded to an integer value after it

is generated and some obvious restrictions (at least 0 and less than 30) have to be added.

To get a better feeling for the algorithm, we present an example of one simulated day inside the department. The example utilizes several technical notations (presented by numbers), which form a state of the person. The first number is the illness state. The possible illness states are presented in Table 2.3. Additionally

State	Notation
Healthy	1
Infected at work	2
Infected at home	3
After illness	4

Table 2.3: Several states of a person

to the illness state, each person can be active (denoted by 1 and meaning that the person can infect someone) or not active (denoted by 0 and meaning that the person can not infect anybody). The activity state is presented as a second number of the state of the person. For instance the notation $[2, 1]$ means that the person is ill while at work (number 2) and has a possibility to infect someone inside the department (number 1). There is a possibility of the third number, which is a group state, meaning that the person was infected by his colleague inside the department and hence is a member of some dependence group. The value of the group state corresponds to the ordinal number of the dependence group, which has been subsequently formed during the reference period. For instance the notation $[2, 0, 5]$ means that the person is ill while at work (number 2), does not have a possibility to infect anyone (number 0) and corresponds to the group number five, which is the fifth group, starting from the beginning of the reference period. If no group number is mentioned, the person does not belong to any dependence group. We call them individuals. Thus, the last dependence group contains the information about the total number of groups during the reference period. Note that in case the person was infected by an individual, they both get the group state which is increased by one, compared to the total number of the dependence groups which have been formed up to that moment. However, if the person was infected by the member of some dependence group, he automatically joins this group by getting the same group state. Therefore, a new group can be started only by the person who was infected independently.

We consider a department with five workers and simulate one working day in three steps. The remaining days are included for illustrative purposes.

Step 1: Initial state of the algorithm where all 5 workers are set to be healthy and hence not active. Obviously none of them belongs to any dependence group, therefore the group state is absent here (see Table 2.4).

Step 2: Personal infection, where each person in the company 'has a chance'

Person	Day 1	Day2	Day 3	Day 4	Day 5	Day 6	Day 7
1	[1,0]	[1,0]	[1,0]	[1,0]	[1,0]	[1,0]	[1,0]
2	[1,0]	[1,0]	[1,0]	[1,0]	[1,0]	[1,0]	[1,0]
3	[1,0]	[1,0]	[1,0]	[1,0]	[1,0]	[1,0]	[1,0]
4	[1,0]	[1,0]	[1,0]	[1,0]	[1,0]	[1,0]	[1,0]
5	[1,0]	[1,0]	[1,0]	[1,0]	[1,0]	[1,0]	[1,0]

Table 2.4: Step 1: Initial state of the algorithm.

to get flu individually. A possible realization of that step is presented in Table 2.5, where two persons (number 2 and number 5) got the infection individually during the first day. One of them (number 2) will stay two days

Person	Day 1	Day2	Day 3	Day 4	Day 5	Day 6	Day 7
1	[1,0]	[1,0]	[1,0]	[1,0]	[1,0]	[1,0]	[1,0]
2	[2,1]	[2,1]	[3,0]	[3,0]	[3,0]	[3,0]	[4,0]
3	[1,0]	[1,0]	[1,0]	[1,0]	[1,0]	[1,0]	[1,0]
4	[1,0]	[1,0]	[1,0]	[1,0]	[1,0]	[1,0]	[1,0]
5	[2,1]	[3,0]	[3,0]	[3,0]	[4,0]	[4,0]	[4,0]

Table 2.5: Step 2: Personal infection.

infected at work (and is active) and will spend four days at home (where he obviously is not active). After his illness he cannot be infected during two months. The second one (number 5) will stay one day infected at work (and is active), three days at home (not active) and two months at work without a possibility to get the infection.

Step 3: Mutual infection, where each person who stays infected at work and is active, 'has a chance' to infect each person in the department (during the first day) with a state [1,0] (healthy and not active). After the person has

Person	Day 1	Day2	Day 3	Day 4	Day 5	Day 6	Day 7
1	[2,1,1]	[2,1,1]	[2,1,1]	[3,0,1]	[3,0,1]	[3,0,1]	[4,0]
2	[2,0,1]	[2,1,1]	[3,0,1]	[3,0,1]	[3,0,1]	[3,0,1]	[4,0]
3	[1,0]	[1,0]	[1,0]	[1,0]	[1,0]	[1,0]	[1,0]
4	[2,1,2]	[3,0,2]	[3,0,2]	[3,0,2]	[3,0,2]	[3,0,2]	[4,0]
5	[2,0,2]	[3,0,2]	[3,0,2]	[3,0,2]	[4,0]	[4,0]	[4,0]

Table 2.6: Step 3: Mutual infection.

used all his infection chances, his activity state is changed to 0. Table 2.6 presents a possible realization of the mutual infection where person 2, who

Person	Day 1	Day2	Day 3	Day 4	Day 5	Day 6	Day 7
1	[2,0,1]	[2,1,1]	[2,1,1]	[3,0,1]	[3,0,1]	[3,0,1]	[4,0]
2	[2,0,1]	[2,1,1]	[3,0,1]	[3,0,1]	[3,0,1]	[3,0,1]	[4,0]
3	[1,0]	[1,0]	[1,0]	[1,0]	[1,0]	[1,0]	[1,0]
4	[2,0,2]	[3,0,2]	[3,0,2]	[3,0,2]	[3,0,2]	[3,0,2]	[4,0]
5	[2,0,2]	[3,0,2]	[3,0,2]	[3,0,2]	[4,0]	[4,0]	[4,0]

Table 2.7: Mutual infection.

got an infection independently, has infected person 1. Both of them got the group state 1, i.e. the first dependence group which has been formed during the reference period. The second person (number 5) has infected the person with number 4 and both of them got the group state 2 (the second dependence group which was formed during the reference period). The persons with numbers 2 and 5 have used their infection chances and hence their activity states are changed from 1 to 0. However, the mutual infection step is not finished yet. Two persons (numbers 4 and 1) which got the infection from their colleagues are still active. They also have a chance to infect each person in the department during the first day. Table 2.7 presents a possible realization of an such event. From the table it is clear that the last person (number 3), who 'had a chance' to get the infection, is healthy. Numbers 4 and 1 have not infected him and became not active.

Step 3 was repeated until all active persons became not active. This is a signal for the algorithm to go to the next day, where steps 1, 2, and 3 are repeated. It can be noted that next day some persons already stay infected at work. Some of them are individuals, some belong to dependence groups. Figure 2.1 presents single realizations of the simulation process when the number of workers in the department is 70 and the number of simulated days is 360. The green color in the picture means that a person is healthy, the blue color represents individuals and the red one illustrates the dependence groups. It is obvious that such a picture is quite pessimistic. In real life the probability of such an outcome is very small (that is why we had to perform a lot of simulations to get such a 'bad' year). Nevertheless, this is only an illustrative example which points out the main message that possible dependencies between the risks could lead to risk underestimation (in November and December (the last 60 days) the red color clearly dominates over the blue one).

Figure 2.1 does not point out the different dependence groups. All the groups of common risk are presented by the red color. Nevertheless, we can determine and calculate the number and sizes of such groups during the simulation process. The number of individuals can be calculated as well. We denote by u the number of simulated years. For each simulated year $t = 1, \dots, u$ we get a single realization of N , denoted by n_t , a single realization of H , denoted by h_t and several realizations for the group sizes G_k , denoted by g_{k_t} , $k_t = 1, \dots, h_t$ (separate realization for

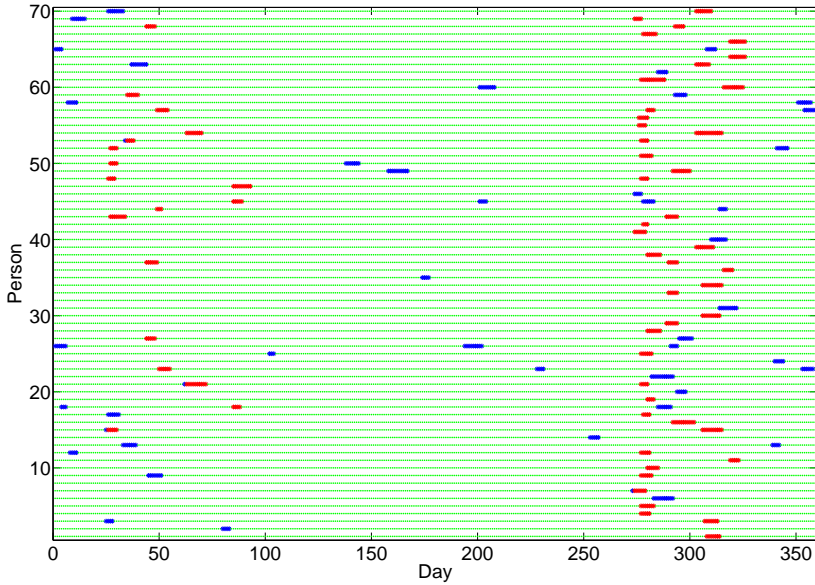


Figure 2.1: Example of local epidemic

each G_k). Precisely these data (n_t , h_t and g_{k_t}) will be analyzed in the later sections of the present chapter.

2.2 Sensitivity analysis

As was mentioned before, this chapter deals with the motivation aspects for the setup of Model 4. In Model 4 (see (1.9)) the distributional forms of N and H are already settled: $N \sim P(\lambda(1 - \epsilon))$ and $H \sim P(\lambda\epsilon/\mu_G)$. Candidates for the distribution of C (and D) are presented in Section 1.2 of Chapter 1. Finally, G is assumed to be $P(L)$ with L still undetermined. All these distributional assumptions should be motivated, and, as far as they are not yet defined, some proposals should be given. Here we concentrate on the distributions of N , H and G , since these concern the new aspect of our model. All the motivations for the distributional assumptions of the N , H and G will be given on the basis of the statistical analysis of the simulated data n_t , h_t and g_{k_t} , where t runs from 1 to u , the number of simulated years. The data set $\{n_t, h_t, g_{k_t}\}$ is an outcome of the simulation algorithm which was explained in detail in Section 2.1. It is obvious that the output of the algorithm depends on the input parameters which were also described in Section 2.1. The influence of the input parameters is very hard

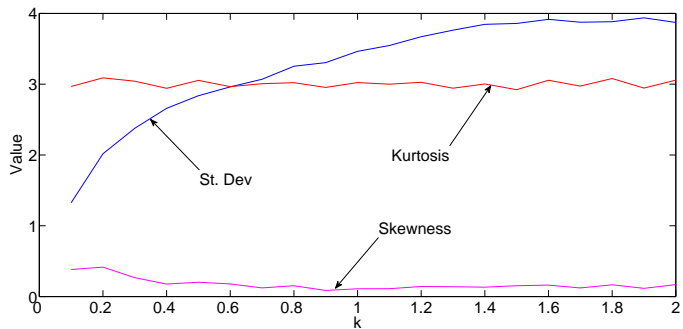
to see or predict in advance. However, this influence is important as we are going to consider the outcome as a possible realization of reality. Should this output turn out to be very sensitive to the input parameters, we cannot rely very well on the results of the statistical analysis, since the input is created by ourselves and does not represent the reality in all detail. Consequently, the algorithm should pass several sensitivity tests, in order to illustrate that general conclusions which are based on its output, are stable and not extremely sensitive to the input.

The main targets of interest are the distributional assumptions of the N , H and G . Therefore, the sensitivity tests which we mentioned will be performed on the basis of these distributions and their main characteristics. The main idea is to check how the distributions and their main characteristics react to different (small) changes of the input parameters. All input parameters are summarized in Table 2.8. The number of values of the input parameters is large and it is quite hard to analyze all possible combinations and variations. Therefore, we will consider only several combinations which can be considered as representative for all reasonable ones. Each of the five assumptions will be considered separately. All the fluctuations of the first set of parameters (probabilities to get an infection) for simplicity will be proportional. This means that all the values in Table 2.1 increase (and decrease) together, proportionally to some parameter k , which will be a multiplier of these values. The values of k will run from 0.1 till 2 with stepsize 0.1. In case $k \in [0.1, 1)$, the considered values are smaller than the originals, in case $k \in (1, 2]$ they are larger and $k = 1$ obviously represents the original values. To start the simulation process we consider a large company of 15 departments with 50 employees each (thus, 750 employees in total). The number of simulated years u is fixed to be 1000. For each value of k we have three samples: n_t , h_t and $g_{k_t t}$, with $t = 1, \dots, u$ and $k_t = 1, \dots, h_t$. Hence, for instance for N we have 20 samples of each 1000 simulated observations. For H and G the situation is similar. To compare the samples, we calculate their characteristics: standard deviation, skewness and kurtosis. Note that the definition of kurtosis which is used in this chapter does not subtract 3 from the result, i.e. the kurtosis here is defined as μ_4/μ_2^2 , where μ_4 and μ_2 are the fourth and second central moments respectively. Later in the thesis we will use the definition of kurtosis where 3 is subtracted.

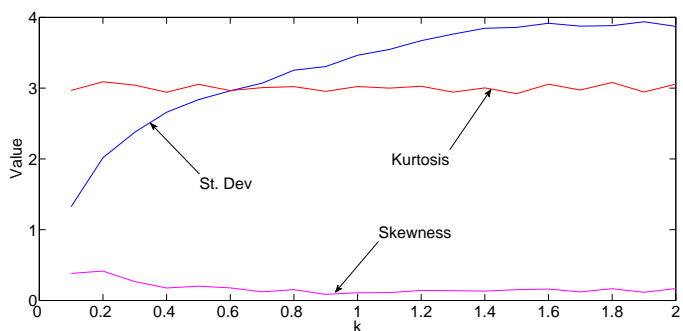
Assumption	Values
Probability to get an infection	Table 2.1
Probability to stay sick at work	Table 2.2
Nr. of days which infected person spend at work	Integers, from 0 to 3
Nr. of days which infected person spend at home	$\text{int}(\min(N(5, 3) , 30))$

Table 2.8: Summary of the input parameters.

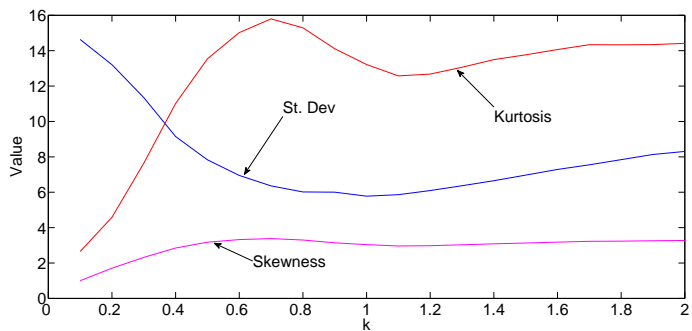
Figure 2.2 represents the characteristics of all the samples for different values of k . So, for instance the kurtosis of N when $k = 0.2$ is based on 1000 simulated observations of N .



(a) Characteristics of N samples



(b) Characteristics of H samples



(c) Characteristics of G samples

Figure 2.2: Characteristics of the simulated samples.

First consider the characteristics of the samples of N (Figure 2.2(a)). The obvious conclusion which can be drawn directly is that skewness and kurtosis are not sensitive to the fluctuation of illness probabilities. Moreover, the skewness is very close to zero while the kurtosis is very close to 3 in all the samples. Exactly the same situation we have while considering the samples of H (see Figure 2.2(b)). The standard deviation is increasing with k in both cases, which is natural since larger probabilities to get flu imply larger variation of individuals, and hence larger variation of the number of groups.

Now consider the characteristics of the samples of G (see Figure 2.2(c)). Here the situation is slightly different. The skewness and kurtosis are not so stable (compared to the previous samples) when k is smaller than 1. The reason for this could be the small sample sizes. When the illness probabilities are small (k smaller than 1), we need more simulations to 'catch' the heavy tail of the underlying distribution. That is why the kurtosis increases very fast at the beginning and stabilizes when k becomes larger. The situation with the skewness is similar, but not as extreme as in the kurtosis case.

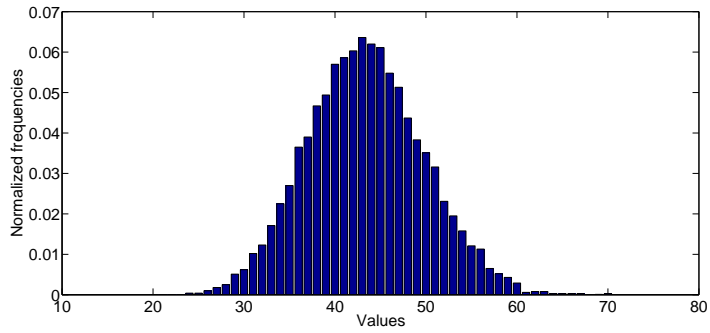
Besides the sample characteristics, we also look more closely at the shapes of the underlying densities, which can be estimated by histograms. Figure 2.3 presents the histograms of n_t , h_t and $g_{k,t}$, when the value of k is equal to 1. These three histograms offer typical examples for the shapes of the underlying densities. Consequently, histograms for the remaining values of k are not presented. These are completely similar and can be predicted easily, using the main characteristics of the samples. The important conclusion here is that the densities do not dramatically change in shape. Hence, we can conclude that in this respect there is little sensitivity to fluctuation of the illness probabilities.

The fluctuations of the second set of parameters (Table 2.2) are impossible to analyze using proportionality since the sum of all the probabilities has to be equal to 1. Therefore, we consider several cases with different representative probabilities. Six different situations are presented in Table 2.9. For each situation

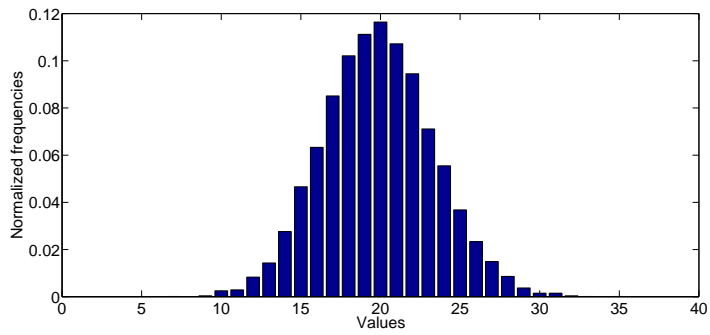
Nr. of days	0	1	2	3
Original prob.	0.2	0.5	0.25	0.05
Situation 1	0.6	0.02	0.1	0.1
Situation 2	0.1	0.3	0.4	0.2
Situation 3	0.1	0.2	0.3	0.4
Situation 4	0.05	0.9	0.03	0.02
Situation 5	0.05	0.03	0.9	0.02
Situation 6	0.9	0.05	0.03	0.02

Table 2.9: Different cases for illness at work probabilities.

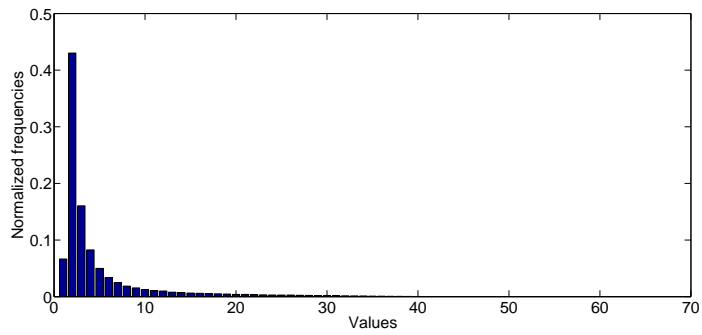
we have performed the same analysis as was done in the proportional variation case. However, we do not present all the results here since they are very close to the ones we got while analyzing the fluctuations of the illness probabilities. The



(a) Histogram of N sample



(b) Histogram of H sample



(c) Histogram of G sample

Figure 2.3: Histograms of the simulated samples.

Situation	Description
1	Proportional variation of illness probabilities when $k = 0.1$
2	Proportional variation of illness probabilities when $k = 1$
3	Proportional variation of illness probabilities when $k = 2$

Table 2.10: Representative situations.

same situation holds with the remaining input parameters, which are the number of days an ill person spends at work and at home. Adding or removing a couple of days does not change the general pattern. The pictures of the characteristics and the histograms stay very close to Figures 2.2 and 2.3. Therefore, the main conclusions which were based on these figures, can be extended for the remaining parameters as well.

Summarizing all the results, we conclude that the algorithm is sufficiently stable for our purposes. The main sample characteristics, together with the shapes of the histograms, are not extremely sensitive to the input. This means that general conclusions about the distributional families of N , H and G (which will be presented in the next section) will remain similar and hence precise knowledge – which we clearly do not have – about the values of the input parameters in reality fortunately is not needed.

2.3 Fitting simulated data

In this section we will try to fit the data which were simulated by the algorithm. As was already mentioned, we consider a large company of 15 departments with 50 employees each (thus, 750 employees in total) and the number of years u (which can be considered as the number of simulations) is fixed to be 1000. In the previous section the algorithm was simulated several times, using different sets of input parameters. For simplicity (and also to save time) in this section the same data sets will be used which were utilized in the sensitivity analysis. For each data set considered we will try to find a distribution which fits the data. We will also utilize the results of the previous section, where the data samples were analyzed nonparametrically. This will help to start the fitting procedure, to determine reasonable candidates among the huge number of available distributions. For instance, we already saw that the distributional shapes of N and H are completely different from the shape of the distribution of G (see Figure 2.3). This fact certainly gives us a hint that these random variables belong to different distributional families.

Since the output of the algorithm is not extremely sensitive to the input, there is no need to include all the data sets (which were considered in Section 2.2) to the fitting procedure. Three situations are chosen to be representative of all possible combinations (see Table 2.10).

2.3.1 Fitting N and H

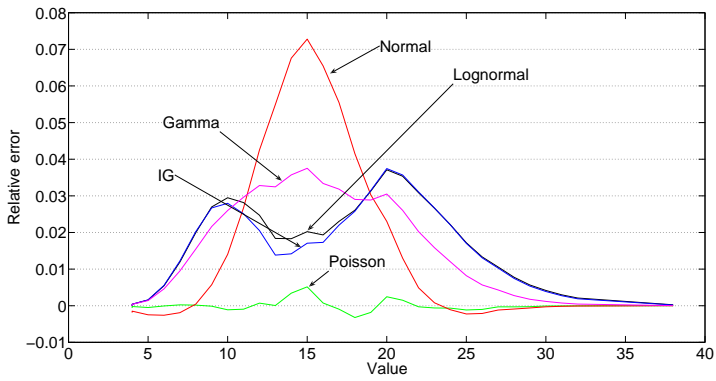
We will start with fitting the simulated data samples n_t and h_t , where $t = 1, \dots, u$. In the previous section several characteristics of these samples were already obtained. Figure 2.2 illustrates that the sample skewness is very close to zero which means that the shape of the distribution is likely to be quite symmetric. The kurtosis, which represents the thickness of the tail of the distribution, is close to 3. These two characteristics give us an idea about reasonable candidates for the fitting procedure. The first candidate is the normal distribution, which has exactly the same characteristics. The second candidate is the Poisson distribution. Skewness and kurtosis of the Poisson distribution are not 0 and 3, but they are very close to these values when the parameter of the distribution is large. Three more candidates considered are the gamma, lognormal and Inverse Gaussian (IG) distributions. According to the information we have so far, these three distributions are not the best candidates. They are not symmetric, which is not in line with the representative histograms 2.3(a) and 2.3(b). However, we include these three distributions for illustration.

After choosing the candidates, we apply the maximum likelihood method for the estimation of the parameters. Using these estimated values, the distributions are fitted to the data. To choose the best distribution out of the five considered, we compare them with the empirical distributions of the underlying data sets n_t and h_t . In each value of the empirical distribution x_i , we calculate a relative error

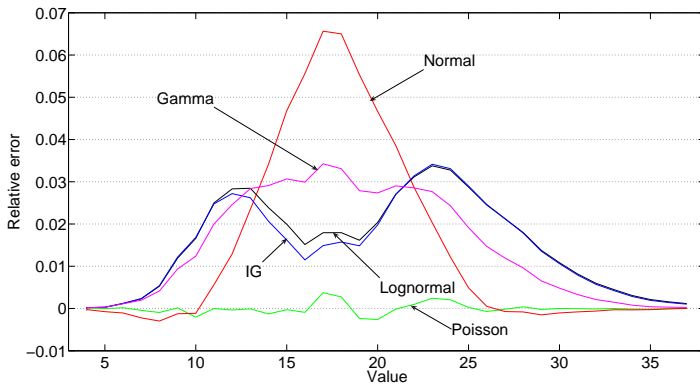
$$\frac{F(x_i) - F_e(x_i)}{F(x_i)},$$

where F_e denotes the empirical distribution and F the candidate distribution. The number of values which are used is not defined in advance since it depends on the simulated data. Basically, the x_i 's are all the different values in the simulated sample, starting from the minimum value and finishing with the maximum one. Figure 2.4 presents the relative errors for the three n_t samples, which are described in Table 2.10. From a first glance it is seen that the relative errors are rather small, at most 0.07. Moreover, it is clear that the Poisson distribution is closer to the empirical distribution almost everywhere. The normal distribution, which was the first candidate due to its skewness and kurtosis characteristics, shows a relatively bad fit in the middle, where the values are close to the mean. Note that the high and low values are fitted well by the normal distribution. As was expected, the gamma, lognormal and IG distributions show a relatively poor fit on the whole range of the values considered. As a result, we conclude that the Poisson distribution fits the data very well. The relative errors are very small on the whole range of the values considered.

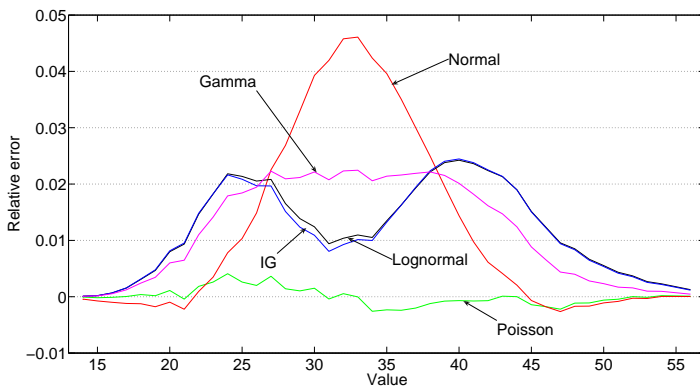
The same procedure, with the same distributional candidates, was performed for the h_t samples. We do not present all the results since they are very close to the results of the analysis of the n_t samples (see Figure 2.4). This is not a surprise since Figure 2.2(b) is very similar to Figure 2.2(a) which is an indication



(a) Situation 1



(b) Situation 2



(c) Situation 3

Figure 2.4: Results of n_t data fitting.

(but certainly not a proof) that these samples come from the same (or a similar) distributional family. The extended analysis shows that this prediction is not far from reality. Therefore, the conclusion about the distribution which fits the data h_t best is the same as the conclusion in the n_t fitting case. The Poisson distribution shows the best (and very good) results in both cases.

2.3.2 Fitting G

Now consider the last set of samples, which are those of the group sizes $g_{k,t}$. By looking to Figure 2.2(c) we see that the main characteristics of the $g_{k,t}$'s differ from the sample characteristics of n_t and h_t . The skewness is not 0, but varying between 3 and 4 and the value of the kurtosis is even higher, reaching 14 when the illness probabilities are high. This suggests that the underlying distribution is not symmetric and has a heavy tail. Additional proof for such a conclusion we get by considering Figure 2.3(c), where the histogram clearly illustrates the asymmetry and heavytailedness. Candidates considered which satisfy the above mentioned requirements are the lognormal, IG , negative binomial and gamma. Of course, in the literature many more distributions with such characteristics exist, but the present ones are the most common. One more distribution which we add for illustrative purposes is the Poisson. Following the same strategy, results of the fitting procedure are presented as a set of relative errors. Like in the previous case, the empirical distribution is compared with all the candidates. Figure 2.5 presents the results when situation 1 (see Table 2.10) is assumed. We do not present the results for the remaining situations since they are very similar to the one considered. From the figure it is clear that the situation is not so straightforward as the one which is presented by Figure 2.4. Among all the candidates we do not have a single leader now. All the candidates (except the Poisson distribution) are very close to each other and it is very hard to distinguish the best one. Additionally it can be noticed that the relative errors are much larger, compared to the n_t fitting case.

Since all candidates are close to each other and the relative errors much larger (compared to the n_t fitting case), the question arises whether they are all reasonable ones. A precise answer would require some criterion of accuracy to be introduced. However, since this is not the main goal of the thesis, we shall not pursue this and simply conclude on qualitative grounds that the accuracy achieved is sufficient for our purposes.

To choose the distribution from the four presented ones, we look more closely at Model 4 (see (1.9)). The random variables G_k , which represent the group sizes, appear in the upper summation index of the second part of the model. It was already mentioned that the general form of G is assumed to be $P(L)$, where $P(\cdot)$ denotes the Poisson distribution and L is a random variable (not yet defined). Therefore, the distribution of G is basically defined by the distribution of L . It is well known that in case L follows gamma distribution, the resulting distribution of G will be negative binomial, which is present among the reasonable candidates for

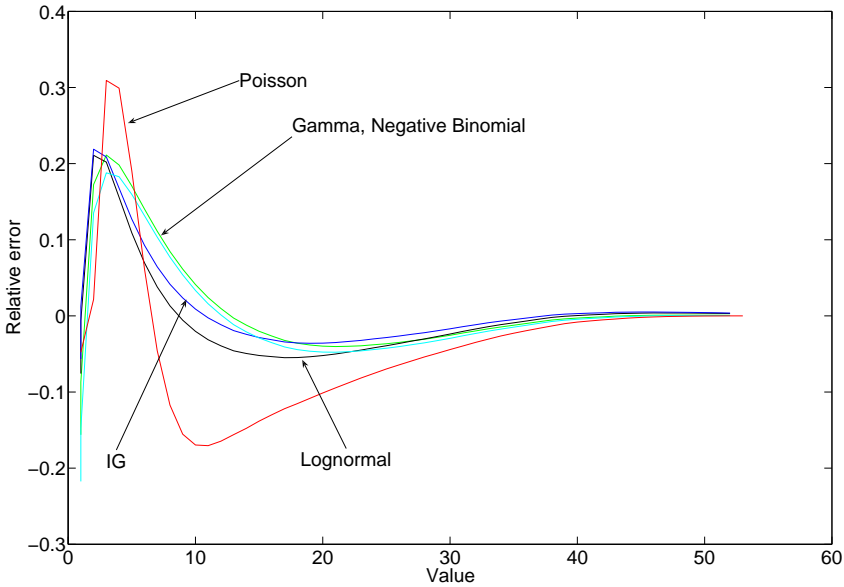


Figure 2.5: Results of $g_{k_{tt}}$ data fitting.

the distribution of G . In this respect, the negative binomial distribution offers an advantage over the remaining candidates. Nevertheless, this does not mean that the negative binomial will be the only distribution considered. More candidates for the distribution of L will be presented in the chapters that follow.

2.4 Conclusions

The main goal of the chapter was to introduce a motivation for the distributional assumptions which are used in Model 4. The main targets of the consideration were the distributions of N , H and G (all the G_k 's are assumed to be identically distributed as G). Representative samples of these random variables were obtained by simulating the algorithm, which was explained in Section 2.1. The obtained samples were analyzed and possible candidates were suggested for each of the random variables considered. All the candidates were fitted to the data, using the maximum likelihood method for parameter estimation. The Poisson distribution showed a very good fit while the samples from N and H were considered. However, the analysis of the samples from G did not point out a leader among the candidates considered. Nevertheless, several important conclusions about the distribution of G were obtained while analyzing the data. These are the positive

skewness (between 3 and 4) and large kurtosis (about 14). The negative binomial distribution was suggested for the distributional assumption of G in view of some of its properties which significantly simplify application of the model.

Chapter 3

The Model

In the Introduction we presented the hierarchical development and a brief description of the models, starting from the independence Model 1 (see (1.4)) and ending with the dependence Model 4 (see (1.9)). All the models, together with their basic assumptions, are summarized in Table 3.1. This chapter deals with

Number	Model	Assumptions
1	$S = \sum_{i=1}^N C_i$	$N \sim P(\lambda)$
2	$S = \sum_{i=1}^N C_i + \sum_{k=1}^H \sum_{j=1}^g D_{jk}$	$N \sim P(\lambda(1 - \epsilon)), H \sim P(\epsilon \frac{\lambda}{g})$
3	$S = \sum_{i=1}^N C_i + \sum_{k=1}^H \sum_{j=1}^{G_k} D_{jk}$	$N \sim P(\lambda(1 - \epsilon)), H \sim P(\epsilon \frac{\lambda}{\mu_G}),$ $G_k \sim P(\mu_G)$ for all k
4	$S = \sum_{i=1}^N C_i + \sum_{k=1}^H \sum_{j=1}^{G_k} D_{jk}$	$N \sim P(\lambda(1 - \epsilon)), H \sim P(\epsilon \frac{\lambda}{\mu_G}),$ $G_k \sim P(L_k)$ for all $k, \mu_L = \mu_G$

Table 3.1: Hierarchical development of the models.

Model 4, the most general version of the dependence model we have considered in the present research. In Chapter 1 the main assumptions (which were further motivated in Chapter 2) and the structure of the model were discussed. Here we will describe Model 4 in more detail, present its main characteristics, discuss the impact of the model parameters and describe the region of interest for the values of the parameters involved. Concerning Model 4, the structure of which is presented in Table 3.1, we remind that all L_k are iid and consequently we can

use L as a representative for all the L_k 's. The same situation holds with G for G_k , C for C_i and D_{jk} (by default (in case nothing is mentioned) all the D_{jk} 's for simplicity are assumed to have the same distribution as the C_i 's).

The difference between Model 4 and Models 2 and 3 is the assumption about the group size G . In Model 3 G was assumed to follow a simple Poisson distribution, in Model 2 the group sizes g were assumed to be fixed, while in Model 4 G has a mixture distribution $P(L)$, where $P(\cdot)$ stands for the Poisson distribution and L is stochastic.

The main reason for such a generalization is hidden in the underlying structure of the data which we need to model. We consider a quite high aggregation level, which definitely makes the underlying data more heterogeneous. An example of such heterogeneity can be the different departments (or even different companies) which form the portfolio considered and divide it into a number of (hopefully homogeneous) parts. Each part has its own specific features (like different expectations of the group sizes), which could have a strong impact on the model assumptions. This overdispersion suggests not to model all the group sizes of the underlying portfolio by one single Poisson distribution. Following standard statistical methodology the overdispersion is modeled by the mixture distribution $P(L)$. A demonstration of that fact we got by simulating the flu epidemic inside a company in Chapter 2. The simulations results showed that the number of individuals N (as well as the number of groups H) can be modeled by the simple Poisson distribution, while the group sizes definitely needed a more heavy tailed distribution.

Using these arguments we came up with the general form of Model 4, which uses more advanced group size modeling. The group sizes are assumed to follow the Poisson distribution with a random parameter, which represents the random group size expectation (each homogeneous part has its own group size expectation). This kind of modeling allows us to overcome the heterogeneity problem.

3.1 Main characteristics of the model

Considering Model 4 we see that the form of the distribution of S will be very complicated, which makes it very hard to analyze its behavior. An even more complicated task would be to analyze the behavior of the distribution of certain functions of S , which is one of the goals of the thesis. This fact was already mentioned in the Introduction, where it was suggested to use an approximation instead of the real distribution.

The basic features which characterize the distribution of S are its mean μ_S , its standard deviation σ_S and its standardized cumulants (the third and fourth are known as skewness and kurtosis). Having such characteristics of the real distribution of S , we can locally approximate its behavior by matching these with the standardized cumulants of some other distribution of a more simple form. The approximation aspect will be covered in detail in Chapters 4 and 5.

In the present section we derive the analytical expressions for the mean μ_S , the standard deviation σ_S and the third and fourth standardized cumulants of S , which will be used later in the approximation formulas. Clearly, the advantage of such explicit expressions will be that the influence of the various components in the model on the moments and cumulants in question is made transparent. First we introduce the notation required. For a single rv X use μ for its mean and σ^2 for its variance. Moreover, for general (central) moments use $\nu_j = E[X^j]$ and $\mu_j = E[(X - \mu)^j]$, $j = 1, 2, \dots$ (hence in particular $\nu_1 = \mu$ and $\mu_2 = \sigma^2$). In addition, let $\gamma = \sigma/\mu$ denote the coefficient of variation (cv). If more than one random variable is involved, the respective indices are added, e.g. μ_X, ν_{3X} , etc. Next, the moment generating function $E[e^{tX}]$ will be denoted by $M(t)$; as is well-known, $\nu_j = M^{(j)}(0)$, $j = 1, 2, \dots$. Similarly, let $K^*(t) = \log M(t)$ denote the cumulant generating function, then the j^{th} cumulant $\kappa_j^* = K^{*(j)}(0)$. In particular, we have that $\kappa_1^* = \mu$, $\kappa_2^* = \sigma^2$, $\kappa_3^* = \mu_3$, $\kappa_4^* = \mu_4 - 3\sigma^4$. Moreover, let K be the corresponding function for the standardized rv $(X - \mu)/\sigma$, then the standardized cumulants κ_j satisfy $\kappa_1 = 0$, $\kappa_2 = 1$, $\kappa_3 = \mu_3/\sigma^3$, $\kappa_4 = \mu_4/\sigma^4 - 3$. Using these notations we can write that

$$\mu_S = \kappa_{1S}^*, \quad \sigma_S^2 = \kappa_{2S}^*, \quad \kappa_{3S} = \frac{\kappa_{3S}^*}{\sigma_S^3}, \quad \kappa_{4S} = \frac{\kappa_{4S}^*}{\sigma_S^4}. \quad (3.1)$$

Therefore, it is very convenient to calculate the mean, variance, skewness and kurtosis of S by using the cumulants of S , which, in their turn, can be expressed as derivatives of the cumulant generating function of S at the point 0. The following lemmas present the expressions for the mean μ_S , the standard deviation σ_S and the third and fourth cumulants of S under the different model assumptions.

Lemma 3.1. *Suppose that Model 4 holds and G, C and D have arbitrary distributions. Then $\mu_S = (1 - \epsilon)\lambda\mu_C + \epsilon\lambda\mu_D$, and*

$$\begin{aligned} \sigma_S^2 &= (1 - \epsilon)\lambda\nu_{2C} + (\epsilon\lambda/\mu_G)\{\nu_{2G}\mu_D^2 + \mu_G\sigma_D^2\}, \\ \kappa_{3S}^* &= (1 - \epsilon)\lambda\nu_{3C} + (\epsilon\lambda/\mu_G)\{\nu_{3G}\mu_D^3 + 3\nu_{2G}\mu_D\sigma_D^2 + \mu_G\kappa_{3D}^*\}, \\ \kappa_{4S}^* &= (1 - \epsilon)\lambda\nu_{4C} + (\epsilon\lambda/\mu_G)\{\nu_{4G}\mu_D^4 + 6\nu_{3G}\mu_D^2\sigma_D^2 + \nu_{2G}(4\mu_D\kappa_{3D}^* + 3\sigma_D^4) + \mu_G\kappa_{4D}^*\}. \end{aligned} \quad (3.2)$$

Proof. Using the definition of Model 4 and the properties of the cumulant generating function we directly have

$$K_S^* = K_N^* \circ K_C^* + K_H^* \circ K_G^* \circ K_D^*. \quad (3.3)$$

As a random variable X which is $Poisson(\theta)$ has $K_X^*(t) = \theta(e^t - 1)$, it follows that (noting that N and H are Poisson)

$$K_S^*(t) = (1 - \epsilon)\lambda\{M_C(t) - 1\} + (\epsilon\lambda/\mu_G)\{M_G(K_D^*(t)) - 1\}. \quad (3.4)$$

The result in (3.2) readily follows from $\kappa_{jS}^* = K_S^{*(j)}(0)$. □

The expressions in (3.2) and (3.4) are already reasonably tractable. By way of illustration, we shall consider some examples. If G is fixed, i.e. $G = \mu_G$ (cf. Reijnen et al. [2005]), $M_G(K_D^*(t))$ in (3.4) boils down to $\{M_D(t)\}^{\mu_G}$ and the ν_{jG} in (3.2) to μ_G^j . When G is *binomial*(g, r) instead (cf. Albers [1999]), we get $M_G(K_D^*(t)) = \{1+r(M_D(t)-1)\}^{\mu_G/r}$ in (3.4). Supposing that μ_G is substantially larger than 1 - which clearly is the case of interest - the ν_{jG} in (3.2) are dominated by μ_G^j (e.g. $\nu_{2G} = \mu_G^2 + \mu_G(1-r)$). Obviously, for $r = 1$, we are back in the fixed case, while for $r \rightarrow 0$ (and $g = \mu_G/r \rightarrow \infty$) we get the Poisson case. In case G is *Poisson*(μ_G), we have $M_G(K_D^*(t)) = \exp\{\mu_G(M_D(t)-1)\}$, while the ν_{jG} in (3.2) are still mainly characterized by μ_G^j (e.g. here $\nu_{2G} = \mu_G^2 + \mu_G$).

As these examples demonstrate, for specific choices of G , we get reasonably explicit results by using (3.2) and (3.4). It just remains to make a choice as well for C and D in that case. As a next step we introduce the overdispersion aspect which was discussed at the beginning of this chapter (see also Table 3.1).

Lemma 3.2. *Suppose that Model 4 holds, C, D and L have arbitrary distributions and the group sizes are distributed as $G \sim \text{Poisson}(L)$. Then $\mu_S = (1-\epsilon)\lambda\mu_C + \epsilon\lambda\mu_D$, and*

$$\begin{aligned} \sigma_S^2 &= (1-\epsilon)\lambda\nu_{2C} + (\epsilon\lambda/\mu_G)\{\nu_{2L}\mu_D^2 + \mu_G\nu_{2D}\}, \\ \kappa_{3S}^* &= (1-\epsilon)\lambda\nu_{3C} + (\epsilon\lambda/\mu_G)\{\nu_{3L}\mu_D^3 + 3\nu_{2L}\mu_D\nu_{2D} + \mu_G\nu_{3D}\}, \\ \kappa_{4S}^* &= (1-\epsilon)\lambda\nu_{4C} + (\epsilon\lambda/\mu_G)\{\nu_{4L}\mu_D^4 + 6\nu_{3L}\mu_D^2\nu_{2D} + \nu_{2L}(4\mu_D\nu_{3D} + 3\nu_{2D}^2) + \mu_G\nu_{4D}\}. \end{aligned} \quad (3.5)$$

Proof. Using the fact that $G|L = l$ is Poisson, we obtain that $E[e^{tG}|L = l] = \exp(l(e^t-1))$. From this result it directly follows that $E[e^{tG}] = E[\exp(L(e^t-1))] = M_L(e^t-1)$. By using these arguments we arrive at the identity

$$K_G^*(t) = K_L^*(e^t-1). \quad (3.6)$$

Using this result and the fact that $K^*(t) = \log M(t)$, the general expression of the cumulant generating function of S takes the form

$$\begin{aligned} K_S^*(t) &= \lambda(1-\epsilon)(M_C(t)-1) + \epsilon\frac{\lambda}{\mu_G}(M_L(e^{K_D^*(t)}-1)-1) \\ &= \lambda(1-\epsilon)(M_C(t)-1) + \epsilon\frac{\lambda}{\mu_G}(M_L(M_D(t)-1)-1) \\ &= \lambda(1-\epsilon)Q_C(t) + \epsilon\frac{\lambda}{\mu_G}(Q_L \circ Q_D(t)), \end{aligned} \quad (3.7)$$

where $Q(t) = M(t) - 1$.

To get the final result we have to differentiate (3.7). The general form of this derivative can be written as

$$\begin{aligned} K_S^{*(j)}(t) &= \lambda(1-\epsilon)Q_C^{(j)}(t) + \epsilon\frac{\lambda}{\mu_G}((Q_L \circ Q_D)^{(j)}(t)) \\ &= \lambda Q_C^{(j)}(t) \left(1 + \epsilon \left[\frac{(Q_L \circ Q_D)^{(j)}(t)}{\mu_G Q_C^{(j)}(t)} - 1 \right] \right). \end{aligned}$$

To obtain (3.5) we evaluate this derivative at 0 for $j = 1, 2, 3, 4$. \square

A final step in reduction of complexity is obtained by letting $C \triangleq D$ (the notation ' \triangleq ' means equality in distribution) as was announced at the beginning of the chapter.

Lemma 3.3. *If in addition to the assumptions of Lemma 3.2, we let $C \triangleq D$, the following general expression holds*

$$\kappa_{jS}^* = \lambda Q_C^{(j)}(0) \left(1 + \epsilon \left[\frac{(Q_L \circ Q_C)^{(j)}(0)}{\mu_G Q_C^{(j)}(0)} - 1 \right] \right). \quad (3.8)$$

In particular, (3.5) can be replaced by $\mu_S = \lambda \mu_C$ and

$$\begin{aligned} \sigma_S^2 &= \lambda \nu_{2C} \{1 + \epsilon \nu_{2L} \mu_C^2 / (\mu_G \nu_{2C})\}, \\ \kappa_{3S}^* &= \lambda \nu_{3C} \{1 + \epsilon (\nu_{3L} \mu_C^3 + 3 \nu_{2L} \mu_C \nu_{2C}) / (\mu_G \nu_{3C})\}, \\ \kappa_{4S}^* &= \lambda \nu_{4C} \{1 + \epsilon (\nu_{4L} \mu_C^4 + 6 \nu_{3L} \mu_C^2 \nu_{2C} + \nu_{2L} [4 \mu_C \nu_{3C} + 3 \nu_{2C}^2]) / (\mu_G \nu_{4C})\}. \end{aligned} \quad (3.9)$$

Proof. The proof follows directly from Lemma 3.2, applying $C \triangleq D$. \square

3.2 Suggested distributional assumptions

Most of the distributional assumptions of Model 4 are already fixed. The number of individual claims N and the number of groups H have a Poisson distribution, the probability mass function of which is denoted by

$$f_P(x) = \frac{\alpha^x e^{-\alpha}}{x!}, \quad (3.10)$$

where α is a fixed parameter, which in our case is equal to $\lambda(1 - \epsilon)$ for N and $\epsilon\lambda/\mu_G$ for H . For the motivation of this choice we refer to Chapters 1 and 2. The distribution of the group sizes G is also assumed to have a Poisson basis. Nevertheless, in view of the discussions in Sections 1.3, 2.3 and at the beginning of this chapter, we introduce the overdispersion aspect, which is reflected by the random parameter L of the underlying Poisson distribution. Hence, the only assumptions which still have to be added are those about the distribution of the claim size C and the distribution of L .

Several distributions which are widely used in the insurance practice are suggested as possible candidates for the distributional assumption of C . These distributions together with their basic characteristics (mean μ , variance $\sigma^2 = \mu_2$, skewness $\mu_3/\mu_2^{3/2}$ and kurtosis $\mu_4/\mu_2^2 - 3$, where μ_j , $j = 2, \dots, 4$ are the central moments) are presented in Table 3.2 below (*IG* stands for Inverse Gaussian). These three distributions will be used as proposed assumptions for C . The choice for distributional assumptions about L will be handled with more care since the

Distribution	Density	Moments	Characteristics
Lognormal	$f_{LN}(x) = \frac{1}{\beta\sqrt{2\pi x}} \exp\left(-\frac{(\ln x - \alpha)^2}{2\beta^2}\right)$	$\nu_1 = e^{\alpha + \beta^2/2}$ $\nu_2 = e^{2(\alpha + \beta^2)}$ $\nu_3 = e^{3\alpha + 9\beta^2/2}$ $\nu_4 = e^{4\alpha + 8\beta^2}$	Mean = $e^{\alpha + \beta^2/2}$ Variance = $e^{\beta^2 + 2\alpha}(e^{\beta^2} - 1)$ Skewness = $\sqrt{e^{\beta^2} - 1}(2 + e^{\beta^2})$ Kurtosis = $e^{4\beta^2} + 2e^{3\beta^2} + 3e^{2\beta^2} - 6$
Gamma	$f_{\gamma}(x) = x^{\alpha-1} \frac{\beta^{\alpha} e^{-\beta x}}{\Gamma(\alpha)}$	$\nu_1 = \alpha/\beta$ $\nu_2 = \alpha/\beta^2 + \alpha^2/\beta^2$ $\nu_3 = (\alpha^3 + 3\alpha^2 + 2\alpha)/\beta^3$ $\nu_4 = (\alpha^4 + 6\alpha^3 + 11\alpha^2 + 6\alpha)/\beta^4$	Mean = α/β Variance = α/β^2 Skewness = $2/\sqrt{\alpha}$ Kurtosis = $6/\alpha$
IG	$f_{IG}(x) = \frac{\alpha}{\sqrt{2\pi\beta x^3}} \exp\left(-\frac{(x-\alpha)^2}{2\beta x}\right)$	$\nu_1 = \alpha$ $\nu_2 = \alpha(\alpha + \beta)$ $\nu_3 = \alpha(3\alpha\beta + 3\beta^2 + \alpha^2)$ $\nu_4 = \alpha(15\alpha\beta^2 + \alpha^3 + 15\beta^3 + 6\beta\alpha^2)$	Mean = α Variance = $\alpha\beta$ Skewness = $3\sqrt{\beta/\alpha}$ Kurtosis = $15\beta/\alpha$

Table 3.2: Proposed distributions for C and L

modeling approach used in the present research is quite new and hardly covered in the literature. A detailed analysis of the distributional assumptions of L is presented in Section 3.4, where reasonable candidates are chosen from Table 3.2. Note that parameters of these distributions will have indexes C or L (like α_C will be the parameter for C and α_L will be the parameter for L).

3.3 Model parameters

In this section we are going to discuss the set of parameters of Model 4 which later will be used as input. It is important to note the difference between the model parameters and the parameters of the underlying distributions. The distributional parameters serve as mechanical tools to steer the model and they rarely have any economical interpretation. For instance, if the claim size $C \sim \text{Lognormal}(\alpha_C, \beta_C)$, the parameter α_C does not have an economical background. It is not the expected value or the standard deviation of the claim size, it is simply the parameter of the distribution. These parameters can be considered as an intermediate step between the input of the model and the model application.

In contrast with the distributional parameters, the model parameters (input) should represent the insurance portfolio characteristics and have to be invariant w.r.t. the distributional assumptions. These features are convenient for the practical application of the model. Therefore, we will try to organize the input from the distributional characteristics (like mean or standard deviation), and not from the distributional parameters. An example of a model parameter can be μ_C , which is the expected claim amount, having a clear economical interpretation.

Unlike the distributional parameters which are fixed by the definition of the distribution, the model (input) parameters have to be chosen by the model users. They should have all the characteristics described above, be convenient in use and define the model completely. As a basis we use the set of model parameters which was presented in Reijnen [2003] and later in Reijnen et al. [2005] (for details see the Introduction). Some of these parameters will be used directly in the form in which they appeared in the papers, while others will be modified w.r.t. specific features of Model 4.

As was already mentioned, approximations will often be used in our calculations. The main characteristics of the approximations are the mean, variance and standardized cumulants of S , which were derived in Section 3.1. Naturally, the mean, variance and standardized cumulants are functions of the model parameters. Therefore, the impact of the model parameters on the approximations and hence on the final result can be predicted through the formulas (3.1)-(3.9). The following sections present brief discussions about a reasonable choice of the input parameters and their impact on the outcome.

3.3.1 Parameters μ_C , ϵ and μ_G

Parameters which remain the same as in the work of Albers [1999], Reijnen [2003] and Reijnen et al. [2005] are μ_C , ϵ and μ_G . These parameters clearly satisfy all the conditions described above and can directly play the role of the input parameters of Model 4. Parameter μ_C is the first parameter of the claim size distribution which defines the claim size expectation. Note that μ_C is a dummy parameter: the problem remains invariant under scale transforms. Therefore in illustrative examples of the thesis the input parameter μ_C will be set at some fixed representative value.

Parameter ϵ is one of the "dependence" parameters, which defines the expected percentage (in decimals) of "special claims" inside the portfolio. It arises in the model in a linear way and that is also the way it comes out in the formulas of the cumulants (3.2)-(3.9) (of course, in the formulas of the standardized cumulants it appears in a slightly more complicated way). Its main role is to make very explicit that out of the average claim amount μ_S indeed only a fraction ϵ has to be attributed to dependent aspects of the model.

Parameter μ_G stands for the 'lumpiness' of the dependence part (which is defined by ϵ) which formally is the expected group size. It is clear that the larger μ_G , the bigger the problems caused by the dependence.

3.3.2 Parameter λ

In the present thesis parameter λ represents the total expected number of claims (simple and special). For the first time it appeared in Albers et al. [2008], replacing two parameters, called n and p in Section 1.3, which represent the size of the insurance portfolio and the probability of the occurrence of a claim, respectively. In our case we obviously have $\lambda = np$, which is more intuitive and convenient.

Speaking about the influence of λ , it can be concluded that as λ becomes larger, all approximations become more accurate. This fact can be seen from the formulas of the standardized cumulants $\kappa_{jS} = \kappa_{jS}^* / \sigma_S^j$, $j = 3, 4$ (see (3.2)-(3.9)). It is clear that these κ_{jS} are proportional to $\lambda^{1-j/2}$, and thus become small as λ increase, and it is well known that distributions with small skewness and kurtosis can be more accurately approximated by moments-based approximations. The increase of λ basically means the increase of the size of the underlying portfolio n , since p seems to be more or less given. From this we conclude that it will really help (from the approximation point of view) to consider a higher aggregation level.

3.3.3 Parameters γ_C and γ_G

In order to define the distribution of the claim size we need one more parameter. Having expectation μ_C as a first parameter, the standard deviation σ_C was used in Albers [1999] as a second one. Albers [1999] used the normal distribution

for modeling the claim size. In that case μ_C and σ_C is a natural and convenient choice. In the present research the claim sizes are assumed to have a more realistic distribution and the choice of Albers [1999] loses its advantages. In our research we use the coefficient of variation γ_C instead. The main advantage of using

$$\gamma_C = \frac{\sigma_C}{\mu_C}$$

is that this quantity is scale invariant, which, among others, allows us to avoid using large numbers. As soon as the distribution of C no longer is assumed to be normal, we often have to deal with higher moments of the underlying distribution. In that case one more advantage of using γ_C is the relatively simple expressions for the standardized cumulants (see Lemma 3.4). The following lemma presents some results which can be useful while applying (3.2)-(3.9).

Lemma 3.4. *For the evaluation of the ν_{jL} and ν_{jC} from (3.2)-(3.9) we have in the*

(i) *general case:*

$$\frac{\nu_2}{\mu^2} = 1 + \gamma^2, \frac{\nu_3}{\mu^3} = 1 + 3\gamma^2 + \gamma^3 \kappa_3, \frac{\nu_4}{\mu^4} = 1 + 6\gamma^2 + 4\gamma^3 \kappa_3 + (\kappa_4 + 3)\gamma^4.$$

(ii) *Gamma case: $\text{Gamma}(\alpha, \alpha/\mu)$ (see Table 3.2)*

$$\gamma = \alpha^{-1/2}, \kappa_3 = 2\gamma, \kappa_4 = 6\gamma^2, \frac{\nu_j}{\mu^j} = \prod_{i=1}^{j-1} \left(1 + \frac{i}{\alpha}\right).$$

(iii) *Inverse-Gaussian case: $\text{IG}(\alpha, \beta)$*

$$\gamma = \sqrt{\frac{\beta}{\alpha}}, \kappa_3 = 3\gamma, \kappa_4 = 15\gamma^2,$$

$$\frac{\nu_3}{\mu^3} = 1 + 3\gamma^2 + 3\gamma^4, \frac{\nu_4}{\mu^4} = 1 + 6\gamma^2 + 15\gamma^4 + 15\gamma^6.$$

(iv) *lognormal case: $\text{LN}(\alpha, \beta)$, writing $\delta = \exp(\beta^2) = \mu^2 \exp(-2\alpha)$*

$$\gamma = (\delta - 1)^{1/2}, \kappa_3 = (\delta - 1)^{1/2}(\delta + 2),$$

$$\kappa_4 = \delta^4 + 2\delta^3 + 3\delta^2 - 6, \frac{\nu_k}{\mu^k} = \delta^{k(k-1)/2}, k = 1, 2, \dots$$

Proof. The general case (i) follows from straightforward evaluation. In the $\text{Gamma}(\alpha, \beta)$ case (where β can be written as α/μ) we have $\mu = \alpha/\beta, \sigma^2 = \alpha/\beta^2$ and $\gamma = \alpha^{-1/2}$ (see also Table 3.2). The cumulants κ_j follow likewise, while for the ν_j we have $\nu_j = \prod_{i=0}^{j-1} \{(\alpha + i)/\beta\} = \nu_1^j \prod_{i=1}^{j-1} (1 + i/\alpha)$ and (ii) follows. The $\text{IG}(\alpha, \beta)$ -case is dealt with in a similar manner and produces (iii). For the $\text{LN}(\alpha, \beta)$ -case, we note that $\nu_k = \exp(k\alpha + k^2\beta^2/2)$ and hence $\mu = \nu_1 = \delta^{1/2}e^\alpha$ and $\nu_k/\mu^k = \delta^{k(k-1)/2}$, as stated in (iv). Hence $\gamma = (\nu_2/\mu^2 - 1)^{1/2} = (\delta - 1)^{1/2}$. The results for κ_j are obtained in a similar way. \square

Parameter γ_G is chosen as a second parameter of the distribution of the group size (the first parameter was μ_G , the expected group size). Its basic advantages can be compared to the advantages of the parameter γ_C : scale invariance and simplicity. Given the value of μ_G , the random variable L in addition describes the so-called overdispersion effect, which further increases the dependence effect. This is expressed in the replacement of μ_G^{j-1} by the larger quantity ν_{jL}/μ_G (see Lemma 3.4). For instance in the Gamma case, we obtain that $\nu_{jL}/\mu_G = \mu_G^{j-1} \prod_{i=1}^{j-1} (1 + i/\alpha)$, which indeed indicates a significant further increase for α small.

Parameter	Description
μ_C	Expected claim size
γ_C	Coefficient of variation of the claim size
λ	Total expected number of claims
ϵ	Percentage of the special claims
μ_G	Expected group size
γ_G	Coefficient of variation of the group size

Table 3.3: Input parameters of Model 4

Table 3.3 summarizes all parameters of the model, which will be used as input.

3.4 Values of the model parameters

In this section we are going to define the region of interest for the parameters of Model 4 and the range of their values which will be used later as representative in the testing procedures and in a number of numerical examples. As starting point will be used the set of parameters values of the models which were presented in Reijnen [2003], Albers [1999] and Reijnen et al. [2005]. Some of these values will be used directly in the form in which they appeared in the papers, while others will be modified w.r.t. specific features of Model 4. In the previous section we presented the definition and discussion of the input parameters. All of them are summarized in Table 3.3. The region of interest and the representative values will be defined for each of these parameters.

We will start with the choice of the portfolio size which will lead to the choice of the values of λ . In Albers [1999] the m and q were taken in such a way that $\lambda = mq$ gives values between 12 and 30. Already for such small values, the approximations worked nicely. But of course, the author considered the extremely well-behaved case of normally distributed claim amounts. In Reijnen [2003] more realistic claim size distributions were utilized and the value $\lambda = 80$ (i.e. $q = 0.008, m = 10000$) was used as the total expected number of claims. Here we consider an even more complicated case with a mixture distribution for the random group sizes G . In that case, by fixing $\lambda = 80$ and $\epsilon = 0.03$, the expected number of special claims is only 2.4. However, if we want to analyze large group sizes like $\mu_G = 20$ (maybe

even larger), the expected number of such groups is only 0.12. Our assumption for the number of groups is Poisson. So the occurrence probabilities for the number of groups of size 20 are (see (3.10) with $\lambda = 0.12$)

$$f_P(1) = 0.1,$$

$$f_P(5) = 1.8 \times 10^{-7}.$$

This means that to see a case with 5 large groups we have to perform at least 10^7 simulations, which seems to be too much. Consequently, in the present study we will use larger values of λ : $\lambda \geq 400$.

One more reason to increase the value of λ is the introduction of the random expectation of the group size. The group sizes of Model 4 are assumed to have a mixture distribution which can dramatically increase the kurtosis of S , which can be unrealistic from the practical point of view and certainly will complicate the analysis of the dependence effect (relative increase of the value of the considered risk measure after introducing the dependency structure). When λ is increasing, the kurtosis of S decreases and the approximations work well. Therefore, our proposal for the region of interest for λ is 400 as a minimum value. The maximum is not defined since the larger λ , the better the accuracy of the approximations. Therefore, $\lambda = 400$ is chosen as a representative value for the testing procedures.

Speaking about testing the accuracy of the approximations, the effect of ϵ is expected to be approximately linear both in the true value (denoted by T) and the approximated value (denoted by X) and hence the relative error (which will be used for the accuracy judgement) $|X - T|/T$ does not depend (heavily) on ϵ , implying that if we compare several values of ϵ , the results about the accuracy will be similar. Therefore, the region of interest for ϵ is chosen to be $\epsilon \leq 0.05$ with a representative value $\epsilon = 0.03$.

The region of interest for the parameters of the claim size distribution is mostly determined by γ_C . Parameter μ_C defines only the calculation scale of the quantities we are interested in. Therefore, we fix μ_C at one value 100000, like it was done in Reijnen [2003], Reijnen et al. [2005] and Albers [1999]. Given μ_C it remains to indicate ranges for γ_C . We introduce the region of interest for γ_C as $0.05 \leq \gamma_C \leq 2.5$ with representative values 0.4 and 1.6. Such a choice is based on the work of Reijnen [2003] and Reijnen et al. [2005], where the skewness of C played an important role in the rule of thumb, the rule which provides an accurate approximation for the situation considered there.

The final parameters for which the region of interest has to be defined are μ_G and γ_G , the parameters of the group size. Parameter μ_G already appeared in the papers of Albers [1999] and Reijnen et al. [2005]. In the present thesis the region of interest for it remains the same, namely $5 \leq \mu_G \leq 20$ with representative values 5, 10 and 20.

Remark 3.1. The range of interest for the parameter μ_G varies between 5 and 20. Hence, G will as a rule be at least equal to 2. However, a value of G equal to 1

is possible. In that case we do not really have a group and it will not be recognized as such. Nevertheless, for most of the theory developed in the present thesis this will cause no problem: the results continue to hold for general G . In view of that we will often (when it is convenient) switch between two parametrizations: (μ_G, γ_G) and (μ_L, γ_L) . By definition of G the relation between the two forms of parametrization is simply given by

$$\mu_G = E[E[G|L]] = \mu_L,$$

$$\gamma_G^2 = \text{Var} \left[\frac{G}{\mu_L} \right] = \mu_L^{-2} (\text{Var}[E[G|L]] + E[\text{Var}[G|L]]) = \gamma_L^2 + \mu_L^{-1}.$$

On the other hand, in practice a value of G equal to 1 will occur only rarely and we may ignore it without making large mistakes. \square

The parameter γ_G is new. It did not appear in the earlier papers. Therefore, to define reasonable values of interest we perform a somewhat deeper analysis. The decision will be based on a number of numerical examples of the distribution of L and its main characteristics. The distributions which are suggested as candidates for a reasonable distribution of L are summarized in Table 3.2. Several situations with different values of γ_G will be checked for each of the distributions considered. Based on that analysis, reasonable situations will be included in the region of interest for the parameter γ_G . Note that when it is convenient, the results will be presented through the parameter γ_L , which is closely related to γ_G (see Remark 3.1).

3.4.1 Situations when $L \sim \text{Gamma}(\alpha_L, \beta_L)$

The first candidate for the distribution of L is the Gamma distribution with parameters

$$\begin{aligned} \alpha_L &= \gamma_L^{-2} = (\gamma_G^2 - \mu_G^{-1})^{-1}, \\ \beta_L &= (\gamma_L^2 \mu_G)^{-1} = (\gamma_G^2 \mu_G - 1)^{-1}. \end{aligned}$$

Figures 3.2 and 3.1 illustrate densities for the different situations which are summarized in Table 3.4.

It can be seen that the situations where μ_G is small (not large enough) and γ_L is large (situations 2,3,6) are not very useful for our purposes. They are too extreme in the sense that the probabilities for small group sizes (< 1) are quite high. We need discrete numbers, therefore as a result we get 0 or 1 too often. This can lead to the domination of the groups with sizes 0 and 1 which is not realistic from practical point of view, see also Remark 3.1. The density for large group sizes is also not very realistic in these situations. When we speak about groups which are larger than 5 we get almost uniformly distributed probabilities. This produces groups like 8 and 100 with almost the same probability. Maybe such situations can find an application, but these are really extreme cases.

Situation	μ_G	σ_L	γ_L	γ_G	α_L	β_L	Skewness	Kurtosis
1	5	5	1	1.09	1	0.2	2	6
2	5	10	2	2.04	0.25	0.05	4	24
3	5	20	4	4.02	0.0625	0.0125	8	96
4	10	5	0.5	0.59	4	0.4	1	1.5
5	10	10	1	1.04	1	0.1	2	6
6	10	20	2	2.02	0.25	0.025	4	24
7	20	5	0.25	0.34	16	0.8	0.5	0.375
8	20	10	0.5	0.55	4	0.2	1	1.5
9	20	20	1	1.02	1	0.05	2	6

Table 3.4: Different situations for the Gamma distribution

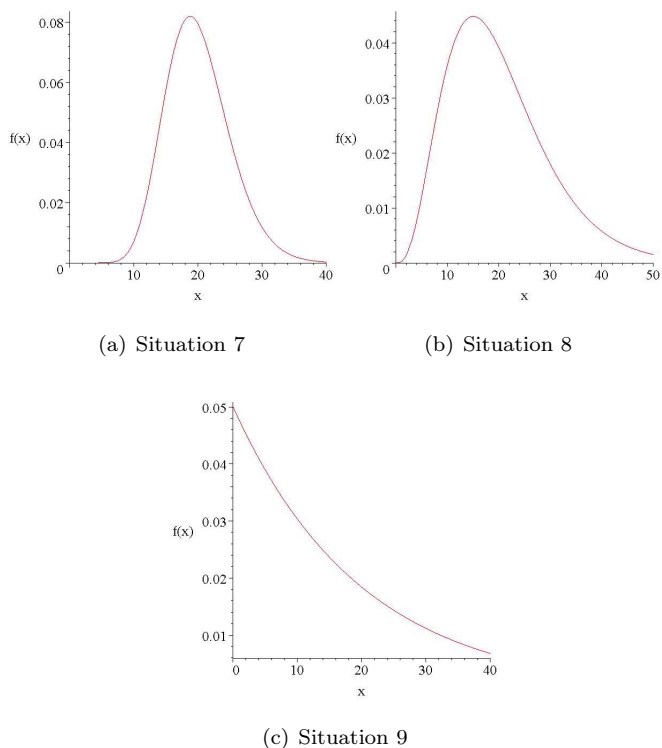


Figure 3.1: Gamma densities for the different situations (7-9)

To get nicer pictures it is necessary to increase the group size. Situations like 4,7 or 8 are nicer in the sense that the tail is not so extremely heavy and the very small values do not have so much density. The third type of situation is when

the tail is not so extremely heavy, but the very small values do have substantial density. These are situations like 1,5,9.

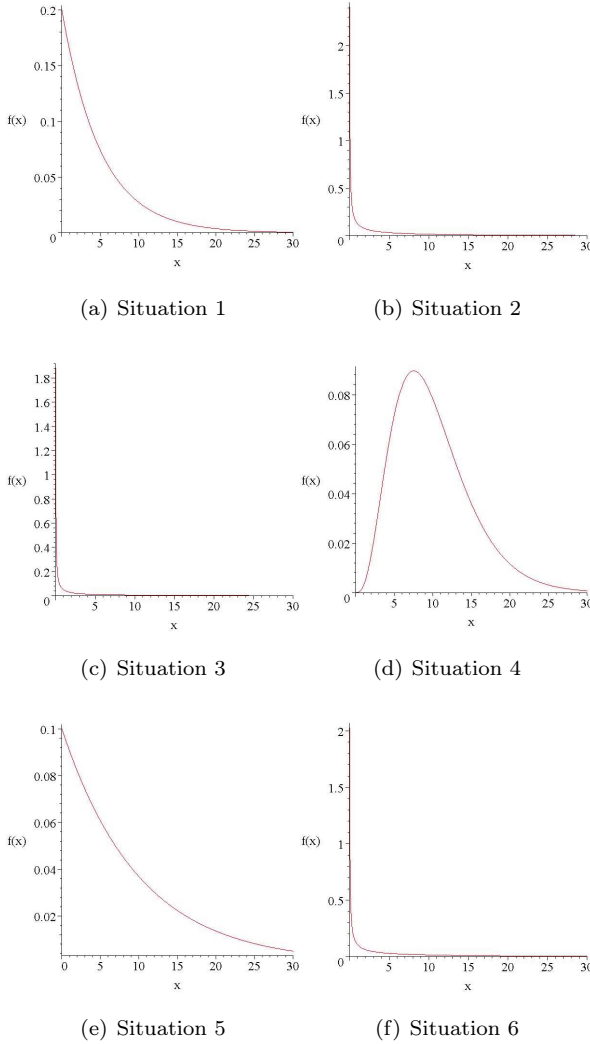


Figure 3.2: Gamma densities for the different situations (1-6)

It is possible to divide the situations considered here into three groups: when $\gamma_L < 1$, $\gamma_L > 1$ and $\gamma_L = 1$. Situations considered here with $\gamma_L \leq 1$ are acceptable when $L \sim \text{Gamma}$ but the situations with $\gamma_L > 1$ have to be strictly motivated. We can consider $\gamma_L = 1.5$ as a boundary case, but $\gamma_L = 4$ is certainly out of the

question. In general $\gamma_L \leq 1.5$ would be the proposed region of interest for γ_L when $L \sim \text{Gamma}$ with representative values 0.75 and 1.5.

3.4.2 Situations when $L \sim IG(\alpha_L, \beta_L)$

The next distribution we are going to investigate is the Inverse Gaussian with

$$\alpha_L = \mu_G$$

and

$$\beta_L = \gamma_L^{-2} \mu_G = (\gamma_G^2 - \mu_G^{-1}) \mu_G.$$

Here, like in the $L \sim \text{Gamma}$ case we consider several situations with different μ_G and γ_L values. Parameters values for all the situations, together with the densities, are found in Table 3.5 and Figures 3.4 and 3.3.

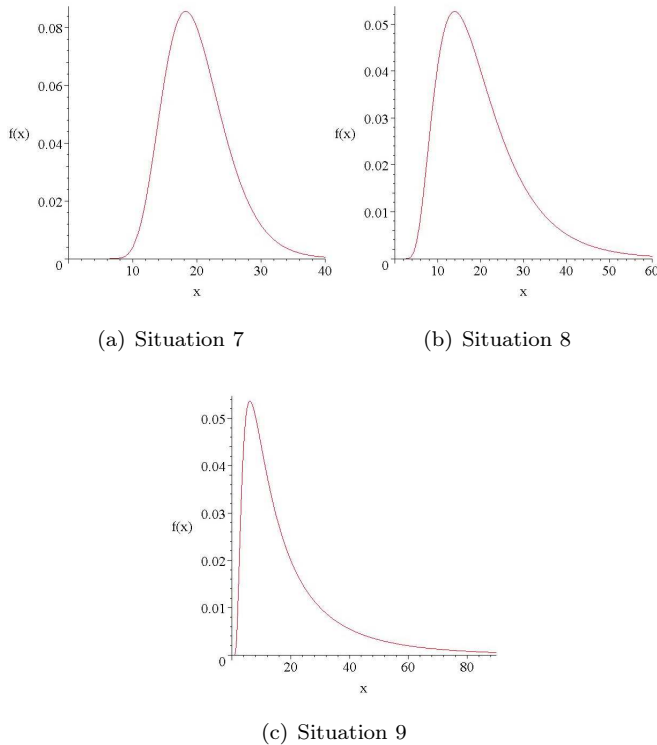


Figure 3.3: Inverse Gaussian densities for the different situations (7-9)

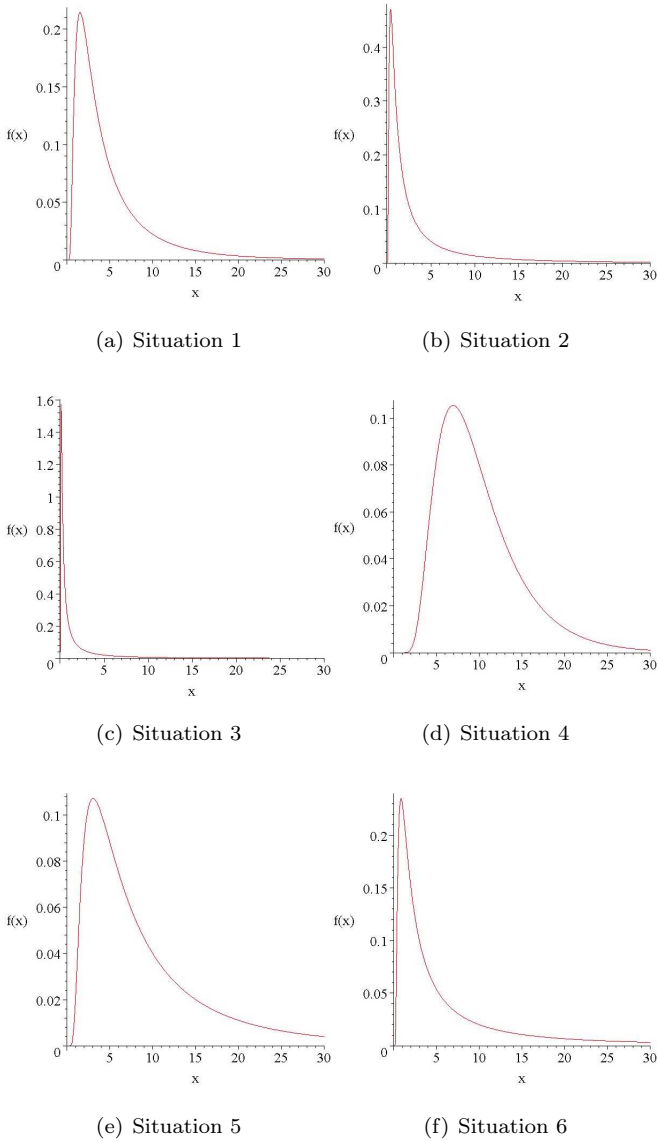


Figure 3.4: Inverse Gaussian densities for the different situations (1-6)

From the figures we conclude that the Inverse Gaussian distribution is more useful for our purposes. When we speak about the distribution for the group size we need something more skewed to the right and more (but not extremely) heavy tailed. We do not need too much mass at the beginning of the density. In the

$L \sim Gamma$ case the skewness is not large enough and we get situations when too much density is concentrated at the very beginning, where values are smaller than 1.

Situation	μ_G	σ_L	γ_L	γ_G	α_L	β_L	Skewness	Kurtosis
1	5	5	1	1.09	5	5	3	15
2	5	10	2	2.04	5	1.25	6	60
3	5	20	4	4.02	5	0.3125	12	240
4	10	5	0.5	0.59	10	40	1.5	3.75
5	10	10	1	1.04	10	10	3	15
6	10	20	2	2.02	10	2.5	6	60
7	20	5	0.25	0.34	20	320	0.75	0.9375
8	20	10	0.5	0.55	20	80	1.5	3.75
9	20	20	1	1.02	20	20	3	15

Table 3.5: Different situations for the Inverse Gaussian distribution

In that sense the Inverse Gaussian distribution produces much better results (compare situations 1,4,5,6,8,9). For example in situation 5 of the $L \sim IG$ case the density is very low from 0 to 1 and in the $L \sim Gamma$ case it is already high at 0. This is important since we get completely different density shapes. The very extreme situation like 3 is similar to the $L \sim Gamma$ situation 2. Also the situations with small γ_L and large μ_G are close to each other (compare situations 7). Therefore, situations when $\gamma_L \leq 1$ are acceptable like in the $L \sim Gamma$ case. Situations with $\gamma_L = 2$ are acceptable as well but $\gamma_L = 4$ case is still too extreme. Therefore, the proposed region of interest for γ_L is $\gamma_L \leq 2.5$ when $L \sim IG$ with the representative values $\{0.75, 1.5, 2, 2.5\}$.

3.4.3 Situations when $L \sim Lognormal(\alpha_L, \beta_L)$

The last distribution we are going to analyze as a candidate for the mixing distribution is the lognormal. Parameters α_L and β_L can be expressed as

$$\alpha_L = \log(\mu_G) - 0.5 \log(1 + \gamma_L^2) = \log(\mu_G) - 0.5 \log(1 + \gamma_G^2 - \mu_G^{-1}),$$

$$\beta_L = \sqrt{\log(1 + \gamma_L^2)} = \sqrt{\log(1 + \gamma_G^2 - \mu_G^{-1})}.$$

Table 3.6 contains parameter values for the different situations.

Situation	μ_L	σ_L	γ_L	γ_G	α_L	β_L	Skewness	Kurtosis
1	5	5	1	1.09	1.2629	0.8326	4	38
2	5	10	2	2.04	0.80472	4.2686	14	944
3	5	20	4	4.02	0.1928	1.6832	76	94189
4	10	5	0.5	0.59	2.1910	0.4724	1.6	5
5	10	10	1	1.04	1.956	0.8326	4	38
6	10	20	2	2.02	1.4979	1.2686	14	944
7	20	5	0.25	0.34	2.9654	0.2462	0.77	1.1
8	20	10	0.5	0.55	2.8841	0.4724	1.6	5
9	20	20	1	1.02	2.6492	0.8326	4	38

Table 3.6: Different situations for the lognormal distribution

$\epsilon \leq 0.05$
$0.05 \leq \gamma_G \leq 2.5$
$\lambda \geq 400$
$5 \leq \mu_G \leq 20$
$\gamma_L \leq 1.5$ for $L \sim \text{Gamma}$
$\gamma_L \leq 2.5$ for $L \sim \text{IG}$

Table 3.7: Restricted region of parameters which is formulated as the region of interest.

It is quite clear already from the table that the lognormal distribution is too extreme for our purposes. Situations with small μ_G and large γ_L produce huge kurtoses. We certainly do not have much mass at the beginning of the distribution, but now we have another extreme. The tail is definitely too heavy. We are not presenting all the densities for the lognormal distribution since their shapes are very close to the Inverse Gaussian case. The only difference is that the tail of the distribution is much heavier. As an example we compare situation 3 of the $L \sim \text{IG}$ case with situation 3 of the $L \sim \text{Lognormal}$ case (see Figure 3.5). Therefore, the conclusion is not to use the lognormal distribution as a mixing distribution for the group size random variable.

All the parameters of Model 4 have been discussed one by one, including the range of interest for each of them, see Table 3.7. Table 3.8 summarizes all the representative parameters values under the different model assumptions.

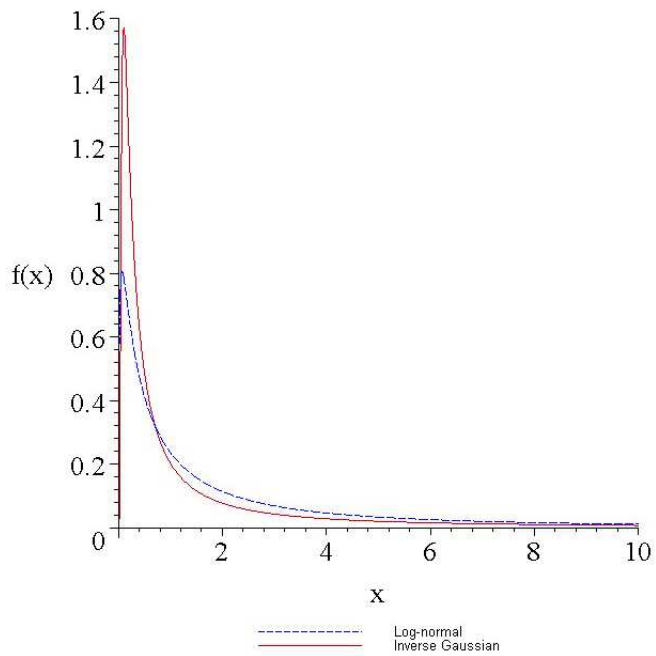


Figure 3.5: Comparison of the Inverse Gaussian situation 3 with the lognormal situation 3

	$L \sim \text{Gamma}(\alpha_L, \beta_L)$	$L \sim \text{IG}(\alpha_L, \beta_L)$
$C \sim \text{Lognormal}(\alpha_C, \beta_C)$	11 $\lambda = 400$ $\epsilon = 0.03$ $\{\mu_C, \gamma_C\} =$ $= \{[100000, 0.4],$ $[100000, 1.6]\}$ $\{\mu_G, \gamma_L\} =$ $= \{[5, 0.75], [5, 1.5], [10, 0.75],$ $[10, 1.5], [20, 0.75], [20, 1.5]\}$	12 $\lambda = 400$ $\epsilon = 0.03$ $\{\mu_C, \gamma_C\} =$ $= \{[100000, 0.4],$ $[100000, 1.6]\}$ $\{\mu_G, \gamma_L\} =$ $= \{[5, 0.75], [5, 1.5], [5, 2], [5, 2.5],$ $[10, 0.75], [10, 1.5], [10, 2], [10, 2.5],$ $[20, 0.75], [20, 1.5], [20, 2], [20, 2.5]\}$
$C \sim \text{IG}(\alpha_C, \beta_C)$	21 $\lambda = 400$ $\epsilon = 0.03$ $\{\mu_C, \gamma_C\} =$ $= \{[100000, 0.4],$ $[100000, 1.6]\}$ $\{\mu_G, \gamma_L\} =$ $= \{[5, 0.75], [5, 1.5], [10, 0.75]$ $[10, 1.5], [20, 0.75], [20, 1.5]\}$	22 $\lambda = 400$ $\epsilon = 0.03$ $\{\mu_C, \gamma_C\} =$ $= \{[100000, 0.4],$ $[100000, 1.6]\}$ $\{\mu_G, \gamma_L\} =$ $= \{[5, 0.75], [5, 1.5], [5, 2], [5, 2.5],$ $[10, 0.75], [10, 1.5], [10, 2], [10, 2.5],$ $[20, 0.75], [20, 1.5], [20, 2], [20, 2.5]\}$
$C \sim \text{Gamma}(\alpha_C, \beta_C)$	31 $\lambda = 400$ $\epsilon = 0.03$ $\{\mu_C, \gamma_C\} =$ $= \{[100000, 0.4],$ $[100000, 1.6]\}$ $\{\mu_G, \gamma_L\} =$ $= \{[5, 0.75], [5, 1.5], [10, 0.75]$ $[10, 1.5], [20, 0.75], [20, 1.5]\}$	32 $\lambda = 400$ $\epsilon = 0.03$ $\{\mu_C, \gamma_C\} =$ $= \{[100000, 0.4],$ $[100000, 1.6]\}$ $\{\mu_G, \gamma_L\} =$ $= \{[5, 0.75], [5, 1.5], [5, 2], [5, 2.5],$ $[10, 0.75], [10, 1.5], [10, 2], [10, 2.5],$ $[20, 0.75], [20, 1.5], [20, 2], [20, 2.5]\}$

Table 3.8: Parameters values under the different assumptions

Chapter 4

Quantities of interest and calculation methods

Chapter 3 describes in detail the model for the aggregated sum S which will be the main point of consideration in the present thesis. Having the model we can analyze any insurance (or reinsurance) contract which is based on S . This chapter introduces several insurance (or reinsurance) products (which are naturally functions of S), the behavior of which will be analyzed in later chapters under the dependence assumption. For each insurance product presented here we will suggest several calculation methods. Each method has its own advantages and disadvantages, which basically depend on the distributional assumptions of the model.

When we are interested in the impact of small dependencies on tail events, the Stop-Loss contract is a good tool for such an analysis. This contract was used as a basic tool in the work of Albers [1999] and later in Reijnen [2003] and Reijnen et al. [2005]. The idea of the contract is simple. Denoting by S the total risk amount, the Stop-Loss contract covers the part of that risk which is above some retention level a . Therefore, the total risk S is divided into two parts: retained risk $S - (S - a)^+$ and Stop-Loss risk $(S - a)^+$ (here $(S - a)^+ = \max(S - a, 0)$). Considering reinsurance as an example, the reinsurer in that case gets the risk $(S - a)^+$ and the remaining part $S - (S - a)^+$ is left for the insurance company which buys the contract. An important characteristic of the Stop-Loss contract is the net Stop-Loss premium (from now on simply SLP) which is often considered as a basic part of the price of the contract. It is defined as the expectation $E[(S - a)^+]$ of the Stop-Loss risk and this quantity is the first quantity of interest which will be investigated in the present thesis.

The SLP is also used as risk measure for ordering of risks, see e.g. Dhaene et al. [2006], where several results on Stop-Loss ordering can be found. It is clear that in reality SLP cannot be the total price of the Stop-Loss contract

because the reinsurers position becomes very risky with an expected profit 0. Some positive loading definitely has to be added to the *SLP*. The value of such loading depends on many factors and it has to be acceptable for both sides: the insurance company and the reinsurer. It can be some fixed value or it can depend on some risk characteristic like mean or standard deviation. For the discussion of different kinds of reinsurance premiums we refer to Kaas et al. [2001]. That is why the second quantity which is considered to be important is the variance of the Stop-Loss contract and the retained risk.

The last quantity which will be investigated is the Value at Risk (*VaR*) of the aggregated claim amount. *VaR* is a quite popular risk measure in the financial world. For given p , the *VaR* is defined as $VaR = s = F_S^{-1}(p)$, where $F_S(s) = P(S \leq s)$ and therefore it can be presented as an upper limit for the aggregated claim given some probability (risk) value.

In total there are three ways which will be used as calculation tools: convolutions, approximations and simulations. Direct Monte-Carlo simulations do not have any specific underlying conditions and the accuracy of the final result only depends on the number of simulations and the random number generator. However, since the form of the model is quite complicated, the method of direct simulation can be very time consuming. Our research showed that when the distributional assumptions of the model are not simple (in the sense of random number generation), the needed precision can be obtained only after several days of computer work. This fact forced us to think about improvement of the simulation speed. An alternative simulation method which is presented here is the method of hybrid simulations.

The accuracy of the convolutions method is limited only by the value of the summation index and probably some internal numerical methods which may have to be applied, but the method has two main drawbacks: it can be very time consuming (depending on the model assumptions) and it can be applied in practice only when the assumed distributions of L and C are closed under convolution. From Table 3.2 only the Gamma and IG satisfy this condition.

The method of approximations was described in detail in the thesis of Reijnen [2003]. A short overview of the used approximation techniques is presented in the Introduction. A more detailed description can be found in Section 4.1.4. Unlike the simulation or convolution case, the accuracy of the approximations has strict limits and depends on the choice of the underlying parameters. The approximations which are used here are the IG and Gamma-IG. The remaining approximations (Edgeworth expansions, Normal power, Gamma) which were tested in Reijnen et al. [2005] are not included. These approximations did not show good results when Model 3 was assumed and therefore they are not expected to be acceptable while working with Model 4. For each of the quantities considered, several possible calculation methods will be suggested and discussed. It should be remembered that we do not only want to calculate the quantities numerically, but are also or sometimes mainly interested in their behavior depending on the underlying parameters, see Section 1.5. We will start with the *SLP*.

4.1 Net Stop-Loss premium

The *SLP* is defined as $E[(S - a)^+]$ where S is the aggregated claim amount and a is the retention level. To calculate such an expectation we need the density of S . Therefore, the calculation of the *SLP* (as well as the remaining quantities) using the convolution method is based on the calculation of the true density of S . The main idea of the approximation method is to replace the density of S by another density of a simpler form. This allows us to obtain more simple integrals as well, which (hopefully) can be solved analytically. The simulation method uses realizations of the random variables which appear in the model. Therefore, the random number generator plays a major role here. As was mentioned at the beginning of the chapter, two simulation methods will be suggested as possible *SLP* calculation tools. Depending on the distributional assumptions in the model, one or the other method will be used in our analysis. The first simulation method is the direct (or simple) Monte-Carlo simulation which is described in Section 4.1.2. An alternative simulation method presented here is the method of hybrid simulations. All the theoretical details, as well as the results on testing the accuracy of the method, are presented in Section 4.1.3. The sections that follow present three calculation methods for the *SLP*.

4.1.1 Convolutions

The basic structure of the underlying model is the (random) summation of certain random variables. Therefore a very straightforward way to get the density of S is the convolution method. As was mentioned before, to apply this method, the underlying distributions of C and L have to be closed under convolution. From Table 3.2 the Gamma and IG satisfy this condition, and therefore only these distributions are going to be considered in the present section.

The general form of Model 4 is

$$S = \sum_{i=1}^N C_i + \sum_{k=1}^H \sum_{j=1}^{G_k} D_{jk}.$$

Assuming that $C \triangleq D$ we can write it in the form

$$S = \sum_{i=1}^W C_i, \tag{4.1}$$

where

$$W = N + \sum_{k=1}^H G_k \tag{4.2}$$

with the assumptions described in Section 1.3. We consider several situations where C and L are assumed to be Gamma or IG.

Theorem 4.1. *Suppose that Model 4 (see Table 3.1) holds. Then, if $C \sim \text{Gamma}(\alpha_C, \beta_C)$ and $L \sim \text{Gamma}(\alpha_L, \beta_L)$ (see Table 3.2), the needed SLP can be calculated as*

$$E[(S - a)^+] = \sum_{w=1}^{\infty} P(W = w) \times \left\{ \frac{w\alpha_c}{\beta_c} [1 - F_\gamma(a; w\alpha_c + 1, \beta_c)] - a[1 - F_\gamma(a; w\alpha_c, \beta_c)] \right\} \quad (4.3)$$

with

$$P(W = w) = (\lambda(1 - \epsilon))^w e^{-\lambda(\epsilon/\mu_G + 1 - \epsilon)} \left(\frac{1}{w!} + \sum_{h=1}^{\infty} \sum_{j=0}^w \left[\frac{(\epsilon\lambda/\mu_G)^h}{h!} \times \frac{\beta_L^{h\alpha_L} \Gamma(h\alpha_L + j)}{(\lambda(1 - \epsilon))^j (w - j)! \Gamma(h\alpha_L) j! (\beta_L + 1)^{h\alpha_L + j}} \right] \right), \quad (4.4)$$

where F_γ denotes the Gamma cdf.

Proof. The Gamma distribution is closed under convolution, hence, conditionally on $W = w$, we get that

$$\sum_{i=1}^w C_i \sim \text{Gamma}(w\alpha_c, \beta_c)$$

and the density of S takes a form

$$f_S(s) = \sum_{w=1}^{\infty} P(W = w) s^{w\alpha_c - 1} \frac{\beta_c^{w\alpha_c} e^{-\beta_c s}}{\Gamma(w\alpha_c)}.$$

Note that the summation index starts from 1 instead of 0. Theoretically the probability $P(W = 0)$ is not 0, but zero convolution is defined to be equal to 0, resulting in $P(W = 0) \times 0 = 0$. Therefore, the index zero is skipped in the calculation.

Using the above formulas we get

$$\begin{aligned} E[(S - a)^+] &= \int_a^\infty (s - a) \sum_{w=1}^{\infty} P(W = w) s^{w\alpha_c - 1} \frac{\beta_c^{w\alpha_c} e^{-\beta_c s}}{\Gamma(w\alpha_c)} ds \\ &= \sum_{w=1}^{\infty} P(W = w) \int_a^\infty (s - a) s^{w\alpha_c - 1} \frac{\beta_c^{w\alpha_c} e^{-\beta_c s}}{\Gamma(w\alpha_c)} ds. \end{aligned}$$

We define

$$\begin{aligned}
 I &:= \int_a^\infty (s-a) s^{w\alpha_c-1} \frac{\beta_c^{w\alpha_c} e^{-\beta_c s}}{\Gamma(w\alpha_c)} ds \\
 &= \int_a^\infty s s^{w\alpha_c-1} \frac{\beta_c^{w\alpha_c} e^{-\beta_c s}}{\Gamma(w\alpha_c)} ds - a \int_a^\infty s^{w\alpha_c-1} \frac{\beta_c^{w\alpha_c} e^{-\beta_c s}}{\Gamma(w\alpha_c)} ds \\
 &= I_1 - aI_2
 \end{aligned} \tag{4.5}$$

and calculate separately

$$\begin{aligned}
 I_1 &= \int_a^\infty s^{(w\alpha_c+1)-1} \frac{\beta_c^{w\alpha_c} e^{-\beta_c s}}{\Gamma(w\alpha_c)} ds \\
 &= \frac{w\alpha_c}{\beta_c} \int_a^\infty s^{(w\alpha_c+1)-1} \frac{\beta_c^{w\alpha_c+1} e^{-\beta_c s}}{\Gamma(w\alpha_c+1)} ds \\
 &= \frac{w\alpha_c}{\beta_c} \left(1 - \int_0^a s^{(w\alpha_c+1)-1} \frac{\beta_c^{w\alpha_c+1} e^{-\beta_c s}}{\Gamma(w\alpha_c+1)} ds \right) \\
 &= \frac{w\alpha_c}{\beta_c} [1 - F_\gamma(a; w\alpha_c+1, \beta_c)]
 \end{aligned} \tag{4.6}$$

and (using the same arguments)

$$I_2 = 1 - F_\gamma(a; w\alpha_c, \beta_c), \tag{4.7}$$

which gives us the general form (4.3).

To get the final result we need to calculate the probabilities $P(W = w)$. The assumption $L \sim \text{Gamma}(\alpha_L, \beta_L)$ leads to $G \sim \text{NB}(\alpha_L, \beta_L)$, the probability mass function of which is given by

$$\begin{aligned}
 P(G = k; \alpha_L, \beta_L) &= \int_0^\infty f_P(k; l) f_\gamma(l; \alpha_L, \beta_L) dl \\
 &= \int_0^\infty \frac{l^k}{k!} e^{-l} \frac{l^{\alpha_L-1} e^{-l\beta_L} \beta_L^{\alpha_L}}{\Gamma(\alpha_L)} dl \\
 &= \frac{\beta_L^{\alpha_L}}{\Gamma(\alpha_L) k!} \int_0^\infty l^{\alpha_L-1+k} \exp(-l\beta_L - l) dl \\
 &= \frac{\beta_L^{\alpha_L}}{\Gamma(\alpha_L) k! (\beta_L + 1)^{\alpha_L+k}} \\
 &\quad \times \int_0^\infty (l(\beta_L + 1))^{\alpha_L+k-1} \exp(-l(\beta_L + 1)) d(l(\beta_L + 1)) \\
 &= \frac{\beta_L^{\alpha_L}}{\Gamma(\alpha_L) k! (\beta_L + 1)^{\alpha_L+k}} \Gamma(\alpha_L + k).
 \end{aligned}$$

The negative binomial distribution is closed under convolution. Therefore, we get that $\sum_{k=1}^h G_k \sim NB(h\alpha_L, \beta_L)$. Hence, conditionally on $H = h$ we have to consider the convolution of Poisson and negative binomial random variables. This convolution law is known in the actuarial literature as the Delaporte distribution. In our case we have $N \sim P(\lambda(1 - \epsilon))$ which gives

$$P(N = n) = e^{-\lambda(1-\epsilon)} \frac{(\lambda(1 - \epsilon))^n}{n!}$$

and $\sum_{k=1}^h G_k \sim NB(h\alpha_L, \beta_L)$ which gives

$$P\left(\sum_{k=1}^h G_k = j\right) = \frac{\beta_L^{h\alpha_L} \Gamma(h\alpha_L + j)}{\Gamma(h\alpha_L) j! (\beta_L + 1)^{h\alpha_L + j}}.$$

Hence, we get

$$\begin{aligned} P(W = w | H = h) &= \sum_{j=0}^w P\left(\sum_{k=1}^h G_k = j\right) P(N = w - j) \\ &= \sum_{j=0}^w \frac{\beta_L^{h\alpha_L} \Gamma(h\alpha_L + j) e^{-\lambda(1-\epsilon)} (\lambda(1 - \epsilon))^{w-j}}{\Gamma(h\alpha_L) j! (\beta_L + 1)^{h\alpha_L + j} (w - j)!} \end{aligned}$$

and the total probability is

$$\begin{aligned} P(W = w) &= \sum_{h=1}^{\infty} \left[P(H = h) \sum_{j=0}^w \frac{\beta_L^{h\alpha_L} \Gamma(h\alpha_L + j)}{\Gamma(h\alpha_L) j! (\beta_L + 1)^{h\alpha_L + j}} \right. \\ &\quad \left. \times e^{-\lambda(1-\epsilon)} \frac{(\lambda(1 - \epsilon))^{w-j}}{(w - j)!} \right] + P(H = 0) P(N = w). \end{aligned}$$

According to our model $H \sim P(\epsilon \frac{\lambda}{\mu_G})$, which makes

$$P(H = h) = e^{-\epsilon \lambda / \mu_G} \frac{(\epsilon \lambda / \mu_G)^h}{h!}$$

and the final formula for $P(W = w)$ takes a form

$$\begin{aligned}
 P(W = w) &= \sum_{h=1}^{\infty} \sum_{0 \leq j \leq w} \left[e^{-\epsilon\lambda/\mu_G} \frac{(\epsilon\lambda/\mu_G)^h}{h!} \frac{\beta_L^{h\alpha_L} \Gamma(h\alpha_L + j)}{\Gamma(h\alpha_L) j! (\beta_L + 1)^{h\alpha_L + j}} \right. \\
 &\quad \left. \times e^{-\lambda(1-\epsilon)} \frac{(\lambda(1-\epsilon))^{w-j}}{(w-j)!} \right] + P(H = 0)P(N = w) \\
 &= e^{-\epsilon\lambda/\mu_G} e^{-\lambda(1-\epsilon)} (\lambda(1-\epsilon))^w \sum_{h=1}^{\infty} \sum_{0 \leq j \leq w} \left[\frac{(\epsilon\lambda/\mu_G)^h}{h!} \frac{\beta_L^{h\alpha_L} \Gamma(h\alpha_L + j)}{\Gamma(h\alpha_L) j! (\beta_L + 1)^{h\alpha_L + j}} \right. \\
 &\quad \left. \times \frac{1}{(\lambda(1-\epsilon))^j (w-j)!} \right] + e^{-\lambda(\epsilon/\mu_G + 1 - \epsilon)} \frac{(\lambda(1-\epsilon))^w}{w!},
 \end{aligned}$$

which completes the proof. \square

Now we have everything we need for the calculation of $E[(S - a)^+]$ when C and L are assumed to be Gamma distributed. Formulas (4.3) and (4.4) can be used to get the numerical results. At this point several difficulties have to be mentioned which the reader will certainly encounter while applying (4.3) and (4.4). From a first glance, these seem nice formulas which can be used directly to obtain $E[(S - a)^+]$. However, several unpleasant parts should be mentioned which blow up the "direct application" illusion.

- The most obvious one is the summation till infinity at two places (for w and for h). It is clear that in practice it is impossible to perform such a summation. Calculation has to be stopped at some point and this point will obviously determine the calculation accuracy.
- Note that in formula (4.3) the expression $P(W = w)$ is multiplied by some function of w (and many other parameters). Certain values of w can lead to very small values of $P(W = w)$ (even of order 10^{-10}). But the second factor in this multiplication can be very large with the same value of w (let say of order 10^{15}). This means that the accuracy of $P(W = w)$ has to be at least up to 15 digits: if it is not, we can get huge errors. The problem is that it is not known in advance which accuracy is required. It has to be defined for each w separately.
- One more difficulty in calculation is hidden in the summation of formula (4.4). Here we have to deal with factorials like $j!$ and gamma functions like $\Gamma(h\alpha_L + j)$. The problem is that j can reach large values like 400 or 600, which makes it necessary to deal with such numbers as 400! or 600! (as well as $\Gamma(400)$ which can be written as 399!). The largest number in $C++$ which modern computer can operate with is 1.7×10^{308} , but 400! is much larger than that and hence it is necessary to tackle this problem as well.

The points above are just some examples of difficulties in application. Appendix A1 contains a detailed overview of all possible problems and their solutions. There we will present as a result the complete implementation algorithm for formulas (4.3) and (4.4).

Theorem 4.2. *Suppose that Model 4 (see Table 3.1) holds. Then, if $C \sim \text{Gamma}(\alpha_C, \beta_C)$ and $L \sim \text{IG}(\alpha_L, \beta_L)$ (see Table 3.2), the needed SLP can be calculated using (4.3) with*

$$P(W = w) = (\lambda(1 - \epsilon))^w e^{-\lambda(\epsilon/\mu_G + 1 - \epsilon)} \left(\frac{1}{w!} + \sum_{h=1}^{\infty} \sum_{j=0}^w \sum_{k=0}^{j-1} \frac{(\epsilon\lambda/\mu_G)^h}{h!} \left(\frac{\beta_L}{2h\alpha_L} \right)^k \right. \\ \left. \times \frac{e^{-(h\alpha_L/\beta_L)\{(1+2\beta_L)^{1/2}-1\}} (h\alpha_L)^j (j-1+k)! (1+2\beta_L)^{-(j+k)/2}}{j!(j-1-k)!k!(\lambda(1-\epsilon))^j (w-j)!} \right). \quad (4.8)$$

Proof. Since $C \sim \text{Gamma}$, (4.3) follows directly from the proof of Theorem 4.1. The only step we have to prove is the formula for the probabilities $P(W = w)$. Now we have $L \sim \text{IG}(\alpha_L, \beta_L)$. In that case we have to deal with the Poisson-IG mixture. According to the model assumptions, the group sizes $G \sim P(L)$, for which the probability generating function P_G is given by

$$P_G(z) = \exp(-(\alpha_L/\beta_L)\{[1 - 2\beta_L(z-1)]^{1/2} - 1\}).$$

To get the probabilities $P(G = n)$ it is necessary to differentiate this function n times. Omitting all the calculations, the probabilities can be written as

$$p_n = P(G = n) = p_0 \frac{\alpha_L^n}{n!} \sum_{k=0}^{n-1} \frac{(n-1+k)!}{(n-1-k)!k!} \left(\frac{\beta_L}{2\alpha_L} \right)^k (1+2\beta_L)^{-(n+k)/2}, \quad (4.9)$$

where

$$p_0 = \exp(-(\alpha_L/\beta_L)\{(1+2\beta_L)^{1/2} - 1\}).$$

So far we have discussed the single Poisson-IG mixture, but in our case we have to deal with the sum $\sum_{k=1}^H G_k$ where each $G_k \sim \text{Poisson} - \text{IG}(\alpha_L, \beta_L)$. Fortunately the Poisson-IG mixtures are closed under convolution, therefore conditionally on $H = h$ we have that $\sum_{k=1}^h G_k \sim \text{Poisson} - \text{IG}(h\alpha_L, \beta_L)$. We are interested in $P(W = w)$ with W defined in (4.2). Therefore, conditionally on $H = h$ we have to consider the convolution of Poisson and Poisson-IG random

variables. This leads to

$$\begin{aligned}
 P(W = w|H = h) &= \sum_{j=0}^w P\left(\sum_{k=1}^h G_k = j\right) P(N = w - j) \\
 &= \sum_{j=0}^w \exp(-(h\alpha_L/\beta_L)\{(1 + 2\beta_L)^{1/2} - 1\}) \frac{(h\alpha_L)^j}{j!} \\
 &\quad \times \sum_{k=0}^{j-1} \frac{(j-1+k)!}{(j-1-k)!k!} \left(\frac{\beta_L}{2h\alpha_L}\right)^k (1 + 2\beta_L)^{-(j+k)/2} \\
 &\quad \times \exp(-\lambda(1 - \epsilon)) \frac{(\lambda(1 - \epsilon))^{w-j}}{(w-j)!} \quad (4.10)
 \end{aligned}$$

and the total probability thus is

$$\begin{aligned}
 P(W = w) &= P(H = 0)P(W = w|H = 0) + \sum_{h=1}^{\infty} P(H = h)P(W = w|H = h) \\
 &= P(H = 0)P(N = w) + \sum_{h=1}^{\infty} \exp(-\epsilon\lambda/\mu_G) \frac{(\epsilon\lambda/\mu_G)^h}{h!} \\
 &\quad \times \sum_{0 \leq j \leq w} \exp(-(h\alpha_L/\beta_L)\{(1 + 2\beta_L)^{1/2} - 1\}) \frac{(h\alpha_L)^j}{j!} \\
 &\quad \times \sum_{k=0}^{j-1} \frac{(j-1+k)!}{(j-1-k)!k!} \left(\frac{\beta_L}{2h\alpha_L}\right)^k (1 + 2\beta_L)^{-(j+k)/2} \\
 &\quad \times \exp(-\lambda(1 - \epsilon)) \frac{(\lambda(1 - \epsilon))^{w-j}}{(w-j)!}, \quad (4.11)
 \end{aligned}$$

from which (4.8) follows directly. \square

Possible difficulties to be faced while applying (4.8) are discussed in Appendix A2. Here we present an important result which can be helpful while calculating $p_n = P(G = n)$ which appears in Theorem 4.2.

Remark 4.1. Practical application of (4.9) can be very time consuming because of the huge factorials which have to be calculated during the calculation process. An alternative method is to use the recursive formula which appeared in Willmot [1986]. By defining

$$p_1 = \alpha_L(1 + 2\beta_L)^{-1/2}p_0,$$

the needed probabilities can be calculated using the identity

$$(1 + 2\beta_L)n(n-1)p_n = 2\beta_L(n-1)(n-3/2)p_{n-1} + \alpha_L^2 p_{n-2}. \quad (4.12)$$

Formula (4.9) is useful if we would like to have the explicit expression for the probability $P(W = w)$, but it is definitely better to use (4.12) in numerical calculations. \square

Theorem 4.3. *Suppose that Model 4 (see Table 3.1) holds. Then, if $C \sim IG(\alpha_C, \beta_C)$ and $L \sim \text{Gamma}(\alpha_L, \beta_L)$ (see Table 3.2), the needed SLP can be calculated as*

$$E[(S - a)^+] = \sum_{w=1}^{\infty} P(W = w) \times [\Phi(x_3)(w\alpha_C - a) + e^{x_1}\Phi(-x_2)(w\alpha_C + a)], \quad (4.13)$$

where $\Phi(\cdot)$ denotes the standard normal cdf, $P(W = w)$ is calculated using (4.4) and x_1, x_2, x_3 are defined as

$$\begin{aligned} x_1 &= 2 \frac{w\alpha_C}{\beta_C} > 0, \\ x_2 &= \frac{a + w\alpha_C}{\sqrt{\beta_C a}} > 0, \\ x_3 &= \frac{w\alpha_C - a}{\sqrt{\beta_C a}}. \end{aligned} \quad (4.14)$$

Proof. The Inverse Gaussian distribution is closed under convolution. Hence, conditionally on $W = w$ we have

$$\sum_{i=1}^w C_i \sim IG(w\alpha_c, \beta_c)$$

and the density of S equals

$$\begin{aligned} f_S(s) &= \sum_{w=1}^{\infty} P(W = w) f_{IG}(s; w\alpha_c, \beta_c) \\ &= \sum_{w=1}^{\infty} P(W = w) \frac{w\alpha_c}{\sqrt{2\pi\beta_c s^3}} \exp(-(s - w\alpha_c)^2 / (2\beta_c s)). \end{aligned}$$

Having the density of S we can calculate

$$\begin{aligned} E[(S - a)^+] &= \sum_{w=1}^{\infty} P(W = w) \\ &\quad \times \int_a^{\infty} (s - a) \frac{w\alpha_c}{\sqrt{2\pi\beta_c s^3}} \exp(-(s - w\alpha_c)^2 / (2\beta_c s)) ds. \end{aligned} \quad (4.15)$$

Since $L \sim \text{Gamma}$ we can use (4.4) to get $P(W = w)$ (see the proof of Theorem 4.1). The integration part can be calculated using some numerical algorithm, or it is possible to rewrite it using the standard normal distribution. To do it we denote

$$I := \int_a^\infty (s - a) f_{IG}(s; w\alpha_C, \beta_C) ds$$

and split it in the form

$$\begin{aligned} I &= \int_a^\infty s f_{IG}(s; w\alpha_C, \beta_C) ds - a \int_a^\infty f_{IG}(s; w\alpha_C, \beta_C) ds \\ &= E[IG(w\alpha_C, \beta_C)] - \int_0^a s f_{IG}(s; w\alpha_C, \beta_C) ds \\ &\quad - a \left(1 - \int_0^a f_{IG}(s; w\alpha_C, \beta_C) ds \right), \end{aligned}$$

where $IG(w\alpha_C, \beta_C)$ is a random variable with the Inverse Gaussian distribution and parameters $w\alpha_C$ and β_C . Under our parametrization $E[IG(w\alpha_C, \beta_C)] = w\alpha_C$. The integral $\int_0^a f_{IG}(s; w\alpha_C, \beta_C) ds$ represents the Inverse Gaussian cdf at point a and can be written using the standard normal cdf. It is not that obvious, but the second integral $\int_0^a s f_{IG}(s; w\alpha_C, \beta_C) ds$ can also be represented as a linear combination of two normal cdf's. Using the identities

$$\frac{d \left(\Phi \left(\frac{s-\alpha}{\sqrt{\beta s}} \right) + e^{2\alpha/\beta} \Phi \left(\frac{-s-\alpha}{\sqrt{\beta s}} \right) \right)}{ds} = f_{IG}(s; \alpha, \beta), \quad (4.16)$$

$$\frac{d \left(-\alpha \Phi \left(\frac{\alpha-s}{\sqrt{\beta s}} \right) - \alpha e^{2\alpha/\beta} \Phi \left(\frac{-s-\alpha}{\sqrt{\beta s}} \right) \right)}{ds} = s f_{IG}(s; \alpha, \beta), \quad (4.17)$$

we can write the needed integrals as

$$\begin{aligned} &\int_0^a f_{IG}(s; w\alpha_C, \beta_C) ds \\ &= \left[\Phi \left(\frac{s - w\alpha_C}{\sqrt{\beta_C s}} \right) + e^{2w\alpha_C/\beta_C} \Phi \left(\frac{-s - w\alpha_C}{\sqrt{\beta_C s}} \right) \right] \Big|_{s=0}^{s=a} \\ &- \left[\Phi \left(\frac{s - w\alpha_C}{\sqrt{\beta_C s}} \right) + e^{2w\alpha_C/\beta_C} \Phi \left(\frac{-s - w\alpha_C}{\sqrt{\beta_C s}} \right) \right] \Big|_{s=0} \\ &= \Phi \left(\frac{a - w\alpha_C}{\sqrt{\beta_C a}} \right) + e^{2w\alpha_C/\beta_C} \Phi \left(\frac{-a - w\alpha_C}{\sqrt{\beta_C a}} \right) \end{aligned}$$

and

$$\begin{aligned}
& \int_0^a s f_{IG}(s; w\alpha_C, \beta_C) ds \\
&= \left[-w\alpha_C \Phi\left(\frac{w\alpha_C - s}{\sqrt{\beta_C s}}\right) - w\alpha_C e^{2w\alpha_C/\beta_C} \Phi\left(\frac{-s - w\alpha_C}{\sqrt{\beta_C s}}\right) \right] \Bigg|_{s=a} \\
&\quad - \left[-w\alpha_C \Phi\left(\frac{w\alpha_C - s}{\sqrt{\beta_C s}}\right) - w\alpha_C e^{2w\alpha_C/\beta_C} \Phi\left(\frac{-s - w\alpha_C}{\sqrt{\beta_C s}}\right) \right] \Bigg|_{s=0} \\
&= -w\alpha_C \Phi\left(\frac{w\alpha_C - a}{\sqrt{\beta_C a}}\right) - w\alpha_C e^{2w\alpha_C/\beta_C} \Phi\left(\frac{-a - w\alpha_C}{\sqrt{\beta_C a}}\right) + w\alpha_C. \quad (4.18)
\end{aligned}$$

The final integral takes a form

$$\begin{aligned}
I &= w\alpha_C + w\alpha_C \Phi\left(\frac{w\alpha_C - a}{\sqrt{\beta_C a}}\right) + w\alpha_C e^{2w\alpha_C/\beta_C} \Phi\left(\frac{-a - w\alpha_C}{\sqrt{\beta_C a}}\right) - w\alpha_C \\
&\quad - a \left(1 - \Phi\left(\frac{a - w\alpha_C}{\sqrt{\beta_C a}}\right) - e^{2w\alpha_C/\beta_C} \Phi\left(\frac{-a - w\alpha_C}{\sqrt{\beta_C a}}\right) \right) \\
&= \Phi\left(\frac{w\alpha_C - a}{\sqrt{\beta_C a}}\right) (w\alpha_C - a) + e^{2w\alpha_C/\beta_C} \Phi\left(\frac{-a - w\alpha_C}{\sqrt{\beta_C a}}\right) (w\alpha_C + a). \quad (4.19)
\end{aligned}$$

By defining

$$\begin{aligned}
x_1 &= 2\frac{w\alpha_C}{\beta_C} > 0, \\
x_2 &= \frac{a + w\alpha_C}{\sqrt{\beta_C a}} > 0, \\
x_3 &= \frac{w\alpha_C - a}{\sqrt{\beta_C a}},
\end{aligned}$$

we get the final formula (4.13). \square

Formula (4.13) looks very nice, but again we have to point out numerical difficulties the user will certainly encounter. For the detailed explanation of the problem and its solution see Appendix A3.

Combining the previous results, we get

Theorem 4.4. *Suppose that Model 4 (see Table 3.1) holds. Then, if $C \sim IG(\alpha_C, \beta_C)$ and $L \sim IG(\alpha_L, \beta_L)$ (see Table 3.2), the needed SLP can be calculated using (4.13) and (4.8).*

Proof. Since $C \sim IG$, the general form of the SLP can be written in the form (4.13) (see Theorem 4.3). As soon as $L \sim IG$, the probabilities $P(W = w)$ have a form (4.8) according to Theorem 4.2. \square

4.1.2 Direct Monte-Carlo Simulations

Direct (or simple) Monte-Carlo simulation is another numerical tool which we are going to use for the SLP calculation. The advantage of this method is that it does not require any specific assumptions (like closedness under convolution in the method of convolutions) for the random variables of the model. However, it also has a drawback. In case the distributional assumptions of the model are complicated (in the sense of generation of the random variables involved), the method can be very time consuming. It can take several days of computer work to get the needed precision. Therefore, the method will be used when the distributional assumptions of the model are reasonably simple. Considering the list of distributions we are using in our research (see Table 3.2), the lognormal, Gamma and Poisson random variables can be generated reasonably fast, while the generation of IG random variables takes more time, which significantly complicates the simulation process. Moreover, for getting more insight in the structure of $E[(S-a)^+]$ Monte-Carlo simulation is less helpful than a reasonably simple analytic formula, see also Section 1.5. Therefore, the method will be used only when it is really necessary, i.e. when the remaining methods can not be used due to some reason. Considering our situation, we suggest not to use the simple Monte-Carlo simulations in case the IG distribution is involved in the model.

The direct (or simple) Monte-Carlo simulation of the SLP has a quite simple structure. The main step is to generate the $\mathbb{S}_{N_S} = \{S_1, S_2, \dots, S_{N_S}\}$ sample using the assumptions of Model 4 (see Chapter 3 and Table 3.1). Here N_S represents the number of simulations. Having this sample we fix the retention level a and calculate $(\mathbb{S}_{N_S} - a)^+ = \{(S_1 - a)^+, (S_2 - a)^+, \dots, (S_{N_S} - a)^+\}$. The simulated SLP is given by the right hand side of the expression

$$E[(S-a)^+] \approx \overline{(\mathbb{S}_{N_S} - a)^+} = \frac{1}{N_S} \sum_{i=1}^{N_S} (S_i - a)^+. \quad (4.20)$$

Obviously, the " \approx " sign in (4.20) can be replaced by the " $=$ " only when $\frac{N_S}{N_S} = \infty$. Otherwise, it remains a major question how close $E[(S-a)^+]$ and $\overline{(\mathbb{S}_{N_S} - a)^+}$ are. It is clear that the larger N_S , the higher the accuracy of the simulations, but it is very hard to say something specific about the simulations accuracy for given N_S . This type of question is very important since half of the analysis will be based on the simulations (the second half will be based on the convolutions), and we have to be sure that the accuracy of the simulations is reasonable for our purposes. For the evaluation of the accuracy of simulations we use the theory of confidence intervals. The idea is to construct intervals which contain $E[(S-a)^+]$ with some (quite high, like 0.95) confidence. Using this theory we can control the length of the confidence intervals by increasing (or decreasing) N_S .

Denote the sample mean by $\bar{X} = \overline{(\mathbb{S}_{N_S} - a)^+}$, the population mean by $\mu = E[(S-a)^+]$, the population standard deviation by $\sigma = St.dev[(S-a)^+]$ and the

sample standard deviation by $S_X = St.dev[(S_{N_S} - a)^+]$, which is defined by

$$S_X^2 = \frac{1}{N_S - 1} \sum_{i=1}^{N_S} ((S_i - a)^+ - \bar{X})^2 = \frac{1}{N_S - 1} \sum_{i=1}^{N_S} [(S_i - a)^+]^2 - \frac{N_S}{N_S - 1} \bar{X}^2. \quad (4.21)$$

As is well known from standard theory,

$$\left[\bar{X} - \frac{\Phi^{-1}\left(1 - \frac{\alpha}{2}\right) S_X}{\sqrt{N_S}}, \bar{X} + \frac{\Phi^{-1}\left(1 - \frac{\alpha}{2}\right) S_X}{\sqrt{N_S}} \right]. \quad (4.22)$$

is an approximate $(1 - \alpha)$ confidence interval for N_S large (which obviously is the case).

Hence, if we denote the desired length of the confidence interval by Len , the number of simulations which is enough to reach Len is given by

$$N_S = \left(2\Phi^{-1}\left(1 - \frac{\alpha}{2}\right) \frac{S_X}{Len} \right)^2. \quad (4.23)$$

However, we do not define the absolute length of the confidence interval which appears in (4.23). The relative length of the confidence interval (which will be defined in Chapter 5) will be used instead. Therefore the values of \bar{X} and S_X will be defined during the simulation process. A pilot study is performed before the main simulation procedure. The idea is to perform 10000 simulations to get the first approximation of \bar{X} and S_X and then continue simulating step by step (say by 100 simulations), updating the values of \bar{X} and S_X , until the needed precision (which is defined by the relative length of the confidence interval) is reached. In later chapters this theory will be used to control the precision of the numerical output while using the Monte-Carlo simulation technique.

4.1.3 Hybrid Monte-Carlo Simulations

So far we have described the direct Monte-Carlo simulation method. The main disadvantage of the method is the calculation time. Sometimes a large number of simulations has to be performed to get the needed length of the confidence interval. Complicated distributional assumptions of the model make this process very time consuming. In this section an alternative simulation method is presented which is much faster than the direct Monte-Carlo method. We suggest to use this alternative method when the IG distribution is involved in the model.

Our calculation target is $E[(S - a)^+]$. This means that we have to deal with the tail of the distribution of S . Therefore, when we use a direct Monte-Carlo simulation method, the vast majority of generated values of S are less than a . This means that we do not use these at all. To obtain the final result we use only a few percents of the total number of simulated values. It would be much faster to simulate (somehow) only the values from the tail of the distribution.

Indeed, there are a lot of importance sampling methods which suggest to change the underlying distribution, sample random variates from the needed region and get the answer as a weighted average. But unfortunately such methods are not applicable in our situation since we do not know the true distribution of S , which certainly has to play a role in the procedure of defining the weights.

Instead, we suggest to use a method which can be called the hybrid simulation method. It is a mixture of the analytical formulas and the direct Monte-Carlo simulation (therefore we called it hybrid).

According to the law of total expectation we can write that

$$E[(S - a)^+] = E[E[(S - a)^+ | W]].$$

Here S is defined by (4.1) and W by (4.2). The outside expectation is the expectation w.r.t W . Hence, using the direct Monte-Carlo method, we can write that

$$\begin{aligned} E[E[(S - a)^+ | W]] &\approx \frac{1}{N_S} \sum_{i=1}^{N_S} E[(S - a)^+ | W = w_i] \\ &= \frac{1}{N_S} \left[\sum_{w_i > 30} E[(S - a)^+ | W = w_i] + \sum_{w_i \leq 30} E[(S - a)^+ | W = w_i] \right]. \end{aligned} \quad (4.24)$$

Here we divided all the simulated values of W into two parts. The first part contains the realizations of W which are > 30 and the second part contains the realizations of W which are ≤ 30 . Given that $W = w_i$, S takes a form of a simple sum of random variables which (according to the Central Limit Theorem) can be approximated by a normal random variable as soon as the summation index w_i is large enough. In our situation $E[W] = \lambda$ is typically chosen as 400, see Table 3.8. This means that the values of W will usually be much larger than 30. In the situation we consider, see again Table 3.8, $P(W \leq 30)$ turns out to be approximately of order 10^{-100} , which makes it almost impossible to encounter the situation where a realization of W is ≤ 30 . Therefore, we ignore the second term in (4.24). Now consider the part of $E[(S - a)^+ | W = w_i]$ with $w_i > 30$. Let

$$\tilde{S} = \frac{S - w_i \mu_C}{\sqrt{w_i} \sigma_C} \quad (4.25)$$

and

$$\tilde{a} = \frac{a - w_i \mu_C}{\sqrt{w_i} \sigma_C}. \quad (4.26)$$

In view of (4.26) we can write that

$$E[(S - a)^+ | W = w_i] = \sqrt{w_i} \sigma_C E[(\tilde{S} - \tilde{a})^+ | W = w_i].$$

According to the Central Limit Theorem, we may approximate \tilde{S} by Z with $Z \sim N(0, 1)$ for $w_i > 30$ and get

$$\begin{aligned} E[(\tilde{S} - \tilde{a})^+ | W = w_i] &\approx E[(Z - \tilde{a})^+] \\ &= \int_{\tilde{a}}^{\infty} (z - \tilde{a})\phi(z)dz = \phi(\tilde{a}) - \tilde{a}(1 - \Phi(\tilde{a})). \end{aligned} \quad (4.27)$$

Combining (4.24) and (4.27) gives

$$E[(S - a)^+] \approx \frac{1}{N_S} \sum_{i=1}^{N_S} \sqrt{w_i} \sigma_C [\phi(\tilde{a}) - \tilde{a}(1 - \Phi(\tilde{a}))], \quad (4.28)$$

where ϕ denotes the standard normal pdf, Φ denotes the standard normal cdf and N_S is the number of simulations.

Note that (4.28) allows us to deal directly with the tail of the distribution of S . This means that all the simulated values will be used to get the final result. The convergence rate in that case should be higher. The simulation itself will be much faster since now we generate only W instead of the whole W, C_1, \dots, C_W .

Hence this is a definite improvement, but we should not forget that by using this method we add one more source of error, namely a systematic one. This means that the method does not converge to the true value anymore and some bias will remain. Replacing $(S_i - a)^+$ in \bar{X} and S_X^2 from (4.21) by

$$\sqrt{w_i} \sigma_C [\phi(\tilde{a}) - \tilde{a}(1 - \Phi(\tilde{a}))], \quad (4.29)$$

we get a confidence interval for the hybrid Monte-Carlo simulation. It is very hard to be more explicit in general, since matters depend on the underlying parameters. Intuitively it is clear that for large values of a we can get situations where the standard normal approximation is not the best one possible. Numerical analysis shows that the systematic error (when a is large, μ_G and γ_L is small) can be quite large in some situations. If a is large, we have to deal with the tail of the distribution. Small values of μ_G and γ_L lead to realizations of W which are probably not large enough to make the tail of the normalized sum close to the standard normal tail. In such situations it is better to use a more refined approximation. The Edgeworth expansions were tested as an alternative to the simple standard normal approximation. By matching more moments we may come closer to the tail of the approximated distribution. Without going into the calculation details, the needed expectation becomes

$$\begin{aligned} E[(\tilde{S} - \tilde{a})^+ | W = w_i] &\approx \phi(\tilde{a}) - \tilde{a}(1 - \Phi(\tilde{a})) \\ &+ \phi(\tilde{a}) \left(\frac{\kappa_3}{6} \tilde{a} + \frac{\kappa_4}{24} (\tilde{a}^2 - 1) + \frac{\kappa_3^2}{72} (\tilde{a}^4 - 6\tilde{a}^2 + 3) \right). \end{aligned} \quad (4.30)$$

Here κ_3 and κ_4 are the cumulants of the standardized sum \tilde{S} given $W = w_i$ (and not the cumulants of the total sum which are defined in Section 3.1!), that is

$$\kappa_3 = \mu_{3C}\sigma_C^{-3}w_i^{-1/2} \text{ and } \kappa_4 = (\mu_{4C}\sigma_C^{-4} - 3)w_i^{-1}.$$

The final formula for the hybrid simulation of *SLP* then becomes

$$E[(S - a)^+] \approx \frac{1}{N_S} \sum_{i=1}^{N_S} \sqrt{w_i} \sigma_C \left[\phi(\tilde{a}) - \tilde{a}(1 - \Phi(\tilde{a})) + \phi(\tilde{a}) \left(\frac{\kappa_3}{6} \tilde{a} + \frac{\kappa_4}{24} (\tilde{a}^2 - 1) + \frac{\kappa_3^2}{72} (\tilde{a}^4 - 6\tilde{a}^2 + 3) \right) \right]. \quad (4.31)$$

Numerical analysis of this method shows a low systematic error and very fast calculation speed. Figure 4.1 compares the desired direct and hybrid Monte-Carlo simulation methods. The comparison method is based on the *SLP* calculation. For each situation (with different underlying parameters), simulations were performed until the needed confidence level was reached. Figure 4.1(b) presents the number of simulations which was needed to reach the fixed confidence level. It can be noted that the number of hybrid simulations is always less than the number of direct simulations (especially for the large retention levels), which illustrates the faster convergence rate. Figure 4.1(a) illustrates the relative errors (convolution results play the role of the true values) of the two methods. It is seen that they are close to each other, which means that the accuracy level of the hybrid simulations method is practically the same as the one of the direct simulations.

The time gain (in hours) is presented in Figure 4.1(c). It can be seen that for large retention levels the time gain can reach 2 hours for a single situation. To give an example of the absolute time gain, we can mention that to obtain the above pictures more than 14 hours was spent on the direct simulations while the hybrid simulations did the same job during 11 minutes. So, the total time gain was about 14 hours. The more general numerical analysis showed that the hybrid simulation method is approximately 20 times faster than the direct Monte-Carlo.

4.1.4 Approximations

A brief description of the approximation techniques which are used in the present thesis is given in Section 1.5. These are the approximations which were used in Albers [1999], Reijnen [2003] and Reijnen et al. [2005]. Using the conclusions of these papers, the most useful approximations are the IG and the Gamma-IG in the sense that they showed the best results when Models 2 or 3 were assumed. For this reason only these two approximations will be considered as candidates for the approximations in Model 4.

The IG approximation was introduced in Chaubey et al. [1998]. The needed density $f_S(s)$ (the density of the aggregate sum S) was approximated by the IG density $f_{IG}(s - x_{0IG})$ (see Table 3.2). The underlying parameters α_{IG}, β_{IG} and x_{0IG} are chosen in such a way that the mean, variance and skewness of S are

equal to the mean, variance and skewness of the IG random variable with density $f_{IG}(s - x_{0IG})$. As a result we get the values of the parameters needed expressed in terms of $\mu_S, \sigma_S, \kappa_{3S}$

$$\alpha_{IG} = 3\sigma_S/\kappa_{3S}, \beta_{IG} = \kappa_{3S}\sigma_S/3 \text{ and } x_{0IG} = \mu_S - 3\sigma_S/\kappa_{3S}. \quad (4.32)$$

Our calculation target in this section is the *SLP* which is defined as

$$E[(S - a)^+] = \int_a^\infty (s - a)f_S(s)ds. \quad (4.33)$$

Using the IG approximation, the above integral is replaced by

$$\int_a^\infty (s - a)f_{IG}(s - x_{0IG})ds,$$

which can be written as, cf. (4.13),

$$\begin{aligned} e^{2\alpha_{IG}/\beta_{IG}} \Phi\left(-\frac{a - x_{0IG} + \alpha_{IG}}{\sqrt{\beta_{IG}(a - x_{0IG})}}\right) (\alpha_{IG} + a - x_{0IG}) \\ + \Phi\left(\frac{\alpha_{IG} - a + x_{0IG}}{\sqrt{\beta_{IG}(a - x_{0IG})}}\right) (\alpha_{IG} - a + x_{0IG}). \end{aligned} \quad (4.34)$$

The Gamma approximation was introduced in Seal [1977] and the density $f_S(s)$ is approximated there by the Gamma density $f_\gamma(s - x_{0G})$ (see Table 3.2) with parameters

$$\alpha_G = (2/\kappa_{3S})^2, \beta_G = 2/(\kappa_{3S}\sigma_S) \text{ and } x_{0G} = \mu_S - 2\sigma_S/\kappa_{3S}. \quad (4.35)$$

By using this approximation, the integral in (4.33) becomes

$$\int_a^\infty (s - a)f_\gamma(s - x_{0G})ds,$$

which can be written as, cf. (4.3),

$$\frac{\alpha_G}{\beta_G} [1 - F_\gamma(a - x_{0G}; \alpha_G + 1, \beta_G)] - (a - x_{0G}) [1 - F_\gamma(a - x_{0G}; \alpha_G, \beta_G)], \quad (4.36)$$

where F_γ is the Gamma cdf.

The Gamma-IG approximation is a mixture of the Gamma and the IG approximation (see Chaubey et al. [1998]). The density $f_S(s)$ is approximated in the following way

$$f_S(s) \approx f_{G-IG}(s) = \omega f_\gamma(s - x_{0G}) + (1 - \omega) f_{IG}(s - x_{0IG}).$$

The approximation has seven parameters. Three of these (α_{IG}, β_{IG} and x_{0IG}) come from the IG approximation, another three (α_G, β_G and x_{0G}) from the Gamma approximation and they are as before. The last one (ω) is used to fit the fourth cumulant. In other words, ω is chosen in such a way that

$$\omega\kappa_{4G} + (1 - \omega)\kappa_{4IG} = \kappa_{4S},$$

which defines

$$\omega = \frac{\kappa_{4IG} - \kappa_{4S}}{\kappa_{4IG} - \kappa_{4G}} = \frac{\frac{5}{3}\kappa_{3S}^2 - \kappa_{4S}}{\frac{5}{3}\kappa_{3S}^2 - \frac{3}{2}\kappa_{3S}^2}. \quad (4.37)$$

Note that the necessary condition for f_{G-IG} to be a density is $\omega \in [0, 1]$ or, equivalently, $3\kappa_{3S}^2/2 \leq \kappa_{4S} \leq 5\kappa_{3S}^2/3$. This is a quite strict limitation which, in our case, is often not satisfied. Nevertheless, we still can use f_{G-IG} outside this range as a simple function (not a density) which approximates the density needed. A similar trick was used in Albers [1999], where the Edgeworth expansions showed very good approximation results while not being density functions. In fact, on the interval in which we are interested ($S > a = \mu_S + k\sigma_S$ with $0 \leq k \leq 3$), often $f_{G-IG}(s)$ behaves like a density, in the sense that it is positive on this interval.

4.2 Variance of a Stop-Loss contract

The second quantity of interest which is considered in the present research is the variance of the Stop-Loss contract. It is clear that in reality *SLP* cannot be the total price of the Stop-Loss contract because the reinsurers position becomes very risky with the expected profit 0. Some positive loading definitely has to be added to the *SLP*. The value of such loading depends on many factors and it has to be acceptable for both sides: the insurance company and the reinsurer. It can be some fixed value or it can depend on some risk characteristic like mean or standard deviation. For the discussion of different kinds of insurance or reinsurance premiums we refer to Kaas et al. [2001]. That is why the second quantity which is considered to be important is the variance of the Stop-Loss contract. Since the variance can be expressed in terms of the first two moments, it suffices to consider only the second one; the first moment was already considered in the previous section in the form of *SLP*. Therefore, $E[((S - a)^+)^2]$ will be the calculation target of this section. Just like in the previous section, we are going to consider three calculation techniques: convolutions, simulations and approximations.

4.2.1 Convolutions

The general expression for the calculation of the second moment can be written as

$$E[((S - a)^+)^2] = \sum_{w=1}^{\infty} P(W = w) \int_a^{\infty} (s - a)^2 f_C^{*w}(s; \alpha_c, \beta_c) ds, \quad (4.38)$$

with W as given in (4.2). Here f_C^{*w} stands for the w -fold convolution of the density f_C . Just like in the *SLP* case, the second moment calculation can be divided into two parts: calculation of $P(W = w)$ and calculation of the integral $\int_a^\infty (s - a)^2 f_C^{*w}(s; \alpha_c, \beta_c) ds$.

Theorem 4.5. *Suppose that Model 4 (see Table 3.1) holds. Then, for the calculation of $E[(S - a)^+]$ we have the following results:*

- (i) *in case $C \sim \text{Gamma}(\alpha_C, \beta_C)$ and $L \sim \text{Gamma}(\alpha_L, \beta_L)$ (see Table 3.2) use (4.38) with $P(W = w)$ defined in (4.4) and*

$$\begin{aligned} \int_a^\infty (s - a)^2 f_C^{*w}(s; \alpha_c, \beta_c) ds &= \frac{w\alpha_c(w\alpha_c + 1)}{\beta_c^2} [1 - F_\gamma(a; w\alpha_c + 2, \beta_c)] \\ &\quad - 2a \frac{w\alpha_c}{\beta_c} [1 - F_\gamma(a; w\alpha_c + 1, \beta_c)] \\ &\quad + a^2 [1 - F_\gamma(a; w\alpha_c, \beta_c)]. \end{aligned} \quad (4.39)$$

- (ii) *in case $C \sim \text{Gamma}(\alpha_C, \beta_C)$ and $L \sim \text{IG}(\alpha_L, \beta_L)$ use (4.38) with $P(W = w)$ defined in (4.8) and $\int_a^\infty (s - a)^2 f_C^{*w}(s; \alpha_c, \beta_c) ds$ defined in (4.39).*

- (iii) *in case $C \sim \text{IG}(\alpha_C, \beta_C)$ and $L \sim \text{Gamma}(\alpha_L, \beta_L)$ use (4.38) with $P(W = w)$ defined in (4.4) and*

$$\int_a^\infty (s - a)^2 f_C^{*w}(s; \alpha_c, \beta_c) ds = \int_a^\infty (s - a)^2 f_{\text{IG}}(s; w\alpha_C, \beta_C) ds. \quad (4.40)$$

- (iv) *in case $C \sim \text{IG}(\alpha_C, \beta_C)$ and $L \sim \text{IG}(\alpha_L, \beta_L)$ use (4.38) with $P(W = w)$ defined in (4.8) and $\int_a^\infty (s - a)^2 f_C^{*w}(s; \alpha_c, \beta_c) ds$ defined in (4.40).*

Proof. Assuming that $L \sim \text{Gamma}(\alpha_L, \beta_L)$ or $L \sim \text{IG}(\alpha_L, \beta_L)$, the calculation of $P(W = w)$ can be performed directly, using formulas (4.4) or (4.8) respectively (see Theorems 4.1 and 4.2). In case $C \sim \text{IG}(\alpha_C, \beta_C)$ identity (4.40) follows directly from the closedness under convolution property of the IG distribution. In case $C \sim \text{Gamma}(\alpha_C, \beta_C)$, the integral $\int_a^\infty (s - a)^2 f_C^{*w}(s; \alpha_c, \beta_c) ds$ can be split as

$$\int_a^\infty (s - a)^2 f_C^{*w}(s; \alpha_c, \beta_c) ds = I_3 - 2aI_1 + a^2I_2,$$

where I_1, I_2 are defined in (4.5) and

$$\begin{aligned}
 I_3 &= \int_a^\infty s^2 f_\gamma(s; w\alpha_c, \beta_c) ds \\
 &= \int_a^\infty s^{(w\alpha_c+2)-1} \frac{\beta_c^{w\alpha_c} e^{-\beta_c s}}{\Gamma(w\alpha_c)} ds \\
 &= \frac{w\alpha_c(w\alpha_c+1)}{\beta_c^2} \int_a^\infty s^{(w\alpha_c+2)-1} \frac{\beta_c^{w\alpha_c+2} e^{-\beta_c s}}{\Gamma(w\alpha_c+2)} ds \\
 &= \frac{w\alpha_c(w\alpha_c+1)}{\beta_c^2} \left(1 - \int_0^a s^{(w\alpha_c+2)-1} \frac{\beta_c^{w\alpha_c+2} e^{-\beta_c s}}{\Gamma(w\alpha_c+2)} ds\right) \\
 &= \frac{w\alpha_c(w\alpha_c+1)}{\beta_c^2} [1 - F_\gamma(a; w\alpha_c+2, \beta_c)]. \quad (4.41)
 \end{aligned}$$

Application of (4.6), (4.7) and (4.41) directly gives (4.39). \square

4.2.2 Direct Monte-Carlo Simulations

The main idea of the direct Monte-Carlo simulations is described in detail in Section 4.1.2. Here we present the confidence interval calculation approach, which is different from the *SLP* case.

Now we are interested in the variance of the Stop-Loss contract or, what is more useful in practice, in its standard deviation. Just like before we generate the $\mathbb{S}_{N_S} = \{S_1, S_2, \dots, S_{N_S}\}$ sample using Model 4 assumptions (see Table 3.1). Here N_S represents the total number of simulations. Having this sample we calculate $(\mathbb{S}_{N_S} - a)^+ = \{(S_1 - a)^+, (S_2 - a)^+, \dots, (S_{N_S} - a)^+\}$ and the simulated standard deviation of the Stop-Loss contract is

$$\begin{aligned}
 St.dev[(S - a)^+] &\approx \sqrt{\frac{1}{N_S - 1} \sum_{i=1}^{N_S} ((S_i - a)^+ - \overline{(\mathbb{S}_{N_S} - a)^+})^2} \\
 &= \sqrt{\frac{1}{N_S - 1} \sum_{i=1}^{N_S} [(S_i - a)^+]^2 - \frac{N_S}{N_S - 1} (\overline{(\mathbb{S}_{N_S} - a)^+})^2}, \quad (4.42)
 \end{aligned}$$

where

$$\overline{(\mathbb{S}_{N_S} - a)^+} = \frac{1}{N_S} \sum_{i=1}^{N_S} (S_i - a)^+.$$

To construct a confidence interval for $St.dev[(S - a)^+]$ we divide N_S into T parts: $N_{S_1}, N_{S_2}, \dots, N_{S_T}$. It is clear that in that case $\sum_{i=1}^T N_{S_i} = N_S$. Next, we

calculate the sequence $S_{X_1}, S_{X_2}, \dots, S_{X_T}$, which is defined through

$$\begin{aligned} S_{X_j} &= \frac{1}{N_{S_j} - 1} \sum_{i=1}^{N_{S_j}} ((S_{ij} - a)^+ - \overline{(S_{N_{S_j}} - a)^+})^2 \\ &= \frac{1}{N_{S_j} - 1} \sum_{i=1}^{N_{S_j}} [(S_{ij} - a)^+]^2 - \frac{N_{S_j}}{N_{S_j} - 1} (\overline{(S_{N_{S_j}} - a)^+})^2, \end{aligned} \quad (4.43)$$

for $j = 1 \dots T$ and can be presented as a realization of a random variable Y . Then $E[Y] = Var[(S - a)^+]$. Denote the sample mean by $\bar{Y} = \frac{1}{T} \sum_{j=1}^T S_{X_j}$. As soon as the sample mean \bar{Y} is considered, we can use the confidence interval theory of Section 4.1.2, where the confidence interval was calculated for the mean. As a result we get the confidence interval for the variance of the Stop-Loss contract as

$$\left[\bar{Y} - \frac{\Phi^{-1}\left(1 - \frac{\alpha}{2}\right) S_Y}{\sqrt{T}}, \bar{Y} + \frac{\Phi^{-1}\left(1 - \frac{\alpha}{2}\right) S_Y}{\sqrt{T}} \right], \quad (4.44)$$

with S_Y defined as

$$S_Y = \sqrt{\frac{1}{T-1} \sum_{i=1}^T (S_{X_i} - \bar{Y})^2} = \sqrt{\frac{1}{T-1} \sum_{i=1}^T S_{X_i}^2 - \frac{T}{T-1} (\bar{Y})^2}.$$

Having a confidence interval for $Var[(S - a)^+]$, we automatically get a confidence interval for $St.dev[(S - a)^+]$ as

$$\left[\sqrt{\bar{Y} - \frac{\Phi^{-1}\left(1 - \frac{\alpha}{2}\right) S_Y}{\sqrt{T}}}, \sqrt{\bar{Y} + \frac{\Phi^{-1}\left(1 - \frac{\alpha}{2}\right) S_Y}{\sqrt{T}}} \right], \quad (4.45)$$

Remark 4.2. Note that the method we use for the construction of the confidence interval for the standard deviation is not the only one which is available in practice. An alternative approach is to utilize the asymptotic normality of the sample standard deviation itself. Knowing that

$$S_X^2 \sim AN\left(\sigma^2, \frac{\mu_4 - \sigma^4}{N_S}\right),$$

we directly obtain

$$\frac{(S_X^2 - \sigma^2)\sqrt{N_S}}{\sqrt{\mu_4 - \sigma^4}} \sim AN(0, 1)$$

and the confidence interval for the variance takes the form

$$\left[S_X^2 - \frac{\Phi^{-1}\left(1 - \frac{\alpha}{2}\right) \sqrt{\mu_4 - \sigma^4}}{\sqrt{N_S}}, S_X^2 + \frac{\Phi^{-1}\left(1 - \frac{\alpha}{2}\right) \sqrt{\mu_4 - \sigma^4}}{\sqrt{N_S}} \right].$$

The fourth central moment μ_4 and the variance σ^2 are replaced by their estimates and the asymptotic normality continue to hold. The confidence interval for the standard deviation can be obtained by taking the square root in the both sides of the interval.

We do not have any specific reason to use one or another method. The one we use in practice nicely fits our simulation algorithm, therefore it was chosen as a calculation tool for the confidence intervals. \square

4.2.3 Hybrid Monte-Carlo Simulations

In this section the alternative simulation method for the variance of a Stop-Loss contract will be discussed. The idea and the motivation of the method was already presented in Section 4.1.3. Therefore, here we present only the main results.

Section 4.1.3 offers formulas for the calculation of *SLP* (see (4.28) and (4.31)). Therefore, only the second moment of the Stop-Loss contract will be considered here; the variance can be obtained using the identity $Var[(S - a)^+] = E[((S - a)^+)^2] - (E[(S - a)^+])^2$. Hence, the calculation target now is $E[\max(S - a, 0)^2] = E[((S - a)^+)^2]$ which, according to the idea of the method, can be approximated by

$$E[((S - a)^+)^2] = E[E[((S - a)^+)^2|W]] \approx \frac{1}{N_S} \sum_{i=1}^{N_S} E[((S - a)^+)^2|W = w_i]. \quad (4.46)$$

We are going to utilize several identities which already appeared in the previous sections.

Lemma 4.6. *Suppose that \tilde{a} is a constant. Then, denoting the standard normal pdf and cdf by $\phi(\cdot)$ and $\Phi(\cdot)$, we have the following identities*

$$\int_{\tilde{a}}^{\infty} \phi(z) dz = 1 - \Phi(\tilde{a}), \quad (4.47)$$

$$\int_{\tilde{a}}^{\infty} z \phi(z) dz = \phi(\tilde{a}), \quad (4.48)$$

$$\int_{\tilde{a}}^{\infty} z^2 \phi(z) dz = \tilde{a} \phi(\tilde{a}) + 1 - \Phi(\tilde{a}), \quad (4.49)$$

$$\int_{\tilde{a}}^{\infty} (z - \tilde{a})^2 \phi(z) dz = (\tilde{a}^2 + 1)(1 - \Phi(\tilde{a})) - \tilde{a} \phi(\tilde{a}). \quad (4.50)$$

Moreover, writing $\phi^{(j)}$ for the j^{th} derivative of ϕ , we have for $j \geq 3$

$$\int_{\tilde{a}}^{\infty} (z - \tilde{a})^2 \phi^{(j)}(z) dz = -2\phi^{(j-3)}(\tilde{a}) = (-1)^j 2\phi(\tilde{a}) H_{j-3}(\tilde{a}), \quad (4.51)$$

where H_j is the j^{th} Hermite polynomial.

Proof. Using

$$\begin{aligned}\Phi^{(1)}(z) &= \phi(z), \\ \phi^{(1)}(z) &= -z\phi(z), \\ \phi^{(2)}(z) &= (z^2 - 1)\phi(z),\end{aligned}$$

the results in (4.47)-(4.50) are straightforward. By partial integration we obtain for $j \geq 3$

$$\begin{aligned}\int_{\tilde{a}}^{\infty} (z - \tilde{a})^2 \phi^{(j)}(z) dz &= \int_{\tilde{a}}^{\infty} (z - \tilde{a})^2 d\phi^{(j-1)}(z) \\ &= (z - \tilde{a})^2 \phi^{(j-1)}(z) \Big|_{\tilde{a}}^{\infty} - \int_{\tilde{a}}^{\infty} 2(z - \tilde{a}) \phi^{(j-1)}(z) dz \\ &= - \int_{\tilde{a}}^{\infty} 2(z - \tilde{a}) d\phi^{(j-2)}(z) = -2(z - \tilde{a}) \phi^{(j-2)}(z) \Big|_{\tilde{a}}^{\infty} \\ &+ \int_{\tilde{a}}^{\infty} 2\phi^{(j-2)}(z) dz = 2 \int_{\tilde{a}}^{\infty} d\phi^{(j-3)}(z) = 2\phi^{(j-3)}(z) \Big|_{\tilde{a}}^{\infty} \\ &= -2\phi^{(j-3)}(\tilde{a}) = -2(-1)^{j-3} \phi(\tilde{a}) H_{j-3}(\tilde{a}) \\ &= (-1)^j 2\phi(\tilde{a}) H_{j-3}(\tilde{a}),\end{aligned}$$

thus completing the proof of (4.51). \square

To get the expectation we have to integrate $\int_a^{\infty} (s - a)^2 f_S(s) ds$. Using the definitions of \tilde{S} (see (4.25)) and \tilde{a} (see (4.26)), we can write that

$$E[(S - a)^{2+} | W = w_i] = w_i \sigma_C^2 E[(\tilde{S} - \tilde{a})^{2+} | W = w_i].$$

According to the Central Limit Theorem, we can approximate \tilde{S} by $Z \sim N(0, 1)$ when w_i is sufficiently large. Hence, we can write that

$$\begin{aligned}E[(\tilde{S} - \tilde{a})^{2+} | W = w_i] &\approx E[(Z - \tilde{a})^{2+} | W = w_i] \\ &= \int_{\tilde{a}}^{\infty} (z - \tilde{a})^2 \phi(z) dz = (\tilde{a}^2 + 1)(1 - \Phi(\tilde{a})) - \tilde{a}\phi(\tilde{a}).\end{aligned}\quad (4.52)$$

Direct application of (4.46) and (4.52) yields

$$E[(S - a)^{2+}] \approx \frac{1}{N_S} \sum_{i=1}^{N_S} w_i \sigma_C^2 [(\tilde{a}^2 + 1)(1 - \Phi(\tilde{a})) - \tilde{a}\phi(\tilde{a})].\quad (4.53)$$

Such a calculation method has all the advantages which were explained in Section 4.1.3. These are a very fast convergence rate and high accuracy (compared

to the direct Monte-Carlo simulations) when the retention level is not very high. When we have to deal with the tail of the distribution of S we can get situations where the systematic error is quite high. In that case it is better to use Edgeworth expansions instead of the standard normal approximation. In that case $f_{\tilde{S}}(s)$ in (4.52) is replaced not by $\phi(s)$ but by

$$\phi(s) - \kappa_3 \frac{\phi^{(3)}(s)}{6} + \kappa_4 \frac{\phi^{(4)}(s)}{24} + \kappa_3^2 \frac{\phi^{(6)}(s)}{72}.$$

Here κ_3 and κ_4 are the cumulants of the normalized sum \tilde{S} given $W = w_i$. In that case, using (4.51), we obtain

$$\begin{aligned} E[(\tilde{S} - \tilde{a})^{2+} | W = w_i] & \\ & \approx \int_{\tilde{a}}^{\infty} (s - \tilde{a})^2 \left(\phi(s) - \kappa_3 \frac{\phi^{(3)}(s)}{6} + \kappa_4 \frac{\phi^{(4)}(s)}{24} + \kappa_3^2 \frac{\phi^{(6)}(s)}{72} \right) ds \\ & = (\tilde{a}^2 + 1)(1 - \Phi(\tilde{a})) - \tilde{a}\phi(\tilde{a}) + \phi(\tilde{a}) \left[\frac{\kappa_3}{3} + H_1(\tilde{a}) \frac{\kappa_4}{12} + H_3(\tilde{a}) \frac{\kappa_3^2}{36} \right], \end{aligned} \quad (4.54)$$

from which directly follows

$$\begin{aligned} E[(S - a)^{2+}] & \approx \frac{1}{N_S} \sum_{i=1}^{N_S} w_i \sigma_C^2 \left[(\tilde{a}^2 + 1)(1 - \Phi(\tilde{a})) \right. \\ & \quad \left. - \tilde{a}\phi(\tilde{a}) + \phi(\tilde{a}) \left[\frac{\kappa_3}{3} + H_1(\tilde{a}) \frac{\kappa_4}{12} + H_3(\tilde{a}) \frac{\kappa_3^2}{36} \right] \right]. \end{aligned} \quad (4.55)$$

4.2.4 Approximations

Approximation techniques which are used in the calculation of the second moment of the Stop-Loss contract are described in Section 4.1.4. These are three approximations: Gamma, IG and Gamma-IG. Since the idea of the approximation is based on the approximation of the underlying density, all the parameters involved stay the same (see (4.35) for the parameters of the Gamma approximation, (4.32) for the IG approximation and (4.37) for the last parameter of the Gamma-IG approximation). The following Lemma gives the expressions for the approximation of the second moment of the Stop-Loss contract when different approximations are applied.

Lemma 4.7. *Suppose that Model 4 (see Table 3.1) holds. Then, for the approximation of $E[(S - a)^+{}^2]$ we have the following results:*

- (i) in case the Gamma approximation is used, the needed expectation can be obtained as

$$\begin{aligned} & \int_a^\infty (s-a)^2 f_G(s-x_{0G}; \alpha_G, \beta_G) ds \\ &= \frac{\alpha_G(\alpha_G+1)}{\beta_G^2} [1 - F_\gamma(a-x_{0G}; \alpha_G+2, \beta_G)] \\ & \quad - 2(a-x_{0G}) \frac{\alpha_G}{\beta_G} [1 - F_\gamma(a; \alpha_G+1, \beta_G)] \\ & \quad + (a-x_{0G})^2 [1 - F_\gamma(a-x_{0G}; \alpha_G, \beta_G)], \quad (4.56) \end{aligned}$$

where parameters $\alpha_G, \beta_G, x_{0G}$ are defined in (4.35).

- (ii) in case the IG approximation is used, the needed expectation can be obtained as $\int_a^\infty (s-a)^2 f_{IG}(s-x_{0IG}; \alpha_{IG}, \beta_{IG}) ds$, where parameters $\alpha_{IG}, \beta_{IG}, x_{0IG}$ are defined in (4.32).
- (iii) in case the Gamma-IG approximation is used, the needed expectation can be obtained as $\int_a^\infty (s-a)^2 f_{G-IG}(s) ds$, where

$$f_{G-IG}(s) = \omega f_\gamma(s-x_{0G}) + (1-\omega) f_{IG}(s-x_{0IG})$$

and parameters $\alpha_G, \beta_{IG}, x_{0G}, \alpha_{IG}, \beta_{IG}, x_{0IG}, \omega$ are defined in (4.35), (4.32) and (4.37).

Proof. The detailed description of the approximation techniques is given in Section 4.1.4. Therefore (ii) and (iii) follow directly from the definition of the approximations. The result presented in (i) easily follows from (4.39). \square

4.3 Value at Risk of the aggregated claim amount

The last quantity which will be investigated is the VaR of the aggregated claim amount. VaR is a quite popular risk measure in the financial world. For given p , the VaR is defined as $VaR = s = F_S^{-1}(p)$, where $F_S(s) = P(S \leq s)$ and therefore it can be presented as an upper limit for the aggregated claim given some probability (risk) value. As was already mentioned, three calculation methods of VaR will be discussed in this section.

Remark 4.3. The method of hybrid simulation is not applicable here in the form it was presented in Sections 4.1.3 and 4.2.3. The calculation of VaR does not involve the calculation of an expectation. Therefore, direct application of the law of total expectation does not help here. Nevertheless, the idea of the method can still be used in a slightly modified form. By conditioning on $W = w$, the distribution of S takes a quite simple form in case the distribution of C is closed under convolution. Or, if it is not, we can approximate the distribution

of S by some suitable distribution, like it was done in Sections 4.1.3 and 4.2.3. Note that the Normal and Edgeworth approximations are not the only ones which can be used in the method. In general, any reasonable approximation (including the approximations which are described in Section 4.1.4) can be easily applied here. In case the approximation applied is closed under convolution (Normal or IG), the distribution of S takes the form of that approximation and S becomes a simple random variable. Hence in both cases, whether the approximation was used or not, we do not have to simulate the whole sequence W, C_1, C_2, \dots , which will significantly improve the simulation speed. However, this method was not tested on the VaR and is not used in our research. \square

4.3.1 Convolutions

From a first glance, a very straightforward way to get the VaR seems to be the convolution method. The aggregated claim S is given by (4.1) with W defined in (4.2). This gives us the opportunity to write $F_S(s)$ as

$$F_S(s) = \sum_{w=1}^{\infty} P(W = w) F_C^{*w}(s), \quad (4.57)$$

where F_C^{*w} stands for the w -fold convolution of the distribution F_C . In case the distribution of C is closed under convolution it is possible to get the closed form expression for $F_C^{*w}(s)$. However, our calculation target now is the VaR , which can be presented as the solution (w.r.t. s) of

$$\sum_{w=1}^{\infty} P(W = w) F_C^{*w}(s) = p, \quad (4.58)$$

with the given risk level p . The probability $P(W = w)$ itself has a quite complicated structure (see Section 4.1.1). Therefore, there is practically no possibility to obtain the analytical solution for this problem. The numerical approach is also not a good idea since the iteration procedure will take too much time. Therefore, the general suggestion is not to use the convolution techniques in the calculation of VaR .

4.3.2 Direct Monte-Carlo Simulations

The second calculation method we would like to discuss is the direct Monte-Carlo simulation. This method does not have any limitations and the accuracy of the result depends only on the number of simulations and the random number generator. The idea is simple. Having the number of simulations N_S and the risk level p as an input, we simulate S N_S times according to the chosen parameters of Model 4 (see Table 3.8). As a result we obtain the sequence S_1, S_2, \dots, S_{N_S} . The next step is to order that sequence in order to get $S_{(1)} \leq S_{(2)} \leq \dots \leq S_{(N_S)}$ with

the minimum value $S_{(1)}$ and the maximum value $S_{(N_S)}$. Under such a setup, the simulated VaR can be found as the $[N_S \times p]$ -th member of the ordered sequence (square brackets are introduced here as the rounding operator). In other words $S_{([N_S \times p])}$ will be the simulated VaR of the given risk level p .

To get the confidence interval it is necessary to reorganize the calculation algorithm. Having the total number of simulations N_S , we divide it into T parts, $N_{S_1}, N_{S_2}, \dots, N_{S_T}$, clearly with the property $\sum_{i=1}^T N_{S_i} = N_S$. Next, we generate the matrix

$$\begin{pmatrix} S_{11} & S_{12} & \dots & S_{1N_{S_1}} \\ S_{21} & S_{22} & \dots & S_{2N_{S_2}} \\ \dots & \dots & \dots & \dots \\ S_{T1} & S_{T2} & \dots & S_{TN_{S_T}} \end{pmatrix},$$

according to the chosen parameters of Model 4. Each row of the matrix we order in ascending order to get the matrix of the form

$$\begin{pmatrix} S_{(11)} & \leq & S_{(12)} & \leq & \dots & \leq & S_{(1N_{S_1})} \\ S_{(21)} & \leq & S_{(22)} & \leq & \dots & \leq & S_{(2N_{S_2})} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ S_{(T1)} & \leq & S_{(T2)} & \leq & \dots & \leq & S_{(TN_{S_T})} \end{pmatrix}.$$

Having the risk level p , we are interested in the vector

$$(S_{(1[N_{S_1} \times p])}, S_{(2[N_{S_2} \times p])}, \dots, S_{(T[N_{S_T} \times p])}) = (V_1, V_2, \dots, V_T),$$

which can be presented as a realization of a random variable V . The simulated value at risk in that case will be the sample mean

$$\bar{V} = \frac{1}{T} \sum_{i=1}^T V_i.$$

When the number of simulations is large, the value of $E[V]$ will be close to $F_S^{-1}(p)$ and therefore we use it as an approximation of the VaR . As soon as we consider $E[V]$ as the calculation target, we can use the theory of confidence intervals which was presented earlier in this chapter. As a result we get the confidence interval for the VaR of the aggregated claim amount as

$$\left[\bar{V} - \frac{\Phi^{-1}\left(1 - \frac{\alpha}{2}\right) S_V}{\sqrt{T}}, \bar{V} + \frac{\Phi^{-1}\left(1 - \frac{\alpha}{2}\right) S_V}{\sqrt{T}} \right], \quad (4.59)$$

where S_V is defined as

$$S_V = \sqrt{\frac{1}{T-1} \sum_{i=1}^T (V_i - \bar{V})^2} = \sqrt{\frac{1}{T-1} \sum_{i=1}^T V_i^2 - \frac{T}{T-1} (\bar{V})^2}.$$

Remark 4.4. Just like in the case of the standard deviation (see Remark 4.2) there is an alternative method to obtain the confidence interval for VaR . The alternative approach is to utilize the asymptotic normality of the sample quantile. Denote by ξ_p the p -quantile and by $\hat{\xi}_{pn}$ the sample p -quantile. Then, knowing that

$$\hat{\xi}_{pn} \sim AN \left(\xi_p, \frac{p(p-1)}{f^2(\xi_p)N_S} \right),$$

we directly obtain

$$\frac{(\hat{\xi}_{pn} - \xi_p)f(\xi_p)\sqrt{N_S}}{\sqrt{p(p-1)}} \sim AN(0, 1)$$

and the confidence interval for the VaR takes the form

$$\left[\hat{\xi}_{pn} - \frac{\Phi^{-1} \left(1 - \frac{\alpha}{2} \right) \sqrt{p(p-1)}}{f(\xi_p)\sqrt{N_S}}, \hat{\xi}_{pn} + \frac{\Phi^{-1} \left(1 - \frac{\alpha}{2} \right) \sqrt{p(p-1)}}{f(\xi_p)\sqrt{N_S}} \right].$$

In practice the quantile ξ_p and the value of $f(\xi_p)$ have to be replaced by their estimates. Again, asymptotic normality continues to hold in case of consistent estimation. It is natural to take the sample quantile as an estimate of ξ_p and the density f can be replaced by some approximation, e.g. the kernel estimate. \square

4.3.3 Approximations

One more calculation method of VaR described here is approximation. All the approximations which are utilized in the present thesis are described in Section 4.1.4. However, only the IG approximation will be used in the calculation of VaR . The IG approximation combines two essential features: relative simplicity and relative accuracy (for the accuracy testing see Chapter 5). The alternative approximations are Gamma and Gamma-IG. The Gamma approximation has a quite simple form, but it is not sufficiently accurate for the approximation of the tail of the distribution of S . The Gamma-IG is accurate but it has a very complicated form which is not suitable for further analysis.

As was described in Section 4.1.4, the approximation of $F_S(s)$ is obtained by matching the mean, variance and skewness of S with the mean, variance and skewness of the IG random variable (see (4.32)). As a result we obtain $F_{IG}(s)$ with the parameters given by (4.32) as approximation of $F_S(s)$. For given risk level p , the approximation of VaR is defined as $F_{IG}^{-1}(p)$, which is obtained by solving

$$F_{IG}(s) - p = 0$$

w.r.t. s . It would be a quite complicated task to solve such an equation analytically. Fortunately, we do not need it. All the further analysis will require only the single values of VaR when all the underlying parameters are given. Therefore, we define VaR as a given function $F_{IG}^{-1}(p)$ (analogically as the definition of the inverse of the standard normal distribution $\Phi^{-1}(p)$) and use mathematical software (like MAPLE) to obtain the numerical results.

Chapter 5

Accuracy of approximations

The essential target of interest for most insurance (or reinsurance) companies is the aggregated claim amount S during some reference period, let us say one year. It is clear that nobody knows what will be the total claim amount at the end of the next year. Therefore S should be modeled by some random variable or, what is more natural, by the sum of random variables which can represent different claim sizes. Different techniques of classical modeling were described in Section 1.1. The more advanced modelling which is used in the present thesis was introduced and described in Chapter 3.

Having a model for S , it is very important to determine its distribution. By using this distribution it is possible to calculate different types of premiums, price contracts and to evaluate and control insurance portfolios. Unfortunately the distribution of S usually has a very complicated structure and often can not be written in a closed form expression. Monte-Carlo simulation can be the solution, but it is a time-consuming method and does not easily reveal the properties of the distribution of S under different circumstances. That is why it is very important to have good approximation methods, which can be applied to get approximations for different insurance products and risk measures. The topic of the importance of approximations was also covered in Section 1.5.

The goal of this chapter is to determine good approximations (and the best among them) for the distribution of the aggregated claim amount S assuming the overdispersion model which was presented in Chapter 3 (see Model 4 in Table 3.1). Similar work was performed in Reijnen et al. [2005], where five approximations (Normal power, Edgeworth expansions, Gamma, IG and Gamma-IG) were tested in detail, assuming different models for S and different underlying distributions. As a result, a relatively simple rule of thumb, which specifies the best approximation for a given range of parameters, was constructed. Here we will try to follow the same strategy and come up with a similar rule of thumb, which will serve as an approximation guide when Model 4 is applied. Using the conclusions of Reijnen et al. [2005], the most useful approximations are the IG

and the Gamma-IG in the sense that they showed the best results when Models 2 or 3 (see Table 3.1) were assumed. For this reason only these two approximations will be considered as candidates for the approximations of the distribution of S when Model 4 is assumed.

In Reijnen et al. [2005] the whole analysis was based on the SLP , which is indeed a good tool as soon as we are interested in the tail of the distribution. In this chapter, besides the SLP , two more quantities will be considered: the variance of the Stop-Loss contract and the VaR , which is often used as a risk measure. These two quantities often play an important role in the premium calculation, evaluation and control of the insurance portfolio. Therefore, we have to be sure that our approximations can still be applied for the calculation of these quantities. The chapter is based on the technical report of Lukocius et al. [2007].

5.1 Retention level and criterion

In Chapter 3 the values of the model parameters which form the region of interest were introduced (see Table 3.7). In order to simplify the analysis, a number of representative values was presented for the region of interest, for each of the parameters considered (see Table 3.8). Therefore, all the testing procedures will be based on Table 3.8 and the conclusions will be given for the entire region of interest.

Nevertheless, there are two quantities which still have to be chosen. These are not the parameters of the model but belonging to the risk measures considered. Therefore, they were not defined in Section 3.4. The first one is the retention level a which appears in the SLP and the variance of the Stop-Loss contract. The second one is the risk level p in the definition of VaR .

It is hard to define the region of interest for the retention level in an absolute sense. Different parameters of the model change the calculation scale and the region of interest defined by the absolute retention level has to be changed according to that scale. Therefore, we will use the approach which was used in Reijnen [2003] and Reijnen et al. [2005]. The retention level was introduced there as a linear combination of μ_{S_1} and σ_{S_1} by

$$a = \mu_{S_1} + k\sigma_{S_1}.$$

Using such a definition we, instead of varying a directly, consider different values of k which fall in the interval $[1, 3]$ and in this way we avoid the scale problem. Here we denote by S_1 the aggregated sum of Model 1 (see Table 3.1). Therefore, the region of interest for the retention level a is given by the values of $k \in [1, 3]$, with as representative values 1, 2, 3. Do note that the scaling of a is based on Model 1, regardless of the actual model we are considering. This is necessary to allow meaningful comparison among the models: if different a 's would be used (e.g. $a = \mu_{S_4} + k\sigma_{S_4}$ as well), the comparison would become severely biased.

In contrast with the retention level, the risk level p can be defined in an absolute sense, without an intermediate step. Since we are interested in the tail of the distribution of S , the probabilities 0.9, 0.95, 0.99, 0.9975 and 0.9995 are proposed as representative values for the region of interest $0.9 \leq p \leq 0.9995$.

All decisions about the accuracy of the approximations will be based on a criterion which appeared for the first time in Reijnen [2003]. It was used later in Reijnen et al. [2005] in order to determine not so much the best approximation among the ones considered (as these can all be bad), but good ones. The idea here is to compare the approximations not merely to each other, but to check also whether or not the approximations satisfy a given absolute criterion. In this way, we can distinguish for example the following two, quite different scenarios. In the first, the winner may outperform its competitors, but nevertheless still be unsatisfactory in the sense that the criterion is not met. No acceptable solution exists here. In the second scenario the winner as well outperforms the competition, but now all the methods satisfy the absolute criterion. Hence we have at best a luxury problem here: several suitable solutions exist.

Since all the approximations match only a few moments of the distribution of S , it is obvious that the accuracy of the approximations depends on the retention level a (in case the SLP or the variance of the Stop-Loss contract is approximated) or the risk level p (in case the approximation target is the VaR). The larger the values of a or p , the farther we go into the tail of the distribution (where presumably matching of the moments has less influence on the behavior) and the relative difference between the approximated and the exact distribution will become larger as well. This also results in a larger relative error in the approximated SLP , variance of the Stop-Loss contract and VaR . Therefore Reijnen et al. [2005] chose a criterion which becomes more liberal when a increases. It was chosen in such a way that the absolute value of the relative error has to be smaller than 2.5% for $k = 0$, smaller than 30% for $k = 3$ and proportionally in between when the true SLP can be calculated by using the convolution method. In case the approximated $SLPs$ have to be compared with simulations, the criterion was chosen as 5% at $k = 0$ and 60% at $k = 3$. At first sight it may seem quite liberal to allow such large relative errors like 60%. Nevertheless, in Albers [1999] was already demonstrated that ignoring small dependencies can cause errors of e.g. 500%. Therefore in comparison, the criterion introduced still makes sense.

In the present thesis the criterion is chosen in a similar way. The linear criterion function starting at 2.5% for $k = 0$ (hence allowing 11.7% at $k = 1$) and giving 30% at $k = 3$ is defined by

$$C_{SLP}(k) = 0.3 \frac{k}{3} + \left(1 - \frac{k}{3}\right) 0.025. \quad (5.1)$$

Note that we use the same criterion in both cases: when the SLP or the variance of the Stop-Loss contract can be calculated by using the convolution method and when these quantities have to be simulated. The reason for that is hidden in the simulation process where (as was explained in Sections 4.1.2 and 4.2.2) the

accuracy of the simulated values is controlled more carefully compared to Reijnen et al. [2005] where a fixed number of simulations (100000) was used in all the cases considered. The criterion for the value at risk $C_{VaR}(p)$ is introduced in a similar way as

$$C_{VaR}(p) = \frac{(p-0.5)0.3}{0.4995} + \left(1 - \frac{p-0.5}{0.4995}\right) 0.025, \quad (5.2)$$

giving 2.5% at the 0.5 quantile (hence allowing 25% at $p = 0.9$) and 30% at the 0.9995 quantile.

Remark 5.1. Note that from a first glance it seems rather strange to use the same criterion for SLP , variance of the Stop-Loss contract and VaR . The criterion becomes milder in the tail, which is a natural assumption for SLP since the absolute values become smaller and hence not practically significant. The situation with the variance and VaR is different. The values of these quantities do not decrease with k and p . Moreover, it is obvious that VaR is strictly increasing with p . However, the criterion is constructed in view of the behavior of the dependence effect which becomes very huge in the tail. Therefore, our choice, which gives us more freedom from the approximation point of view, still makes sense. \square

Next follows the description of the accuracy testing process. For each situation from Table 3.8 the approximated value, which is denoted by X , will be compared with the true value, which is denoted by T . Then, the relative error

$$\frac{|X - T|}{T}$$

will be compared with the criterion at the calculation point. We say that the approximation satisfies the criterion at the given point if the relative error is less or equal than the criterion at that point. The convolutions are going to play the role of true values in cases 21, 22, 31 and 32 (see Table 3.8). In cases 11 and 12 the approximations will be compared with the simulated values. The simulation process requires more explanation because of the confidence interval approach which defines the accuracy of the simulations.

For each simulated value a 95% confidence interval will be constructed and its length will be based on the criterion at that point. Hence, we do not define the length of the confidence interval in an absolute sense. One of the reasons for that is the calculation scale, which heavily depends on the underlying parameters of the model. The situation is similar the definition of to that of the retention level, which was presented and motivated earlier in this section. Another reason is hidden in the definition of the criterion. It is obvious that the accuracy of the simulated value at the fixed point has to be in line with the criterion at the same point. In other words, if at some point the criterion is strict, the simulated value, which plays the role of the true value, has to be sufficiently accurate as well. In practice we replace the true value T by \bar{X} , which is the center of the

confidence interval, and the relative error of such a replacement should not be large, compared to the criterion at that point. First consider the *SLP* and the variance of the Stop-Loss contract.

According to the definition above, the approximated value X satisfies the criterion if

$$\bar{X}(1 - C_{SLP}(k)) \leq X \leq \bar{X}(1 + C_{SLP}(k)).$$

The confidence interval is defined as, cf. Section 4.1.2,

$$\bar{X}(1 - \theta C_{SLP}(k)) \leq T \leq \bar{X}(1 + \theta C_{SLP}(k)).$$

Hence, we can make \bar{X} as close to T as we want by decreasing the value of θ which obviously depends on the number of simulations. We are interested in

$$\frac{|X - T|}{T} \leq \frac{(1 + \theta)C_{SLP}(k)}{1 - \theta C_{SLP}(k)}$$

and we want to have $(1 + \theta)C_{SLP}(k)/\{1 - \theta C_{SLP}(k)\}$ as close to $C_{SLP}(k)$ as possible. It is clear that equality can be achieved only when $\theta = 0$, which is possible only for an infinite number of simulations. This means that we can not eliminate the error of the replacement, but we can control it. Let us say that we are satisfied if the error of the replacement is not more than 0.5% at $k = 0$, not more than 5% at $k = 3$, and proportional in between (hence 2% at $k = 1$). To achieve this, we introduce

$$E_{SLP}(k) = \frac{k}{3}0.05 + \left(1 - \frac{k}{3}\right)0.005$$

and define

$$\frac{(1 + \theta)C_{SLP}(k)}{1 - \theta C_{SLP}(k)} = C_{SLP}(k) + E_{SLP}(k),$$

which defines θ (which is actually a function of k) as

$$\theta(k) = \frac{E_{SLP}(k)}{C_{SLP}(k)\{1 + C_{SLP}(k) + E_{SLP}(k)\}}.$$

Hence, for each value of k , the length of the confidence interval has to be at most

$$\frac{2\bar{X}E_{SLP}(k)}{1 + C_{SLP}(k) + E_{SLP}(k)}$$

to achieve the precision prescribed above. As was mentioned before, see Section 4.1.2, in practice the length of the confidence interval is not known in advance since \bar{X} and S_X is defined during the simulation process. Therefore, some pilot study has to be performed before starting the main simulation procedure. The proposal is to perform 10000 simulations to get a first approximation of \bar{X} and S_X , and then to continue simulating step by step (let say by 100 simulations,

updating \bar{X} and S_X) until the needed accuracy (which is defined by the length of the confidence interval) is reached.

We act in a similar way when the calculation target is VaR . The auxiliary error function $E_{VaR}(p)$ is defined as

$$E_{VaR}(k) = \frac{(p - 0.5)}{0.49} 0.05 + \left(1 - \frac{p - 0.5}{0.49}\right) 0.005$$

and for given value of p the length of the confidence interval has to be not larger than

$$\frac{2\bar{X}E_{VaR}(p)}{1 + C_{VaR}(p) + E_{VaR}(p)}$$

to satisfy the condition of the replacement error. Note that here \bar{X} obviously denotes the center of the confidence interval of VaR .

5.2 Testing

We start with testing of the accuracy of the approximations in the case where the net Stop-Loss premium is the calculation target. For each situation from Table 3.8 the approximated value will be compared with the true value on the basis of criterion (5.1). We are interested in the parameter range in which the approximations (or at least one approximation) work well. The result will be presented as the rule of thumb for Model 4.

As a starting point we take the rule of thumb which was presented in Reijnen et al. [2005]. This rule has the form

$$\begin{aligned} 0 \leq \kappa_{3C} \leq 5 \text{ and } 0 \leq \kappa_{4S} \leq 1.5 &\Rightarrow \text{Gamma-IG approximation} \\ 5 < \kappa_{3C} < 15 \text{ or } 1.5 < \kappa_{4S} < 50 &\Rightarrow \text{IG approximation} \end{aligned} \quad (5.3)$$

and was originally presented for the model described in (1.6). Its close analogue can be found in Table 3.1 as Model 2. Here κ_{3C} denotes the skewness of the claim size and κ_{4S} denotes the kurtosis of the aggregate sum S . One of the main conclusions of Reijnen et al. [2005] was the importance of these two quantities. It was shown that they have a major influence on the risk measures (which are calculated using the model) and hence to the accuracy of the approximations. Therefore, it is interesting to know whether the rule still works for Model 4. To see this, Table 5.1 was constructed. All the situations from Table 3.8 are divided into two parts according to the equations from (5.3). Then, the relative approximation error is calculated and compared with the criterion. The numbers in the table show the numbers of situations where the relative approximation error satisfies the criterion and the number in brackets shows the total number of situations involved. The "Total" column tells in how many situations at least one approximation satisfies the criterion.

Assumption 11 $L \sim \text{Gamma}, C \sim \text{Lognormal}$	Satisfy the criterion		
	Total	IG	G-IG
$0 \leq \kappa_{3C} \leq 5$ and $0 \leq \kappa_{4S} \leq 1.5$	12(12)	11(12)	12(12)
$5 < \kappa_{3C} < 15$ or $1.5 < \kappa_{4S} < 50$	24(24)	19(24)	23(24)

Assumption 12 $L \sim \text{IG}, C \sim \text{Lognormal}$	Satisfy the criterion		
	Total	IG	G-IG
$0 \leq \kappa_{3C} \leq 5$ and $0 \leq \kappa_{4S} \leq 1.5$	9(9)	7(9)	9(9)
$5 < \kappa_{3C} < 15$ or $1.5 < \kappa_{4S} < 50$	51(60)	24(60)	47(60)

Assumption 21 $L \sim \text{Gamma}, C \sim \text{IG}$	Satisfy the criterion		
	Total	IG	G-IG
$0 \leq \kappa_{3C} \leq 5$ and $0 \leq \kappa_{4S} \leq 1.5$	27(27)	27(27)	27(27)
$5 < \kappa_{3C} < 15$ or $1.5 < \kappa_{4S} < 50$	9(9)	8(9)	9(9)

Assumption 22 $L \sim \text{IG}, C \sim \text{IG}$	Satisfy the criterion		
	Total	IG	G-IG
$0 \leq \kappa_{3C} \leq 5$ and $0 \leq \kappa_{4S} \leq 1.5$	27(27)	27(27)	27(27)
$5 < \kappa_{3C} < 15$ or $1.5 < \kappa_{4S} < 50$	40(42)	20(42)	39(42)

Assumption 31 $L \sim \text{Gamma}, C \sim \text{Gamma}$	Satisfy the criterion		
	Total	IG	G-IG
$0 \leq \kappa_{3C} \leq 5$ and $0 \leq \kappa_{4S} \leq 1.5$	27(27)	27(27)	27(27)
$5 < \kappa_{3C} < 15$ or $1.5 < \kappa_{4S} < 50$	9(9)	8(9)	9(9)

Assumption 32 $L \sim \text{IG}, C \sim \text{Gamma}$	Satisfy the criterion		
	Total	IG	G-IG
$0 \leq \kappa_{3C} \leq 5$ and $0 \leq \kappa_{4S} \leq 1.5$	27(27)	27(27)	27(27)
$5 < \kappa_{3C} < 15$ or $1.5 < \kappa_{4S} < 50$	40(42)	20(42)	39(42)

Table 5.1: Results of the accuracy testing when the net Stop-Loss premium is the approximation target.

The first conclusion which can be drawn from the table is that we should not use the rule of thumb of Model 2 for Model 4. According to the rule (5.3) we have to switch to the IG approximation as soon as $1.5 < \kappa_{4S} < 50$, but Table 5.1 clearly shows that the Gamma-IG approximation is better (more often satisfies the criterion) than the IG one in all the situations considered.

The general advice will be to use the Gamma-IG approximation when dealing with the *SLP* calculation in the region of interest (which is expressed in terms of the distributional assumptions considered as defined in Table 3.2 and the values of the underlying parameters as defined in Table 3.7). In view of this result, a couple of interesting remarks can be made. These remarks can be presented as

answers to questions like 'Why do the approximations work well even for large cumulants when Model 4 is assumed?' or 'Why do large kurtosis values not harm the Gamma-IG approximation?' In the first place it should be noted that we consider a restricted region in the parameter space. In particular, as has been explained in Section 3.4, λ is larger than in Reijnen et al. [2005], and this means that we have more observations. Furthermore, the range of a is from μ_{S_1} to $\mu_{S_1} + 3\sigma_{S_1}$. If we compare σ_{S_1} and σ_{S_4} , it is seen that σ_{S_1} is much smaller than σ_{S_4} and hence we are not that far in the tail of S_4 . In general, when considering relative error, approximations work better for the central part than in the tail.

Assumption 11 $L \sim \text{Gamma}, C \sim \text{Lognormal}$	Satisfy the criterion		
	Total	IG	G-IG
$0 \leq \kappa_{3C} \leq 5$ and $0 \leq \kappa_{4S} \leq 1.5$	12(12)	12(12)	12(12)
$5 < \kappa_{3C} < 15$ or $1.5 < \kappa_{4S} < 50$	24(24)	24(24)	24(24)

Assumption 12 $L \sim \text{IG}, C \sim \text{Lognormal}$	Satisfy the criterion		
	Total	IG	G-IG
$0 \leq \kappa_{3C} \leq 5$ and $0 \leq \kappa_{4S} \leq 1.5$	9(9)	9(9)	9(9)
$5 < \kappa_{3C} < 15$ or $1.5 < \kappa_{4S} < 50$	60(60)	59(60)	60(60)

Assumption 21 $L \sim \text{Gamma}, C \sim \text{IG}$	Satisfy the criterion		
	Total	IG	G-IG
$0 \leq \kappa_{3C} \leq 5$ and $0 \leq \kappa_{4S} \leq 1.5$	27(27)	27(27)	27(27)
$5 < \kappa_{3C} < 15$ or $1.5 < \kappa_{4S} < 50$	9(9)	9(9)	9(9)

Assumption 22 $L \sim \text{IG}, C \sim \text{IG}$	Satisfy the criterion		
	Total	IG	G-IG
$0 \leq \kappa_{3C} \leq 5$ and $0 \leq \kappa_{4S} \leq 1.5$	27(27)	26(27)	27(27)
$5 < \kappa_{3C} < 15$ or $1.5 < \kappa_{4S} < 50$	42(42)	42(42)	39(42)

Assumption 31 $L \sim \text{Gamma}, C \sim \text{Gamma}$	Satisfy the criterion		
	Total	IG	G-IG
$0 \leq \kappa_{3C} \leq 5$ and $0 \leq \kappa_{4S} \leq 1.5$	27(27)	27(27)	27(27)
$5 < \kappa_{3C} < 15$ or $1.5 < \kappa_{4S} < 50$	9(9)	9(9)	9(9)

Assumption 32 $L \sim \text{IG}, C \sim \text{Gamma}$	Satisfy the criterion		
	Total	IG	G-IG
$0 \leq \kappa_{3C} \leq 5$ and $0 \leq \kappa_{4S} \leq 1.5$	27(27)	26(27)	27(27)
$5 < \kappa_{3C} < 15$ or $1.5 < \kappa_{4S} < 50$	42(42)	40(42)	41(42)

Table 5.2: Results of the accuracy testing when the variance of the Stop-Loss contract is the approximation target.

The next step concerns the testing of the calculation accuracy for the variance of the Stop-Loss contract. Just as in the *SLP* case, the whole analysis will be based on the parameter values from Table 3.8 and the conclusion will be given for the region of interest from Table 3.7. The structure of the testing procedure is the same as well. Under the assumptions 21, 22, 31, 32, the convolution results will play the role of the true values and under the assumptions 11 and 12, the 'true' values will be simulated. The confidence interval approach stays the same as for the *SLP* which was discussed in Section 5.1. Table 5.2, just like Table 5.1, is based on the rule of thumb for Model 2 (see (5.3)). Moreover, the conclusion which we draw from Table 5.2 is completely in line with the conclusions for the *SLP* under Model 4. This means that the general advice for the approximation of the variance is the Gamma-IG approximation. But we have to admit that the IG approximation is almost as good as the Gamma-IG one when the variance of the Stop-Loss contract is the calculation target. Therefore, the rule of thumb (5.3) is technically applicable here, but it can be simplified to one line and we will do it.

Like was mentioned in the previous chapter, only the IG approximation is used in the present thesis for the calculation of *VaR*. Therefore, we only have to check whether this approximation is reliable in the region of interest which is considered here. Consequently, Figure 5.1 presents approximation results only for the IG case. The scaling on the *x*-axis organized as follows: the quantiles between the numbers which are shown always corresponds to the left number. For instance the quantiles between 0.9 and 0.95 are equal to 0.9. The upper part of the figure presents the relative errors under the situations 11, 21 and 31. It can be seen that in all the cases the relative error satisfies the chosen criterion. This is a nice result, which is in line with our requirements. The lower part of the figure contains the relative errors of the 12, 22 and 32 situations. Here we have a bit larger relative errors, but all of them satisfies the chosen criterion too. The large errors correspond to the very extreme cases where the dependence effect is expected to be huge. Therefore, the conclusion can be that the IG approximation is sufficiently accurate and thus can be used in the Value at Risk calculation.

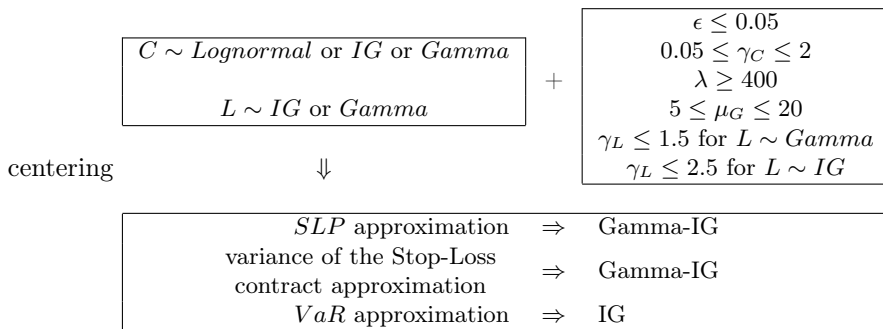
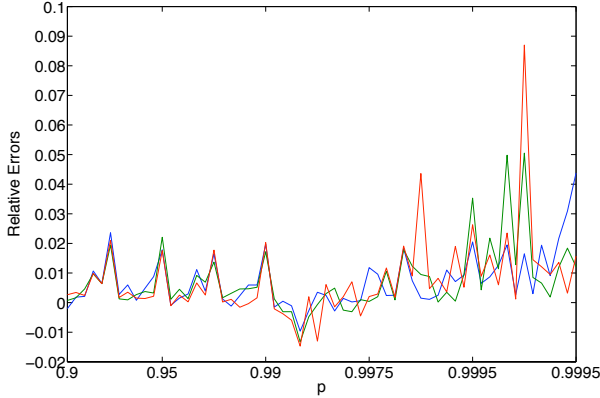
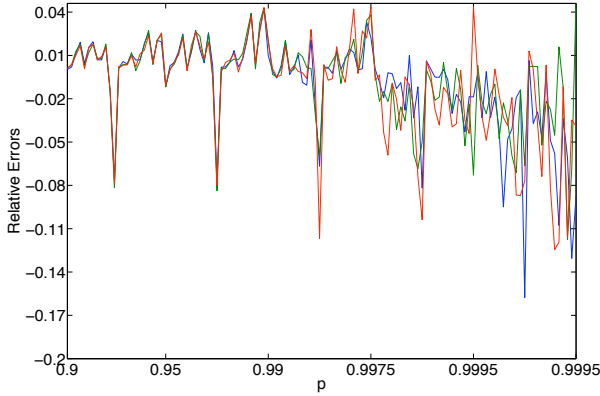


Table 5.3: The rule of thumb for Model 4.

As a conclusion of this section we are going to present the rule of thumb for Model 4. The general suggestion would be to use the Gamma-IG approximation for the *SLP* and the variance of the Stop-Loss contract approximation and the IG approximation for the *VaR* approximation when the aggregated sum is modelled by Model 4. This rule is valid in the limited region of the parameters of interest which is presented in Table 3.7 and restricted to the distributions under consideration which are presented in Table 3.2. The schematic overview of the rule of thumb for Model 4 is presented in Table 5.3.



(a) Situations 11, 21 and 31



(b) Situations 12, 22, 32

Figure 5.1: Relative errors of the Value at Risk using the Inverse Gaussian approximation

Chapter 6

Dependence Effect

In this chapter we present an explicit example of the dependence effect. The example concerns a concrete line of the insurance business, the so-called workers compensation insurance. The underlying data of this branch of insurance perfectly fit our model (dependence structure is present) and, moreover, it was suggested by the users committee of our STW-project as a potential source of problems due to dependence. However, the needed structure is not present in the available data. Therefore, it will serve only as a basis for the construction of the data set needed. Hence, real data will not be used in this chapter, but compared to Chapter 2 (where the dependence problem was invented by ourselves and all data were simulated) here we go a step further, since real data are the starting point. This contributes to the reality aspect of the example.

Before considering the real data, it is useful to know which data we need to implement Model 4 (or Model 2 or 3). Section 6.1 is focused on this aspect. Several examples of the data will give a feeling about the structure and content of the data which we need for the model setup. Next follows the section which gives some information about the workers compensation insurance in general. We do not present any specific information about the Dutch workers compensation structure, but we give the idea and the basic information which should hold for all countries. Section 6.2 considers the real data which we got from Bert Teeuwen, one of the members of the users committee. Here we explain why these data cannot be used directly for our purposes and give an example of the extended data which would satisfy all the requirements. The last part of the chapter is the analysis of two scenarios. The first scenario assumes that the dependence is completely ignored and the second scenario deals with different levels of dependence. The chapter is closed by presenting the conclusions about the potentially serious consequences of the fact that the dependence was ignored.

6.1 The data we need

In this section we discuss the type of the data which we need for the application of Model 4 (and actually Models 2 and 3 as well). We start with the summary of all the parameters of the model which have to be estimated from the data. We do not present any estimation algorithms here. This aspect will be covered in Chapter 7. In this chapter we (for simplicity) assume that the estimates are given and the estimation effect can be ignored. Model 4 has the form

$$S = \sum_{i=1}^N C_i + \sum_{k=1}^H \sum_{j=1}^{G_k} C_{jk},$$

where $N \sim P(\lambda(1 - \epsilon))$, $H \sim P(\epsilon\lambda/\mu_G)$, $G_k \sim P(L_k)$ for all $k = 1, 2, \dots$. Moreover, we assume that $L \sim \text{Gamma}(\alpha_L, \beta_L)$ (or $IG(\alpha_L, \beta_L)$) and $C \sim \text{Lognormal}(\alpha_C, \beta_C)$ (or $IG(\alpha_C, \beta_C)$, or $\text{Gamma}(\alpha_C, \beta_C)$).

All parameters which have to be estimated from the data are summarized in Table 6.1.

Parameter	Explanation
α_C	First parameter of the claim amount distribution
β_C	Second parameter of the claim amount distribution
λ	Total expected number of claims
ϵ	Expected percentage of the special claims
α_L	First parameter of the expected group size distribution
β_L	Second parameter of the expected group size distribution

Table 6.1: Summary of the model parameters

Note that parameters α_L, β_L can be replaced by α_G, β_G . In practice the parametrization with α_G, β_G or μ_G, γ_G maybe preferred, but in this chapter it is not important which parametrization is used. All general conclusions hold for both parametrizations. The next chapter will cover this aspect in more detail.

These are all the parameters we need to setup Model 4. To get a better feeling about the data which we need for the model implementation, we present a brief discussion about the parameters. For more information see Section 3.3. We start with the "easiest" ones. The category of the "easiest" parameters considered here consists of λ and the pair $\{\alpha_C, \beta_C\}$. Parameter λ is just the expected number of all claims. To get its estimate we do not need any information about the dependence (groups) structure. The same situation holds for the pair $\{\alpha_C, \beta_C\}$. These are the parameters of the claim amount distribution and while we assume that all the claims ("special" and "simple") follow the same distribution, it is quite easy to estimate them as well (because we do not need any information about the dependence either). The most difficult part is the estimation of ϵ and $\{\alpha_L, \beta_L\}$. For the estimation of such parameters the data has to include information about

the dependence groups. If such information is not directly included, it has to be possible to derive it from available extra information about the data. Therefore, it is very important to have data with information on the level of individuals. If the data are aggregated, they are not flexible anymore and we cannot obtain any information about the dependence structure.

To get a better feeling about the requirements for the data, we present several examples of data sets. The first example is the "perfect" data set, where all the necessary information is included (see Table 6.2). It is "perfect" in the sense that

Claim amount	Group code
75	0
40	0
5	1
9	1
92	1
44	0
50	0
27	0
88	2
42	2
68	0
26	3
27	3
27	3
96	3
35	0
...	...
85	0
92	43
12	43

Table 6.2: Example of the "Perfect" data set

we directly have all the information about the groups and the claim amounts. By using the "Group code" field we can get all the information about the groups: the number of groups and the group sizes during the reference period. As can be easily understood, 0 in the table denotes an individual claim and the numbers 1, 2, ... define different groups of dependence. In this example there are 43 groups with sizes 3, 2, 4, ..., 2, respectively. It is easily seen that we have everything in order to start the estimation procedure.

The main problem is that usually a field "Group code" is not available in the data set. In that case it should be possible to create this type of field from the existing data and available extra information about the data. We present an example of such a possibility. In Table 6.3 the "Group code" field is created

Nr	Date	Name	Claim amount	Group code	Incident	Place	Claim type	Risk class
1	d1	n1	c1	0	i7	p1	t4	r1
2	d2	n2	c2	0	i2	p2	t2	r2
3	d3	n3	c3	d3.i3.p3	i3	p3	t1	r3
4	d3	n4	c4	d3.i3.p3	i3	p3	t2	r3
5	d3	n5	c5	d3.i3.p3	i3	p3	t1	r2
6	d4	n6	c6	0	i4	p4	t8	r2
7	d4	n7	c7	0	i1	p1	t3	r2
8	d6	n8	c8	0	i6	p4	t3	r1
9	d7	n9	c9	d7.i7.p2	i7	p2	t5	r4
10	d7	n10	c10	d7.i7.p2	i7	p2	t4	r4
11	d8	n11	c11	0	i2	p1	t2	r1
12	d9	n12	c12	d9.i9.p4	i9	p4	t3	r1
13	d9	n13	c13	d9.i9.p4	i9	p4	t2	r1
14	d9	n14	c14	d9.i9.p4	i9	p4	t3	r2
15	d9	n15	c15	d9.i9.p4	i9	p4	t3	r1
16	d9	n16	c16	0	i3	p3	t2	r2
...
18	d10	n18	c18	0	i2	p1	t4	r3
19	d11	n19	c19	d11.i7.p2	i7	p2	t2	r3
20	d11	n20	c20	d11.i7.p2	i7	p2	t2	r3

Table 6.3: Example of creating the "Group code" field from the available information.

from the "Date", "Incident" and "Place" fields (in general, it can be a much more complicated structure). Individual claims are denoted by 0, as in the previous example. In case several accidents of the same type happened at the same date and at the same place, we assume that they form one dependence group. If, for example, in some laboratory there was an explosion on June 21, the people which have received compensation for this accident, are in the same group. But this does not mean that all of them have the same type of claim. These people could suffer completely different injuries. One of them could get a light shock and another a serious injury with need for rehabilitation (hence, it is quite natural to assume independence between the claim sizes inside the groups). This also does not mean that all of them are from the same occupation class. Probably in most situations this is the case, but this assumption has to be strictly motivated. Because of this, the workers compensation data which we got from Bert Teeuwen cannot be applied directly for the estimation of the parameters. The aggregation there is on a yearly basis for the different risk classes. Even if the total groups would belong to the same occupation classes, there is a yearly aggregation, which hides the information about the number of groups and the group sizes. Therefore, the fields "Claim type" and "Risk class" cannot be used for the creation of the "Group code" field. As an example consider claims 18, 19, 20. It is seen that claims 19 and 20 belong to one group and this group belongs to the occupation class $r3$. Claim 18 also belongs to the same occupation class, but it cannot be in the same group since the incident and place of the incident are different. Hence, if we make the aggregation on the risk class we lose information about the groups of dependence.

6.1.1 Workers compensation insurance

This section gives a general overview of the workers compensation insurance. The main idea can be easily understood from the name. Each company which employs people has to compensate their health injuries in case these occurred at the working place. This procedure usually goes through an insurance company which, instead of the company itself, covers the underlying risk (or part of it). Naturally, the insurance company requires a premium for the coverage of the risk. The premium is calculated by using rates which obviously depend on the nature of the business. Different types of businesses usually are classified by different classification codes. Each (or a combination of) classification code(s) corresponds to some risk class which can be understood as a risk level of the business considered.

Risk is determined by two factors: the frequency of on-the-job injury and the severity of injury. Severity is measured by both medical payments and indemnity benefits (payments made directly to the injured employee to compensate for losses suffered as a result of an accident). For example, roofers have the highest risk level, and office clerks have the lowest. Obviously the risks of a roofer are much different and quite a bit higher than those of an office clerk. Therefore, workers

compensation rates are much higher for roofing companies than for administrative companies.

As was already mentioned, the workers compensation insurance provides a good basis for the illustration of the dependence effect. As an example we can consider some construction company which buys workers compensation insurance service. Quite often, employees of the construction company work in groups and one accident can effect more than one member of the group. This is a classical example of the dependence we consider in the present research. In real life the dependencies can have a more complicated structure, but the main idea stays the same: several claims can arise at once.

Consequently, the main idea of the workers compensation insurance is quite simple. The total population is divided into risk classes according to some features which probably are different for different countries and/or insurance companies. Then, the rates are assigned to each risk class and the premium is calculated by multiplying the total annual payroll for each risk class by the rate. The most important things for us are the fact that the premium has to be assigned to each risk class separately and that separate risk classes are formed from a quite large number of companies (for example the risk class "Restaurants" can include up to 1000 units). This information should be sufficient for our purposes.

6.2 Workers compensation data

Consider some large insurance company which offers workers compensation insurance. Assume that the company has a quite large portfolio with a large amount of the risk classes involved. For each risk class the company determines the rate which is used in the calculation of the premium. Suppose that the company is interested in purchasing the reinsurance contract (assume that it is a Stop-Loss reinsurance) for its huge portfolio (or for a part of it). Assume that we are the company which is willing to sell such a contract, i.e. we are ready to cover the part of the risk of the insurance company which is above some retention level a . For the description of the Stop-Loss contract we refer to Section 1.4.

Remark 6.1. It was mentioned before that this chapter can be considered as an explicit and realistic example. To make the exposition as expressive as possible we present the analysis from a personal point of view. In other words, as was just mentioned, we imagine ourselves as reinsurer and the data considered in this chapter will be looked upon as coming from the insurance company which is going to conclude the reinsurance contract with us. \square

The natural and most important task for us (as the reinsurer) is to analyze and price the reinsurance contract. To calculate the Stop-Loss premium we need the workers compensation historical data from the company. In this section we consider the workers compensation data which we obtained through the users committee and which plays the role of data we (as the reinsurer) got from the

insurance company. A part of these data can be found in Table 6.4. Inspection of

Risk class	Year	Payroll	Loss
1	1	21798086	538707
1	2	22640528	439184
1	3	22572010	1059775
1	4	24789710	560013
1	5	25876764	1004997
1	6	28033613	1097314
1	7	22525887	609833
2	1	12004031	270222
2	2	12713178	229566
2	3	13596610	596850
2	4	14811727	196539
2	5	12774073	134248
2	6	20245789	489312
2	7	24242468	418218
3	1	50216515	769208
3	2	56099793	649707
3	3	58109747	503919
3	4	67807105	675466
3	5	73852437	545745
3	6	84208474	1562266
3	7	83604216	931762

Table 6.4: Part of the workers compensation data

the data set learns that the portfolio of the underlying insurance company consists of 124 risk classes (Table 6.4 presents only 3 of them) and the premium has to be assigned for each of them. Examples of risk classes are restaurants, construction industry, transportation, etc.

For simplicity we consider only one risk class for which the Stop-Loss premium has to be assigned. For the remaining classes the procedure is similar. Possible differences can be ascribed to the specific features of the class considered. Let us consider class number 2 and assume that this is the construction industry (note that here starts the artificial part since we do not have any information on how the risk classes are organized in general and whether the risk class 'construction industry' exists at all). For class 2 (just as for the other classes), we have 7 years of historical data, which is illustrated in Table 6.4. From a first glance it is clear that the data do not contain all the information we need for the model setup. There even is no exact information about the total number of claims per year. There is no information about the total claim amount either. Only the total loss is available, which probably equals $TotalClaimAmount - TotalPremium$. It is

not hard to guess that the data we have is the highest aggregation level of the company's database. However, to price the contract and to perform a reasonable analysis we need the whole database. That is why we are going to expand the data to a complete database structure. The expansion will be based on Table 6.3. The main goal is to obtain data which allow the estimation of the parameters of Model 4.

We decided to consider only one risk class, hence the information takes the form of Table 6.5.

Risk class	Year	Payroll	Loss
2	1	12004031	270222
2	2	12713178	229566
2	3	13596610	596850
2	4	14811727	196539
2	5	12774073	134248
2	6	20245789	489312
2	7	24242468	418218

Table 6.5: Information for risk class 2.

It was decided to consider construction industry as a representative example of the risk class number 2. Let us say that in the company's portfolio there are 100 construction companies, which together form the risk class 2. The next information level should contain information about each company for all the years considered. We also need the number of claims and the claim amounts. So, this table should also include this information (let us say the total number of claims and the total claim amounts per company for each year). The payrolls and losses for each company can be presented as well. Next follows the lowest information level table which contains the information about the individual claim amounts, claim types, places, etc. It should look similar to Table 6.3, where all the information needed is present. Figure 6.1 presents the possible structure of such a database.

The three tables in Figure 6.1 are connected by the keys used to one single database. It is not hard to guess that the key 2.1.N1 means the second risk class, the first year and the first company. The remaining notations are similar. For example, $p.N1.1$ defines the payroll of the first company of the second class. It is also clear that $\sum_{i=1}^{100} p.N_i.1 = 12004031$ and $\sum_{i=1}^{100} l.N_i.3 = 596850$. Unfortunately, the total number of claims, as well as the total claim amount, cannot be connected to the first table (since we simply do not have detailed information about these quantities). As it can be seen, the Payroll and Loss fields are not used directly in this chapter. We keep them for the illustration of the connection between the tables. The third (the one at the bottom of Figure 6.1) table has a structure as was described in Section 6.1. It completely satisfies all our requirements and in the remaining part of the chapter we will act as if we got

such information directly from the workers compensation insurance company.

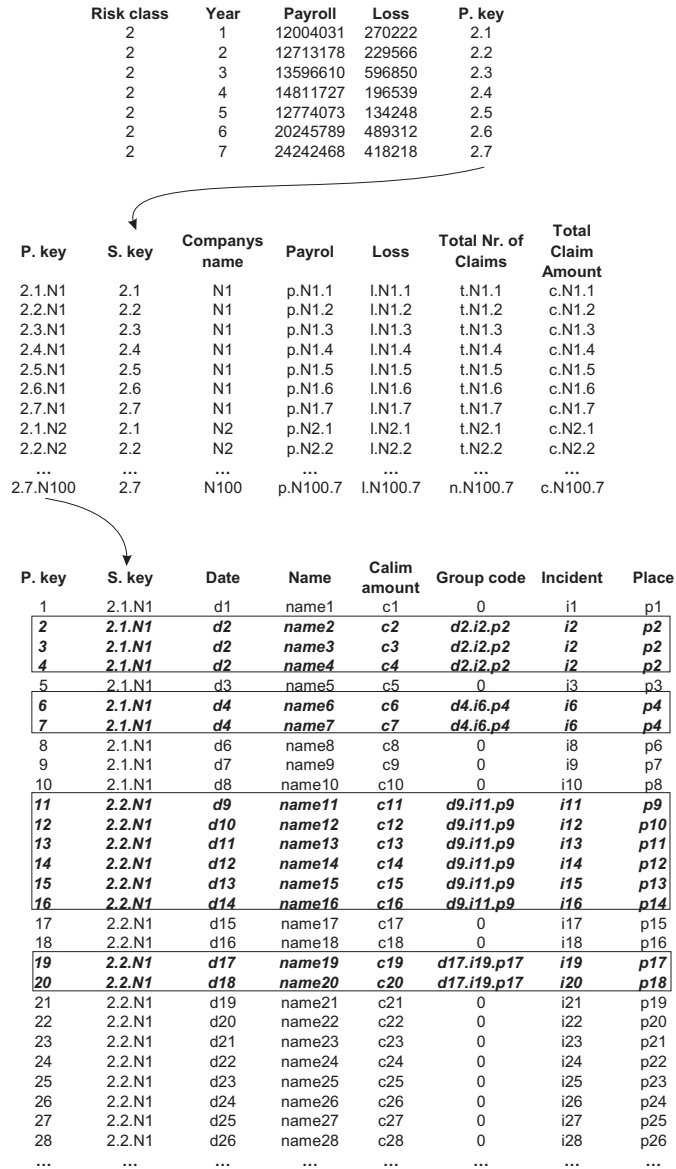


Figure 6.1: Possible structure of the company's database

6.3 Two scenarios

Having the data, we can start the pricing procedure. As was already mentioned we are going to cover the aggregate risk of the insurance company which is above some retention level a . Denoting by S the total aggregate claim of the company during some period of time (let us say one year), we obtain the risk $(S - a)^+$ and the risk $S - (S - a)^+$ is left for the insurance company which buys the Stop-Loss contract from us. It is clear that for such a contract the insurance company has to pay some premium, which we denote by P . Our profit (which obviously is not certain in advance) in that case becomes $P - (S - a)^+$. Naturally, S will be modeled by a sum of random variables which makes S a random variable itself. In case we define $P = E[(S - a)^+]$, our position becomes quite risky (standard deviation is positive) without any expected profit. Therefore, some additional value must be added to the SLP . There are a lot of ways to define the loading of the premium. In general this choice depends on many factors like risk averseness, character of the risk gained, etc. The well known methods are the premium with safety (or security) loading (where loading to SLP is defined as a fixed constant times the SLP itself), variance principle (where loading to SLP is a fixed constant times the variance of the Stop-Loss contract), standard deviation principle (where loading to SLP is a fixed constant times the standard deviation of the Stop-Loss contract) and quantile premium (which is defined as the value at risk of the Stop-Loss contract at some probability level). Here we decide to determine the premium according to the standard deviation principle, which is defined as

$$P = E[(S - a)^+] + \alpha \sigma_{(S-a)^+},$$

where α is a constant. In that case the standard deviation of our profit stays the same, but the expected profit becomes positive and equal to $\alpha \sigma_{(S-a)^+}$. If α is more than 1 and fixed, the expectation of our profit is larger than its standard deviation and basically depends on the behavior of $(S - a)^+$. That is why it is very important for us to analyze the behavior of $(S - a)^+$ since it basically defines the scale of the profit (or loss). The two main quantities which characterize its behavior are $E[(S - a)^+]$ and $\sigma_{(S-a)^+}$.

The next step after fixing the premium principle is to choose the model for the aggregate sum S . This is the major point since the whole example of the dependence effect is going to be based on this choice. The idea is to consider two scenarios. In the first scenario it is assumed that we utilize the independence model (Model 1) and the second scenario illustrates the situation when the modeling is based on Model 4. Both scenarios are assumed to be completely independent. It can be thought of as two persons doing two different researches and at the end we (as the head of the reinsurance company) compare the results.

6.3.1 The first scenario

The first scenario considers the situation where the aggregate sum S is modeled by Model 1 (see Table 3.1). After the structure of the model is chosen, we should continue with the general data analysis to get an idea about the distributions which are involved in the model. There are many methods available for this step and we are not going into details. We simply act as if this step has been performed and the decision is to model the total number of claims by the Poisson distribution and the single claim amounts by the Gamma distribution. Hence, the aggregated sum is modeled as

$$S = \sum_{i=1}^N C_i,$$

where $N \sim P(\lambda)$ and $C_i \sim \text{Gamma}(\alpha_C, \beta_C)$ for $i = 1, 2, \dots$. When all the distributions have been determined, we can start the estimation procedure. For this we need only the lowest table from the database which is illustrated in Figure 6.1. Having this table, we simply collect the information we need for the estimation purposes. An example of such a summary is presented in Table 6.6. The table

Risk class	Year	Total Nr. of claims	$\hat{\mu}_C$	$\hat{\sigma}_C$
2	1	394	100514	4940
2	2	379	99750	5093
2	3	346	99679	5031
2	4	461	100233	4996
2	5	404	100033	4690
2	6	388	99710	4765
2	7	450	99730	4974

Table 6.6: Information summary for Model 1

presents aggregated information for each of 7 years. This information includes total number of claims per year, estimated expected claim size and estimated standard deviation of the claim sizes which occurred during the year. The final estimates can be obtained from these numbers. In this particular example the estimates are $\hat{\lambda} = 403$, $\hat{\mu}_C = 99959$, $\hat{\sigma}_C = 4930$, from which the estimates of the parameters of the distribution $\hat{\alpha}_C, \hat{\beta}_C$ can be obtained easily.

However, for simplicity we assume that the estimated parameters of the model are

$$\hat{\lambda} = 400, \hat{\alpha}_C = 400, \hat{\beta}_C = 0.004,$$

which directly yield $\hat{\mu}_C = 10^5$ and $\hat{\sigma}_C = 5000$, which are quite close to the values from the example presented above. Since the estimation aspect is not covered in this chapter, we can simply act as if these estimates are the true values. This actually is all the information we need to price the reinsurance contract under

Model 1. Next, it is up to the insurance company to decide about the retention level. It is clear that the higher the retention level, the cheaper the contract.

As long as we believe in our model, we can offer to the company several contracts with different retention levels (and hence with different prices). Additionally we can provide information about the Value at Risk for the total claim amount and standard deviation of the Stop-Loss contract and the retained risk. Having this information, the company's managers can decide which option is most suitable for them. Table 6.7 contains an example of such information. Company's managers can compare the prices for the different retention levels and directly see the probabilities of having a total claim amount larger than these levels ($1 - p$).

If the company chooses contract 1 its risk stops at the retention level 40 million (40 m) and the probability that the total claim amount exceeds this level is quite high ($1 - p = 0.5$). But this contract is quite expensive. The cheapest contract is the contract number 10, but the chance that it will be realized during the year is quite small (about 0.001). In that case the company takes an additional 6 m risk for its own account (46 instead of 40 m). Of course, it remains up to the company to choose the contract. Our aim is to ask the suitable premium which would secure our position.

6.3.2 The second scenario

Here we believe that there are some dependencies among the claims. This means that there is a possibility to have two types of claims: the ones which come separately (simple claims) and the ones which come together as a group (special claims). In that case we believe that Model 4 (see Table 3.1) would be the best tool to describe the behavior of the total claim amount. After such a decision is made, we have to analyze the data more closely to decide which distributions could be assumed for Model 4. Assume we arrived at the following assumptions for the total claim amount:

$$S = \sum_{i=1}^N C_i + \sum_{k=1}^H \sum_{j=1}^{G_k} C_{jk},$$

where

$$N \sim P(\lambda(1 - \epsilon)),$$

$$H \sim P\left(\epsilon \frac{\lambda}{\mu_G}\right),$$

$$G_k \sim P(L), \text{ for all } k = 1, 2, \dots$$

$$L \sim \text{Gamma}(\alpha_L, \beta_L) \text{ and } C \sim \text{Gamma}(\alpha_C, \beta_C).$$

Having all the assumptions, we can start the estimation. We use the same database, but now we definitely need more information. Just as in the previous case we collect the information we need from the lowest table of the database. An

Contract	k	a	μ_S	σ_S	$\mu_{S-(S-a)^+}$	$\sigma_{S-(S-a)^+}$	$\mu_{(S-a)^+}$	$\sigma_{(S-a)^+}$	p	VaR
1	0.00	4000000	4000000	2002498	39201174	1157635	798825	1180523	0.50	399999999
2	0.33	40667499	4000000	2002498	39488841	1378152	511158	951700	0.60	40508403
3	0.67	41334998	4000000	2002498	39693866	1571252	306133	732363	0.70	41054727
4	1.00	42002498	4000000	2002498	39829142	1725682	170857	537915	0.75	41358306
5	1.33	42669997	4000000	2002498	39911449	1838200	88550	377459	0.80	41697251
6	1.67	43337497	4000000	2002498	39957501	1912717	42498	253418	0.85	42093526
7	2.00	44004996	4000000	2002498	39981151	1957481	18848	163041	0.90	42593961
8	2.33	44672496	4000000	2002498	39992287	1981837	7712	100650	0.95	43339445
9	2.67	45339995	4000000	2002498	39997091	1993830	2908	59672	0.99	44720037
10	3.00	46007495	4000000	2002498	39998989	1999176	1010	33992	0.999	46350204
$a = \mu_S + k\sigma_S$										
$SLP = E[(S - a)^+]$										
$SLP_2 = E[((S - a)^+)^2]$										
$\sigma_{(S-a)^+} = \sqrt{SLP_2 - SLP^2}$										
$VaR = F_S^{-1}(p)$										

Table 6.7: Information for the company according to Model 1

Risk class	Year	T	$\hat{\mu}_C$	$\hat{\sigma}_C$	H	$\hat{\mu}_L$	$\hat{\sigma}_L$	N
2	1	394	100514	4940	2	22	15	358
2	2	379	99750	5093	0	0	0	379
2	3	346	99679	5031	1	6	0	340
2	4	461	100233	4996	2	11	1	435
2	5	404	100033	4690	0	0	0	404
2	6	388	99710	4765	1	21	0	367
2	7	450	99730	4974	2	24	13	402

$\hat{\mu}_C$: estimated expectation of the claim size per year
 $\hat{\sigma}_C$: estimated standard deviation of the claim size per year
 $\hat{\mu}_L$: estimated expectation of the group size per year
 $\hat{\sigma}_L$: estimated standard deviation of L per year
 T : total number of claims per year
 H : total number of groups per year
 N : total number of individual claims per year

Table 6.8: Information summary for Model 4

example of the summarized information can be found in Table 6.8. Just as in the first scenario the table presents aggregated information for each of 7 years. This information includes total number of claims per year, total number of individual claims per year, estimated expected claim size, estimated standard deviation of the claim sizes which occurred during the year, number of dependence groups which occurred during the year, estimated expectation of the group size due to different companies and estimated standard deviation of L due to different companies. The final estimates can be obtained from these numbers. In this particular example the estimates are $\hat{\lambda} = 403, \hat{\mu}_C = 99959, \hat{\sigma}_C = 4930, \hat{\epsilon} = 0.048, \hat{\mu}_L = 17.1, \hat{\sigma}_L = 12.68$, from which the estimates of the parameters of the distribution $\hat{\alpha}_C, \hat{\beta}_C, \hat{\alpha}_L, \hat{\beta}_L$ can be obtained easily.

However, for simplicity we assume that the estimated parameters of the model are

$$\hat{\lambda} = 400, \hat{\epsilon} = 0.03, \hat{\alpha}_L = 1, \hat{\beta}_L = 0.05, \hat{\alpha}_C = 400, \hat{\beta}_C = 0.004.$$

Note that the estimated expectation and expected standard deviation of the claim size just like in the first scenario are $\hat{\mu}_C = 10^5$ and $\hat{\sigma}_C = 5000$. Having these estimates we can get a lot of information about the total claim amount and the contract itself for the reference year (just like we did in the case of Model 1). Table 6.9 presents such information.

Note that here we use exactly the same database as we used in Model 1. The only aspect we changed is the assumptions for the total number of claims during the reference year. The next section presents a discussion how these changes affect the Stop-Loss contract and the Value at Risk.

Contract	k	a	μ_S	σ_S	$\mu_{S-(S-a)^+}$	$\sigma_{S-(S-a)^+}$	$\mu_{(S-a)^+}$	$\sigma_{(S-a)^+}$	p	VaR
1	0.00	4000000	4000000	2968164	38856668	1347225	1143331	2092981	0.50	39466739
2	0.33	40667499	4000000	2968164	39111610	1580174	888389	1883766	0.60	40187796
3	0.67	41334998	4000000	2968164	39314995	1794149	685004	1680358	0.70	41040648
4	1.00	42002498	4000000	2968164	39475227	1985759	524772	1488042	0.75	41551671
5	1.33	42669997	4000000	2968164	39600166	2153839	399833	1310013	0.80	42154846
6	1.67	43337497	4000000	2968164	39696761	2298787	303238	1147846	0.85	42905428
7	2.00	44004996	4000000	2968164	39770915	2422018	229084	1001951	0.90	43925729
8	2.33	44672496	4000000	2968164	39827508	2525523	172491	871950	0.95	45603458
9	2.67	45339995	4000000	2968164	39870486	2611558	129513	756976	0.99	49328619
10	3.00	46007495	4000000	2968164	39902989	2682426	97010	655886	0.999	54480303
$a = \mu_S + k\sigma_S$										
$SLP = E[(S - a)^+]$										
$SLP_2 = E[((S - a)^+)^2]$										
$\sigma_{(S-a)^+} = \sqrt{SLP_2 - SLP^2}$										
$VaR = F_S^{-1}(p)$										

Table 6.9: Information for the company according to Model 4

6.4 Conclusions

In this section we discuss the consequences of ignoring the dependence effect. The two sections above present the results of the reinsurance contract pricing using two different models. The first model (Model 1) assumes that all the claims arrive at independent time points while the second model (Model 4) assumes the dependence structure.

In principle, all results are already available in Tables 6.7 and 6.9. However, for comparison it is more convenient to see them all together in one table. Table 6.10 presents the comparison of all the quantities considered. From that table we directly see that the *SLP* under Model 4 is much higher than the one under Model 1 in a relative sense and this difference becomes huge for high retention levels. This fact definitely has to be a signal for the company and for us (who play a role of reinsurer). It means that the expected risk of selling the Stop-Loss contract is much higher than we expected using Model 1. However, in case the premium is determined according to the standard deviation principle, the dominant quantity in the premium is not the *SLP* but the standard deviation. In spite of the fact that the relative difference in standard deviation according to different models is smaller (but still huge) than in case of the *SLP*, the absolute values are much larger (see Tables 6.7 and 6.9).

Also there is a huge difference in a relative sense between the second moments of the contract (quantity $(S - a)^+$). For the highest retention level it could reach a factor 380 (38000%). This definitely means that the tail of the contract distribution can be much heavier than we expected using Model 1. All this shows that the supposed rare events are not so rare after all and thus can cause a lot of troubles for the reinsurer.

It can be noted that the relative difference in *VaR* between the two models is much smaller than in case of the *SLP* or the standard deviation. However, it is clear that the absolute differences of *VaR* is much larger than the ones of the *SLP* or the standard deviation, which makes the factor 0.17 still very significant. Therefore, the general conclusion is to consider both the relative and the absolute difference. None of these two comparison methods is more important and the final decision should be based on both of them.

Contract	k	a	$\frac{SLP_4}{SLP_1} - 1$	$\frac{SLP_{2,4}}{SLP_{2,1}} - 1$	$\frac{\sigma_{(S_4 - a)^+}}{\sigma_{(S_1 - a)^+}} - 1$	p	$\frac{VaR_4}{VaR_1} - 1$
1	0	4000000	0.43	1.80	0.77	0.5	-0.01
2	0.33	40667499	0.73	2.72	0.97	0.6	-0.007
3	0.67	41334998	1.23	4.23	1.29	0.7	-0.0003
4	1	42002498	2.07	6.82	1.76	0.75	0.004
5	1.33	42669997	3.51	11.48	2.47	0.8	0.01
6	1.67	43337497	6.13	20.35	3.52	0.85	0.019
7	2	44004996	11.15	38.22	5.14	0.9	0.031
8	2.33	44672496	21.36	76.53	7.66	0.95	0.052
9	2.67	45339995	43.53	164.24	11.68	0.99	0.102
10	3	46007495	95.04	379.10	18.29	0.999	0.175
$a = \mu_S + k\sigma_S$							
$SLP_4(SLP_1)$: SLP according to the Model 4 (1)							
$SLP_{2,4}(SLP_{2,1})$: SLP_2 according to Model 4 (1)							
$S_4(S_1)$: total claim amount according to Model 4 (1)							
$VaR_4(VaR_1)$: VaR according to Model 4 (1)							

Table 6.10: Dependence effects for different quantities

Chapter 7

Estimation Effects

As usual in stochastic modeling, all the parameters which are involved in the model considered, have to be estimated. Replacing the unknown model parameters by their estimates (which are usually obtained from the data) will result in estimation errors. Just as with ignoring the dependence effect, it is too optimistic to act as if the estimation errors are negligible, unless the number of observations is large. This topic, the effect of the estimation step, is exactly the issue which is addressed in this chapter.

The model of consideration (Model 4) was introduced in Chapter 3. The model is too complicated to allow an exact evaluation of the estimation effects in such a way that transparent conclusions can be drawn. Therefore, we use approximations which were introduced in 4.1.4. The accuracy of these approximations has been settled in Chapter 5. In general, two aspects play a role when considering the effect of the estimation step: the accuracy of the estimators and the fluctuation of the risk measure considered as function of the parameters. The set of parameters (just like the model itself) may be divided into two parts, those concerning the ordinary claims and those which are inserted in particular for the special causes. For the first part a lot of data are usually available and these parameters can be estimated quite accurately. However, it is obvious that special claims do not appear very often and hence estimation of the parameters linked up with the common risk part is expected to be much less accurate. As was mentioned before, their influence on the final outcome, even when a rather small part is due to a common risk, is quite large and hence estimation of the parameters connected with the special causes is the most important issue.

In Section 7.1 the needed structure of the observations to obtain estimators is given and the estimators based on them are derived. The local behavior of the risk measures considered is discussed in Section 7.2. In total there are two risk measures which will be presented in this chapter. These are *SLP* and *Var*. Analysis of the estimation effect of *SLP* was already performed in the paper of Albers and Kallenberg [2007], therefore only the main conclusions and results will

be presented here. Regardless the fact that the main analysis will be concentrated on VaR , some important conclusions and theorems which were derived in Albers and Kallenberg [2007] for SLP will be used here as well. Moreover, most theory developed in Albers and Kallenberg [2007] is independent from the specific features of the risk measures. These parts will be presented in Section 7.3 in general form with a later application on SLP and VaR . Sections 7.4 and 7.5 discuss the estimation effect when VaR and SLP are the calculation targets.

7.1 Observations and estimators

In this section we describe the basic structure of the observations and derive the formulas for the estimators. The general data structure which we need for the model implementation was described in detail in Chapter 6. The main conclusion was that each claim should be considered as a pair (X_i, Y_i) , where X_i is the claim amount and Y_i is the group code, 0 for the independent (ordinary) claim and $1, 2, \dots$ for the various dependent claims (due to the common risk). Hence, in case several claims can be considered as a dependent group, all of them must have the same Y part. Assuming that such a structure is given, the following information can be deduced

- n : the number of independent claims
- c_1, \dots, c_n : the claim amounts for the independent claims
- h : the number of group codes for the dependent claims
- g_1, \dots, g_h : the group sizes
- $d_{11}, \dots, d_{g_h h}$: the claim amounts for the dependent claims.

It will typically not be enough to have these data for one year. The reason for that is the scarcity of special claims. For a reasonable estimation of ϵ , μ_G and γ_G we have to consider more years, which will be denoted in what follows by $t = 1, \dots, u$. Therefore, the estimators will be based on $N_t, C_{1t}, \dots, C_{N_t t}, H_t, G_{1t}, \dots, G_{H_t t}, D_{11t}, \dots, D_{G_{H_t t} t}$, for $t = 1, \dots, u$.

For the observed data $n_t, c_{1t}, \dots, c_{n_t t}, h_t, g_{1t}, \dots, g_{h_t t}, d_{11t}, \dots, d_{g_{h_t t} h_t t}$, with $t = 1, \dots, u$, the likelihood equals

$$\exp(-\theta)\theta^{n_{tot}+h_{tot}}p^{h_{tot}}(1-p)^{n_{tot}} \times \left\{ \prod_{t=1}^u \prod_{k=1}^{h_t} P(G = g_{kt}) \right\} \\ \times \prod_{t=1}^u \left[\left\{ \prod_{i=1}^{n_t} f_C(c_{it}) \right\} \left\{ \prod_{k=1}^{h_t} \prod_{j=1}^{g_{kt}} f_C(d_{jkt}) \right\} \right] \times \prod_{t=1}^u \frac{1}{n_t! h_t! u^{n_t+h_t}},$$

where

$$\theta = \theta(\lambda, \epsilon, \mu_G) = u\lambda(1 - \epsilon + \epsilon\mu_G^{-1}),$$

$$p = p(\epsilon, \mu_G) = \frac{\epsilon \mu_G^{-1}}{1 - \epsilon + \epsilon \mu_G^{-1}},$$

$$n_{tot} = \sum_{t=1}^u n_t \text{ and } h_{tot} = \sum_{t=1}^u h_t.$$

For simplicity we will often write n and h instead of n_{tot} and h_{tot} . Maximizing the likelihood w.r.t. λ for given ϵ and μ_G gives $\hat{\theta} = n + h$ and hence

$$\hat{\lambda} = \hat{\lambda}(\epsilon, \mu_G) = \frac{n + h}{u(1 - \epsilon + \epsilon \mu_G^{-1})}.$$

After inserting the above estimate, it can be noted that $\exp(-\hat{\theta})\hat{\theta}^{n+h}$ does not depend on (ϵ, μ_G) anymore. Next, the likelihood is maximized w.r.t. ϵ for given μ_G by taking $\hat{p} = h/(n + h)$ and hence

$$\hat{\epsilon} = \hat{\epsilon}(\mu_G) = \frac{h}{h + n\mu_G^{-1}}.$$

Inserting this we can note that $\hat{p}^h(1 - \hat{p})^n$ does not depend on μ_G . It is seen that we end up with the likelihood of the G 's times the likelihood of the C 's and D 's. This means that we can proceed with estimating the parameters of the distribution of G using only the G -observations and, separately, estimating the parameters of the distribution of C using the C and D observations.

Assuming that L follows the Gamma distribution, the distribution of G will be a negative binomial. Although in general the number of observations from this negative binomial distribution, $\sum_{t=1}^u H_t$, will be not very large, the expectation of G is as a rule not small, say between 5 and 20. Under these circumstances, Saha and Paul [2005] show that moment estimators are a good alternative to maximum likelihood estimators.

Both when L has a Gamma distribution and when L has an IG distribution, G has a distribution with two parameters. Moment estimators do not depend on the parametrization. It is convenient to take as parametrization for G its expectation μ_G and its coefficient of variation γ_G (see Remarks 3.1 and 7.1). The moment estimates of the expectation and coefficient of variation are

$$\hat{\mu}_G = \bar{g} = \frac{1}{h} \sum_{t=1}^u \sum_{k=1}^{h_t} g_{kt},$$

$$\hat{\gamma}_G = \frac{1}{\bar{g}} \sqrt{g^2 - \bar{g}^2}, \text{ where } \bar{g}^2 = \frac{1}{h} \sum_{t=1}^u \sum_{k=1}^{h_t} g_{kt}^2.$$

Inserting $\hat{\mu}_G$ in $\hat{\epsilon}$, and writing $g_{tot} = \sum_{t=1}^u \sum_{k=1}^{h_t} g_{kt}$, we get

$$\hat{\epsilon} = \frac{h}{h + n\bar{g}^{-1}} = \frac{h\bar{g}}{h\bar{g} + n} = \frac{g_{tot}}{g_{tot} + n_{tot}},$$

which indeed is the "natural" (observed fraction special claims) estimate of ϵ . Inserting $\hat{\epsilon} = h\bar{g}/(h\bar{g} + n)$, $\hat{\mu}_G = \bar{g}$ in $\hat{\lambda}$ gives

$$\hat{\lambda} = \frac{h\bar{g} + n}{u} = \frac{g_{tot} + n_{tot}}{u},$$

which also is the "natural" (observed total number of claims divided by the number of years) estimate of λ . For simplicity we denote

$$\bar{h} = \frac{1}{u} \sum_{t=1}^u h_t = \frac{h}{u}, \bar{n} = \frac{1}{u} \sum_{t=1}^u n_t = \frac{n}{u},$$

to get

$$\hat{\lambda} = \bar{h}\bar{g} + \bar{n}.$$

For the estimation of the parameters of the claim size distribution C we have much more observations, compared to the group size G . Therefore we definitely can use moment estimators here as well. As parametrization we once more take the expectation μ_C and the coefficient of variation γ_C . Denoting

$$\overline{c+d} = \frac{1}{n_{tot} + g_{tot}} \left[\sum_{t=1}^u \sum_{i=1}^{n_t} c_{it} + \sum_{t=1}^u \sum_{k=1}^{h_t} \sum_{j=1}^{g_{kt}} d_{jkt} \right],$$

and

$$\overline{c^2 + d^2} = \frac{1}{n_{tot} + g_{tot}} \left[\sum_{t=1}^u \sum_{i=1}^{n_t} c_{it}^2 + \sum_{t=1}^u \sum_{k=1}^{h_t} \sum_{j=1}^{g_{kt}} d_{jkt}^2 \right],$$

we get

$$\hat{\mu}_C = \overline{c+d},$$

and

$$\hat{\gamma}_C = \frac{\sqrt{\overline{c^2 + d^2} - (\overline{c+d})^2}}{\overline{c+d}}.$$

As a section summary we present our estimators. These are

$$\begin{aligned} \hat{\mu}_G &= \bar{G}, \hat{\mu}_C = \overline{C+D}, \\ \hat{\lambda} &= \frac{G_{tot} + N_{tot}}{u}, \hat{\epsilon} = \frac{G_{tot}}{G_{tot} + N_{tot}}, \\ \hat{\gamma}_G &= \frac{\sqrt{\bar{G}^2 - G^2}}{\bar{G}}, \hat{\gamma}_C = \frac{\sqrt{\overline{C^2 + D^2} - \overline{C+D}^2}}{\overline{C+D}}, \end{aligned} \tag{7.1}$$

where u is the number of years and

$$H_{tot} = \sum_{t=1}^u H_t, N_{tot} = \sum_{t=1}^u N_t, G_{tot} = \sum_{t=1}^u \sum_{k=1}^{H_t} G_{kt},$$

$$\begin{aligned}\bar{G} &= \frac{1}{H_{tot}} \sum_{t=1}^u \sum_{k=1}^{H_t} G_{kt}, \quad \overline{G^2} = \frac{1}{H_{tot}} \sum_{t=1}^u \sum_{k=1}^{H_t} G_{kt}^2, \\ \overline{C+D} &= \frac{\sum_{t=1}^u \sum_{i=1}^{N_t} C_{it} + \sum_{t=1}^u \sum_{k=1}^{H_t} \sum_{j=1}^{G_{kt}} D_{jkt}}{N_{tot} + G_{tot}}, \\ \overline{C^2+D^2} &= \frac{\sum_{t=1}^u \sum_{i=1}^{N_t} C_{it}^2 + \sum_{t=1}^u \sum_{k=1}^{H_t} \sum_{j=1}^{G_{kt}} D_{jkt}^2}{N_{tot} + G_{tot}}.\end{aligned}$$

Remark 7.1. Obviously, we can replace the parameters μ_G and γ_G and their estimators $\hat{\mu}_G$ and $\hat{\gamma}_G$ by the parameters μ_L and γ_L and the corresponding estimators $\hat{\mu}_L$ and $\hat{\gamma}_L$. Since $\mu_G = \mu_L$ and $\sigma_G^2 = \mu_L + \sigma_L^2$, implying that $\gamma_L = \mu_G^{-1} \sqrt{\sigma_G^2 - \mu_G}$, we get

$$\begin{aligned}\hat{\mu}_L &= \bar{G}, \\ \hat{\gamma}_L &= \frac{\sqrt{\overline{G^2} - \bar{G}^2 - \bar{G}}}{\bar{G}}.\end{aligned}\tag{7.2}$$

As long as γ_L is not equal to 0 or close to it, there is no problem with $\hat{\gamma}_L$. However, when $\gamma_L = 0$ (or close to 0) it may easily happen that $\overline{G^2} - \bar{G}^2 - \bar{G} < 0$ and hence a problem arises with application of (7.2). Note that the case $\gamma_L = 0$ corresponds to a fixed parameter of the Poisson distribution of G , a situation which we also want to take into account. In view of the problems with (7.2), indeed it is more convenient to use the parametrization μ_G, γ_G (see also Remark 3.1). \square

7.2 Local behavior of the risk measures

The main goal of the chapter is to analyze the effect of the parameters estimation for some insurance products. As two candidates we consider VaR and SLP . The influence of the estimators on the considered insurance products (measures) certainly depends on their behavior as a function of the parameters $\mu_C, \gamma_C, \mu_G, \gamma_G, \epsilon, \lambda$, which are going to be estimated, as well as on the accuracy of the estimators. For instance, if SLP is a flat function of the parameters $\mu_C, \gamma_C, \mu_G, \gamma_G, \epsilon, \lambda$ and the estimators are accurate, the small changes due to the estimation will have not much effect. So, these two points have to be considered: how fluctuate the quantities considered (VaR and SLP) and how accurate are the estimators. In this section we discuss the first point, the second one will be covered in the later sections.

7.2.1 Local behavior of VaR

We will start with analyzing the behavior of VaR . It is defined as the inverse of the cumulative distribution function of S , the aggregated claim amount. Hence

$$VaR = F_S^{-1}(p),$$

where p is a fixed risk level. Therefore, it is quite clear that under the model considered VaR has a very complicated structure and some approximation (simplification) techniques have to be applied.

We are going to apply two simplification steps. In the first step, we approximate the distribution of S by the IG distribution, applying the so-called Inverse-Gaussian approximation, which was described in detail in Section 4.1.4, see also Section 4.3.3. The following lemma describes the first simplification step.

Lemma 7.1. *After introducing a temporary notation $\theta = (\mu_C, \gamma_C, \mu_G, \gamma_G, \epsilon, \lambda)$, fixing some probability level p and denoting by $VaRapp(\theta)$ the approximated VaR value at p , we can write that*

$$VaRapp(\theta) = \mu_S(\theta) + t(\kappa_{3S}(\theta))\sigma_S(\theta),$$

where

$$\mu_S(\theta) = \lambda\mu_C, \tag{7.3}$$

$$\sigma_S(\theta) = \sqrt{\lambda}\mu_C\sqrt{1 + \gamma_C^2 + \epsilon(1 + \gamma_G^2 - 1/\mu_G)\mu_G}, \tag{7.4}$$

and $t(\kappa_{3S}(\theta))$ is calculated as a solution of

$$g(\kappa_{3S}(\theta), t(\kappa_{3S}(\theta))) = 0, \tag{7.5}$$

where the function $g(\cdot, \cdot)$ is defined by the IG approximation, i.e.

$$g(x, y) = \Phi\left(\frac{y}{\sqrt{xy/3 + 1}}\right) + e^{2(3/x)^2}\Phi\left(\frac{-y - 6/x}{\sqrt{xy/3 + 1}}\right) - p.$$

Proof. Using the Inverse-Gaussian approximation, the real density $f_S(s)$ is approximated by the Inverse-Gaussian density $f_{IG}(s - x_{0IG})$, with the appropriate parameters $\alpha_{IG} = 3\sigma_S/\kappa_{3S}$, $\beta_{IG} = \kappa_{3S}\sigma_S/3$ and $x_{0IG} = \mu_S - 3\sigma_S/\kappa_{3S}$, see Section 4.1.4 and Table 3.2. Hence, the needed distribution function $F_S(s)$ is approximated by $F_{IG}(s - x_{0IG})$, which, using the identity (4.16), can be expressed in terms of the standard normal distribution function as

$$\Phi\left(\frac{s - x_{0IG} - \alpha_{IG}}{\sqrt{\beta_{IG}(s - x_{0IG})}}\right) + e^{2\alpha_{IG}/\beta_{IG}}\Phi\left(\frac{-(s - x_{0IG}) - \alpha_{IG}}{\sqrt{\beta_{IG}(s - x_{0IG})}}\right).$$

Replacing s (which is basically the needed value of VaR at probability p) by

$\mu_S(\theta) + t(\kappa_{3S}(\theta))\sigma_S(\theta)$ and substituting α_{IG}, β_{IG} and x_{0IG} , we obtain

$$\begin{aligned} F_{IG}(s - x_{0IG}) - p &= \Phi\left(\frac{t(\kappa_{3S}(\theta))}{\sqrt{\kappa_{3S}(\theta)t(\kappa_{3S}(\theta))/3 + 1}}\right) \\ &+ e^{2(3/\kappa_{3S}(\theta))^2} \Phi\left(\frac{-t(\kappa_{3S}(\theta)) - 6/\kappa_{3S}(\theta)}{\sqrt{\kappa_{3S}(\theta)t(\kappa_{3S}(\theta))/3 + 1}}\right) - p \\ &= g(\kappa_{3S}(\theta), t(\kappa_{3S}(\theta))). \end{aligned}$$

Hence, solving $g(\kappa_{3S}(\theta), t(\kappa_{3S}(\theta))) = 0$ w.r.t. $t(\kappa_{3S}(\theta))$ and substituting (7.3) and (7.4) gives us the needed approximation of VaR , $VaRapp(\theta) = \mu_S(\theta) + t(\kappa_{3S}(\theta))\sigma_S(\theta)$. The formulas (7.3) and (7.4) are obtained from Lemma 3.3 and Lemma 3.4(i) using that D is distributed as C and that $\mu_L = \mu_G, \gamma_L^2 = \gamma_G^2 - 1/\mu_G$. \square

The above explained simplification step definitely makes life easier, but the resulting function is still rather complicated. Therefore, we apply a one step Taylor expansion on the approximation around the true value $(\mu_{C0}, \gamma_{C0}, \mu_{G0}, \gamma_{G0}, \epsilon_0, \lambda_0)$ of the parameters. This linearization will be called $VaRapp1$ and its general form is

$$\begin{aligned} VaRapp1(\mu_C, \gamma_C, \mu_G, \gamma_G, \epsilon, \lambda) &= VaRapp(\mu_{C0}, \gamma_{C0}, \mu_{G0}, \gamma_{G0}, \epsilon_0, \lambda_0) \\ &+ (\mu_C - \mu_{C0}) \frac{\partial}{\partial \mu_C} VaRapp(\mu_{C0}, \gamma_{C0}, \mu_{G0}, \gamma_{G0}, \epsilon_0, \lambda_0) \\ &+ \dots + (\lambda - \lambda_0) \frac{\partial}{\partial \lambda} VaRapp(\mu_{C0}, \gamma_{C0}, \mu_{G0}, \gamma_{G0}, \epsilon_0, \lambda_0). \end{aligned} \quad (7.6)$$

For the final set up we need the partial derivatives of $VaRapp$. The general form of the $VaRapp$ derivative w.r.t. the parameter θ_i (one of the components of the vector θ) is

$$\frac{\partial VaRapp(\theta)}{\partial \theta_i} = \frac{\partial \mu_S(\theta)}{\partial \theta_i} + t'(\kappa_{3S}(\theta)) \frac{\partial \kappa_{3S}(\theta)}{\partial \theta_i} \sigma_S(\theta) + t(\kappa_{3S}(\theta)) \frac{\partial \sigma_S(\theta)}{\partial \theta_i}.$$

The partial derivatives of μ_S, σ_S and κ_{3S} are straightforward. The general expressions for μ_S and σ_S are given in (7.3) and (7.4), and are not complicated. Assuming that C and L follow different Gamma distributions, the closed form expression for κ_{3S} can be written as, see also Lemma 3.3 and 3.4,

$$\kappa_{3S} = \frac{\kappa_{3S}^*}{\sigma_S^3},$$

with

$$\kappa_{3S}^* = \lambda \nu_{3C} \left\{ 1 + \epsilon \left[\mu_G^2 \left(\frac{\nu_{3L}}{\mu_G^3} \right) \left(\frac{\mu_C^3}{\nu_{3C}} \right) + 3\mu_G \left(\frac{\nu_{2L}}{\mu_G^2} \right) \left(\frac{\mu_C \nu_{2C}}{\nu_{3C}} \right) \right] \right\},$$

and

$$\begin{aligned}\nu_{2C} &= (\gamma_C^2 + 1)\mu_C^2, \\ \nu_{2L} &= (\gamma_G^2 - 1/\mu_G + 1)\mu_G^2, \\ \nu_{3C} &= \mu_C^3(1 + 3\gamma_C^2 + 2\gamma_C^4), \\ \nu_{3L} &= \mu_G^3(1 + 3(\gamma_G^2 - 1/\mu_G) + 2(\gamma_G^2 - 1/\mu_G)^2).\end{aligned}$$

The only problem is the derivative of t as a function of κ_{3S} . At this place we apply the implicit function theorem. Denoting

$$g_x(x, y) = \frac{\partial g(x, y)}{\partial x} \quad \text{and} \quad g_y(x, y) = \frac{\partial g(x, y)}{\partial y},$$

the required derivative takes the form

$$t'(\kappa_{3S}) = -\frac{g_x(\kappa_{3S}, t(\kappa_{3S}))}{g_y(\kappa_{3S}, t(\kappa_{3S}))}.$$

We do not present the explicit expressions for these derivatives since these are quite complicated, but they can be easily implemented in any mathematical software.

The accuracy of the IG approximation in the considered range of the underlying parameters was analyzed in detail in Section 5.2. Table 7.1 gives an impression of the accuracy of *VaRapp1*. Here C and L each have a (different) Gamma-distribution and for the true value of the parameters we have the following representative choice: $(\mu_{C0}, \gamma_{C0}, \mu_{G0}, \gamma_{G0}, \epsilon_0, \lambda_0) = (100000, 0.7, 15, 0.8, 0.03, 400)$, implying $\gamma_{L0} = 0.76$. For this fixed choice we have $VaRapp(100000, 0.7, 15, 0.8, 0.03, 400) = 43904712, 47861121, 49877491, 52037384$ for $p = 0.9, 0.99, 0.9975, 0.9995$ respectively. For convenience, we also present the values of

$$\gamma_L = \sqrt{\gamma_G^2 - \mu_G^{-1}}.$$

Table 7.1 indicates that the approximation *VaRapp1* is sufficiently accurate for our purposes. Therefore, precisely this approximation will be used in the next sections. Having in mind that $VaRapp1(\mu_{C0}, \gamma_{C0}, \mu_{G0}, \gamma_{G0}, \epsilon_0, \lambda_0) = VaRapp(\mu_{C0}, \gamma_{C0}, \mu_{G0}, \gamma_{G0}, \epsilon_0, \lambda_0)$, we can note that Table 7.1 also gives interesting information on the error in $VaRapp(\hat{\mu}_C, \hat{\gamma}_C, \hat{\mu}_G, \hat{\gamma}_G, \hat{\epsilon}, \hat{\lambda}) - VaRapp(\mu_{C0}, \gamma_{C0}, \mu_{G0}, \gamma_{G0}, \epsilon_0, \lambda_0)$ due to replacing *VaRapp* by *VaRapp1*.

VaRapp1 is basically a linearization of the *VaRapp* function around the fixed point $(\mu_{C0}, \gamma_{C0}, \mu_{G0}, \gamma_{G0}, \epsilon_0, \lambda_0)$. As soon as we believe that *VaRapp*, *VaRapp1* are good approximations of *VaR*, we can hope that the coefficients $\frac{\partial}{\partial \mu_C} VaRapp(\mu_{C0}, \gamma_{C0}, \mu_{L0}, \gamma_{L0}, \epsilon_0, \lambda_0) \dots \frac{\partial}{\partial \lambda} VaRapp(\mu_{C0}, \gamma_{C0}, \mu_{L0}, \gamma_{L0}, \epsilon_0, \lambda_0)$ determining *VaRapp1* will give some impression of the *VaR* fluctuation at the given point.

To get some impression about the order of magnitude of these coefficients we have calculated these at $(\mu_{C0}, \gamma_{C0}, \mu_{L0}, \gamma_{L0}, \epsilon_0, \lambda_0) = (100000, 0.7, 15, 0.8, 0.03, 400)$ using the same distributional assumptions. The results are given in Table 7.2.

$(\mu_C, \gamma_C, \mu_G, \gamma_G, \epsilon, \lambda, p)$	γ_L	$VaRapp$	$VaRapp1$	rel. error	abs. error
(100000, 0.5, 10, 0.6, 0.05, 400, 0.9)	0.51	43580956	43705461	0.003	124506
(110000, 0.3, 12, 1, 0.04, 450, 0.99)	0.96	58612875	57918884	-0.012	-693991
(90000, 0.9, 18, 0.7, 0.05, 450, 0.9975)	0.66	51831945	52660432	0.016	828487
(150000, 0.2, 10, 1.1, 0.02, 400, 0.9995)	1.05	74789728	75942791	0.015	1153063
(70000, 1, 20, 1, 0.03, 400, 0.9)	0.97	31306076	31490797	0.006	184720
(120000, 0.1, 10, 0.6, 0.03, 450, 0.99)	0.51	61480780	60310048	-0.019	-1170733
(200000, 0.8, 20, 0.5, 0.04, 400, 0.9975)	0.45	101314280	100700660	-0.006	-613620
(150000, 0.05, 10, 1.1, 0.05, 400, 0.9995)	1.05	79526453	79580667	0.001	54214

Table 7.1: Accuracy of approximation $VaRapp1$

p	$\frac{\partial}{\partial \mu_C} VaRapp$	$\frac{\partial}{\partial \gamma_C} VaRapp$	$\frac{\partial}{\partial \mu_G} VaRapp$	$\frac{\partial}{\partial \gamma_G} VaRapp$	$\frac{\partial}{\partial \epsilon} VaRapp$	$\frac{\partial}{\partial \lambda} VaRapp$
0.9	439	$12\mu_{C0}$	$0.4946\mu_{C0}$	$7.5\mu_{C0}$	$219\mu_{C0}$	$1.04\mu_{C0}$
0.99	479	$19.2\mu_{C0}$	$1.6824\mu_{C0}$	$28\mu_{C0}$	$553\mu_{C0}$	$1.09\mu_{C0}$
0.9975	499	$21.8\mu_{C0}$	$2.4441\mu_{C0}$	$41.4\mu_{C0}$	$749\mu_{C0}$	$1.1\mu_{C0}$
0.9995	520	$23.8\mu_{C0}$	$3.3590\mu_{C0}$	$57.7\mu_{C0}$	$975\mu_{C0}$	$1.12\mu_{C0}$

Table 7.2: Coefficients of $VaRapp1$ at $(\mu_{C0}, \gamma_{C0}, \mu_{G0}, \gamma_{G0}, \epsilon_0, \lambda_0) = (100000, 0.7, 15, 0.8, 0.03, 400)$ for $p = 0.9, 0.99, 0.9975, 0.9995$

7.2.2 Local behavior of SLP

This section has exactly the same structure and main goals as the previous one. The only difference is that now, instead of VaR , we analyze the behavior of the SLP . Moreover, this work has already been done in Albers and Kallenberg [2007], hence we will be brief here. The definition of SLP is

$$E(S - a)^+.$$

Here a is the retention level which depends on μ_S and σ_S . The larger μ_S and σ_S , the larger retention a will be chosen. This is a natural approach, which is used in the entire thesis. Defining k by $a = \mu_S + k\sigma_S$, or

$$k = \frac{a - \mu_S}{\sigma_S},$$

we will assume that k is chosen in advance, determining the retention a in "standard units". That means that in our approach k does not depend on the parameters, while a does depend on the parameters $\mu_C, \gamma_C, \mu_G, \gamma_G, \epsilon, \lambda$ through μ_S and σ_S .

In order to get insight into the fluctuation of

$$E(S - a)^+ = \sigma_S E\left(\frac{S - \mu_S}{\sigma_S} - k\right)^+,$$

we have to simplify $\sigma_S E(\sigma_S^{-1}(S - \mu_S) - k)^+$ somewhat, because otherwise no conclusions can be drawn. Just as in the previous section we apply two simplification steps. The first step concerns the approximation for the distribution of S . Using the conclusion of Chapter 5, the best candidates are the IG or the Gamma-IG approximation. In spite of the fact that the Gamma-IG approximation showed better results (compared to the IG approximation), the IG approximation will be used in this chapter as a basic approximation. It has a clear advantage of simplicity and, together with the reasonable accuracy (it still matches our chosen criterion), it is considered as a handy approximation tool. All the theoretical results will be proven on the basis of the IG approximation, thus denoting by $SLPapp$ the first-step approximation of SLP .

The second simplification step, as in the VaR simplification case, will be a one step Taylor expansion on the approximation around the true value $(\mu_{C0}, \gamma_{C0}, \mu_{G0}, \gamma_{G0}, \epsilon_0, \lambda_0)$ of the parameters. Similarly to the previous section we call this function $SLPapp1$, which is given by

$$\begin{aligned} SLPapp1(\mu_C, \gamma_C, \mu_G, \gamma_G, \epsilon, \lambda) &= SLPapp(\mu_{C0}, \gamma_{C0}, \mu_{G0}, \gamma_{G0}, \epsilon_0, \lambda_0) \\ &+ (\mu_C - \mu_{C0}) \frac{\partial}{\partial \mu_C} SLPapp(\mu_{C0}, \gamma_{C0}, \mu_{G0}, \gamma_{G0}, \epsilon_0, \lambda_0) \\ &+ \dots + (\lambda - \lambda_0) \frac{\partial}{\partial \lambda} SLPapp(\mu_{C0}, \gamma_{C0}, \mu_{G0}, \gamma_{G0}, \epsilon_0, \lambda_0). \end{aligned} \quad (7.7)$$

Table 7.3 gives an impression of the accuracy of SLP_{app1} . Again, C and L have (different) Gamma-distributions and for the true value of the parameters we use the same choice: $(\mu_{C0}, \gamma_{C0}, \mu_{G0}, \gamma_{G0}, \epsilon_0, \lambda_0) = (100000, 0.7, 15, 0.8, 0.03, 400)$, implying $\gamma_{L0} = 0.76$. For this fixed choice we have $SLP_{app1}(100000, 0.7, 15, 0.8, 0.03, 400) = 1164042, 292282, 56003, 9086$ for $k = 0, 1, 2, 3$ respectively. For convenience, we also present the values of $\gamma_L = \sqrt{\gamma_G^2 - \mu_G^{-1}}$.

To get some impression about the order of magnitude of the coefficients in SLP_{app1} (partial derivatives of SLP w.r.t. the different parameters) we have calculated these at $(\mu_{C0}, \gamma_{C0}, \mu_{G0}, \gamma_{G0}, \epsilon_0, \lambda_0) = (100000, 0.7, 15, 0.8, 0.03, 400)$ using the same distributional assumptions. The results are given in Table 7.4. It can be seen that the coefficients of the λ -term are very small, compared to the remaining terms, and the coefficients of the ϵ -term are large. This means that the function is not very sensitive to changes in λ and the other way around for ϵ .

7.3 Asymptotic behavior of the estimators

In this section we discuss the asymptotic behavior of the estimators which later will be used in the analysis of the estimation effect and construction of the confidence bounds for the risk measures considered. The asymptotics will be considered w.r.t. λ , which will be assumed to tend to infinity. That seems to be the natural way, because λ is the total expected number of claims, that is the expected number of observations. The other parameters are assumed to be fixed.

The theory in this section will be based on the paper of Albers and Kallenberg [2007], where the same theory was presented for SLP . However, all the theorems which are presented in the paper can be formulated in general form, without any connection to the risk measures. To make the chapter self contained, some theorems (without proofs) from Albers and Kallenberg [2007] will be repeated here.

Our estimates are

$$\hat{\mu}_C, \hat{\gamma}_C, \hat{\mu}_G, \hat{\gamma}_G, \hat{\epsilon}, \hat{\lambda}$$

These are functions of the vector

$$\left(\overline{C+D}, \overline{C^2+D^2}, \overline{G}, \overline{G^2}, \overline{H}, \overline{N} \right),$$

the components of which were presented in Section 7.1. Therefore, to get the asymptotic behavior of the estimators, we start with the asymptotic behavior of the vector above.

$(\mu_C, \gamma_C, \mu_G, \gamma_G, \epsilon, \lambda, p)$	γ_L	SLP_{app}	SLP_{app1}	rel. error	abs. error
(10000, 0.5, 10, 0.6, 0.05, 400, 0.9)	0.51	1089187	1131613	0.04	42429
(110000, 0.3, 12, 1, 0.04, 450, 0.99)	0.96	339776	332509	0.02	7267
(90000, 0.9, 18, 0.7, 0.05, 450, 0.9975)	0.66	64051	67969	0.06	3918
(150000, 0.2, 10, 1.1, 0.02, 400, 0.9995)	1.05	13180	15544	0.18	2364
(70000, 1, 20, 1, 0.03, 400, 0.9)	0.97	957230	1009965	0.06	52735
(120000, 0.1, 10, 0.6, 0.03, 450, 0.99)	0.51	275809	272302	0.01	3508
(200000, 0.8, 20, 0.5, 0.04, 400, 0.9975)	0.45	114474	115798	0.01	1324
(150000, 0.05, 10, 1.1, 0.05, 400, 0.9995)	1.05	18330	18904	0.03	575

Table 7.3: Accuracy of approximation SLP_{app1}

k	$\frac{\partial}{\partial \mu_C} SLP_{app}$	$\frac{\partial}{\partial \gamma_C} SLP_{app}$	$\frac{\partial}{\partial \mu_G} SLP_{app}$	$\frac{\partial}{\partial \gamma_G} SLP_{app}$	$\frac{\partial}{\partial \epsilon} SLP_{app}$	$\frac{\partial}{\partial \lambda} SLP_{app}$
0	11.6404	3.8817 μ_{C0}	0.1047 μ_{C0}	1.3173 μ_{C0}	61.9452 μ_{C0}	0.0150 μ_{C0}
1	2.9228	0.7076 μ_{C0}	0.0632 μ_{C0}	1.0253 μ_{C0}	21.6362 μ_{C0}	0.0032 μ_{C0}
2	0.5600	-0.0210 μ_{C0}	0.0343 μ_{C0}	0.6532 μ_{C0}	6.4573 μ_{C0}	0.0003 μ_{C0}
3	0.0909	-0.0459 μ_{C0}	0.0116 μ_{C0}	0.2336 μ_{C0}	1.5790 μ_{C0}	-0.0001 μ_{C0}

Table 7.4: Coefficients of SLP_{app1} at $(\mu_{C0}, \gamma_{C0}, \mu_{G0}, \gamma_{G0}, \epsilon_0, \lambda_0) = (100000, 0.7, 15, 0.8, 0.03, 400)$ for $k = 0, 1, 2, 3$

Theorem 7.2. *Assume that $\lambda \rightarrow \infty$ and that $u, \mu_C, \gamma_C, \mu_G, \gamma_G, \epsilon$ are fixed. Let*

$$\begin{aligned} X_{1\lambda} &= \left\{ \frac{\overline{C+D}}{\mu_C} - 1 \right\} \frac{\sqrt{u\lambda}}{\gamma_C}, \\ X_{2\lambda} &= \left\{ \frac{\overline{C^2+D^2}}{\mu_C^2} - (1 + \gamma_C^2) \right\} \frac{\sqrt{u\lambda}}{\gamma_C}, \\ X_{3\lambda} &= \left\{ \frac{\overline{G}}{\mu_G} - 1 \right\} \sqrt{\frac{\epsilon u \lambda}{\mu_G}}, \\ X_{4\lambda} &= \left\{ \frac{\overline{G^2}}{\mu_G} - \mu_G(1 + \gamma_G^2) \right\} \sqrt{\frac{\epsilon u \lambda}{\mu_G}}, \\ X_{5\lambda} &= \left\{ \frac{\overline{H}\mu_G}{\epsilon\lambda} - 1 \right\} \sqrt{\frac{\epsilon u \lambda}{\mu_G}}, \\ X_{6\lambda} &= \left\{ \frac{\overline{N}}{\lambda(1-\epsilon)} - 1 \right\} \sqrt{u\lambda(1-\epsilon)}. \end{aligned}$$

Then, as $\lambda \rightarrow \infty$,

$$(X_{1\lambda}, X_{2\lambda}, X_{3\lambda}, X_{4\lambda}, X_{5\lambda}, X_{6\lambda}) \rightarrow (U_1, U_2, U_3, U_4, U_5, U_6)$$

with

$$\begin{aligned} (U_1, U_2) &\sim N\left(0, 0, \frac{1}{2 + \gamma_C \kappa_{3C}}, \frac{2 + \gamma_C \kappa_{3C}}{\gamma_C^2(\kappa_{4C} + 2) + 4\gamma_C \kappa_{3C} + 4}\right), \\ (U_3, U_4) &\sim N\left(0, 0, \frac{\gamma_G^2}{\mu_G \gamma_G^2(2 + \gamma_G \kappa_{3G})}, \frac{\mu_G \gamma_G^2(2 + \gamma_G \kappa_{3G})}{\mu_G^2 \gamma_G^2 \{\gamma_G^2(\kappa_{4G} + 2) + 4\gamma_G \kappa_{3G} + 4\}}\right), \\ U_5 &\sim N(0, 1), U_6 \sim N(0, 1) \end{aligned}$$

and $(U_1, U_2), (U_3, U_4), U_5, U_6$ independent.

Proof. For the proof see Theorem 5.1 in Albers and Kallenberg [2007]. □

Remark 7.2. Theorem 7.2 can be applied to $G \sim P(L)$ with parametrization μ_L, γ_L . The necessary condition for that is the final fourth moment of L . We rewrite $X_{3\lambda}$ and $X_{4\lambda}$ as

$$\begin{aligned} X_{3\lambda} &= \left\{ \frac{\overline{G}}{\mu_L} - 1 \right\} \sqrt{\frac{\epsilon u \lambda}{\mu_L}}, \\ X_{4\lambda} &= \left\{ \frac{\overline{G^2}}{\mu_L} - \mu_L(1 + \gamma_L^2) - 1 \right\} \sqrt{\frac{\epsilon u \lambda}{\mu_L}} \end{aligned}$$

and use the formula $\gamma_G^2 = \gamma_L^2 + \mu_L^{-1}$. The asymptotic normality in that case is described by (U_3, U_4) , which is distributed as

$$N \left(\begin{array}{cc} \gamma_L^2 + \mu_L^{-1} & \mu_L \gamma_L^2 (2 + \gamma_L \kappa_{3L}) \\ 0, 0, \mu_L \gamma_L^2 (2 + \gamma_L \kappa_{3L}), & +2 + 3\gamma_L^2 + \mu_L^{-1} \\ +2 + 3\gamma_L^2 + \mu_L^{-1} & \mu_L^2 \gamma_L^2 \{ \gamma_L^2 (\kappa_{4L} + 2) + 4\gamma_L \kappa_{3L} + 4 \} \\ & +2\mu_L (3\gamma_L^3 \kappa_{3L} + 8\gamma_L^2 + 2) + 6 + 7\gamma_L^2 + \mu_L^{-1} \end{array} \right).$$

Obviously, in $X_{5\lambda}$ we can replace μ_G by μ_L . \square

The following theorem gives the asymptotic behavior of linear combination of the estimators. Later these results will be used in the analysis of the estimation effects on the risk measures.

Theorem 7.3. *Assume that $\lambda \rightarrow \infty$ and that $u, \mu_C, \gamma_C, \mu_G, \gamma_G, \epsilon$ are fixed and all positive with $\epsilon < 1$. Let c_1, c_2, \dots, c_6 be deterministic functions of $\mu_C, \gamma_C, \mu_G, \gamma_G, \epsilon$ and λ . Define*

$$Z_1 = c_1 \frac{\hat{\mu}_C - \mu_C}{\mu_C} + c_2 (\hat{\gamma}_C - \gamma_C),$$

$$Z_2 = c_3 (\hat{\mu}_G - \mu_G) \sqrt{\epsilon} + c_4 (\hat{\gamma}_G - \gamma_G) \sqrt{\epsilon} + c_5 \left(\frac{\hat{\epsilon} - \epsilon}{\epsilon} \right) \sqrt{\epsilon} + c_6 \frac{\hat{\lambda} - \lambda}{\lambda}.$$

Then, as $\lambda \rightarrow \infty$,

$$\left(\frac{Z_1}{\tau_1}, \frac{Z_2}{\tau_2} \right) \sqrt{u\lambda} \rightarrow (V_1, V_2)$$

with V_1, V_2 independent and $V_1 \sim N(0, 1)$ and $V_2 \sim N(0, 1)$ with

$$\tau_1^2 = \gamma_C^2 \{ c_1^2 + c_1 c_2 (\kappa_{3C} - 2\gamma_C) + c_2^2 (\gamma_C^2 + \kappa_{4C}/4 + 1/2 - \gamma_C \kappa_{3C}) \}$$

and

$$\begin{aligned} \tau_2^2 = & c_3^2 \mu_G^3 \gamma_G^2 \\ & + c_4^2 \mu_G \gamma_G^2 (\gamma_G^2 - \gamma_G \kappa_{3G} + \kappa_{4G}/4 + 1/2) \\ & + c_5^2 (1 - \epsilon) \{ \mu_G (1 - \epsilon) (1 + \gamma_G^2) + \epsilon \} \\ & + c_6^2 \{ \mu_G \epsilon (1 + \gamma_G^2) + 1 - \epsilon \} \\ & + c_3 c_4 \mu_G^2 \gamma_G^2 (\kappa_{3G} - 2\gamma_G) \\ & + 2c_3 c_5 (1 - \epsilon) \mu_G^2 \gamma_G^2 \\ & + 2c_3 c_6 \sqrt{\epsilon} \mu_G^2 \gamma_G^2 \\ & + c_4 c_5 \mu_G \gamma_G^2 (1 - \epsilon) (\kappa_{3G} - 2\gamma_G) \\ & + c_4 c_6 \mu_G \gamma_G^2 \sqrt{\epsilon} (\kappa_{3G} - 2\gamma_G) \\ & + 2c_5 c_6 \sqrt{\epsilon} (1 - \epsilon) \{ \mu_G (1 + \gamma_G^2) - 1 \}. \end{aligned}$$

Proof. We start with considering Z_1 . By using (5.2) and (5.3) in Albers and Kallenberg [2007] we have

$$\frac{\hat{\mu}_C - \mu_C}{\mu_C} \sqrt{u\lambda} = X_{1\lambda} \gamma_C \quad (7.8)$$

and

$$(\hat{\gamma}_C - \gamma_C) \sqrt{u\lambda} = \frac{1}{2} X_{2\lambda} - X_{1\lambda} (1 + \gamma_C^2) + O_P(\lambda^{-1/2}) \quad (7.9)$$

as $\lambda \rightarrow \infty$. Now, using (7.8) and (7.9), we can write that

$$\begin{aligned} \frac{Z_1}{\tau_1} \sqrt{u\lambda} &= \tau_1^{-1} \left\{ c_1 X_{1\lambda} \gamma_C + c_2 \left(\frac{1}{2} X_{2\lambda} - X_{1\lambda} (1 + \gamma_C^2) + O_P(\lambda^{-1/2}) \right) \right\} \\ &= \tau_1^{-1} \left\{ c_1 (X_{1\lambda} - U_1) \gamma_C + c_2 \left(\frac{1}{2} (X_{2\lambda} - U_2) - (X_{1\lambda} - U_1) (1 + \gamma_C^2) \right) \right\} \\ &\quad + \tau_1^{-1} \left\{ c_1 U_1 \gamma_C + c_2 \left(\frac{1}{2} U_2 - U_1 (1 + \gamma_C^2) \right) \right\} + \frac{c_2}{\tau_1} O_P(\lambda^{-1/2}). \end{aligned}$$

Next we show that $|c_1/\tau_1|$ and $|c_2/\tau_1|$ are bounded above as functions of λ . This simply follows from the fact that $\tau_1^2 = \text{Var}[c_1 X + c_2 Y]$ with $X = U_1 \gamma_C$ and $Y = U_2/2 - U_1 (1 + \gamma_C^2)$ and thus $\tau_1^2 \geq (1 - \rho(X, Y)^2) \max\{\text{Var}[c_1 X], \text{Var}[c_2 Y]\}$. Since X and Y do not depend on λ (and therefore $\text{Var}[X]$, $\text{Var}[Y]$ and $\rho(X, Y)$ also do not depend on λ) the result immediately follows.

Application of Theorem 7.2 now yields

$$\frac{Z_1}{\tau_1} \sqrt{u\lambda} = \tau_1^{-1} \left\{ c_1 U_1 \gamma_C + c_2 \left(\frac{1}{2} U_2 - U_1 (1 + \gamma_C^2) \right) \right\} + o_P(1)$$

and hence $\tau_1^{-1} Z_1 \sqrt{u\lambda}$ converges in distribution to a $N(0, 1)$ random variable.

In the proof of Theorem 5.2 of Albers and Kallenberg [2007] it has been shown that

$$(\hat{\mu}_G - \mu_G) \sqrt{\epsilon u \lambda} = X_{3\lambda} \mu_G^{3/2}, \quad (7.10)$$

$$\begin{aligned} (\hat{\gamma}_G - \gamma_G) \sqrt{\epsilon u \lambda} &= \frac{1}{2} \gamma_G^{-1} \mu_G^{-1/2} \{ X_{4\lambda} - 2\mu_G (1 + \gamma_G^2) X_{3\lambda} \} + O_P(\lambda^{-1/2}), \quad (7.11) \end{aligned}$$

$$\begin{aligned} \left(\frac{\hat{\epsilon} - \epsilon}{\epsilon} \right) \sqrt{\epsilon u \lambda} &= (1 - \epsilon) \mu_G^{1/2} (X_{5\lambda} + X_{3\lambda}) - X_{6\lambda} (\epsilon(1 - \epsilon))^{1/2} + O_P(\lambda^{-1/2}) \quad (7.12) \end{aligned}$$

and

$$\frac{\hat{\lambda} - \lambda}{\lambda} \sqrt{u\lambda} = (\epsilon\mu_G)^{1/2} (X_{5\lambda} + X_{3\lambda}) + (1 - \epsilon)^{1/2} X_{6\lambda} + O_P(\lambda^{-1/2}) \quad (7.13)$$

as $\lambda \rightarrow \infty$. Now consider $Z_2\tau_2^{-1}\sqrt{u\lambda}$. Using (7.10)-(7.13) we can write that $\tau_2^{-1}Z_2\sqrt{u\lambda}$ is equal to

$$\begin{aligned} & \frac{c_3}{\tau_2} (X_{3\lambda} - U_3) \mu_G^{3/2} + \frac{c_4}{\tau_2} \left(\frac{1}{2} \gamma_G^{-1} \mu_G^{-1/2} \{ (X_{4\lambda} - U_4) - 2\mu_G (1 + \gamma_G^2) (X_{3\lambda} - U_3) \} \right) \\ & + \frac{c_5}{\tau_2} \left((1 - \epsilon) \mu_G^{1/2} ((X_{5\lambda} - U_5) + (X_{3\lambda} - U_3)) - (X_{6\lambda} - U_6) (\epsilon (1 - \epsilon))^{1/2} \right) \\ & + \frac{c_6}{\tau_2} \left((\epsilon\mu_G)^{1/2} ((X_{5\lambda} - U_5) + (X_{3\lambda} - U_3)) + (1 - \epsilon)^{1/2} (X_{6\lambda} - U_6) \right) \\ & + \tau_2^{-1} \left[c_3 U_3 \mu_G^{3/2} + c_4 \left(\frac{1}{2} \gamma_G^{-1} \mu_G^{-1/2} \{ U_4 - 2\mu_G (1 + \gamma_G^2) U_3 \} \right) \right. \\ & + c_5 \left((1 - \epsilon) \mu_G^{1/2} (U_5 + U_3) - U_6 (\epsilon (1 - \epsilon))^{1/2} \right) \\ & \left. + c_6 \left((\epsilon\mu_G)^{1/2} (U_5 + U_3) + (1 - \epsilon)^{1/2} U_6 \right) \right] \\ & + \left(\frac{c_4}{\tau_2} + \frac{c_5}{\tau_2} + \frac{c_6}{\tau_2} \right) O_P(\lambda^{-1/2}). \end{aligned}$$

By a similar argument as before it follows that $|c_3/\tau_2|, |c_4/\tau_2|, |c_5/\tau_2|$ and $|c_6/\tau_2|$ are bounded above as functions of λ . Note that τ_2^2 is of the form $Var[c_3X_1 + c_4X_2 + c_5X_3 + c_6X_4]$ and thus $\tau_2^2 \geq (1 - \rho_i^{*2}) Var[c_{2+i}X_i], i = 1, \dots, 4$, where ρ_i^{*2} is the multiple correlation coefficient of X_i with the other X_j 's, which does not depend on λ .

Application of Theorem 7.2 gives that $\tau_2^{-1}Z_2\sqrt{u\lambda}$ converges to a $N(0, 1)$ random variable. Since Z_1 is a function of $X_{1\lambda}$ and $X_{2\lambda}$, and Z_2 a function of $X_{3\lambda}, \dots, X_{6\lambda}$, the asymptotic independence of $(X_{1\lambda}, X_{2\lambda})$ and $(X_{3\lambda}, \dots, X_{6\lambda})$ completes the proof. \square

Remark 7.3. Theorem 7.3 can be applied to $G : P(L)$ with parametrization μ_L, γ_L (provided that the fourth moment of L is finite), replacing $c_3(\hat{\mu}_G - \mu_G) \times \sqrt{\epsilon} + c_4(\hat{\gamma}_G - \gamma_G)\sqrt{\epsilon}$ by $c_3(\hat{\mu}_L - \mu_L)\sqrt{\epsilon} + c_4(\hat{\gamma}_L - \gamma_L)\sqrt{\epsilon}$ and τ_2^2 by

$$\begin{aligned}
\tau_2^2 &= c_3^2(\mu_L^3\gamma_L^2 + \mu_L^2) \\
&+ c_4^2 \left\{ \mu_L\gamma_L^2 \left(\gamma_L^2 - \gamma_L\kappa_{3L} + \frac{\kappa_{4L}}{4} + \frac{1}{2} \right) - \gamma_L^2 + \gamma_L\kappa_{3L} + 1 + \frac{1}{2}\mu_L^{-1}(1 + \gamma_L^{-2}) \right\} \\
&+ c_5^2(1 - \epsilon)\{\mu_L(1 - \epsilon)(1 + \gamma_L^2) + 1\} \\
&+ c_6^2\{\mu_L\epsilon(1 + \gamma_L^2) + 1\} \\
&+ c_3c_4\mu_L^2\gamma_L^2(\kappa_{3L} - 2\gamma_L) \\
&+ 2c_3c_5(1 - \epsilon)(\mu_L^2\gamma_L^2 + \mu_L) \\
&+ 2c_3c_6\sqrt{\epsilon}(\mu_L^2\gamma_L^2 + \mu_L) \\
&+ c_4c_5\mu_L\gamma_L^2(1 - \epsilon)(\kappa_{3L} - 2\gamma_L) \\
&+ c_4c_6\mu_L\gamma_L^2\sqrt{\epsilon}(\kappa_{3L} - 2\gamma_L) \\
&+ 2c_5c_6\mu_L\sqrt{\epsilon}(1 - \epsilon)(1 + \gamma_L^2).
\end{aligned}$$

□

The following results present the asymptotic behavior of some functions of the estimators considered.

Lemma 7.4. *Let us denote $\theta = (\mu_C, \gamma_C, \mu_G, \gamma_G, \epsilon, \lambda)$ with $\theta_i = \mu_C, \gamma_C, \mu_G, \gamma_G, \epsilon, \lambda$ for $i = 1, 2, 3, 4, 5, 6$, respectively. Denote also $\hat{\theta} = (\hat{\mu}_C, \hat{\gamma}_C, \hat{\mu}_G, \hat{\gamma}_G, \hat{\epsilon}, \hat{\lambda})$, $\Omega = \{\theta | \theta \in R_+^6, \epsilon < 1\}$. We assume that Model 4 holds and that C and D are identically distributed with a Gamma, Inverse Gaussian or lognormal distribution. In that case*

- (a) $\mu_S(\theta), \sigma_S(\theta), \kappa_{3S}(\theta), \kappa_{4S}(\theta) \in C^1(\Omega)$, where $C^1(\Omega)$ is used to denote the set of differentiable functions whose derivative is continuous in Ω ,
- (b) $\frac{\mu_S(\hat{\theta})}{\mu_S(\theta)} \xrightarrow{P} 1, \frac{\sigma_S(\hat{\theta})}{\sigma_S(\theta)} \xrightarrow{P} 1, \frac{\kappa_{3S}(\hat{\theta})}{\kappa_{3S}(\theta)} \xrightarrow{P} 1$ as $\lambda \rightarrow \infty$,
- (c) $\frac{\frac{\partial}{\partial \theta_i} \mu_S(\hat{\theta})}{\frac{\partial}{\partial \theta_i} \mu_S(\theta)} \xrightarrow{P} 1, \frac{\frac{\partial}{\partial \theta_i} \sigma_S(\hat{\theta})}{\frac{\partial}{\partial \theta_i} \sigma_S(\theta)} \xrightarrow{P} 1, \frac{\frac{\partial}{\partial \theta_i} \kappa_{3S}(\hat{\theta})}{\frac{\partial}{\partial \theta_i} \kappa_{3S}(\theta)} \xrightarrow{P} 1$ as $\lambda \rightarrow \infty$.

Proof. Using Lemma 3.1 and (3.1), we directly obtain

$$\begin{aligned}
\mu_S &= \lambda f_1(\mu_C), \\
\sigma_S &= \lambda^{1/2} f_2(\mu_C, \gamma_C, \mu_G, \gamma_G, \epsilon), \\
\kappa_{3S} &= \lambda^{-1/2} f_3(\mu_C, \gamma_C, \mu_G, \gamma_G, \epsilon),
\end{aligned} \tag{7.14}$$

where

$$\begin{aligned} f_1(\mu_C) &= \mu_C, \\ f_2(\mu_C, \gamma_C, \mu_G, \gamma_G, \epsilon) &= [(1 - \epsilon)\nu_{2C} + (\epsilon/\mu_G)\{\nu_{2G}\mu_C^2 + \mu_G\sigma_C^2\}]^{1/2}, \\ f_3(\mu_C, \gamma_C, \mu_G, \gamma_G, \epsilon) &= \frac{(1 - \epsilon)\nu_{3C} + (\epsilon/\mu_G)\{\nu_{3G}\mu_C^3 + 3\nu_{2G}\mu_C\sigma_C^2 + \mu_G\kappa_{3C}^*\}}{f_2^3(\mu_C, \gamma_C, \mu_G, \gamma_G, \epsilon)}. \end{aligned}$$

Because for the distributions of C considered here quantities like ν_{3C}, κ_{3C}^* are nice functions of (μ_C, γ_C) and similarly for G , we obtain the result given in **(a)**. Since the estimators $\hat{\mu}_C, \hat{\gamma}_C, \hat{\mu}_G, \hat{\gamma}_G, \hat{\epsilon}$ are consistent estimators of the corresponding parameters as $\lambda \rightarrow \infty$, we directly obtain

$$\begin{aligned} \frac{f_1(\hat{\mu}_C)}{f_1(\mu_C)} &\xrightarrow{P} 1, \\ \frac{f_2(\hat{\mu}_C, \hat{\gamma}_C, \hat{\mu}_G, \hat{\gamma}_G, \hat{\epsilon})}{f_2(\mu_C, \gamma_C, \mu_G, \gamma_G, \epsilon)} &\xrightarrow{P} 1, \\ \frac{f_3(\hat{\mu}_C, \hat{\gamma}_C, \hat{\mu}_G, \hat{\gamma}_G, \hat{\epsilon})}{f_3(\mu_C, \gamma_C, \mu_G, \gamma_G, \epsilon)} &\xrightarrow{P} 1 \end{aligned}$$

as $\lambda \rightarrow \infty$. Because (see (7.13)) $\hat{\lambda}/\lambda = 1 + O_P(\lambda^{-1/2})$ and hence

$$\frac{\hat{\lambda}}{\lambda} \xrightarrow{P} 1, \quad (7.15)$$

(b) follows directly. Using **(a)**, (7.15) and the consistency of the estimators, we obtain **(c)**. \square

Lemma 7.5. *Suppose that the assumptions and notations of Lemma 7.4 hold. Then*

(a) $SLPapp(\theta) = \sigma_S(\theta) h(\kappa_{3S}(\theta))$ with

$$h(y) = \int_k^\infty \frac{x - k}{\sqrt{2\pi} (1 + \frac{1}{3}xy)^3} \exp\left\{-\frac{x^2}{2(1 + \frac{1}{3}xy)}\right\} dx,$$

(b) $\lim_{y \rightarrow 0} h(y) = \int_k^\infty (x - k) \phi(x) dx = \phi(k) - k\Phi(-k),$

(c) $\lim_{y \rightarrow 0} h'(y) = \frac{1}{6}k\phi(k),$

(d) $\frac{SLPapp(\hat{\theta})}{SLPapp(\theta)} \xrightarrow{P} 1$ as $\lambda \rightarrow \infty,$

(e) $\frac{\frac{\partial}{\partial \hat{\theta}_i} SLPapp(\hat{\theta})}{\frac{\partial}{\partial \theta_i} SLPapp(\theta)} \xrightarrow{P} 1$ as $\lambda \rightarrow \infty.$

Proof. Taking $a = \mu_S + k\sigma_S$ and inserting (4.32) in the IG approximation gives **(a)**. Direct calculation of the limit of $h(y)$ gives **(b)**. Direct calculation gives

$$h'(y) = - \int_k^\infty \frac{(1/2)x(x-k)}{\sqrt{2\pi(1+(1/3)xy)^5}} \exp\left\{-\frac{x^2}{2(1+(1/3)xy)}\right\} dx \\ + \int_k^\infty \frac{(1/6)x^3(x-k)}{\sqrt{2\pi(1+(1/3)xy)^7}} \exp\left\{-\frac{x^2}{2(1+(1/3)xy)}\right\} dx$$

and hence

$$\lim_{y \rightarrow 0} h'(y) = \int_k^\infty ((1/6)x^3 - (1/2)x)(x-k)\phi(x) dx = (1/6)k\phi(k),$$

which is the result presented in **(c)**. By (7.14) we have $\lim_{\lambda \rightarrow \infty} \kappa_{3S}(\theta) = 0$ and hence in view of Lemma 7.4**(b)** $\kappa_{3S}(\hat{\theta}) \xrightarrow{P} 0$ as $\lambda \rightarrow \infty$. Since

$$\frac{SLP_{app}(\hat{\theta})}{SLP_{app}(\theta)} = \frac{\sigma_S(\hat{\theta})h(\kappa_{3S}(\hat{\theta}))}{\sigma_S(\theta)h(\kappa_{3S}(\theta))},$$

application of Lemma 7.4**(b)** together with **(b)** of the present lemma yields

$$\frac{SLP_{app}(\hat{\theta})}{SLP_{app}(\theta)} \xrightarrow{P} \frac{\phi(k) - k\Phi(-k)}{\phi(k) - k\Phi(-k)} = 1,$$

which is the result presented in **(d)**. We have

$$\frac{\partial}{\partial \theta_i} SLP_{app}(\theta) = \left\{ \frac{\partial}{\partial \theta_i} \sigma_S(\theta) \right\} h(\kappa_{3S}(\theta)) + \sigma_S(\theta) h'(\kappa_{3S}(\theta)) \frac{\partial}{\partial \theta_i} \kappa_{3S}(\theta).$$

It is easily seen from (7.14) that the second term is of lower order, that is

$$\lim_{\lambda \rightarrow \infty} \frac{\frac{\partial}{\partial \theta_i} SLP_{app}(\theta)}{\left\{ \frac{\partial}{\partial \theta_i} \sigma_S(\theta) \right\} h(\kappa_{3S}(\theta))} = 1 \quad (7.16)$$

and similarly

$$\frac{\frac{\partial}{\partial \theta_i} SLP_{app}(\hat{\theta})}{\left\{ \frac{\partial}{\partial \theta_i} \sigma_S(\hat{\theta}) \right\} h(\kappa_{3S}(\hat{\theta}))} \xrightarrow{P} 1. \quad (7.17)$$

Moreover, by Lemma 7.4**(c)**

$$\frac{\frac{\partial}{\partial \theta_i} \sigma_S(\hat{\theta})}{\frac{\partial}{\partial \theta_i} \sigma_S(\theta)} \xrightarrow{P} 1. \quad (7.18)$$

Combining (7.16), (7.17) and (7.18) together with **(b)** from the present lemma gives

$$\frac{\frac{\partial}{\partial \theta_i} SLPapp(\hat{\theta})}{\frac{\partial}{\partial \theta_i} SLPapp(\theta)} = \frac{\frac{\frac{\partial}{\partial \theta_i} SLPapp(\hat{\theta})}{\left\{ \frac{\partial}{\partial \theta_i} \sigma_S(\hat{\theta}) \right\} h(\kappa_{3S}(\hat{\theta}))}}{\frac{\frac{\partial}{\partial \theta_i} SLPapp(\theta)}{\left\{ \frac{\partial}{\partial \theta_i} \sigma_S(\theta) \right\} h(\kappa_{3S}(\theta))}} \left\{ \frac{\partial}{\partial \theta_i} \sigma_S(\hat{\theta}) \right\} h(\kappa_{3S}(\hat{\theta})) \xrightarrow{P} 1,$$

which completes the proof of the lemma. \square

Lemma 7.6. *Suppose that the assumptions and notations of Lemma 7.4 hold. Then*

(a) $VaRapp(\theta) = \mu_S(\theta) + t(\kappa_{3S}(\theta)) \sigma_S(\theta)$, where $t(x)$ is the solution of $g(x, t(x)) = 0$ with

$$g(x, y) = \Phi\left(\frac{y}{\sqrt{xy/3+1}}\right) + e^{2(3/x)^2} \Phi\left(\frac{-y-6/x}{\sqrt{xy/3+1}}\right) - p,$$

(b) $\lim_{x \rightarrow 0} g(x, y) = \Phi(y) - p$,

(c) $\lim_{\lambda \rightarrow \infty} t(\kappa_{3S}(\theta)) = \Phi^{-1}(p)$,

(d) $\lim_{\substack{x \rightarrow 0 \\ y \rightarrow \Phi^{-1}(p)}} g_x(x, y) = -\frac{1}{6} \phi(\Phi^{-1}(p)) \left\{ (\Phi^{-1}(p))^2 - 1 \right\}$,
 $\lim_{\substack{x \rightarrow 0 \\ y \rightarrow \Phi^{-1}(p)}} g_y(x, y) = \phi(\Phi^{-1}(p))$,

(e) $\lim_{\lambda \rightarrow \infty} t'(\kappa_{3S}(\theta)) = \frac{1}{6} \left\{ (\Phi^{-1}(p))^2 - 1 \right\}$,

(f) $\lim_{\lambda \rightarrow \infty} \frac{VaRapp(\theta)}{\mu_S(\theta)} = 1$,

(g) $\lim_{\lambda \rightarrow \infty} \frac{\frac{\partial}{\partial \theta_i} VaRapp(\theta)}{\frac{\partial}{\partial \theta_i} \mu_S(\theta)} = 1$,

(h) $\frac{VaRapp(\hat{\theta})}{VaRapp(\theta)} \xrightarrow{P} 1$,

(j) $\frac{\frac{\partial}{\partial \theta_i} VaRapp(\hat{\theta})}{\frac{\partial}{\partial \theta_i} VaRapp(\theta)} \xrightarrow{P} 1$.

Proof. Using Lemma 7.1 we directly obtain **(a)**. Obviously

$$\lim_{x \rightarrow 0} \Phi\left(\frac{y}{\sqrt{xy/3+1}}\right) - p = \Phi(y) - p. \quad (7.19)$$

Using

$$\lim_{y \rightarrow \infty} \Phi(-y) y / \phi(y) = 1 \quad (7.20)$$

it is seen that

$$\begin{aligned} & \lim_{x \rightarrow 0} e^{2(3/x)^2} \Phi\left(\frac{-y-6/x}{\sqrt{xy/3+1}}\right) \\ &= \lim_{x \rightarrow 0} \frac{e^{2(3/x)^2} \phi\left(\frac{y+6/x}{\sqrt{xy/3+1}}\right) \sqrt{xy/3+1}}{y+6/x} \\ &= \lim_{x \rightarrow 0} \frac{e^{-\frac{y^2}{2(1+xy/3)}} \sqrt{xy/3+1}}{y+6/x} = 0, \end{aligned}$$

which together with (7.19) gives **(b)**. The result in **(c)** follows from the definition of $t(\kappa_{3S}(\theta))$, **(b)** of the present lemma and $\lim_{\lambda \rightarrow \infty} \kappa_{3S}(\theta) = 0$. Now consider

$$\begin{aligned} g_x(x, y) &= \phi\left(\frac{y}{\sqrt{xy/3+1}}\right) \frac{\partial}{\partial x} \left(\frac{y}{\sqrt{xy/3+1}}\right) \\ &\quad + e^{18/x^2} \Phi\left(\frac{-y-6/x}{\sqrt{xy/3+1}}\right) \frac{\partial}{\partial x} \left(\frac{18}{x^2}\right) \\ &\quad + e^{18/x^2} \phi\left(\frac{-y-6/x}{\sqrt{xy/3+1}}\right) \frac{\partial}{\partial x} \left(\frac{-y-6/x}{\sqrt{xy/3+1}}\right). \end{aligned}$$

Using direct calculation

$$\lim_{x \rightarrow 0} \phi\left(\frac{y}{\sqrt{xy/3+1}}\right) \frac{\partial}{\partial x} \left(\frac{y}{\sqrt{xy/3+1}}\right) = -\frac{\phi(y) y^2}{6}. \quad (7.21)$$

Using

$$\Phi(-y) = \phi(y) y^{-1} (1 - y^{-2} + O(y^{-4}))$$

as $y \rightarrow \infty$, we obtain

$$\begin{aligned} & e^{18/x^2} \Phi\left(\frac{-y-6/x}{\sqrt{xy/3+1}}\right) \frac{\partial}{\partial x} \left(\frac{18}{x^2}\right) + e^{18/x^2} \phi\left(\frac{-y-6/x}{\sqrt{xy/3+1}}\right) \frac{\partial}{\partial x} \left(\frac{-y-6/x}{\sqrt{xy/3+1}}\right) \\ &= e^{18/x^2} \phi\left(\frac{y+6/x}{\sqrt{xy/3+1}}\right) \left[\frac{\sqrt{xy/3+1}}{y+6/x} \left(1 - \frac{xy/3+1}{(y+6/x)^2} + O(x^4)\right) \right. \\ &\quad \times \left. \frac{\partial}{\partial x} \left(\frac{18}{x^2}\right) - \frac{\partial}{\partial x} \left(\frac{y+6/x}{\sqrt{xy/3+1}}\right) \right] = \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2(1+xy/3)}} \left[\frac{36(1+xy/3)^{3/2}}{(xy+6)^3} \right. \\ &\quad \left. + \frac{3xy^3/2}{(xy+6)(xy+3)\sqrt{3xy+9}} + O(x^2) \right] \rightarrow \frac{1}{6} \phi(y) \end{aligned}$$

as $x \rightarrow 0$. Together with (7.21) we obtain

$$\lim_{x \rightarrow 0} g_x(x, y) = -\frac{1}{6} \phi(y) (y^2 - 1)$$

and

$$\begin{aligned} g_y(x, y) &= \phi \left(\frac{y}{\sqrt{xy/3 + 1}} \right) \frac{\partial}{\partial y} \left(\frac{y}{\sqrt{xy/3 + 1}} \right) \\ &\quad + e^{18/x^2} \phi \left(\frac{-y - 6/x}{\sqrt{xy/3 + 1}} \right) \frac{\partial}{\partial y} \left(\frac{-y - 6/x}{\sqrt{xy/3 + 1}} \right) \\ &= \phi \left(\frac{y}{\sqrt{xy/3 + 1}} \right) \frac{\partial}{\partial y} \left(\frac{y}{\sqrt{xy/3 + 1}} \right) \\ &\quad - \phi \left(\frac{y}{\sqrt{xy/3 + 1}} \right) \frac{\partial}{\partial y} \left(\frac{y + 6/x}{\sqrt{xy/3 + 1}} \right) \\ &= -\phi \left(\frac{y}{\sqrt{xy/3 + 1}} \right) \frac{\partial}{\partial y} \left(\frac{6/x}{\sqrt{xy/3 + 1}} \right) \\ &= \phi \left(\frac{y}{\sqrt{xy/3 + 1}} \right) (xy/3 + 1)^{-3/2} \rightarrow \phi(y) \text{ as } x \rightarrow 0, \end{aligned}$$

from which **(d)** follows directly. We know that

$$t'(\kappa_{3S}(\theta)) = -\frac{g_x(\kappa_{3S}(\theta), t(\kappa_{3S}(\theta)))}{g_y(\kappa_{3S}(\theta), t(\kappa_{3S}(\theta)))}.$$

Using (7.14) and **(c)** of the present lemma, we can write that

$$\lim_{\lambda \rightarrow \infty} t'(\kappa_{3S}(\theta)) = \lim_{\substack{x \rightarrow 0 \\ y \rightarrow \Phi^{-1}(p)}} -\frac{g_x(x, y)}{g_y(x, y)}.$$

Using **(d)** of the present lemma, we directly obtain **(e)**. By definition

$$VaRapp(\theta) = \mu_S(\theta) + t(\kappa_{3S}(\theta)) \sigma_S(\theta).$$

Using (7.14) and **(c)** of the present lemma, we obtain

$$\lim_{\lambda \rightarrow \infty} \frac{t(\kappa_{3S}(\theta)) \sigma_S(\theta)}{\mu_S(\theta)} = 0,$$

from which **(f)** follows directly. We know that

$$\frac{\partial VaRapp(\theta)}{\partial \theta_i} = \frac{\partial \mu_S(\theta)}{\partial \theta_i} + t'(\kappa_{3S}(\theta)) \frac{\partial \kappa_{3S}(\theta)}{\partial \theta_i} \sigma_S(\theta) + t(\kappa_{3S}(\theta)) \frac{\partial \sigma_S(\theta)}{\partial \theta_i}.$$

Hence, using (7.14), (c) and (e) of the present lemma, we obtain

$$\lim_{\lambda \rightarrow \infty} \frac{t'(\kappa_{3S}(\theta)) \frac{\partial \kappa_{3S}(\theta)}{\partial \theta_i} \sigma_S(\theta) + t(\kappa_{3S}(\theta)) \frac{\partial \sigma_S(\theta)}{\partial \theta_i}}{\frac{\partial \mu_S(\theta)}{\partial \theta_i}} = 0,$$

from which (g) follows directly. The results in (h) and (j) follow directly from (f) and (g) of the present lemma and (b) and (c) of Lemma 7.4. \square

Next we apply Theorem 7.3 in order to get an idea of the impact of the estimators on the risk measures considered. As was mentioned before, the two risk measures will be analyzed in simplified form. This form can be written as a general linear function

$$\begin{aligned} RMapp1(\mu_C, \gamma_C, \mu_G, \gamma_G, \epsilon, \lambda) &= RMapp(\mu_{C0}, \gamma_{C0}, \mu_{G0}, \gamma_{G0}, \epsilon_0, \lambda_0) \\ &+ (\mu_C - \mu_{C0}) \frac{\partial}{\partial \mu_C} RMapp(\mu_{C0}, \gamma_{C0}, \mu_{G0}, \gamma_{G0}, \epsilon_0, \lambda_0) \\ &+ \dots + (\lambda - \lambda_0) \frac{\partial}{\partial \lambda} RMapp(\mu_{C0}, \gamma_{C0}, \mu_{G0}, \gamma_{G0}, \epsilon_0, \lambda_0), \end{aligned}$$

where RM means risk measure and $RMapp$ denotes its (first step cf. Section 7.2.1) approximation. Hence, by replacing RM with VaR or SLP we get formulas (7.6) and (7.7), respectively. This is a convenient way to present the results in general form. Later the concrete risk measures VaR and SLP will be presented as illustrative examples.

The error due to estimation, divided by μ_{C0} , equals

$$\begin{aligned} &\mu_{C0}^{-1} \left\{ RMapp1(\hat{\mu}_C, \hat{\gamma}_C, \hat{\mu}_G, \hat{\gamma}_G, \hat{\epsilon}, \hat{\lambda}) - RMapp(\mu_{C0}, \gamma_{C0}, \mu_{G0}, \gamma_{G0}, \epsilon_0, \lambda_0) \right\} \\ &= \left(\frac{\hat{\mu}_C - \mu_{C0}}{\mu_{C0}} \right) \frac{\partial}{\partial \mu_C} RMapp(\mu_{C0}, \gamma_{C0}, \mu_{G0}, \gamma_{G0}, \epsilon_0, \lambda_0) \\ &+ (\hat{\gamma}_C - \gamma_C) \mu_{C0}^{-1} \frac{\partial}{\partial \gamma_C} RMapp(\mu_{C0}, \gamma_{C0}, \mu_{G0}, \gamma_{G0}, \epsilon_0, \lambda_0) \\ &+ \dots + \left(\frac{\hat{\lambda} - \lambda_0}{\lambda_0} \right) \lambda_0 \mu_{C0}^{-1} \frac{\partial}{\partial \lambda} RMapp(\mu_{C0}, \gamma_{C0}, \mu_{G0}, \gamma_{G0}, \epsilon_0, \lambda_0). \end{aligned}$$

The asymptotic distribution of

$$\mu_{C0}^{-1} \left\{ RMapp1(\hat{\mu}_C, \hat{\gamma}_C, \hat{\mu}_G, \hat{\gamma}_G, \hat{\epsilon}, \hat{\lambda}) - RMapp(\mu_{C0}, \gamma_{C0}, \mu_{G0}, \gamma_{G0}, \epsilon_0, \lambda_0) \right\} \sqrt{u \lambda_0}$$

is obtained by application of Theorem 7.3 with

$$\begin{aligned}
c_1 &= \frac{\partial}{\partial \mu_C} RMapp(\mu_{C0}, \gamma_{C0}, \mu_{G0}, \gamma_{G0}, \epsilon_0, \lambda_0), \\
c_2 &= \mu_{C0}^{-1} \frac{\partial}{\partial \gamma_C} RMapp(\mu_{C0}, \gamma_{C0}, \mu_{G0}, \gamma_{G0}, \epsilon_0, \lambda_0), \\
c_3 &= \epsilon_0^{-1/2} \mu_{C0}^{-1} \frac{\partial}{\partial \mu_G} RMapp(\mu_{C0}, \gamma_{C0}, \mu_{G0}, \gamma_{G0}, \epsilon_0, \lambda_0), \\
c_4 &= \epsilon_0^{-1/2} \mu_{C0}^{-1} \frac{\partial}{\partial \gamma_G} RMapp(\mu_{C0}, \gamma_{C0}, \mu_{G0}, \gamma_{G0}, \epsilon_0, \lambda_0), \\
c_5 &= \epsilon_0^{1/2} \mu_{C0}^{-1} \frac{\partial}{\partial \epsilon} RMapp(\mu_{C0}, \gamma_{C0}, \mu_{G0}, \gamma_{G0}, \epsilon_0, \lambda_0), \\
c_6 &= \mu_{C0}^{-1} \lambda_0 \frac{\partial}{\partial \lambda} RMapp(\mu_{C0}, \gamma_{C0}, \mu_{G0}, \gamma_{G0}, \epsilon_0, \lambda_0).
\end{aligned} \tag{7.22}$$

The result is a normal distribution with expectation 0 and variance $\tau_1^2 + \tau_2^2$. This variance should give an idea of the error due to estimation.

7.4 Estimation effect on *SLP*

This section discusses the estimation effect when *SLP* is the underlying risk measure. All the results which are presented in this section were demonstrated in Albers and Kallenberg [2007]. Therefore, we will be brief here. We will start with an example of the calculation of τ_1^2 and τ_2^2 . This should give an impression of the estimation effect and point out the most important parameters from the estimation point of view. We calculate τ_1^2 and τ_2^2 for $(\mu_{C0}, \gamma_{C0}, \mu_{G0}, \gamma_{G0}, \epsilon_0, \lambda_0) = (100000, 0.7, 15, 0.8, 0.03, 400)$ and $k = (a - \mu_S)/\sigma_S = 1$. The distributions of C and L are assumed to be Gamma with (obviously) different input parameters. In Section 7.2.2 we already saw that $SLPapp(100000, 0.7, 15, 0.8, 0.03, 400) = 292282$. The values of c_1, \dots, c_6 are easily obtained from Table 7.4. We get

$$\begin{aligned}
c_1^2 \gamma_{C0}^2 &= 4.19 \\
c_1 c_2 \gamma_{C0}^2 (\kappa_{3C0} - 2\gamma_{C0}) &= 0 \\
c_2^2 \gamma_{C0}^2 (\gamma_{C0}^2 + \kappa_{4C0}/4 + 1/2 - \gamma_{C0} \kappa_{3C0}) &= 0.18
\end{aligned}$$

and therefore

$$\tau_1^2 = 4.37.$$

Using that L has a Gamma distribution, direct calculation gives $\kappa_{3G} = 2\gamma_G - \mu_G^{-1}\gamma_G^{-1}$ and $\kappa_{4G} = 6\gamma_G^2 - 6\mu_G^{-1} + \mu_G^{-2}\gamma_G^{-2}$. We obtain

$$\begin{aligned}
c_3^2\mu_{G0}^3\gamma_{G0}^2 &= 287.43 \\
\frac{1}{2}c_4^2(\mu_{G0}\gamma_{G0}^4 - \gamma_{G0}^2 + \mu_{G0}^{-1}/2 + \mu_{G0}\gamma_{G0}^2) &= 265.21 \\
c_5^2(1 - \epsilon_0)\{\mu_{G0}(1 - \epsilon_0)(1 + \gamma_{G0}^2) + \epsilon_0\} &= 325.47 \\
c_6^2\{\mu_{G0}\epsilon_0(1 + \gamma_{G0}^2) + 1 - \epsilon_0\} &= 2.84 \\
-c_3c_4\mu_{G0}\gamma_{G0} &= -25.91 \\
2c_3c_5(1 - \epsilon_0)\mu_{G0}^2\gamma_{G0}^2 &= 381.9 \\
2c_3c_6\sqrt{\epsilon_0}\mu_{G0}^2\gamma_{G0}^2 &= 23.46 \\
-c_4c_5\gamma_{G0}(1 - \epsilon_0) &= -17.21 \\
-c_4c_6\gamma_{G0}\sqrt{\epsilon_0} &= -1.06 \\
2c_5c_6\sqrt{\epsilon_0}(1 - \epsilon_0)\{\mu_{G0}(1 + \gamma_{G0}^2) - 1\} &= 38.31
\end{aligned}$$

and hence

$$\tau_2^2 = 1280.43.$$

This example is really illuminating. One of the conclusions of Albers and Kallenberg [2007] was that the estimation error for SLP is dominated by the estimation of the parameters related to the common risk, that is by estimating μ_G, γ_G and ϵ . The reason is that we have a lot of observations for estimating μ_C and γ_C . Typical values for u and λ are values like 7 and 400, respectively. That means about 2800 observations to estimate the parameters of the common distribution of C and D . Due to this large number of observations, these estimators are very accurate. The same situation occurs with λ , as can be seen from the various terms contributing to τ_2^2 . The terms in which estimating λ is involved, that is the terms where c_6 appears, are much smaller than the other terms. Hence, in the sequel, parameters μ_C, γ_C and λ are assumed to be known in case SLP is the underlying risk measure. Using this assumption we formulate the following theorem.

Theorem 7.7. *Let $(\mu_{C0}, \gamma_{C0}, \mu_{G0}, \gamma_{G0}, \epsilon_0, \lambda_0)$ be the true value of the parameters. Then*

$$SLP(\hat{\mu}_C, \hat{\gamma}_C, \hat{\mu}_G, \hat{\gamma}_G, \hat{\epsilon}, \hat{\lambda}) \approx SLP_{app1}(\mu_{C0}, \gamma_{C0}, \hat{\mu}_G, \hat{\gamma}_G, \hat{\epsilon}, \lambda_0)$$

and

$$\begin{aligned}
&\mu_{C0}^{-1} \left\{ SLP_{app1}(\mu_{C0}, \gamma_{C0}, \hat{\mu}_G, \hat{\gamma}_G, \hat{\epsilon}, \lambda_0) \right. \\
&\quad \left. - SLP_{app1}(\mu_{C0}, \gamma_{C0}, \mu_{G0}, \gamma_{G0}, \epsilon_0, \lambda_0) \right\} \tau^{-1} \sqrt{u\lambda_0\epsilon_0} \rightarrow V
\end{aligned}$$

as $\lambda_0 \rightarrow \infty$, with $V \sim N(0, 1)$, in which

$$\begin{aligned}
\tau^2 = & c_3^2 \mu_{G0}^3 \gamma_{G0}^2 \\
& + c_4^2 \mu_{G0} \gamma_{G0}^2 (\gamma_{G0}^2 - \gamma_{G0} \kappa_{3G0} + \kappa_{4G0}/4 + 1/2) \\
& + c_5^2 (1 - \epsilon_0) \{ \mu_{G0} (1 - \epsilon_0) (1 + \gamma_{G0}^2) + \epsilon_0 \} \\
& + c_3 c_4 \mu_{G0}^2 \gamma_{G0}^2 (\kappa_{3G0} - 2\gamma_{G0}) \\
& + 2c_3 c_5 (1 - \epsilon_0) \mu_{G0}^2 \gamma_{G0}^2 \\
& + c_4 c_5 \mu_{G0} \gamma_{G0}^2 (1 - \epsilon_0) (\kappa_{3G0} - 2\gamma_{G0}),
\end{aligned} \tag{7.23}$$

where

$$\begin{aligned}
c_3 &= \mu_{C0}^{-1} \frac{\partial}{\partial \mu_G} SLP_{app}(\mu_{C0}, \gamma_{C0}, \mu_{G0}, \gamma_{G0}, \epsilon_0, \lambda_0), \\
c_4 &= \mu_{C0}^{-1} \frac{\partial}{\partial \gamma_G} SLP_{app}(\mu_{C0}, \gamma_{C0}, \mu_{G0}, \gamma_{G0}, \epsilon_0, \lambda_0), \\
c_5 &= \epsilon_0 \mu_{C0}^{-1} \frac{\partial}{\partial \epsilon} SLP_{app}(\mu_{C0}, \gamma_{C0}, \mu_{G0}, \gamma_{G0}, \epsilon_0, \lambda_0).
\end{aligned} \tag{7.24}$$

Proof. The limiting result follows directly from Theorem 7.3, because

$$\begin{aligned}
& \mu_{C0}^{-1} \{ SLP_{app1}(\mu_{C0}, \gamma_{C0}, \hat{\mu}_G, \hat{\gamma}_G, \hat{\epsilon}, \lambda_0) - SLP_{app1}(\mu_{C0}, \gamma_{C0}, \mu_{G0}, \gamma_{G0}, \epsilon_0, \lambda_0) \} \\
& = c_3 (\hat{\mu}_G - \mu_{G0}) + c_4 (\hat{\gamma}_G - \gamma_{G0}) + c_5 \frac{\hat{\epsilon} - \epsilon_0}{\epsilon_0}
\end{aligned}$$

with c_3, c_4 and c_5 given by (7.24). Note that in the formulation of the theorem a different norming is used (i.e. $\sqrt{u\lambda_0\epsilon_0}$ instead of $\sqrt{u\lambda_0}$), because the expected number of special claims equals $u\lambda_0\epsilon_0$. The same norming is kept in Theorem 7.9 as well. \square

The following theorem gives the asymptotic behavior of τ .

Theorem 7.8. *Suppose that the assumptions and notations of Lemma 7.4 hold. Then*

$$\frac{\hat{\tau}}{\tau} \xrightarrow{P} 1$$

as $\lambda \rightarrow \infty$ with $\tau = \tau(\theta)$ defined in (7.23) and $\hat{\tau} = \tau(\hat{\theta})$.

Proof. Since estimators $\hat{\mu}_C, \hat{\gamma}_C, \hat{\mu}_G, \hat{\gamma}_G$ and $\hat{\epsilon}$ are consistent and (cf. (7.15)) $\hat{\lambda}/\lambda \xrightarrow{P} 1$, it is enough to show that

$$\frac{c_i(\hat{\theta})}{c_i(\theta)} \xrightarrow{P} 1 \tag{7.25}$$

as $\lambda \rightarrow \infty$ for $i = 3, 4, 5$ (see (7.24)) and that the $c_i(\theta)$ are all of the same exact order for $i = 3, 4, 5$. Using the same arguments, (7.25) is true if

$$\frac{\frac{\partial}{\partial \theta_i} SLP_{app}(\hat{\theta})}{\frac{\partial}{\partial \theta_i} SLP_{app}(\theta)} \xrightarrow{P} 1$$

as $\lambda \rightarrow \infty$ for $i = 3, 4, 5$ (i.e. derivatives w.r.t. μ_G, γ_G and ϵ). Therefore, application of Lemma 7.5(e) gives (7.25). It easily follows from (7.24), (7.16), (7.14) and Lemma 7.5(b) that $c_i(\theta)$ is of exact order $\lambda^{1/2}$ for $i = 3, 4, 5$, which completes the proof. \square

Remark 7.4. Theorems 7.7 and 7.8 can be applied to $G : P(L)$ with parametrization μ_L, γ_L (provided that the fourth moment of L is finite), replacing $SLP(\hat{\mu}_C, \hat{\gamma}_C, \hat{\mu}_G, \hat{\gamma}_G, \hat{\epsilon}, \hat{\lambda})$, $SLP_{app1}(\mu_{C0}, \gamma_{C0}, \hat{\mu}_G, \hat{\gamma}_G, \hat{\epsilon}, \lambda_0)$ and $SLP_{app1}(\mu_{C0}, \gamma_{C0}, \mu_{G0}, \gamma_{G0}, \epsilon_0, \lambda_0)$ by $SLP(\hat{\mu}_C, \hat{\gamma}_C, \hat{\mu}_L, \hat{\gamma}_L, \hat{\epsilon}, \hat{\lambda})$, $SLP_{app1}(\mu_{C0}, \gamma_{C0}, \hat{\mu}_L, \hat{\gamma}_L, \hat{\epsilon}, \lambda_0)$ and $SLP_{app1}(\mu_{C0}, \gamma_{C0}, \mu_{L0}, \gamma_{L0}, \epsilon_0, \lambda_0)$, respectively, and τ^2 by

$$\begin{aligned} \tau^2 &= c_3^2(\mu_{L0}^3 \gamma_{L0}^2 + \mu_{L0}^2) \\ &+ c_4^2 \left\{ \mu_{L0} \gamma_{L0}^2 \left(\gamma_{L0}^2 - \gamma_{L0} \kappa_{3L0} + \frac{\kappa_{4L0}}{4} + \frac{1}{2} \right) \right. \\ &\quad \left. - \gamma_{L0}^2 + \gamma_{L0} \kappa_{3L0} + 1 + \frac{1}{2} \mu_{L0}^{-1} (1 + \gamma_{L0}^{-2}) \right\} \\ &+ c_5^2 (1 - \epsilon_0) \{ \mu_{L0} (1 - \epsilon_0) (1 + \gamma_{L0}^2) + 1 \} \\ &+ c_3 c_4 \mu_{L0}^2 \gamma_{L0}^2 (\kappa_{3L0} - 2\gamma_{L0}) \\ &+ 2c_3 c_5 (1 - \epsilon_0) (\mu_{L0}^2 \gamma_{L0}^2 + \mu_{L0}) \\ &+ c_4 c_5 \mu_{L0} \gamma_{L0}^2 (1 - \epsilon_0) (\kappa_{3L0} - 2\gamma_{L0}), \end{aligned} \tag{7.26}$$

where

$$\begin{aligned} c_3 &= \mu_{C0}^{-1} \frac{\partial}{\partial \mu_L} SLP_{app}(\mu_{C0}, \gamma_{C0}, \mu_{L0}, \gamma_{L0}, \epsilon_0, \lambda_0), \\ c_4 &= \mu_{C0}^{-1} \frac{\partial}{\partial \gamma_L} SLP_{app}(\mu_{C0}, \gamma_{C0}, \mu_{L0}, \gamma_{L0}, \epsilon_0, \lambda_0), \\ c_5 &= \epsilon_0 \mu_{C0}^{-1} \frac{\partial}{\partial \epsilon} SLP_{app}(\mu_{C0}, \gamma_{C0}, \mu_{L0}, \gamma_{L0}, \epsilon_0, \lambda_0). \end{aligned} \tag{7.27}$$

\square

7.4.1 Protection approach

So far we have presented a lot of formulas which can help in the analysis of the estimation effect of SLP . The next step is to see the investigation of the

impact of the estimation on the *SLP*, which originally was presented in Albers and Kallenberg [2007]. We start with an example which clearly illustrates this aspect. Let the true values of the parameters be equal to $(\mu_{C0}, \gamma_{C0}, \mu_{G0}, \gamma_{G0}, \epsilon_0, \lambda_0) = (100000, 0.7, 15, 0.8, 0.03, 400)$ and $k = (a - \mu_S)/\sigma_S = 1$. Let C and L each have a (different) Gamma distribution. In Section 7.2.2 we saw that $SLPapp(100000, 0.7, 15, 0.8, 0.03, 400) = 292282$. The natural question which can arise having these numbers is: what is the probability that $SLPapp(\mu_{C0}, \gamma_{C0}, \hat{\mu}_G, \hat{\gamma}_G, \hat{\epsilon}, \lambda_0)$ is smaller than (let us say) 200000, that is an error of more than 92282? To answer such a question we apply Theorem 7.7. Direct calculation gives $\tau^2 = 36.51$ and hence, with Φ the standard normal distribution function,

$$\begin{aligned} P(SLPapp1(\mu_{C0}, \gamma_{C0}, \hat{\mu}_G, \hat{\gamma}_G, \hat{\epsilon}, \lambda_0) < 200000) \\ = P\left(10^{-5} \{SLPapp1(\mu_{C0}, \gamma_{C0}, \hat{\mu}_G, \hat{\gamma}_G, \hat{\epsilon}, \lambda_0) - 292282\} 36.51^{-1/2} \sqrt{12u} \right. \\ \left. < 10^{-5} (200000 - 292282) 36.51^{-1/2} \sqrt{12u}\right) \approx \Phi(-0.53\sqrt{u}). \end{aligned}$$

From the above formula we see that taking only one year, that is $u = 1$, the estimated value is smaller than 200000 with a quite high probability of 30%. This clearly illustrates that indeed one year is not enough. The reason for that is of course lack of observations for μ_G, γ_G and ϵ . The expected number of groups in the presented example is only $\epsilon\lambda/\mu_G = 12/15 = 0.8$ which makes the estimation of these parameters very inaccurate. If we take $u = 7$, the probability reduces from 30% to 8% which is considered to be acceptable.

From the above example we see that the effect of estimation may be quite large and we may want to control it. This can be done by using the protection approach which originally was presented in Albers and Kallenberg [2007]. We simply construct the confidence interval, in which *SLP* falls with some fixed probability. The following theorem deals with such protection.

Theorem 7.9. *Assume that Model 4 holds and that C and D are identically distributed with a Gamma, Inverse Gaussian or lognormal distribution. Let $(\mu_{C0}, \gamma_{C0}, \mu_{G0}, \gamma_{G0}, \epsilon_0, \lambda_0)$ be the true values of the parameters. Then*

$$\begin{aligned} \lim_{\lambda_0 \rightarrow \infty} P(SLPapp(\mu_{C0}, \gamma_{C0}, \mu_{G0}, \gamma_{G0}, \epsilon_0, \lambda_0) < UB(\alpha)) &= 1 - \alpha, \\ \lim_{\lambda_0 \rightarrow \infty} P(SLPapp(\mu_{C0}, \gamma_{C0}, \mu_{G0}, \gamma_{G0}, \epsilon_0, \lambda_0) > LB(\alpha)) &= 1 - \alpha, \\ \lim_{\lambda_0 \rightarrow \infty} P(LB(\alpha/2) < SLPapp(\mu_{C0}, \gamma_{C0}, \mu_{G0}, \gamma_{G0}, \epsilon_0, \lambda_0) < UB(\alpha/2)) &= 1 - \alpha \end{aligned}$$

with

$$\begin{aligned} UB(\alpha) &= SLPapp1(\mu_{C0}, \gamma_{C0}, \hat{\mu}_G, \hat{\gamma}_G, \hat{\epsilon}, \lambda_0) + \Phi^{-1}(1 - \alpha) (\hat{\epsilon}u\lambda_0)^{-1/2} \hat{\tau}\mu_{C0}, \\ LB(\alpha) &= SLPapp1(\mu_{C0}, \gamma_{C0}, \hat{\mu}_G, \hat{\gamma}_G, \hat{\epsilon}, \lambda_0) - \Phi^{-1}(1 - \alpha) (\hat{\epsilon}u\lambda_0)^{-1/2} \hat{\tau}\mu_{C0}, \end{aligned}$$

where $\hat{\tau} = \sqrt{\hat{\tau}^2}$ and $\hat{\tau}^2$ is given in (7.23) and (7.24) with $\mu_{G0}, \gamma_{G0}, \epsilon_0$ replaced by their estimators $\hat{\mu}_G, \hat{\gamma}_G, \hat{\epsilon}$ (also in $c_3, c_4, c_5, \kappa_{3G0}$ and κ_{4G0}).

Proof. Application of Theorem 7.8 and Theorem 6.1 in Albers and Kallenberg [2007] gives the result. \square

Remark 7.5. Theorem 7.9 can be applied to $G : P(L)$ with parametrization μ_L, γ_L (provided that the fourth moment of L is finite), replacing $SLPapp(\mu_{C0}, \gamma_{C0}, \mu_{G0}, \gamma_{G0}, \epsilon_0, \lambda_0)$ and $SLPapp1(\mu_{C0}, \gamma_{C0}, \hat{\mu}_G, \hat{\gamma}_G, \hat{\epsilon}, \lambda_0)$ by $SLPapp(\mu_{C0}, \gamma_{C0}, \mu_{L0}, \gamma_{L0}, \epsilon_0, \lambda_0)$ and $SLPapp1(\mu_{C0}, \gamma_{C0}, \hat{\mu}_L, \hat{\gamma}_L, \hat{\epsilon}, \lambda_0)$, respectively, and $\hat{\tau}^2$ by the estimated version of (7.26) and (7.27). \square

As was already discussed, the contribution of estimating μ_C, γ_C and λ is very small (compared to that of estimating μ_G, γ_G and ϵ) in case of *SLP*. Therefore, in Theorem 7.9 again μ_{C0}, γ_{C0} and λ_0 are assumed to be known. In practice these parameters obviously have to be replaced by their estimates in the formulas of $UB(\alpha)$ and $LB(\alpha)$.

In Figures 7.1-7.3 some examples are presented of the extra amount due to the protection against estimation and the effect of dependence in these situations. Figures 7.1(a)-7.3(a) show the relative difference between the independent case and the dependence one. Here we assume that C and L each have a (different) Gamma distribution. We take $\gamma_{C0} = 0.4$ or 1.2 , $\mu_{G0} = 5, 10$ or 15 , $\gamma_{G0} = 0.5$ or 1 , $\epsilon_0 = 0.03$ and $\lambda_0 = 400$. The relative difference is defined by

$$\begin{aligned} & \frac{SLP - SLP_I}{SLP_I} \\ &= \frac{SLPapp(\mu_{C0}, \gamma_{C0}, \mu_{G0}, \gamma_{G0}, \epsilon_0, \lambda_0) - SLPapp(\mu_{C0}, \gamma_{C0}, \mu_{G0}, \gamma_{G0}, 0, \lambda_0)}{SLPapp(\mu_{C0}, \gamma_{C0}, \mu_{G0}, \gamma_{G0}, 0, \lambda_0)}, \end{aligned}$$

where SLP denotes the (approximated) SLP under dependence (Model 4) and SLP_I the (approximated) SLP under independence (Model 1). For a fair comparison we take both for the independence model and the dependence one the same retentions

$$a = \mu_S + k\sigma_{S_I}$$

with $k = 0, \dots, 3$ and $\sigma_{S_I} = \mu_C \sqrt{\lambda(1 + \gamma_C^2)}$, the standard deviation of S for the independence model (Model 1).

Figures 7.1(b)-7.3(b) show the extra amount due to protection against estimation, also measured in a relative way by taking

$$\frac{UB(\alpha) - SLPapp(\mu_{C0}, \gamma_{C0}, \mu_{G0}, \gamma_{G0}, \epsilon_0, \lambda_0)}{SLPapp(\mu_{C0}, \gamma_{C0}, \mu_{G0}, \gamma_{G0}, \epsilon_0, \lambda_0)}$$

with in $UB(\alpha)$ the estimators $\hat{\mu}_G, \hat{\gamma}_G, \hat{\epsilon}$ and $\hat{\tau}$ replaced by $\mu_{G0}, \gamma_{G0}, \epsilon_0$ and $\sqrt{\tau^2}$, respectively. We take $\alpha = 0.1$ and $u = 7$. Note that both measures do not depend on μ_{C0} .

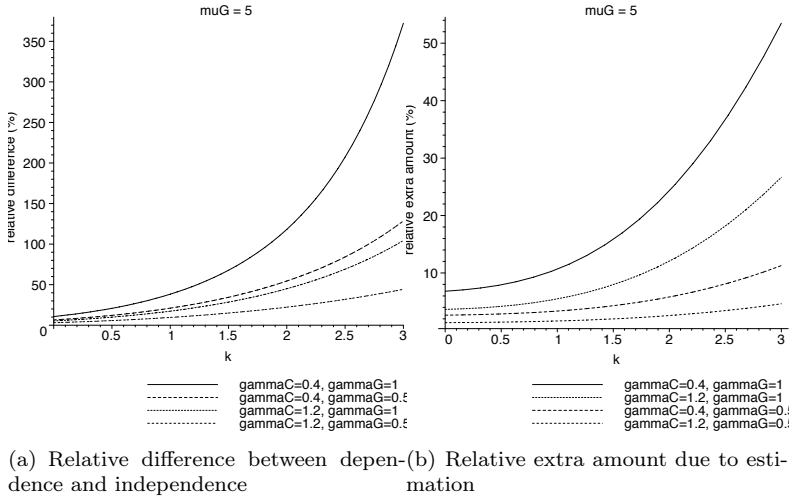
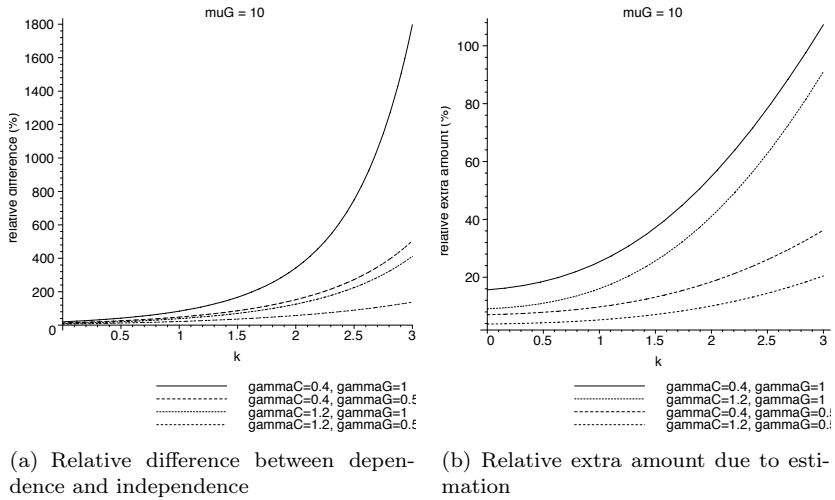


Figure 7.1: Dependence and estimation effects for SLP when $\mu_G = 5$

Note that the order of the displayed cases is slightly different in the figures a and b : for instance, for $\mu_{C0} = 5$ (Figures 7.1(a),(b)) the relative difference between dependence and independence is higher for $\gamma_{C0} = 0.4, \gamma_{G0} = 0.5$ than for $\gamma_{C0} = 1.2, \gamma_{G0} = 1$, while their order w.r.t. the relative extra amount due to protection against estimation is reversed.

Figures 7.1-7.3 affirm that ignoring dependence may lead to very large errors (up to 4300% in Figure 7.3). But also the additional step due to protection against estimation is large (up to 138% in Figure 7.3). A numerical example may illustrate this. Consider again the example with true values of the parameters being equal to $(\mu_{C0}, \gamma_{C0}, \mu_{G0}, \gamma_{G0}, \epsilon_0, \lambda_0) = (100000, 0.7, 15, 0.8, 0.03, 400)$. By taking $k = 1$ we have $a = \mu_S + \sigma_{S_I} = 4 \times 10^7 + 2561250 = 42561250$. If we ignore the dependence we get $SLP_{app}(100000, 0.7, 15, 0.8, 0.03, 400) = 211277$. If we take into account the dependence without protection against estimation we get $SLP_{app}(100000, 0.7, 15, 0.8, 0.03, 400) = 382006$. If we add the protection (taking $\hat{\mu}_G = \mu_{G0} = 15, \hat{\gamma}_G = \gamma_{G0} = 0.8, \hat{\epsilon} = \epsilon_0 = 0.03, \hat{\tau} = \sqrt{\tau^2}$) we get $UB(0.1) = 476596$.

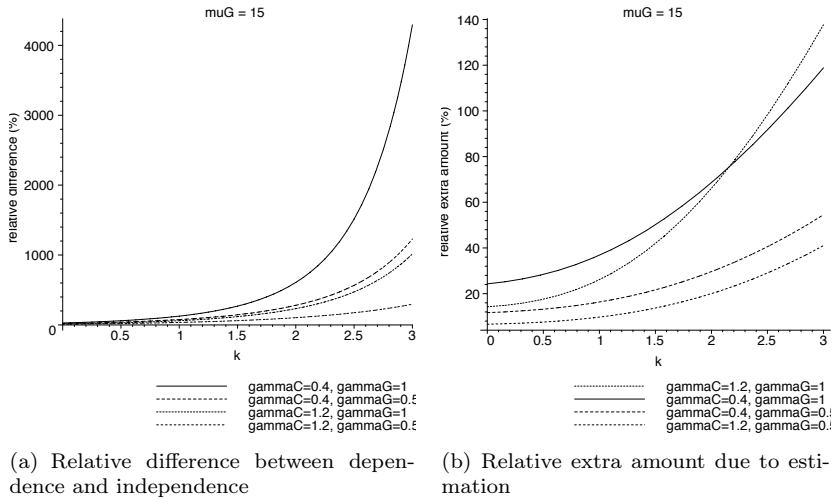
The upper and the lower bounds $UB(\alpha)$ and $LB(\alpha)$ contain the term $\hat{\tau}\mu_{C0}$. It was already mentioned in the previous chapters of the thesis that μ_C is a kind of dummy parameter. However, the expression of $\hat{\tau}$ has a quite complicated structure. For illustrative purposes we show the behavior of τ^2 as a function of ϵ when $(\mu_C, \gamma_C, \mu_G, \gamma_G, \lambda, k) = (100000, 0.7, 15, 0.8, 400, 1)$ are fixed. It is clearly seen in Figure 7.4 that τ^2 tends to 0 when $\epsilon \rightarrow 0$. This result is in line with our expectations since τ^2 is the variance of the asymptotic distribution of the estimation error of μ_G, γ_G and ϵ , which are the parameters of the special part of



(a) Relative difference between dependence and independence

(b) Relative extra amount due to estimation

Figure 7.2: Dependence and estimation effects for SLP when $\mu_G = 10$



(a) Relative difference between dependence and independence

(b) Relative extra amount due to estimation

Figure 7.3: Dependence and estimation effects for SLP when $\mu_G = 15$

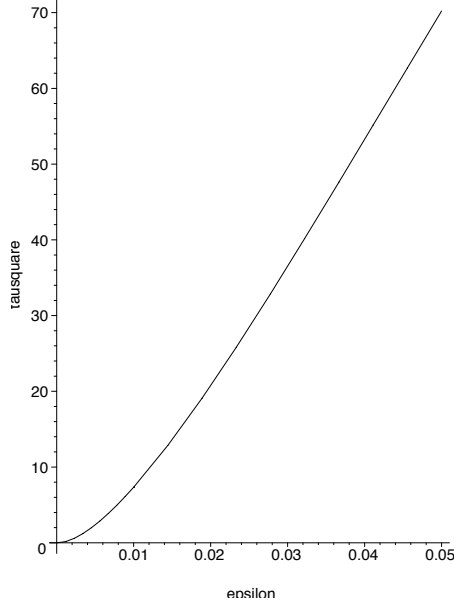


Figure 7.4: Behavior of $\tau^2(\mu_C, \gamma_C, \mu_G, \gamma_G, \epsilon, \lambda, k)$ as $\epsilon \rightarrow 0$ with $(\mu_C, \gamma_C, \mu_G, \gamma_G, \lambda, k) = (100000, 0.7, 15, 0.8, 400, 1)$ and $k = (a - \mu_S)/\sigma_S = 1$

the model. By tending ϵ to 0 we decrease the special part of the model and it disappears when $\epsilon = 0$.

7.5 Estimation effect on VaR

This section deals with the estimation effect when the underlying risk measure is *VaR*. Just as in the previous section, where *SLP* was the underlying risk measure, we start the analysis with an example of the calculation of τ_1^2 and τ_2^2 (see Theorem 7.3). This should give an impression of which parameters are important from the estimation point of view. We fix $(\mu_{C0}, \gamma_{C0}, \mu_{G0}, \gamma_{G0}, \epsilon_0, \lambda_0) = (100000, 0.7, 15, 0.8, 0.03, 400)$ and $p = 0.99$. Assume that *C* and *L* both have (different) Gamma distributions. In that case $VaRapp(\mu_{C0}, \gamma_{C0}, \mu_{G0}, \gamma_{G0}, \epsilon_0, \lambda_0) = 47861121$. Using these assumptions we obtain $\tau_1 = 335$ and $\tau_2 = 1303$. To get a feeling about the most important parameters we present all the calculation terms separately (like it was done in the previous section for *SLP*), starting with the τ_1^2 terms. They are

$$\begin{aligned}
 c_1^2 \gamma_{C0}^2 &= 112244, \\
 \gamma_{C0}^2 c_1 c_2 (\kappa_{3C0} - 2\gamma_{C0}) &= 0, \\
 \gamma_{C0}^2 c_2^2 (\gamma_{C0}^2 + \frac{\kappa_{4C0}}{4} + 0.5 - \gamma_{C0} \kappa_{3C0}) &= 192,
 \end{aligned}$$

resulting in $\tau_1^2 = 112436$. It is clearly seen that the most important value comes from the first term, and it is clearly not the γ_{G0}^2 , which is only 0.49. Hence c_1 is the dominating quantity in the τ_1 calculation. Note that the absolute values of τ_1 and τ_2 are much larger than the ones for *SLP*. This can be explained by the fact that the absolute values of *Var* itself are also larger than the values of *SLP* while considering high quantiles.

The τ_2^2 terms are

$$\begin{aligned} c_3^2 \mu_{G0}^3 \gamma_{G0}^2 &= 203807, \\ c_4^2 \mu_{G0} \gamma_{G0}^2 (\gamma_{G0}^2 - \gamma_{G0} \kappa_{3G0} + \kappa_{4G0}/4 + 1/2) &= 198061, \\ c_5^2 (1 - \epsilon_0) \{ \mu_{G0} (1 - \epsilon_0) (1 + \gamma_{G0}^2) + \epsilon_0 \} &= 212705, \\ c_6^2 \{ \mu_{G0} \epsilon_0 (1 + \gamma_{G0}^2) + 1 - \epsilon_0 \} &= 322550, \\ c_3 c_4 \mu_{G0}^2 \gamma_{G0}^2 (\kappa_{3G0} - 2\gamma_{G0}) &= -18856, \\ 2c_3 c_5 (1 - \epsilon_0) \mu_{G0}^2 \gamma_{G0}^2 &= 259971, \\ 2c_3 c_6 \sqrt{\epsilon_0} \mu_{G0}^2 \gamma_{G0}^2 &= 210567, \\ c_4 c_5 \mu_{G0} \gamma_{G0}^2 (1 - \epsilon_0) (\kappa_{3G0} - 2\gamma_{G0}) &= -12026, \\ c_4 c_6 \mu_{G0} \gamma_{G0}^2 \sqrt{\epsilon_0} (\kappa_{3G0} - 2\gamma_{G0}) &= -9741, \\ 2c_5 c_6 \sqrt{\epsilon_0} (1 - \epsilon_0) \{ \mu_{G0} (1 + \gamma_{G0}^2) - 1 \} &= 330146, \end{aligned}$$

resulting in $\tau_2^2 = 1697184$. Again, like in the *SLP* case, we see that τ_2 is larger than τ_1 . But now the difference between τ_1 and τ_2 is not so extremely large. Now we see that

$$\frac{\tau_2(\text{Var})}{\tau_1(\text{Var})} = 3.9,$$

while

$$\frac{\tau_2(\text{SLP})}{\tau_1(\text{SLP})} = 17.$$

This indicates that the contribution of estimating μ_C and γ_C is larger for *Var* than for *SLP*. Additionally we see that estimation of λ gives a much higher contribution here. Terms in which c_6 is involved are much more important than in the *SLP* case, even more than the remaining terms.

From the above analysis we conclude that the estimation error is dominated by the estimation of μ_G , γ_G , ϵ and λ . Estimation of μ_C and γ_C gives more contribution than in the *SLP* case, but μ_C and γ_C can still be considered as known. Additional arguments for such a conclusion will be presented at the end of the section.

In analogy with Theorem 7.7 we have the following result.

Theorem 7.10. *Let $\text{Var}(\hat{\mu}_C, \hat{\gamma}_C, \hat{\mu}_G, \hat{\gamma}_G, \hat{\epsilon}, \hat{\lambda})$ be the estimator of the Value at Risk and $(\mu_{C0}, \gamma_{C0}, \mu_{G0}, \gamma_{G0}, \epsilon_0, \lambda_0)$ be the true value of the parameters. Then*

$$\text{Var}(\hat{\mu}_C, \hat{\gamma}_C, \hat{\mu}_G, \hat{\gamma}_G, \hat{\epsilon}, \hat{\lambda}) \approx \text{VarRapp1}(\mu_{C0}, \gamma_{C0}, \mu_{G0}, \gamma_{G0}, \hat{\epsilon}, \hat{\lambda})$$

and

$$\mu_{C0}^{-1}\{VaRapp1(\mu_{C0}, \gamma_{C0}, \hat{\mu}_G, \hat{\gamma}_G, \hat{\epsilon}, \hat{\lambda}) - VaRapp1(\mu_{C0}, \gamma_{C0}, \mu_{G0}, \gamma_{G0}, \epsilon_0, \lambda_0)\} \tau_2^{-1} \sqrt{u\lambda_0} \rightarrow V$$

as $\lambda_0 \rightarrow \infty$, with $V \sim N(0, 1)$, where τ_2^2 is as given in Theorem 7.3, with the coefficients c_3, \dots, c_6 defined in (7.22), replacing $RMapp$ by $VaRapp$.

Proof. The result follows directly from Theorem 7.3 □

The following theorem gives the asymptotic behavior of τ_2 .

Theorem 7.11. *Suppose that the assumptions and notations of Lemma 7.4 hold. Then*

$$\frac{\hat{\tau}_2}{\tau_2} \xrightarrow{P} 1$$

as $\lambda \rightarrow \infty$ with $\tau_2 = \tau_2(\theta)$ defined in Theorem 7.3, with c_3, c_4, c_5, c_6 defined in (7.22), replacing $RMapp$ by $VaRapp$, and $\hat{\tau}_2 = \tau_2(\hat{\theta})$.

Proof. Using the fact that the estimators $\hat{\mu}_C, \hat{\gamma}_C, \hat{\mu}_G, \hat{\gamma}_G$ and $\hat{\epsilon}$ are consistent and (cf. (7.15)) $\hat{\lambda}/\lambda \xrightarrow{P} 1$, it is enough to show that

$$\frac{c_i(\hat{\theta})}{c_i(\theta)} \xrightarrow{P} 1 \tag{7.28}$$

as $\lambda \rightarrow \infty$ for $i = 3, 4, 5, 6$ and that the $c_i(\theta)$ are all of the same exact order for $i = 3, 4, 5, 6$. Using the same arguments, (7.28) is true if

$$\frac{\frac{\partial}{\partial \theta_i} VaRapp(\hat{\theta})}{\frac{\partial}{\partial \theta_i} VaRapp(\theta)} \xrightarrow{P} 1$$

as $\lambda \rightarrow \infty$ for $i = 3, 4, 5, 6$. Therefore, application of Lemma 7.6(j) yields (7.28). It follows from (7.22) (with $RMapp$ replaced by $VaRapp$), Lemma 7.6(g) and (7.14) that $c_i(\theta)$ is of exact order λ for $i = 3, 4, 5, 6$, which completes the proof. □

Remark 7.6. Theorems 7.10 and 7.11 can be applied to $G : P(L)$ with parametrization μ_L, γ_L (provided that the fourth moment of L is finite), replacing $VaR(\hat{\mu}_C, \hat{\gamma}_C, \hat{\mu}_G, \hat{\gamma}_G, \hat{\epsilon}, \hat{\lambda})$, $VaRapp1(\mu_{C0}, \gamma_{C0}, \hat{\mu}_G, \hat{\gamma}_G, \hat{\epsilon}, \lambda_0)$ and $VaRapp1(\mu_{C0}, \gamma_{C0}, \mu_{G0}, \gamma_{G0}, \epsilon_0, \lambda_0)$ by $VaR(\hat{\mu}_C, \hat{\gamma}_C, \hat{\mu}_L, \hat{\gamma}_L, \hat{\epsilon}, \hat{\lambda})$, $VaRapp1(\mu_{C0}, \gamma_{C0}, \hat{\mu}_L, \hat{\gamma}_L, \hat{\epsilon}, \lambda_0)$ and $VaRapp1(\mu_{C0}, \gamma_{C0}, \mu_{L0}, \gamma_{L0}, \epsilon_0, \lambda_0)$, respectively, and τ_2^2 by

$$\begin{aligned}
\tau_2^2 &= c_3^2(\mu_{L0}^3\gamma_{L0}^2 + \mu_{L0}^2) \\
&+ c_4^2 \left\{ \mu_{L0}\gamma_{L0}^2 \left(\gamma_{L0}^2 - \gamma_{L0}\kappa_{3L0} + \frac{\kappa_{4L0}}{4} + \frac{1}{2} \right) - \gamma_{L0}^2 + \gamma_{L0}\kappa_{3L0} + 1 + \frac{1}{2}\mu_{L0}^{-1}(1 + \gamma_{L0}^{-2}) \right\} \\
&+ c_5^2(1 - \epsilon_0)\{\mu_{L0}(1 - \epsilon_0)(1 + \gamma_{L0}^2) + 1\} \\
&+ c_6^2\{\mu_{L0}\epsilon_0(1 + \gamma_{L0}^2) + 1\} \\
&+ c_3c_4\mu_{L0}^2\gamma_{L0}^2(\kappa_{3L0} - 2\gamma_{L0}) \\
&+ 2c_3c_5(1 - \epsilon_0)(\mu_{L0}^2\gamma_{L0}^2 + \mu_{L0}) \\
&+ 2c_3c_6\sqrt{\epsilon_0}(\mu_{L0}^2\gamma_{L0}^2 + \mu_{L0}) \\
&+ c_4c_5\mu_{L0}\gamma_{L0}^2(1 - \epsilon_0)(\kappa_{3L0} - 2\gamma_{L0}) \\
&+ c_4c_6\mu_{L0}\gamma_{L0}^2\sqrt{\epsilon_0}(\kappa_{3L0} - 2\gamma_{L0}) \\
&+ 2c_5c_6\mu_{L0}\sqrt{\epsilon_0}(1 - \epsilon_0)(1 + \gamma_{L0}^2).
\end{aligned} \tag{7.29}$$

where

$$\begin{aligned}
c_3 &= \epsilon_0^{-1/2} \mu_{C0}^{-1} \frac{\partial}{\partial \mu_L} VaRapp(\mu_{C0}, \gamma_{C0}, \mu_{L0}, \gamma_{L0}, \epsilon_0, \lambda_0), \\
c_4 &= \epsilon_0^{-1/2} \mu_{C0}^{-1} \frac{\partial}{\partial \gamma_L} VaRapp(\mu_{C0}, \gamma_{C0}, \mu_{L0}, \gamma_{L0}, \epsilon_0, \lambda_0), \\
c_5 &= \epsilon_0^{1/2} \mu_{C0}^{-1} \frac{\partial}{\partial \epsilon} VaRapp(\mu_{C0}, \gamma_{C0}, \mu_{L0}, \gamma_{L0}, \epsilon_0, \lambda_0), \\
c_6 &= \mu_{C0}^{-1} \lambda_0 \frac{\partial}{\partial \lambda} VaRapp(\mu_{C0}, \gamma_{C0}, \mu_{L0}, \gamma_{L0}, \epsilon_0, \lambda_0).
\end{aligned} \tag{7.30}$$

□

7.5.1 Protection approach

Using these results we formulate the protection theorem, the analogue of which was introduced for *SLP* in the previous section. However, the main idea we would like to illustrate by means of an example.

Let the true values of the parameters be equal to $(\mu_{C0}, \gamma_{C0}, \mu_{G0}, \gamma_{G0}, \epsilon_0, \lambda_0) = (100000, 0.7, 15, 0.8, 0.03, 400)$ and $p = 0.99$. As usual we assume that C and L have a (different) Gamma distribution. As we have seen before, $VaRapp(100000, 0.7, 15, 0.8, 0.03, 400) = 47861121$ in that case. Like in the case of *SLP*, we can be interested in the probability that $VaRapp1(\mu_{C0}, \gamma_{C0}, \hat{\mu}_G, \hat{\gamma}_G, \hat{\epsilon}, \hat{\lambda})$ is smaller than 43861121, which means an error of more than 4000000. To calculate such a probability we use the above stated theorem.

$$\begin{aligned}
& P(\text{VaRapp1}(\mu_{C0}, \gamma_{C0}, \hat{\mu}_G, \hat{\gamma}_G, \hat{\epsilon}, \hat{\lambda}) < 43861121) \\
&= P(\mu_{C0}^{-1} \{ \text{VaRapp1}(\mu_{C0}, \gamma_{C0}, \hat{\mu}_G, \hat{\gamma}_G, \hat{\epsilon}, \hat{\lambda}) - 47861121 \} 20 \times 1303^{-1} \sqrt{u} \\
&\quad < \mu_{C0}^{-1} \{ -4000000 \} 20 \times 1303^{-1} \sqrt{u}) \approx \Phi(-0.614\sqrt{u}).
\end{aligned}$$

From this example we can see that by taking $u = 1$, we get an estimated value smaller than 43861121 with a probability of 27%. By increasing the number of years we can reduce this probability. For instance, taking $u = 7$, the mentioned probability is only 5%.

Inspired by this example we formulate the following theorem.

Theorem 7.12. *Assume that Model 4 holds and that C and D are identically distributed with a Gamma, Inverse Gaussian or lognormal distribution. Let $(\mu_{C0}, \gamma_{C0}, \mu_{G0}, \gamma_{G0}, \epsilon_0, \lambda_0)$ be the true value of the parameters. Then*

$$\begin{aligned}
\lim_{\lambda_0 \rightarrow \infty} P(\text{VaRapp}(\mu_{C0}, \gamma_{C0}, \mu_{G0}, \gamma_{G0}, \epsilon_0, \lambda_0) < UB(\alpha)) &= 1 - \alpha, \\
\lim_{\lambda_0 \rightarrow \infty} P(\text{VaRapp}(\mu_{C0}, \gamma_{C0}, \mu_{G0}, \gamma_{G0}, \epsilon_0, \lambda_0) > LB(\alpha)) &= 1 - \alpha, \\
\lim_{\lambda_0 \rightarrow \infty} P(LB(\alpha/2) < \text{VaRapp}(\mu_{C0}, \gamma_{C0}, \mu_{G0}, \gamma_{G0}, \epsilon_0, \lambda_0) < UB(\alpha/2)) &= 1 - \alpha,
\end{aligned}$$

with

$$\begin{aligned}
UB(\alpha) &= \text{VaRapp1}(\mu_{C0}, \gamma_{C0}, \hat{\mu}_G, \hat{\gamma}_G, \hat{\epsilon}, \hat{\lambda}) + \Phi^{-1}(1 - \alpha)(u\hat{\lambda})^{-1/2} \hat{\tau}_2 \mu_{C0}, \\
LB(\alpha) &= \text{VaRapp1}(\mu_{C0}, \gamma_{C0}, \hat{\mu}_G, \hat{\gamma}_G, \hat{\epsilon}, \hat{\lambda}) - \Phi^{-1}(1 - \alpha)(u\hat{\lambda})^{-1/2} \hat{\tau}_2 \mu_{C0},
\end{aligned}$$

where $\hat{\tau}_2 = \sqrt{\hat{\tau}_2^2}$ and $\hat{\tau}_2^2$ is given in Theorem 7.3, with the coefficients c_3, \dots, c_6 defined in (7.22), replacing RMapp by VaRapp1 and $\mu_{G0}, \gamma_{G0}, \epsilon_0, \lambda_0$ by their estimators $\hat{\mu}_G, \hat{\gamma}_G, \hat{\epsilon}, \hat{\lambda}$ (also in $c_3, \dots, c_6, \kappa_{3G0}$ and κ_{4G0}).

Proof. Application of Theorems 7.10 and 7.11 yields

$$\begin{aligned}
& (\hat{\tau}_2 \mu_{C0})^{-1} \left\{ \text{VaRapp1}(\mu_{C0}, \gamma_{C0}, \hat{\mu}_G, \hat{\gamma}_G, \hat{\epsilon}, \hat{\lambda}) \right. \\
& \quad \left. - \text{VaRapp1}(\mu_{C0}, \gamma_{C0}, \mu_{G0}, \gamma_{G0}, \epsilon_0, \lambda_0) \right\} \sqrt{u\lambda_0} \rightarrow U
\end{aligned}$$

with $U \sim N(0, 1)$. Hence, writing temporary $\hat{V} = \text{VaRapp1}(\mu_{C0}, \gamma_{C0}, \hat{\mu}_G, \hat{\gamma}_G, \hat{\epsilon}, \hat{\lambda})$ and noting that $\text{VaRapp}(\mu_{C0}, \gamma_{C0}, \mu_{G0}, \gamma_{G0}, \epsilon_0, \lambda_0) = \text{VaRapp1}(\mu_{C0}, \gamma_{C0},$

$\mu_{G_0}, \gamma_{G_0}, \epsilon_0, \lambda_0$), we obtain, using $\hat{\lambda}/\lambda_0 \xrightarrow{P} 1$ (cf. (7.15)),

$$\begin{aligned} & P(\text{VaRapp}(\mu_{C_0}, \gamma_{C_0}, \mu_{G_0}, \gamma_{G_0}, \epsilon_0, \lambda_0) < UB(\alpha)) \\ &= P\left(\text{VaRapp}(\mu_{C_0}, \gamma_{C_0}, \mu_{G_0}, \gamma_{G_0}, \epsilon_0, \lambda_0) < \hat{V} + \Phi^{-1}(1 - \alpha)(u\hat{\lambda})^{-1/2}\hat{\tau}_2\mu_{C_0}\right) \\ &= P\left((\hat{\tau}_2\mu_{C_0})^{-1} \left\{ \hat{V} - \text{VaRapp}(\mu_{C_0}, \gamma_{C_0}, \mu_{G_0}, \gamma_{G_0}, \epsilon_0, \lambda_0) \right\} \sqrt{u\hat{\lambda}} \right. \\ &\quad \left. > -\Phi^{-1}(1 - \alpha)\right) \rightarrow P(U > -\Phi^{-1}(1 - \alpha)) = 1 - \alpha, \end{aligned}$$

thus giving the first result. The other statements are obtained in a similar way. \square

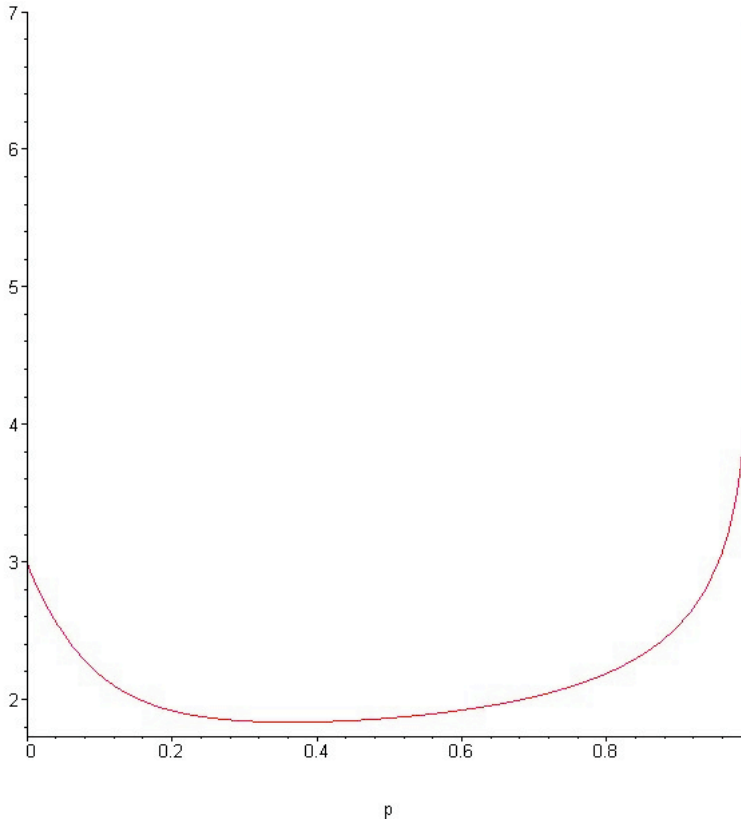


Figure 7.5: τ_2/τ_1 as a function of p

Remark 7.7. Theorem 7.12 can be applied to $G : P(L)$ with parametrization μ_L, γ_L (provided that the fourth moment of L is finite), replacing $\text{VaRapp}(\mu_{C_0},$

$\gamma_{C0}, \mu_{G0}, \gamma_{G0}, \epsilon_0, \lambda_0$) and $VaRapp1\left(\mu_{C0}, \gamma_{C0}, \hat{\mu}_G, \hat{\gamma}_G, \hat{\epsilon}, \hat{\lambda}\right)$ by $VaRapp(\mu_{C0}, \gamma_{C0}, \mu_{L0}, \gamma_{L0}, \epsilon_0, \lambda_0)$ and $VaRapp1\left(\mu_{C0}, \gamma_{C0}, \hat{\mu}_L, \hat{\gamma}_L, \hat{\epsilon}, \hat{\lambda}\right)$, respectively, and $\hat{\tau}_2^2$ by the estimated version of (7.29) and (7.30). \square

Note that in the theorem we assume that μ_{C0} and γ_{C0} are known. At a first glance this can look a bit strange. We saw that $\tau_2/\tau_1 = 3.9$ for VaR . This means that the contribution of estimating μ_C and γ_C is more than in the SLP case, which could suggest to include the influence of μ_C and γ_C . Moreover, Figure 7.5 illustrates that a similar situation (when τ_1 is large, comparing to τ_2) is present in a quite wide range of p . All this can suggest to include τ_1 in the calculation. However, our region of interest are probabilities $p > 0.9$, where the quantity τ_2/τ_1 grows very fast. This makes τ_1 not so significant while considering large quantiles.

To check whether τ_1 has a strong influence on the final result, we calculate $UB(\alpha)$ for $\alpha = 0.1, u = 7, \mu_{C0} = 100000$, and $\hat{\epsilon} = 0.03, \hat{\tau}_2 = 1303, \hat{\lambda} = 400$. Then $\Phi^{-1}(1 - \alpha)(u\hat{\lambda})^{-1/2}\hat{\tau}_2\mu_{C0} = 3155741$ and $VaRapp1(\mu_{C0}, \gamma_{C0}, \hat{\mu}_G, \hat{\gamma}_G, \hat{\epsilon}, \hat{\lambda}) = 47861121$. We see that the $\Phi^{-1}(1 - \alpha)(u\hat{\lambda})^{-1/2}\hat{\tau}_2\mu_{C0}$ part is about 6.6% (in the SLP case we had about 30% assuming the same parameters values) of the $VaRapp1(\mu_{C0}, \gamma_{C0}, \hat{\mu}_G, \hat{\gamma}_G, \hat{\epsilon}, \hat{\lambda})$ value. Including the influence of estimating μ_C and γ_C (taking $\hat{\tau}_1 = 335$) the relation becomes

$$\frac{\Phi^{-1}(1 - \alpha)(u\hat{\lambda})^{-1/2}(\hat{\tau}_2^2 + \hat{\tau}_1^2)^{1/2}\mu_{C0}}{VaRapp1(\mu_{C0}, \gamma_{C0}, \hat{\mu}_G, \hat{\gamma}_G, \hat{\epsilon}, \hat{\lambda})} = 0.068.$$

Hence, the estimation impact of μ_C and γ_C is not large, compared to the basic value of VaR (only 0.2%). Therefore, it still can be neglected, like it was done in the SLP case. However, while in Theorem 7.10 μ_{C0}, γ_{C0} are assumed to be known, in practice these parameters obviously have to be replaced by their estimates in the formulas of $UB(\alpha)$ and $LB(\alpha)$.

In Figures 7.6-7.8 some examples are presented of the extra amount due to the protection against estimation and the effect of dependence in these situations. Figures 7.6(a)-7.8(a) show the relative difference between the independent case and the dependence one. Here we assume that C and L each have a (different) Gamma distribution. We take $\gamma_{C0} = 0.4$ or $1.2, \mu_{G0} = 5, 10$ or $15, \gamma_{G0} = 0.5$ or $1, \epsilon_0 = 0.03$ and $\lambda_0 = 400$. The relative difference is defined by

$$\begin{aligned} & \frac{VaR - VaR_I}{VaR_I} \\ &= \frac{VaRapp(\mu_{C0}, \gamma_{C0}, \mu_{G0}, \gamma_{G0}, \epsilon_0, \lambda_0) - VaRapp(\mu_{C0}, \gamma_{C0}, \mu_{G0}, \gamma_{G0}, 0, \lambda_0)}{VaRapp(\mu_{C0}, \gamma_{C0}, \mu_{G0}, \gamma_{G0}, 0, \lambda_0)}, \end{aligned}$$

where VaR denotes the (approximated) VaR under dependence (Model 4) and VaR_I the (approximated) VaR under independence (Model 1). For a fair comparison we take both for the independence model and the dependence the same

set of quantiles

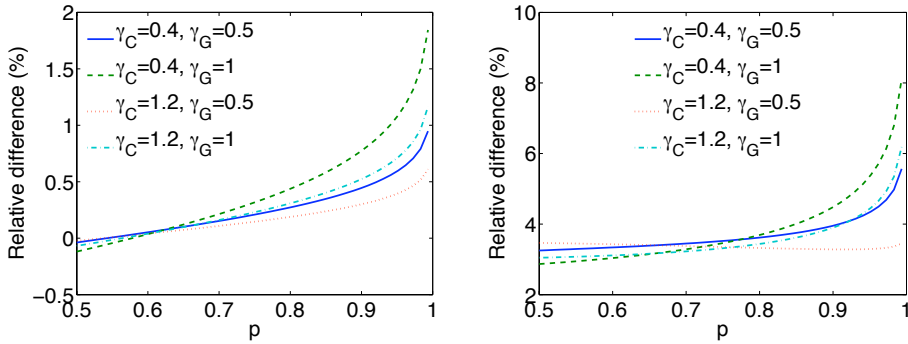
$$0.5 \leq p \leq 0.09995.$$

Figures 7.6(b)-7.8(b) show the extra amount due to protection against estimation, also measured in a relative way by taking

$$\frac{UB(\alpha) - VaRapp(\mu_{C0}, \gamma_{C0}, \mu_{G0}, \gamma_{G0}, \epsilon_0, \lambda_0)}{VaRapp(\mu_{C0}, \gamma_{C0}, \mu_{G0}, \gamma_{G0}, \epsilon_0, \lambda_0)}$$

with in $UB(\alpha)$ (which is obviously the one from Theorem 7.10) the estimators $\hat{\mu}_G, \hat{\gamma}_G, \hat{\epsilon}$ and $\hat{\tau}_2$ replaced by $\mu_{G0}, \gamma_{G0}, \epsilon_0$ and $\sqrt{\tau_2^2}$, respectively. We take $\alpha = 0.1$ and $u = 7$.

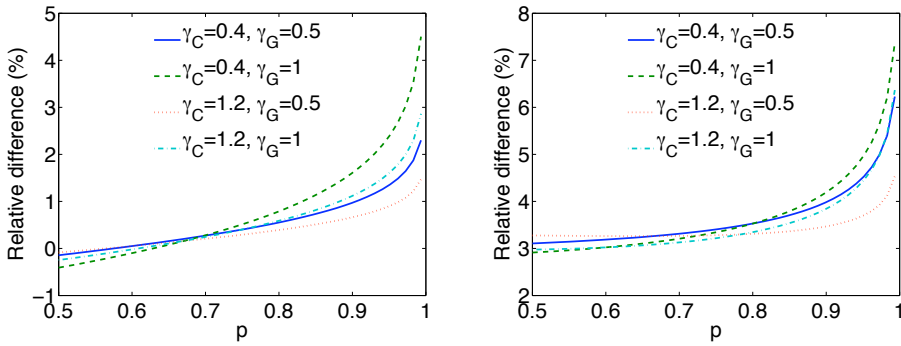
Note that the order of the displayed cases is slightly different in the figures *a* and *b*: for instance, for $\mu_{G0} = 15$ (Figures 7.8(a),(b)) the relative difference between dependence and independence is lower for $\gamma_{C0} = 0.4, \gamma_{G0} = 0.5$ than for $\gamma_{C0} = 1.2, \gamma_{G0} = 1$, while their order w.r.t. the relative extra amount due to protection against estimation is reversed.



(a) Relative difference between dependence and independence (b) Relative extra amount due to estimation

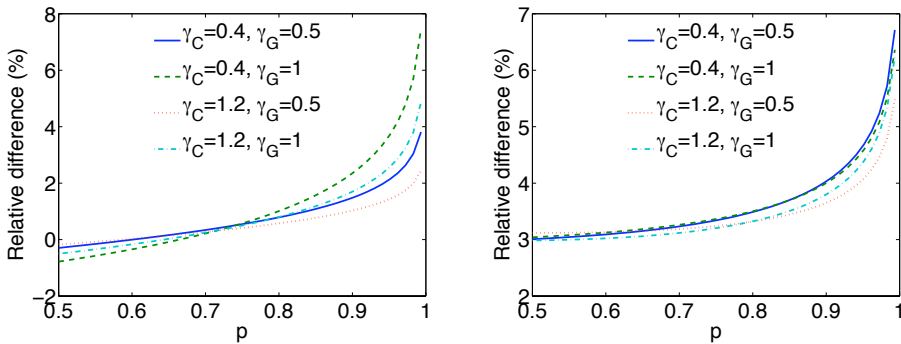
Figure 7.6: Dependence and estimation effects for *VaR* when $\mu_G = 5$

Figures 7.6-7.8 affirm that ignoring dependence may lead to very large absolute errors. But also the additional step due to protection against estimation is large. Sometimes even larger than the dependence effect. Figure 7.6 clearly illustrates this situation. The dependence effect reaches only 2% in the tail while the estimation effect reaches 8% under the same scale. Such situations are not in line with the *SLP* case where the dependence effect is usually much larger than the estimation effect. This can be explained by differences in fluctuation of the underlying risk measures. It is obvious that the estimated parameters are the same in both cases. Therefore, the main role here play the coefficients c_1, \dots, c_6 , which contain partial derivatives of the risk measures w.r.t. estimated parameters.



(a) Relative difference between dependence and independence (b) Relative extra amount due to estimation

Figure 7.7: Dependence and estimation effects for VaR when $\mu_G = 10$



(a) Relative difference between dependence and independence (b) Relative extra amount due to estimation

Figure 7.8: Dependence and estimation effects for VaR when $\mu_G = 15$

We already saw that these coefficients are different for VaR and SLP in absolute and even in relative sense, comparing to the values of the risk measures. For instance we saw that $c_6^2 \{ \mu_{G0} \epsilon_0 (1 + \gamma_{G0}^2) + 1 - \epsilon_0 \} = 2.84$ in case of SLP (which is only 0.00097% of the underlying SLP value) and equals 322550 in case of VaR (which is 0.67% of the underlying VaR value). Clearly, such a great relative difference stems from c_6 , which means that the fluctuation of VaR w.r.t. λ is much larger than the fluctuation of SLP . A similar (maybe less extreme) situation exists with the remaining parameters. This explanation gives an idea why the estimation impact is so (relatively) large in the case of VaR , compared to SLP and shows that it is useful to study the behavior of the risk measures and

the behavior of the estimators separately, as has been done in Sections 7.2 and 7.3.

To conclude we present the following results for the example with the true values of the parameters being equal to $(\mu_{C0}, \gamma_{C0}, \mu_{G0}, \gamma_{G0}, \epsilon_0, \lambda_0) = (100000, 0.7, 15, 0.8, 0.03, 400)$, $p = 0.99$ and C and L each having a (different) Gamma distribution. If we ignore the dependence, we get $VaRapp = 45825083$. If we take into account the dependence without protection against estimation we get $VaRapp(100000, 0.7, 15, 0.8, 0.03, 400) = 47861120$. If we add the protection (taking $\hat{\epsilon} = 0.03, \hat{\lambda} = 400$ and $\hat{\tau}_2 = \tau_2$) we get $UB(0.1) = 51016282$. Again, the protection and the dependence steps are very huge in absolute sense. The relative difference is not large, about 4% for the dependence step and about 6% for the protection. In total the result is an increase of about 10%. Compared to SLP this is not large. For high retentions such an increase for SLP can reach hundreds percents. However, we should not forget that high retentions in the SLP calculation produce low values of SLP . The situation with VaR is the other way around. The values of VaR increase with p and therefore significance of the dependence effect and the estimation effect in terms of relative difference should not be considered on the same scale as with SLP .

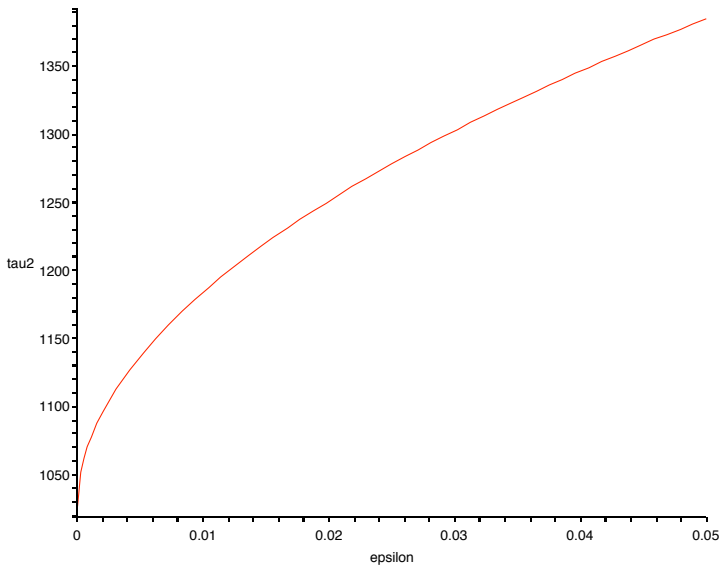


Figure 7.9: Behavior of $\tau_2(\mu_C, \gamma_C, \mu_G, \gamma_G, \epsilon, \lambda, p)$ as $\epsilon \rightarrow 0$ with $(\mu_C, \gamma_C, \mu_G, \gamma_G, \lambda, p) = (100000, 0.7, 15, 0.8, 400, 0.99)$

The upper and the lower bounds $UB(\alpha)$ and $LB(\alpha)$ contain the term $\hat{\tau}_2 \mu_{C0}$.

It was already mentioned in the previous chapters of the thesis that μ_C is a kind of dummy parameter. However, the expression of $\hat{\tau}_2$ has a quite complicated structure. For illustrative purposes we show the behavior of $\hat{\tau}_2$ as a function of ϵ when $(\mu_C, \gamma_C, \mu_G, \gamma_G, \lambda, p) = (100000, 0.7, 15, 0.8, 400, 0.99)$ are fixed. It is seen in Figure 7.9 that $\hat{\tau}_2$ does not tend to 0 when $\epsilon \rightarrow 0$. This is because τ_2 is the standard deviation of the asymptotic distribution of the estimation error of $\mu_G, \gamma_G, \epsilon$ and λ , which are the parameters of the special and simple part of the model. By tending ϵ to 0 we decrease the special part of the model and it disappears when $\epsilon = 0$. However, some influence is left from the simple part which enters to the standard deviation through λ .

Chapter 8

Fire data

This chapter deals with data obtained from Bert Teeuwen (Nationale Nederlanden), one of the members of the users committee of our STW-project. The main goals are analysing the data, trying to fit our model distributions and finding ways to determine dependencies within the data. The main structure which we need the data to have in order to be able to set up the model, has already been described in Chapter 6. A couple of examples presented there gave a basic feeling for possible dependence structures inside the data. It was demonstrated how the dependence groups can be defined using the information available. In this chapter we perform more or less the same procedure, but use a different strategy.

The chapter is divided into several sections. The first section is a description of the data, where the main features and examples of the data are presented. Next follows the section on the number of claims, which is the most important section of the chapter. Here we try to distinguish the small number of special claims from among the huge amount of the simple ones. Then follows the claim sizes section, where we fit the claim size distributions of the model. As soon as we assume the same distribution for simple and special claims, this step can be performed by using standard statistical analysis. The last section discusses the dependence and estimation effects on SLP and VaR on the basis of the estimated parameters.

8.1 Description of the data

We start with the description of the data, presenting the information which can be obtained without deep analysis. The data corresponds to a collection of claims during five years (2003.01.01-2007.12.17). Note that the year 2007 is present only partly, i.e. only 338 days are available for the analysis. The name of the data set is 'fire', which gives an idea about the line of the insurance business to which the data set corresponds. This concerns non-life insurance which covers different types of damages (excluding car accidents). The total number of claims

during these years is 39079. All claims are positive. The information which is presented (fields in the table) is OORZAAK (type of the damage), SCHAJR (year of the claim), SCHAMND (month of the claim), SCHADAG (day of the claim), BETAALD (claim size), NUMMER (number of the claim, starting from 2003.01.01), POSTCODE (area where the claim occurred). In total there are 1799 days for which claims are present. It is easily noted that one particular day contains a huge amount of claims. This is 2007.01.18, which stands for 1409 claims. Such a day can be considered as a catastrophe. There was a windstorm during that day and a huge amount of claims arose. Obviously, here we have a classical example of a dependence which in principle can be covered by our model as well. However, our main interest in this thesis is aimed at small dependencies and showing that these cannot be neglected at all. Therefore, all the claims which occurred on 2007.01.18 will be excluded from our analysis, decreasing the total number of days to 1798. Table 8.1 presents a small sample from the data (when SCHAJR is 2003) as an example. It can be noted that the POSTCODE field contains the numbers of two symbols. These are the two first numbers of the usual postcodes in the Netherlands. In total there are 90 postcodes with numbers

SCHAMND	SCHADAG	OORZAAK	POSTCODE	BETAALD
1	1	84	25	34.01
1	1	31	21	46.19
1	1	21	14	51.5
1	1	89	22	59.94
1	1	89	52	68.65

Table 8.1: Example of the 'fire' data

10, 11, ..., 98, 99. The OORZAAK field contains types of claims, denoted by different codes. In total there are 23 codes which can be divided into 7 groups: (11, 12, 13, 14) corresponds to fire claims which include fire, explosion, melting and singeing, (21, 22, 23, 24, 25) corresponds to different types of theft as burglary, theft, absence, theft form car, small thefts, (31, 32) corresponds to drains (and frost) and precipitation, (41, 42) denotes collision and smoke, (51) is storm, (61) denotes various, (83, 84, 85, 86, 87, 88, 89) is glass.

8.2 Number of claims

This section deals with the most important part of the model, the number of claims during the reference period. As soon as we assume the same distribution for the simple and special claims sizes, the number of claims becomes the major aspect through which dependence is introduced. As was already mentioned, the total number of claims can be divided into two groups: simple (which is modeled by random variable N) and special (which is modeled as a sum $\sum_{k=1}^H G_k$, where

H is the number of dependence groups and G_k are the group sizes). Therefore, the main goal is to recognize and distinguish these two types of claims. To do this, we group the data according to some feature. There are several grouping possibilities: time grouping (where different groups correspond to the different dates), space grouping (where groups are created according to the geographical position), type of the claim grouping (where claims are grouped using the type of claim field) and all combinations of these three types (for instance time and space). Each type of grouping corresponds to different special causes. For instance, in the time grouping the special cause can be the new year, Christmas, or other festive occasions which can lead to a larger number of claims (different damages, fires, etc.), compared to the remaining part of the year. In the space grouping the special cause can be 'bad' neighborhoods, in which numbers of claims are higher, not because of the larger number of clients living there, but because of some special feature of the neighborhood itself (for instance in student areas we can expect larger number of claims than in senior-living sectors). However, to perform the space grouping we need a bit more information. Namely, we need a total number of clients for each postcode. This is necessary to make all the postcodes equally important. This can be achieved by introducing weights which would define significance of the postcode depending on the number of clients which live there.

8.2.1 Time grouping

We start with the time grouping where different groups are formed according to year, month and day. As a result, we get a number of claims for each day, starting with 2003.01.01 and finishing with 2007.12.17, excluding the catastrophe day 2007.01.18. Hence, in total 1798 data points. The main goal now is to divide all the claims into simple and special. We proceed as follows: first we determine the 'suspicious' days, the days where the number of claims is not in line with the overall picture. These days will be considered later as special days (the days which include special claims). The next step is the analysis of only the simple days, trying to fit some distribution there and estimating the underlying parameters. Next, to obtain the number of special claims during the special day we generate (from the distribution estimated before) total number of simple claims for each special day and randomly subtract these claims from the total number of claims during the special day. Assuming that for each special day we have only one group, the expected group size and the standard deviation can be obtained straightforwardly.

To obtain a feeling about the suspicious days we construct the normal Q-Q plot for the daily claims, taking only one year 2003 (hence 365 data points in total) (see Figure 8.1). It can be seen that almost all the data lie on one line which indicates that these are 'more or less' normal data, but on the other hand, there is a breakpoint at the end which starts a new line with a different slope. Therefore, we can suppose that the data come from two different normal distributions.

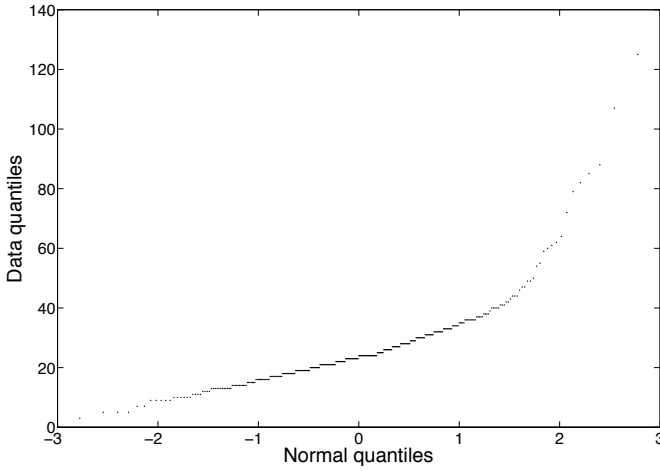


Figure 8.1: Normal Q-Q plot for the daily claims data

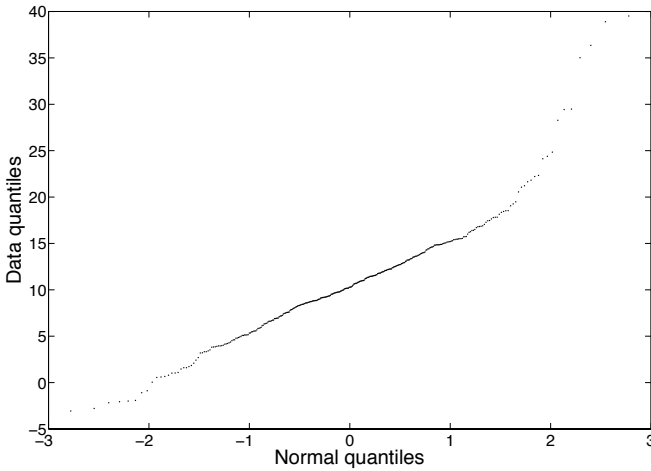


Figure 8.2: Normal Q-Q plot for the random variates of two different normal distributions

Indeed, as an example we present a normal Q-Q plot of a generated sample of 365 random variates, 355 of which are $N(10, 5^2)$ and 10 are $N(30, 5^2)$ (see Figure 8.2). It can be noted that these two pictures are similar in a sense that in both of them there is a breakpoint which intuitively divides all the data into two parts

stemming from different normal distributions. In case of the 'fire' data, it is clear that the suspicious (or special) days have larger numbers of claims. Therefore we assume that the number of claims during the special days come from the distribution with the larger (than during the simple days) mean. To make life more simple, we assume that the number of claims during the simple and special days follow normal distributions with different means. This normality assumption will be checked more carefully later in this chapter.

8.2.1.1 Simple claims

We start with determining the simple days, the days which do not contain any special claims. For this we will use the Q-Q plot data. First of all we construct the normal Q-Q plot of the data considered. Denoting the total number of observations by n , we create a plot with normal quantiles $x_i = \Phi^{-1}(i/(n+1))$ on the x -axis and with the ordered data points Y_i on the y -axis. As a result we get a sequence of observations

$$(x_1, y_1), \dots, (x_n, y_n).$$

The next step is to fit the simple linear regression to all the data points, assuming the usual form

$$Y_i = \beta_0 + \beta_1 x_i + \eta_i,$$

with as estimates for β_0 and β_1 :

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad \text{and} \quad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}.$$

The usual assumption for the errors η_i is the normal distribution $N(0, \sigma^2)$. Since most days are simple and the whole sample is ordered in ascending order, we can hope that the regression line mostly describes the simple data points. Therefore, the next step is to construct the one-sided 95% confidence interval for the mean response when $x = x_0$. This will define the data points which do not belong to the expected majority, as described by the regression. The confidence distance is defined by

$$t_{0.95, n-2} \hat{\sigma} \sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2}},$$

where

$$\hat{\sigma} = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n-2}}$$

and $t_{0.95, n-2}$ is the corresponding quantile of the t -distribution. All the points which are above the confidence bound will be considered as special data points (which obviously are the special days). The next step is to fit the above described linear regression, excluding the points which were determined as special during

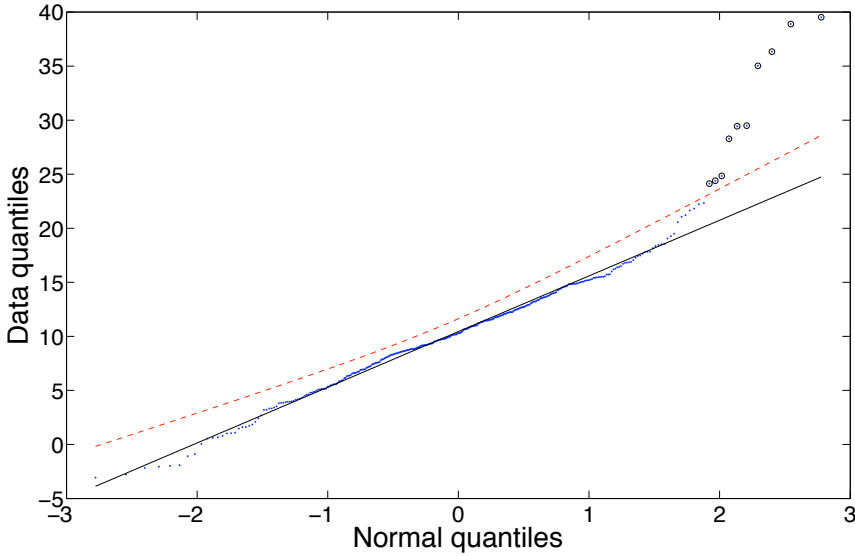


Figure 8.3: Confidence intervals for the regression of the normal Q-Q plot of the random variates of two different normal distributions. Result of two iterations of the method

the previous step. The construction of the confidence bound for that regression can single out more special points, since now the sample of the data considered contains more simple points compared to the special ones and the regression fits them better. These points also have to be excluded from the simple days. The process is continued until all the special days are excluded, i.e. until exclusion of the special point and construction of a new confidence bound do not give any new special points. As a result we get a regression line which fits only the simple days. All the data points which are below the confidence bound are considered to be simple.

As an example, we illustrate the method on the data of the random variates from two different normal distributions: $N(10, 5^2)$ (from which we generate 355 points) and $N(30, 5^2)$ (from which we generate 10 points). The normal Q-Q plot of that data is presented in Figure 8.2. Two iterations have distinguished 10 special points which is in line with the number of outliers which were included in the sample. Figure 8.3 presents the final result: the linear regression which corresponds to the simple points only, the upper confidence bound and 10 outliers which are denoted by circles.

To make sure that this is not a chance result, we perform a brief simulation study. By taking different number of outliers and different mean distances we

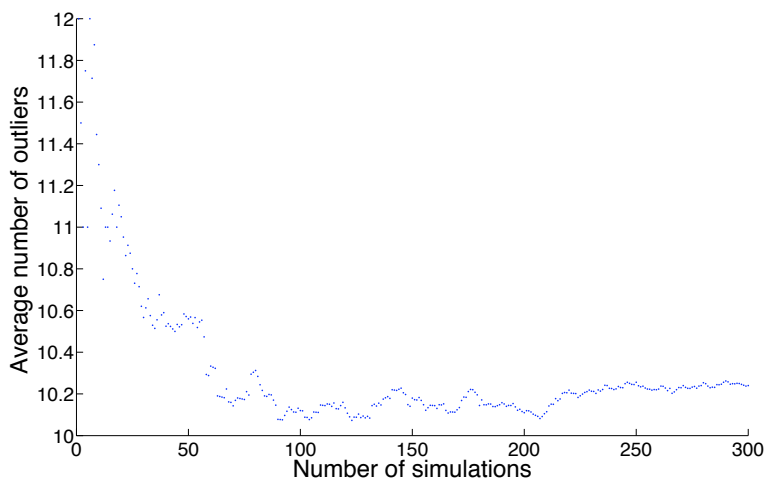


Figure 8.4: Simulation result when the number of $N(30, 5^2)$ outliers is 10 and the number of the simple $N(10, 5^2)$ points is 355

simulate the method and check whether the expected number of outliers is in line with the real amount. An example of such a simulation can be found in Figure 8.4. It can be easily seen that the expected number of outliers converges to a number which is quite close to 10, which is in line with the real number of outliers. Analysing the simulation results resulted in further conclusions, related to the mean distance between the simple and special points. The distance between the means should not be too low, as in that case it is impossible to distinguish the special and simple points by using the above described method. Simulation results showed that in order to 'catch' the outliers, the distance in mean should be not lower than 15 (for fixed $\sigma = 5$) and even then, slight underestimation is possible. The larger the distance in means, the easier the outliers can be distinguished. The good news is that this type of restriction is reasonable for our purposes and hence the method can be applied. We will use it as the main tool in distinguishing the simple and special claims. Of course, it should be noted that the approach stays rather heuristic (or intuitive). The lack of knowledge about the underlying structure undermines the possibilities for a more thorough theoretical analysis.

We proceed as follows. Having the five year of data, we consider each year separately and use the above described method to distinguish the suspicious days, which will be considering as special days, containing the special claims. The results are presented in Table 8.2. The next step is to estimate the distribution of the number of claims for the simple days for each year (special days will be excluded from the analysis). It was already mentioned that this distribution will be assumed to be normal. To illustrate that this is indeed a reasonable assumption,

Year	Total number of days	Nr. of simple days	Nr. of special days
2003	365	341	24
2004	366	330	36
2005	365	333	32
2006	365	343	22
2007	338	313	25

Table 8.2: Number of simple and special days of the fire data during several years

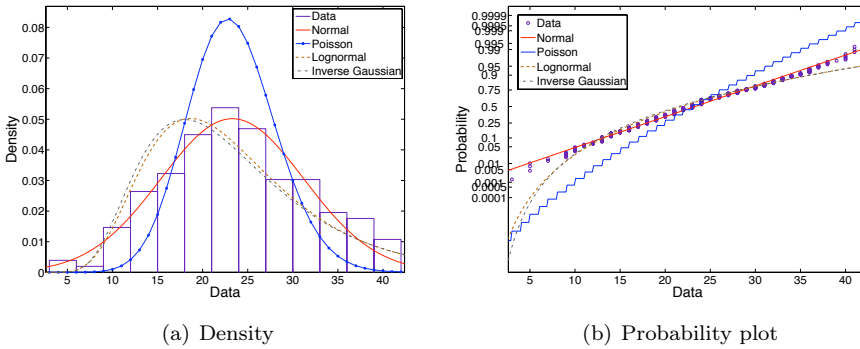


Figure 8.5: Several distributions are fitted to the simple days of the fire data

we compare several fits of the different distributions. As an illustrative example we consider year 2003. Figure 8.5 illustrates the fits of several distributions. It can be noted that the normal distribution is the best one. Moreover, it is also a good one. This can be concluded from the probability plot, Figure 8.5(b). For each year we fit a normal distribution to the data. Parameter estimates can be found in Table 8.3.

8.2.1.2 Special claims

Having obtained the estimated distributions of the number of claims for simple days for different years, we can start distinguishing the numbers of special claims during the suspicious days. The point is that each suspicious day can contain both simple and special claims. It is impossible to distinguish these using the claim size, since we assume that the simple and special claim sizes follow the same distribution. Therefore, to get the number of special claims during the suspicious days, we generate a total number of claims per simple day for each of the suspicious ones and subtract this from the total number of claims.

Having the number of special days for 5 years (see Table 8.2), we start the procedure. Table 8.4 presents the results, which basically are all we need for the

Year	$\hat{\mu}$	$\hat{\sigma}$
2003	23.3	7.95
2004	21.34	7.26
2005	17.95	6.16
2006	16.9	6.33
2007	11.04	5.56

Table 8.3: Parameter estimates of the normal distribution of the simple days for different years

u	H_{tot}	N_{tot}	G_{tot}	\bar{G}	\bar{G}^2
5	139	30212	4969	36	1761

Table 8.4: Time grouping results

estimation of the parameters of the number of claims. Now, using the estimators which were presented in Chapter 7, we calculate the estimates of the parameters

$$\hat{\mu}_G = 36, \hat{\lambda} = 7036, \hat{\epsilon} = 0.14, \hat{\gamma}_G = 0.6.$$

We see that some estimates are larger than we expected at the beginning of the analysis. Moreover, they are out of the region of interest which was introduced in order to check the accuracy of the approximations. The estimate of λ is 7036 instead of 400, which was assumed before. However, this is not a problem since it was shown that the accuracy of the approximations becomes better when λ increases. Here the estimate of λ is 17.59 times larger than 400. For this reason we should not much worry about the larger values of the estimates of μ_G and ϵ , which earlier were assumed to be 20 and 0.03.

One more remark should be made about the estimates. In Chapter 7 it is assumed that both N_t and H_t ($t = 1, \dots, u$) follow a Poisson distribution. However, here it turned out that N_t is fitted well by a normal distribution. In spite of this fact the estimators which were introduced in Section 7.1 can still be used here since these are 'natural' estimators of the parameters.

8.2.2 Space grouping

One more type of grouping we consider here is the space grouping. In the previous section the time grouping approach was described where the special cause was some event in time, for instance Christmas, New Year, or some other festive occasion which can lead to a larger number of claims (different damages, fires, etc.), compared to the remaining part of the year. In the space grouping the special cause can be 'bad' neighbourhoods (or other space areas), in which the number of claims is higher, not because of the larger number of clients living there, but because of some specific feature of the area itself. For instance in

Year	Nr. of suspicious postcodes	Suspicious postcodes
2003	3	(10, 25, 68)
2004	8	(10, 25, 68, 79, 29, 49, 30, 56)
2005	3	(10, 68, 25)
2006	3	(68, 10, 25)
2007	5	(68, 49, 70, 10, 25)

Table 8.5: Suspicious postcodes

student areas we can expect larger number of claims compared to living sectors with mainly seniors.

Having 90 POSTCODES in total (remember that only the first two digits are available), we calculate the total number of claims for each postcode for each year. However, before starting the procedure of distinguishing the special claims, we have to make all the areas equally important. The problem is that the total number of policies for different postcodes is also different and hence the larger number of claims in some postcode does not directly mean that it is special. Therefore, for the analysis we consider not the total number of claims but the fraction of total number of claims to the total number of policies for each postcode.

The procedure of distinguishing special claims stays the same. It was explained in detail in the previous section. The only difference here is the grouping strategy and the introduction of the weights. The suspicious postcodes are presented in Table 8.5. Note that the postcodes 10, 25 and 68 in the table are present in each of the years.

The next step in the procedure is to distinguish the special claims from the total number of claims for the suspicious postcodes. However, we see that the number of the special postcodes (and hence the number of the dependence groups) is small, ranging from 3 to 8. We could expect this, since the total number of postcodes is only 90. This fact significantly complicates the estimation procedure: it is practically impossible to get reasonable estimates of the underlying parameters from only three numbers. We have to change the grouping strategy and the most logical way is to use postcodes which are not aggregated, i.e. to consider smaller areas. However, this information is not available: the postcodes in the data are presented in aggregated form w.r.t. first two numbers and it is impossible to split them. For illustration purposes we could split the postcodes ourselves, according to some strategy. However, this type of approach was already considered in Chapter 6. Here we will prefer to stick to the real data.

Another possibility is to consider mixture grouping. There are two possibilities: space-time grouping and space-claim type (or space-cause) grouping. We do not consider space-cause grouping, since different claim types definitely have different expected group sizes (according to different special causes). This aspect is covered by the overdispersion approach which is covered by the model. Since we still would like to concentrate on space, we consider space-time grouping

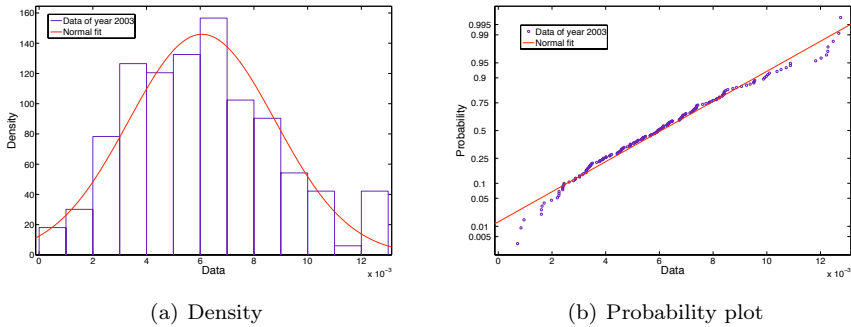


Figure 8.6: Normal distribution is fitted to the number of simple claims in different space-time grouping points

or, more precisely, space-half a year grouping (space and two equal parts of the year). Such a strategy certainly splits the data into more points (compared to only space grouping) and preserves space as the main special cause. Probably one of the parts of the year could be considered as the special and some groups can occur due to the special feature of the part, not the postcode. However, we assume that both parts are equally important which makes space the main special cause.

Table 8.6 presents the results of the first step of the procedure. Here we distinguish suspicious grouping points. Note that the postcodes which were defined as suspicious during the postcode grouping stay suspicious here as well. Moreover, three postcodes which were pointed out due to their occurrence in each year, contain special claims in both parts for all the years except 2007 (probably because we do not have data for the whole year 2007), which indeed makes them essentially independent from the time aspect. This result is in line with our expectations.

The next step is to distinguish the special claims within the total number of claims of the suspicious points. As mentioned before, we consider not the number of claims itself, but the fraction of the total number of claims to the total number of policies per postcode and for this reason normal distributions (instead of Poisson) will be fitted to these fractions. Figure 8.6 illustrates that the normal distribution provides a reasonable assumption for the fraction of simple claims in different grouping points. The fitting results are presented in Table 8.7. Hence, instead of the number of special claims we get the fractions of the special claims to the total number of policies. To get the number of special claims, we should multiply the fractions by the total number of policies of the corresponding postcode. The method of finding the number of special claims was described in detail in the previous section. Here we present only the final result, the estimates

Year	Total	Special	$\hat{\mu}$	$\hat{\sigma}$
2003	267	14	0.42	0.21
2004	264	23	0.4	0.18
2005	263	14	0.35	0.17
2006	263	10	0.33	0.16
2007	257	6	0.3	0.21

Table 8.7: Fitting results during the space-time grouping

of the model parameters which deal with the number of claims. They are

$$\hat{\mu}_G = 107, \hat{\lambda} = 7035, \hat{\epsilon} = 0.2, \hat{\gamma}_G = 0.86.$$

Again, we see that some estimates are larger than we expected. Moreover, as in the time grouping case, they are out of the region of interest which was introduced in order to check the accuracy of the approximations. The estimate of λ is 7035 instead of 400, which was assumed before. However, as was already mentioned in the previous section, this is not a problem since it was shown that the accuracy of the approximations becomes better when λ increases. Here the estimate of λ is 17.59 times larger than 400. For this reason we should not much worry about the larger values of the estimates of μ_G and ϵ , which earlier were assumed to be 20 and 0.03. Note that the estimate of μ_G is much larger than the value obtained for the case of time grouping. This fact can be explained by the grouping strategy. Our assumption is that each grouping point can have only one dependence group. Hence, by considering higher grouping aggregation, we automatically consider larger groups, but the number of groups becomes smaller. This also can be seen from the formula of the expected number of groups, which is $\epsilon\lambda/\mu_G$.

The same (as in the previous section) remark about the estimates is applicable here as well. In Chapter 7 it is assumed that both N_t and H_t ($t = 1, \dots, u$) follow a Poisson distribution. Here N_t is introduced through the fractions and a Poisson distribution becomes a poor assumption. However, it turned out that the fractions are fitted well by a normal distribution. In spite of this fact the estimators which were introduced in Section 7.1 can still be used here since these are 'natural' estimators of the parameters.

8.3 Claim sizes

The goal of this section is to find a reasonable distribution which fits the individual claim sizes of the fire data. Each claim here is considered as a separate data point and the total number of such points is 38990. It was already mentioned that in our analysis we assume that special and simple claim sizes follow the same distribution. Hence, we do not need any information about the dependence

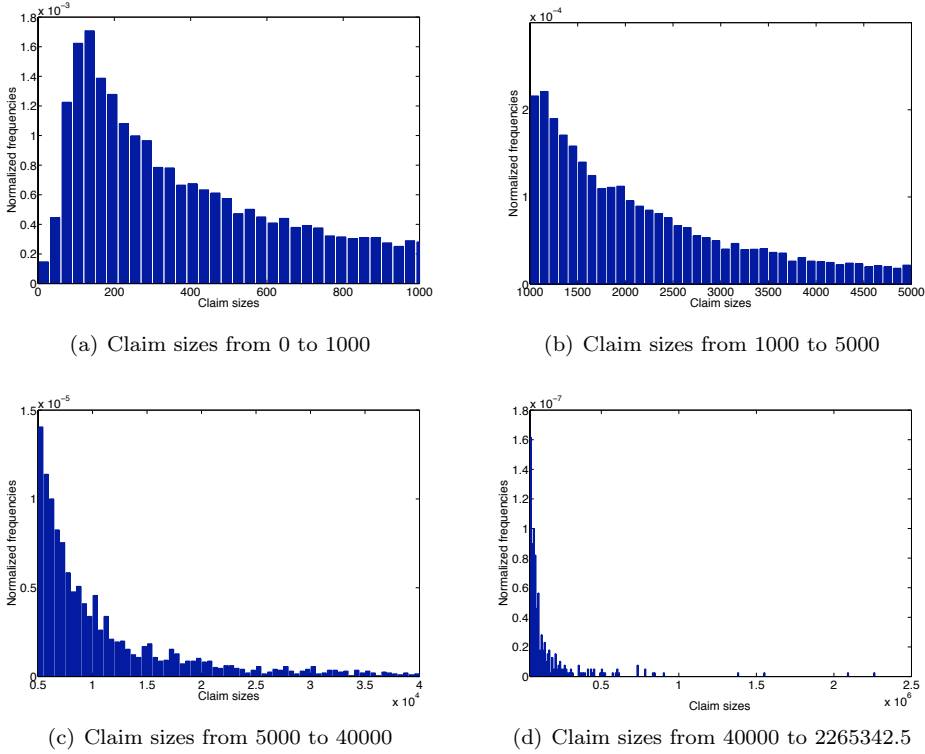


Figure 8.7: Histogram of the fire data

structure and consider all the claim sizes together, as one sample. There is also no need to divide the data according to different years since in the parameters estimation procedure all the years will be aggregated anyway.

We will start with plotting the histogram and the empirical mean excess function of the data. This should give us an impression about the main features of the underlying distribution. The maximum claim size during these five years is 2265342.5 EUR and the minimum one is 0.01 EUR. The scale is too wide to present the histogram of the data in one picture. Therefore we divide it into four parts. The four histograms are presented in Figure 8.7. In general the histogram plot provides a means of assessing the symmetry and variability of the data. If the data are symmetric, then the structure of the histogram plot will be symmetric around a central point such as a mean. The histogram plot also indicates if the data are skewed and the direction of the skewness. In our case we see that the data are not symmetric and are skewed to the right. We also see that the tail of the underlying distribution is heavy. Additional evidence for that we present by using the empirical mean excess function plot (see Figure 8.8). When considering

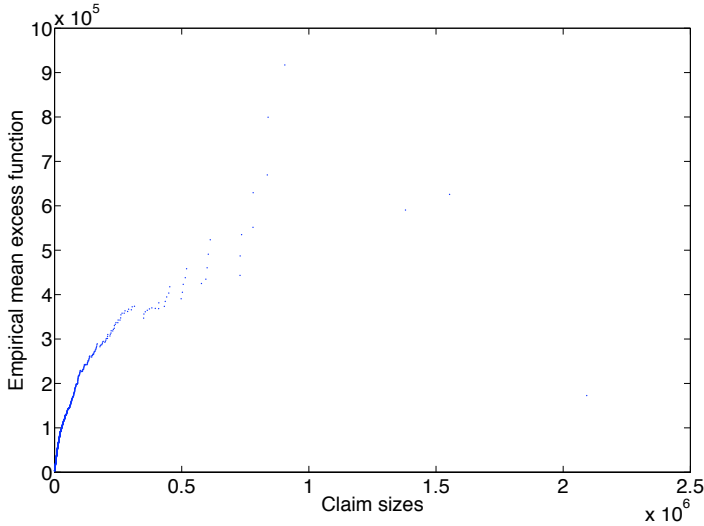


Figure 8.8: Empirical mean excess function of the fire data

the shapes of mean excess functions, the exponential distribution plays a central role. The mean excess function for the exponential distribution is constant. If the distribution of the underlying data is heavier-tailed than the exponential distribution, the mean excess function ultimately increases, when it is lighter-tailed it ultimately decreases. In our case the empirical mean excess function increases and its shape is similar to the mean excess function of the lognormal distribution. Therefore, in the fitting process we should try heavy-tailed distributions which are skewed to the right. There are many distributions with such characteristics. However, we will limit ourselves to only three of them, namely the Gamma, Inverse Gaussian and lognormal (see Table 3.2). The whole model theory and the approximations were developed and tested extensively in the foregoing chapters for these three representative distributions. To introduce more distributions, the whole analysis would have to be performed from the very beginning, which would take too much time. Moreover, these distributions are commonly used as an assumption for the claim sizes in the insurance business and we already saw that the shape of the empirical mean excess function is similar to the mean excess function of the lognormal distribution. All this indicates that we are going in the right direction.

Using the estimators which were presented in Chapter 7, we get the estimates for μ_C and γ_C , which are

$$\hat{\mu}_C = 2894.19, \quad \hat{\gamma}_C = 9.3.$$

The estimates of the parameters of the distributions considered are presented

Distribution	$\hat{\alpha}$	$\hat{\beta}$
Gamma	0.0116	4×10^{-6}
Lognormal	5.74	2.11
Inverse Gaussian	2894.19	250132.37

Table 8.8: Estimates of the parameters of the fitting candidates

Grouping method	$\hat{\lambda}$	$\hat{\epsilon}$	$\hat{\mu}_G$	$\hat{\gamma}_G$	$\hat{\mu}_C$	$\hat{\gamma}_C$
Time	7036	0.14	36	0.6	2894.19	9.3
Space-time	7035	0.2	107	0.86	2894.19	9.3

Table 8.9: Estimates of the model parameters

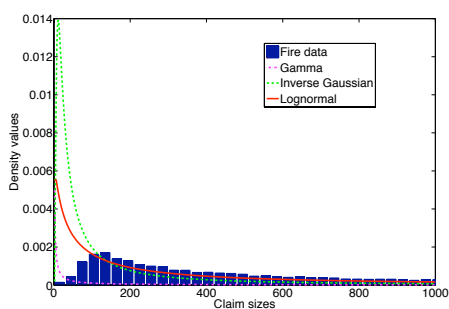
in Table 8.8. To judge the goodness of fit we first check how the estimated densities fit the histogram of the data. Once more, the results are presented as four pictures with different scales (see Figure 8.9). A lot of information can be obtained already from the first picture. Here we basically are interested in the concentration of the density at small values. Figure 8.9(a) shows that the Gamma distribution gives a very bad fit. In fact the parameters are very extreme, leading to e.g. $\int_0^1 f_\gamma(x; \hat{\alpha}, \hat{\beta}) dx = 0.87$, while we have only 11 observations out of 38990 in $[0, 1]$, thus showing that the fit is very bad. The Inverse Gaussian distribution is also not very good. Here we have $\int_0^{100} f_{IG}(x; \hat{\alpha}, \hat{\beta}) dx = 0.57$, i.e. more than half of the density is concentrated at the very beginning (interval $[0, 100]$) where we have only 2769 out of 38990 observations (about 7%). The lognormal is the best one, but here we also have $\int_0^{100} f_{LN}(x; \hat{\alpha}, \hat{\beta}) dx = 0.25$.

Using the above arguments we conclude that lognormal distribution fits the data best, compared to Gamma and Inverse Gaussian distributions.

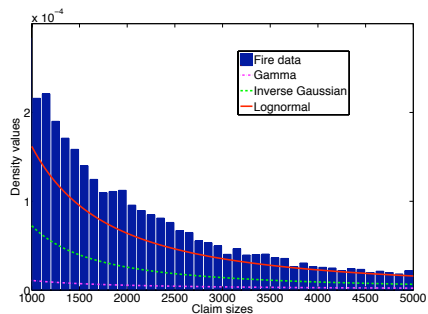
8.4 Examples and discussion

This section deals with the illustration of the dependence and estimation effects on the basis of the risk measures SLP and VaR . The two methods of grouping which were considered in this chapter gave us two different sets of the parameters estimates. For the summary of the estimates we refer to Table 8.9. For each of these sets we will illustrate and discuss the dependence and the estimation effects which were already defined in Chapters 6 and 7.

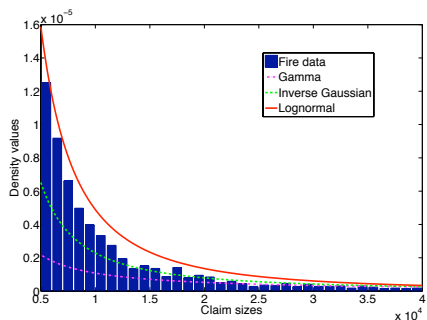
However, before presenting the results we discuss the fact that the distribution of the number of claims which was estimated from the data turned out to be Normal, not Poisson (as was assumed beforehand). This means that in principle Model 4 (as well as Models 1-3) can not be applied immediately. To match the data with the basic assumption of Model 4, we should change the model assumption of the number of claims. Such a replacement is more or less straightforward



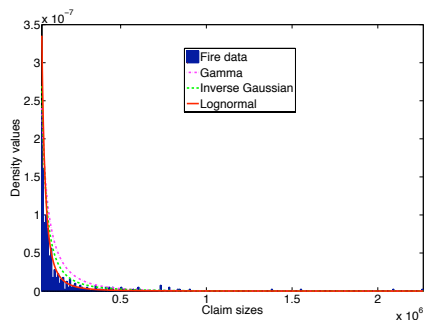
(a) Claim sizes from 0 to 1000



(b) Claim sizes from 1000 to 5000



(c) Claim sizes from 5000 to 40000



(d) Claim sizes from 40000 to 2265342.5

Figure 8.9: Comparison of the estimated densities and the histogram

and does not introduce any new features to the model. However, it means a huge amount of technical work, starting with changing the formulas for the cumulants and finishing with rewriting Chapter 7. All of this program can be done along the same lines as for Model 4. We are not going to do it here, but this is a useful suggestion for the extension of the model. Note that when changing the assumptions of Model 4, we should also change the assumptions of Model 1, in order to keep the models on the same scale. However, the Poisson assumption for the number of claims is very classical in the insurance world. That explains why Poisson was taken as a basic assumption for the number of claims and that is why such an approach is still the one of main interest. After all, the main target of interest is not the aggregated sum itself and even not the underlying risk measures. In the present thesis we are basically concentrated on the dependence and estimation effects, which are relative quantities. Possibly, the change of assumption for the number of claims may have a strong influence on the aggregated sum and the underlying risk measures under both models, but we may hope that the dependence and estimation effects will be similar for both assumptions. Therefore, in this chapter we will continue to use the assumptions of Model 4 which were originally introduced in Chapter 3.

8.4.1 Dependence and estimation effects

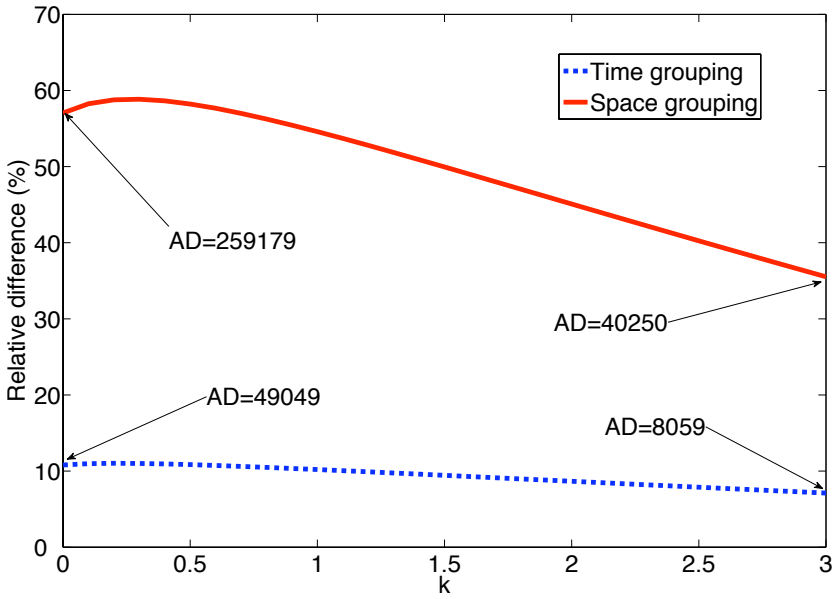
Having two sets of the estimated parameters, we calculate SLP and VaR under the two models (Model 1 (independence model) and Model 4 (dependence model)) for each set. The dependence effect for SLP or VaR is defined as the relative difference of these risk measures w.r.t. the independence and dependence models, while the estimation effect is defined as the relative extra amount which has to be added to the risk measure as a protection against the estimation error. Approximations which were introduced in Chapter 4 will be used as the main tool in the calculation of SLP and VaR . As usual, the retention level for SLP is $a = \mu_{S_1} + k\sigma_{S_1}$, with $k = 0, \dots, 3$. The illustrative range of probabilities for VaR is $p = 0.9, \dots, 0.9995$. Using the conclusion of the previous section the distribution of C is assumed to be lognormal, which is in line with the model assumption which were introduced beforehand. The next step is to choose the assumption for L . Here the choice is limited to the Inverse Gaussian and Gamma distributions. We know that the assumption $L \sim \text{Gamma}(\alpha_L, \beta_L)$ leads to $G \sim \text{NB}(\alpha_L, \beta_L)$, which in principle can be presented as a sum of α_L independent random variables having a geometric distribution with parameter β_L . In that case the Central Limit Theorem states that G is approximately normal when the summation index is large. The analysis of the data showed that the normal distribution fits the number of simple claims best. The original formulation of the model assumes that the number of simple claims and the group sizes follow the same distribution with different levels of dispersion. Therefore $L \sim \text{Gamma}$ should be a reasonable choice which keeps our model and the data in line.

Figure 8.10(a) presents the dependence effect for SLP under the two sets of

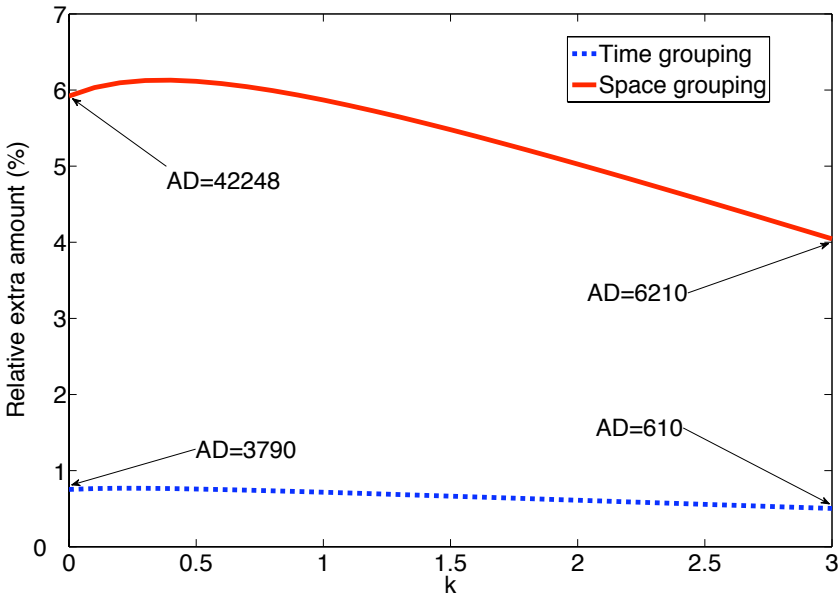
parameter estimates. The two lines here correspond to two different types of grouping, which have led to two different sets of model parameters (see Table 8.9). We see that the dependence effect here is not so huge in a relative sense (about 60% at maximum) if compared to the examples constructed before (see e.g. Figure 7.3(a)) where it reaches almost 4300%. However, it is still very large. Moreover, as was stated before, we should not blindly concentrate on the relative difference. Absolute difference definitely has to be taken into account as well and the following example clearly illustrates that situation. The absolute differences at the most illustrative points are denoted on the figures by AD . At $k = 3$ we have $AD = 40250$ in case of the space grouping correspondingly to a dependence effect of 38%. We compare this result to the example which is illustrated on Figure 7.3(a). There we have $AD = 43077$ at the same retention level and this is in that situation a dependence effect of 4300%. The similar situation holds for the whole picture, including the time grouping situations. We note that the dependence effect under the time grouping is less than in case of the space grouping. However, this fact is not a surprise since it could be predicted from the values of the model parameters (see Table 8.9). Values of ϵ , μ_G and γ_G are larger in case of the space grouping which (according to the discussion of Section 3.3) makes the dependence effect larger as well. We note also that the dependence/estimation effect for SLP is decreasing with the retention level. Up to now almost all the examples have illustrated the opposite situation when the dependence effect increases with the retention level. However, this is certainly not the case in general. Moreover, we already saw (see e.g. Figures 7.1(a), 7.2(a) and 7.3(a)) that growth of the dependence effect is much slower when the ratio of γ_C/γ_G is large. In the mentioned example we had at most $\gamma_C/\gamma_G = 2.4$ which made the growth of the dependence effect slower (compared to the remaining situations) for all the considering values of μ_G . In our case we have $\hat{\gamma}_C/\hat{\gamma}_G = 15.5$ in case of the time grouping and $\hat{\gamma}_C/\hat{\gamma}_G = 10.8$ in case of the space grouping. This is much more than 2.4 and that could be a reason for the negative slope of the dependence/estimation effect.

Figure 8.10(b) illustrates the estimation effect for SLP under the time and space grouping. We see that the estimation effect is about 10 times smaller than the dependence effect. This fact is in line with our expectations. Similar results were presented before (see e.g. Figure 7.1(b)). For high retention levels the estimation effect is not significant compared to the dependence effect, especially in case of the time grouping. However, at lower retentions it definitely should be taken into account.

Figure 8.11 illustrates the dependence and estimation effects on the basis of VaR . Here we see that the dependence effect is smaller in the relative sense, compared to the SLP case. This is not a surprise since similar results were obtained before. However, the absolute difference here runs into the millions, which makes it not so insignificant. As was expected, the dependence effect in case of the time grouping is less, compared to the space grouping. Additional attention should be attracted by the estimation effect which is comparable to

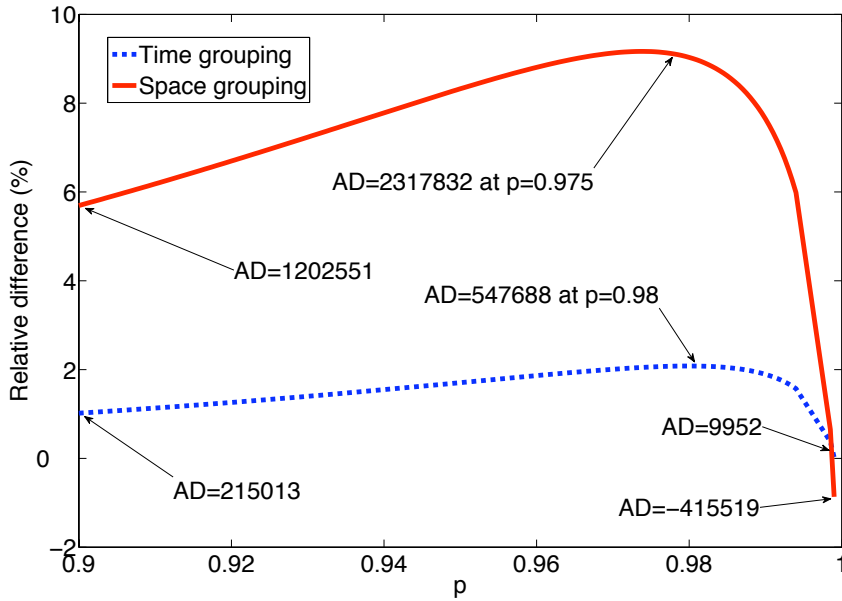


(a) Dependence effect

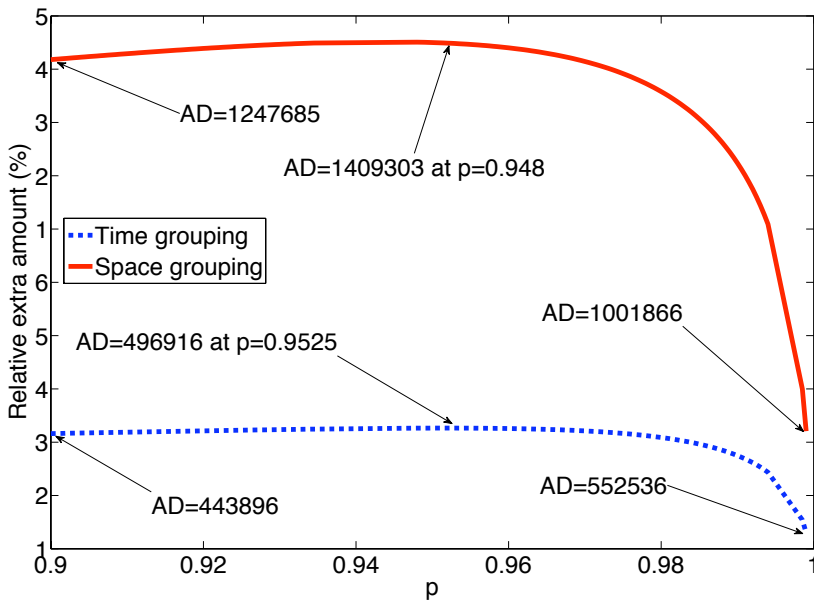


(b) Estimation effect

Figure 8.10: Dependence and estimation effects for *SLP*. Notation 'AD' stands here for 'absolute difference'



(a) Dependence effect



(b) Estimation effect

Figure 8.11: Dependence and estimation effects for *VaR*. Notation 'AD' stands here for 'absolute difference'

(and sometimes even larger than) the dependence effect. Similar situations can be observed in Figures 7.6-7.8. In our case we see that for high risk levels the dependence effect becomes comparably small, even negative, while the estimation effect is still huge, even in the time grouping case.

Chapter 9

Review and discussions

This chapter contains the review of the thesis. Starting from the first chapter, we will briefly discuss the main goals, the most important conclusions and (if necessary) recommendations for future development.

The present thesis considers an advanced actuarial modeling technique which introduces dependence as a natural extension of the existing independence models. The structure of dependence considered here, together with the classical ways of insurance modeling, is briefly introduced in Chapter 1. The (so called) dependence model has gone through several development steps (the main ones we have pointed out by calling them Model 1, Model 2 and Model 3) before reaching the final form (Model 4), which is the main target of investigation of the thesis. Chapter 1 briefly discusses all these development steps, introduces the risk measures, discusses the approximation aspect and different kinds of influences on the risk measures considered. In the present research we mostly concentrate on the distribution of the aggregate sum of claims S , which usually is far from normally distributed, i.e. not symmetric and having a heavy tail. Therefore the risk measures which concern the tail behavior of the distributions involved, are of main interest. The net Stop-Loss premium (SLP), the variance of the Stop-Loss contract and the Value at Risk (VaR) were chosen as representative risk measures.

The form of dependence introduced in the thesis is quite new and differs considerably from the structures available in literature. Therefore, all assumptions and generalizations of the model have to be motivated. This aspect is covered in Chapter 2, which is devoted to an illustrative example of a possible dependence structure. The idea was to create an algorithm which would simulate some concrete dependence pattern, illustrating the needed features which are hidden in the available data sets. Such features are mostly connected to the distributional assumptions of the random variables which appear in the model (see (1.9)). In Model 4 the distributional forms of N and H have been settled in Chapter 1. These are $N \sim P(\lambda(1 - \epsilon))$ and $H \sim P(\lambda\epsilon/\mu_G)$. Candidates for the distribution

of C (and D) are presented in Section 1.2. Finally, G is assumed to be $P(L)$, where L is random. All these distributional assumptions are motivated in Chapter 2, where we mostly concentrate on the distributions of N, H and G , since these concern the new aspect of our model. Representative samples of these random variables were obtained by simulating a flu epidemic inside a hypothetical company (for the algorithm see Section 2.1). The samples obtained were analyzed and possible candidates were suggested for the distributions of each of the random variables considered. The Poisson distribution showed a very good fit while the samples from N and H were considered. However, the analysis of the samples from G did not point out a leader among the candidates considered. Nevertheless, several important conclusions about the distribution of G were obtained while analyzing the data. These are the positive skewness (between 3 and 4) and large kurtosis (about 14). The negative binomial distribution was suggested for the distributional assumption of G in view of some of its properties which significantly simplify application of the model.

Chapter 3 describes Model 4 in more detail, presents its main characteristics, discusses the impact of the model parameters (see Section 3.3) and describes the region of interest for the values of the parameters involved (see Tables 3.8 and 3.7).

Calculation aspects are covered in Chapter 4. There we present several calculation methods for the chosen risk measures. These are convolutions, approximations and simulations. The major calculation tool which is used while analyzing dependence and estimation effects is the approximation. The methods of convolutions and simulations are used in the approximation testing procedure, which is the main topic of Chapter 5. Here we carefully analyze the chosen approximations, introduce the retention level for SLP (and the variance of the Stop-Loss contract) and the representative probabilities for VaR . All decisions about the accuracy of the approximations are based on the chosen criterion, which is defined in (5.1) and (5.2). As a result, we present a rule of thumb for Model 4. The general recommendation is to use the Gamma-IG approximation for the SLP and the variance of the Stop-Loss contract approximation and the IG approximation for the VaR approximation when the aggregated sum S is modelled by Model 4. This rule is valid in the given region of parameters of interest, which is presented in Table 3.7 and restricted to the distributions under consideration, which are presented in Table 3.2. The schematic overview of the rule of thumb for Model 4 is presented in Table 5.3.

In Chapter 6 we present an explicit example of the dependence effect. The example concerns a concrete line of the insurance business, the so-called workers compensation insurance. The underlying data of this branch of insurance perfectly fit our model (dependence structure is present) and, moreover, it was suggested by the users committee of our STW-project as a potential source of problems due to dependence. However, the needed structure is not present in the available data. Therefore, it serves only as a basis for the construction of the data set needed. Hence, real data are not used in this chapter, but compared to Chapter

2 (where the dependence problem was invented by ourselves and all data were simulated) here we go a step further, since real data are the starting point. This contributes to the reality aspect of the example. Another goal of the chapter is to discuss the structure of the data which is needed for the Model 4 implementation. To get a better feeling about the requirements for the data, we present several examples of data sets. The first example is the "perfect" data set, where all the necessary information is included (see Table 6.2). It is "perfect" in the sense that we directly have all the information about the groups and the claim amounts. By using the "Group code" field we can get all the information about the groups: the number of groups and the group sizes during the reference period. The main problem is that usually a field "Group code" is not available in the data set. In that case it should be possible to create this type of field from the existing data and available extra information about the data. We present an example of such a possibility. In Table 6.3 the "Group code" field is created from the "Date", "Incident" and "Place" fields (in general, it can be a much more complicated structure). Individual claims are denoted by 0, as in the previous example. In case several accidents of the same type happened at the same date and at the same place, we assume that they form one dependence group.

All the results are presented in Tables 6.7 and 6.9. Table 6.10 presents the comparison of all the quantities considered. From that table we directly see that the *SLP* under Model 4 is much higher than the one under Model 1 in a relative sense and that this difference becomes huge for high retention levels. It means that the expected risk of selling the Stop-Loss contract is much higher than we expected while believing in Model 1. In case the premium is determined according to the standard deviation principle, the dominant quantity in the premium is not the *SLP* but the standard deviation. In spite of the fact that the relative difference in standard deviation according to different models is smaller (but still huge) than in case of the *SLP*, the absolute values are much larger (see Tables 6.7 and 6.9). Also there is a huge difference in a relative sense between the second moments of the contract (quantity $(S - a)^+$). For the highest retention level it could reach a factor 380 (38000%). This definitely means that the tail of the contract distribution can be much heavier than we expected using Model 1. All this shows that the supposed rare events are not so rare after all and thus can cause a lot of troubles for the (re)insurer.

The relative difference in *Var* between the two models is much smaller than in case of the *SLP* or its standard deviation. However, it is clear that the absolute differences of *Var* are much larger than the ones of the *SLP* or the standard deviation, which makes a factor like 0.17, see Table 6.10, still very significant. Therefore, the general conclusion is to consider both the relative and the absolute difference.

As usual in stochastic modeling, all the parameters which are involved in the model considered, have to be estimated. Replacing the unknown model parameters by their estimates (which are usually obtained from the data) will result in estimation errors. Just as with ignoring the dependence effect, it is too optimistic

to act as if the estimation errors are negligible, unless the number of observations is large. This topic, the effect of the estimation step, is exactly the issue which is addressed in Chapter 7. We start with discussing the observations and defining the estimators of the model parameters, which summary is presented in (7.1).

The influence of the estimators on the considered insurance products (measures) certainly depends on their behavior as a function of the parameters $\mu_C, \gamma_C, \mu_G, \gamma_G, \epsilon, \lambda$, which are going to be estimated, as well as on the accuracy of the estimators. For instance, if SLP is a flat function of the parameters $\mu_C, \gamma_C, \mu_G, \gamma_G, \epsilon, \lambda$ and the estimators are accurate, the small changes due to the estimation will have not much effect. So, these two points have to be considered: how fluctuate the quantities considered (VaR and SLP) and how accurate are the estimators. The first aspect is covered in Section 7.2. Tables 7.4 and 7.2 give some impression about the order of magnitude of the model parameters by considering their partial derivatives at $(\mu_{C0}, \gamma_{C0}, \mu_{L0}, \gamma_{L0}, \epsilon_0, \lambda_0) = (100000, 0.7, 15, 0.8, 0.03, 400)$, assuming that C and L follow (different) Gamma-distributions. The second aspect is covered in Section 7.3. There we discuss the asymptotic behavior of the estimators which later is used in the analysis of the estimation effect and construction of the confidence bounds for the risk measures considered. The asymptotics are considered w.r.t. λ , which is assumed to tend to infinity. That seems to be the natural way, because λ is the total expected number of claims, that is the expected number of observations. The other parameters are assumed to be fixed. To measure the estimation effect we calculate confidence intervals for the risk measures considered. The bounds of these intervals are considered to offer protection against estimation. In Figures 7.1-7.3 some examples are presented of the extra amount due to the protection against estimation and the effect of dependence on the basis of SLP . These figures affirm that ignoring dependence may lead to very large errors (up to 4300% in Figure 7.3). But also the additional step due to protection against estimation is large (up to 138% in Figure 7.3). A numerical example illustrates this. Consider again the example with true values of the parameters being equal to $(\mu_{C0}, \gamma_{C0}, \mu_{G0}, \gamma_{G0}, \epsilon_0, \lambda_0) = (100000, 0.7, 15, 0.8, 0.03, 400)$. By taking $k = 1$ we have $a = \mu_S + \sigma_{S_I} = 4 \times 10^7 + 2561250 = 42561250$. If we ignore the dependence we get $SLP_{app}(100000, 0.7, 15, 0.8, 0.03, 400) = 211277$. If we take into account the dependence without protection against estimation we get $SLP_{app}(100000, 0.7, 15, 0.8, 0.03, 400) = 382006$. If we add the protection (taking $\hat{\mu}_G = \mu_{G0} = 15$, $\hat{\gamma}_G = \gamma_{G0} = 0.8$, $\hat{\epsilon} = \epsilon_0 = 0.03$, $\hat{\tau} = \sqrt{\tau^2}$) we get $UB(0.1) = 476596$.

The estimation effect on VaR is analyzed in the similar way. In Figures 7.6-7.8 some examples are presented of the extra amount due to the protection against estimation and the effect of dependence on the basis of VaR . These figures affirm that ignoring dependence may lead to very large absolute errors. But also the additional step due to protection against estimation is large. Sometimes even larger than the dependence effect. Figure 7.6 clearly illustrates this situation.

The dependence effect reaches only 2% in the tail while the estimation effect reaches 8% under the same scale. Such situations are not in line with the *SLP* case where the dependence effect is usually much larger than the estimation effect. This can be explained by differences in fluctuation of the underlying risk measures. It is obvious that the estimated parameters are the same in both cases. Therefore, the main role here play the coefficients c_1, \dots, c_6 , which contain partial derivatives of the risk measures w.r.t. estimated parameters. These coefficients are different for *VaR* and *SLP* in absolute and even in relative sense, comparing to the values of the risk measures. For instance $c_6^2 \{ \mu_{C0} \epsilon_0 (1 + \gamma_{G0}^2) + 1 - \epsilon_0 \} = 2.84$ in case of *SLP* (which is only 0.00097% of the underlying *SLP* value) and equals 322550 in case of *VaR* (which is 0.67% of the underlying *VaR* value). Clearly, such a great relative difference stems from c_6 , which means that the fluctuation of *VaR* w.r.t. λ is much larger than the fluctuation of *SLP*. A similar (maybe less extreme) situation exists with the remaining parameters. This explanation gives an idea why the estimation impact is so (relatively) large in the case of *VaR*, compared to *SLP* and shows that it is useful to study the behavior of the risk measures and the behavior of the estimators separately, as has been done in Sections 7.2 and 7.3.

To conclude we present the following results for the example with the true values of the parameters being equal to $(\mu_{C0}, \gamma_{C0}, \mu_{G0}, \gamma_{G0}, \epsilon_0, \lambda_0) = (100000, 0.7, 15, 0.8, 0.03, 400)$, $p = 0.99$ and C and L each having a (different) Gamma distribution. If we ignore the dependence, we get $VaRapp = 45825083$. If we take into account the dependence without protection against estimation we get $VaRapp(100000, 0.7, 15, 0.8, 0.03, 400) = 47861120$. If we add the protection (taking $\hat{\epsilon} = 0.03$, $\hat{\lambda} = 400$ and $\hat{\tau}_2 = \tau_2$) we get $UB(0.1) = 51016282$. Again, the protection and the dependence steps are very huge in absolute sense. The relative difference is not large, about 4% for the dependence step and about 6% for the protection. In total the result is an increase of about 10%. Compared to *SLP* this is not large. For high retentions such an increase for *SLP* can reach hundreds of percents. However, we should not forget that high retentions in the *SLP* calculation produce low values of *SLP*. The situation with *VaR* is the other way around. The values of *VaR* increase with p and therefore significance of the dependence effect and the estimation effect in terms of relative difference should not be considered on the same scale as with *SLP*.

Chapter 8 deals with the real data obtained from Bert Teeuwen (Nationale Nederlanden), one of the members of the users committee of our STW project. The main goal is the data analysis, as illustration of the model implementation: in particular finding ways to determine dependencies within the data and illustration of the dependence and estimation effects.

As soon as we assume the same distribution for the simple and special claims sizes, the number of claims becomes the major aspect through which dependence is introduced. As was already mentioned, the total number of claims can be divi-

ded into two groups: simple (which is modeled by the random variable N) and special (which is modeled as a sum $\sum_{k=1}^H G_k$, where H is the number of dependence groups and G_k are the group sizes). Therefore, the significant challenge is to recognize and distinguish these two types of claims.

The number of simple claims is determined as follows. First of all we construct the normal Q-Q plot of the data considered. Denoting the total number of observations by n , we create a plot with normal quantiles $x_i = \Phi^{-1}(i/(n+1))$ on the x -axis and with the ordered data points Y_i on the y -axis. As a result we get a sequence of observations $(x_1, y_1), \dots, (x_n, y_n)$. The next step is to fit the simple linear regression to these data points. Since most points are simple and the whole sample is ordered in ascending order, we can hope that the regression line mostly describes the simple data points. Therefore, the next step is to construct the one-sided 95% confidence interval for the mean response when $x = x_0$. This will define the data points which do not belong to the expected majority, as described by the regression. All the points which are above the confidence bound will be considered as special data points. The next step is to fit the above described linear regression, excluding the points which were determined as special during the previous step. The construction of the confidence bound for that regression can single out more special points, since now the sample of the data considered contains more simple points compared to the special ones and the regression fits them better. These points also have to be excluded from the simple points. The process is continued until all the special points are excluded, i.e. until exclusion of the special points and construction of a new confidence bound do not give any new special points. As a result we get a regression line which fits only the simple points. All the data points which are below the confidence bound are considered to be simple.

Having obtained the estimated distributions of the number of claims for simple groups, we can start distinguishing the numbers of special claims during the suspicious groups. The point is that each suspicious group can contain both simple and special claims. It is impossible to distinguish these using the claim size, since we assume that the simple and special claim sizes follow the same distribution. Therefore, to get the number of special claims in the suspicious group, we generate a total number of claims in a simple group for each of the suspicious ones and subtract this from the total number of claims.

Table 8.9 presents the estimation results of time and space grouping. We see that some estimates are larger than we expected at the beginning of the analysis. Moreover, they are out of the region of interest which was introduced in order to check the accuracy of the approximations. The estimate of λ is 7036 instead of 400, which was assumed before. However, this is not a problem since it was shown that the accuracy of the approximations becomes better when λ increases. Here the estimate of λ is 17.59 times larger than 400. For this reason we do not much worry about the larger values of the estimates of μ_G and ϵ , which earlier

were assumed to be 20 and 0.03. We also note that two different ways of grouping give different estimates of the expected number of groups and the group sizes. Basically it means that the values of SLP and Var depends on the grouping strategy and (as illustrated in Figures 8.10 and 8.11) the difference can be quite large. This is not a very pleasant result since this makes it difficult to give a clear practical advice. It is clear that the reason for this is hidden in the lack of the information about the dependence in the data. Ideally (as was mentioned in the earlier parts of the thesis), the data should have a clear structure and in particular, allow identification of special causes and resulting claims. However, that was not the case in our situation which forced us to develop methods for the identification of the special claims. Hence, this result serves as yet another argument to strive for satisfactory data, from which results can be obtained in a more or less confident way.

One more remark should be made about the fact that the distribution of the number of claims which was estimated from the data turned out to be Normal, not Poisson (as was assumed beforehand). In Chapter 7 it is assumed that both N_t and H_t ($t = 1, \dots, u$) follow a Poisson distribution. However, here it turned out that N_t is fitted well by a normal distribution. In spite of this fact the estimators which were introduced in Section 7.1 can still be used here since these are 'natural' estimators of the parameters. However, this fact can cause more trouble while calculating the risk measures. In principle Model 4 (as well as Models 1-3) cannot be applied immediately. To match the data with the basic assumption of Model 4, we should change the model assumption of the number of claims. Such a replacement is more or less straightforward and does not introduce any new features to the model. However, it means a huge amount of technical work, starting with changing the formulas for the cumulants and finishing with rewriting Chapter 7. All of this program can be done along the same lines as for Model 4. We are not doing it here, but this is a useful suggestion for the extension of the model. Note that when changing the assumptions of Model 4, we should also change the assumptions of Model 1, in order to keep the models on the same scale. However, the Poisson assumption for the number of claims is very classical in the insurance world. That explains why Poisson was taken as a basic assumption for the number of claims and that is why such an approach is still the one of main interest. After all, the main target of interest is not the aggregated sum itself and even not the underlying risk measures. In the present thesis we are basically concentrated on the dependence and estimation effects, which are relative quantities. Possibly, the change of assumption for the number of claims may have a strong influence on the aggregated sum and the underlying risk measures under both models, but we may hope that the dependence and estimation effects will be similar for both assumptions.

As concerns the estimation of the distribution of C , the results here are not perfect as well. The lognormal distribution showed the best fit while compared

with Gamma and Inverse Gaussian, but generally speaking the fit is not perfect (see Section 8.3). However, the lognormal distribution is a very common assumption for the claim sizes in the insurance world. Therefore it was considered as satisfactory and for the parameter estimation a parametric approach was used. Nevertheless, we can avoid the distributional assumption on C by considering a nonparametric approach. While believing in our approximations, we need only $\mu_S, \sigma_S, \kappa_{3S}, \kappa_{4S}$ to get the values of the risk measures considered. Using Lemma 3.3 we can express these quantities in terms of the first four moments of C and L . As soon as we have many observations for C we can apply nonparametric estimators. However, for L we still need to make a distributional assumption (Gamma), because we do not have many observations for estimation in the case of special claims. In that case the calculation method can be considered as semi-parametric. This approach was not considered in the present thesis and hence is presented here as a recommendation for future research and development.

Appendix A

Convolutions calculation algorithms

In this section we will describe in detail the most important calculation difficulties (and their solutions) which occurred when applying the convolution formulas presented in Chapter 4. All examples and remarks will be given on the basis of the C++ programming language. There are several reasons to use C++ for all the calculations.

The main one is computing time. The method of convolutions is based on summation, where each summand itself has a very complicated structure. A huge number of operations has to be performed for each summand separately, which will certainly be very time consuming in the higher level languages (software) like MAPLE or Visual Basic. Even MATLAB did not offer fast calculation speed. To give some impression on C++ calculation speed, the same *SLP* was computed in MAPLE and C++. The MAPLE calculation time was 7 hours while C++ did the same job in 10 seconds. In some situations MAPLE did not come up with an answer within 3 days while C++ managed to do it in 2 minutes. MAPLE has the clear advantage of having a large library of numerical algorithms, but there are no references on which algorithms are used precisely, which also hides the information about the calculation accuracy. In C++ all numerical algorithms used were programmed by ourselves. This was a large and very time consuming job, which unfortunately remains hidden from the reader. But it allows us to control the calculation process by specifying the calculation accuracy, which sometimes can be very important.

Another reason to use the C++ language is the possibility of independent application. If the program is written in MAPLE or MATLAB, the person who wants to use the program, has to have this software. C++ allows to create a separate application which can be used independently or as a function in another software (like Excel). This is important for the non-mathematical organizations

like insurance companies which probably do not have the above mentioned mathematical software. Basically it would be very nice to have C++ speed and Excel "user friendly" appearance in one place.

The structure of this section will be similar to the one used in the body of the thesis. Under the different model assumptions, we have different convolution formulas which have to be programmed in C++. Each of the following sections contains the calculation algorithms under one of the various model assumptions.

A.1 Situation when $C \sim \text{Gamma}$ and $L \sim \text{Gamma}$

This part contains the application algorithm for the formula (A.1) ((4.3) in Chapter 4).

$$\sum_{w=1}^{\infty} ST(w) = \sum_{w=1}^{\infty} P(W = w) \left\{ \frac{w\alpha_c}{\beta_c} [1 - F_\gamma(a; w\alpha_c + 1, \beta_c)] - a[1 - F_\gamma(a; w\alpha_c, \beta_c)] \right\} \quad (\text{A.1})$$

Here F_γ stands for the gamma cumulative distribution function and $P(W = w)$ is defined as (cf. (4.4) in Chapter 4).

$$P(W = w) = (\lambda(1 - \epsilon))^w e^{-\lambda(\epsilon/\mu_G + 1 - \epsilon)} \left(\frac{1}{w!} + \sum_{h=1}^{\infty} \sum_{j=0}^w \left[\frac{(\epsilon\lambda/\mu_G)^h}{h!} \times \frac{\beta_L^{h\alpha_L} \Gamma(h\alpha_L + j)}{(\lambda(1 - \epsilon))^j (w - j)! \Gamma(h\alpha_L) j! (\beta_L + 1)^{h\alpha_L + j}} \right] \right), \quad (\text{A.2})$$

The main algorithm utilizes several numerical routines which have to be defined beforehand. The very first one is the numerical calculation of the incomplete gamma function, which is defined as

$$\Gamma(p, x) = \frac{1}{\Gamma(p)} \int_0^x t^{p-1} e^{-t} dt.$$

This part will be used in the calculation of the gamma cdf which can be written as

$$F_\gamma(x; \alpha, \beta) = \Gamma(\alpha, \beta x).$$

The algorithm itself together with the description of the method precision, can be found in Shea [1988].

Another numerical routine which will be used in the main algorithm, is the calculation of the natural logarithm of the gamma function. This part will be used

for the large factorials (which appear in (A.2)) calculation. The gamma function is defined as

$$\Gamma(x) = \int_0^{\infty} t^{x-1} e^{-t} dt.$$

When $x > 12$, the very accurate algorithm for the calculation of $\log(\Gamma(x))$ is described in Hart et al. [1968]. When $x < 12$ it is better to use the algorithm of Cody and Hillstrom [1967].

Once we have all the numerical algorithms we need, we can start implementing formula (A.1) in practice. The very first place which ruins the direct application illusion is the summation index w . In the original convolution formula summation has to be performed up to infinity. Only in this case the result will be the true value. Otherwise (if the summation index is less than infinity) the result is only an approximation.

After analyzing the convolution formulas we found out that the convergence rate of such summations is quite fast and the summands are obviously decreasing while tending away from $E[W]$. The numerical analysis showed that in case the user is not interested in the decimal precision, the summation procedure can stop as soon as the summand value $ST(w)$ is less than 0.01. So, in the main algorithm there is no fixed value for the summation index w . The summation will be performed as long as the summand value is larger than 0.01.

According to this strategy the first task is to find those w for which $ST(w) > 0.01$. We know that $E(W) = \lambda$, which usually is equal to 400). Hence, 400 is taken as the starting value for w . In case $ST(w) < 0.01$, we check whether $ST(w+1) > 0.01$ or $ST(w-1) > 0.01$. One should continue this procedure until a value w for which $ST(w) > 0.01$ is found.

As soon as $ST(w) > 0.01$ for some w , we can start the main summation procedure. Summation will be performed both left and right. Basically we define $w_l = w$ and $w_r = w + 1$ (continue with $w_l = w_l - 1, w_r = w_r + 1$) and sum all the $ST(w_l), ST(w_r)$ values which are larger than 0.01. This is the convenient way to perform such kind of summation.

Next follow several other important aspects about the calculation of $ST(w)$. Consider the second part of the ST -term, which is defined as

$$ST_R(w) = \frac{w\alpha_c}{\beta_c} [1 - F_\gamma(a; w\alpha_c + 1, \beta_c)] - a[1 - F_\gamma(a; w\alpha_c, \beta_c)].$$

Here we have to deal with the calculation of the gamma cdf. For this we utilize the algorithm of Shea [1988]. The precision aspect is very important here since the value $1 - F_\gamma$ has to be multiplied by the numbers a and $w\alpha_c/\beta_c$. Quite often we deal with the tail of the gamma distribution. The values of the cdf there are very close to 1 and in case the calculation precision of F_γ is low, it will imply that the value $1 - F_\gamma$ can become 0 too early. For example, when 0.999999 is already assumed to be 1, we have $a(1 - F_\gamma) = 0$, whatever the value of a . But in our situation a can be of the order $10^6 - 10^{10}$, which means that F_γ has to be not rounded to 1 at least till 0.9999999999. But this is only an example. In general

the values of a and w do not have an upper bound and thus in the main algorithm the precision of F_γ has to be defined for each w and a . The algorithm of Shea [1988] allows to do this by changing the value of parameter "tol".

The first part of ST involves the calculation of $P(W = w)$ which is defined in (A.2). This is the most difficult part of the algorithm. Clearly, the formula has a quite complicated structure and it also contains an infinity summation "problem". Before defining the limits for the summation index h , several difficulties have to be mentioned which will be faced with while trying to calculate the summands directly. Here we have to deal with factorials like $j!$ and gamma functions like $\Gamma(h\alpha_L + j)$. The problem is that j can reach large values like 400 or 600, which makes it necessary to deal with numbers like 400! or 600! (as well as $\Gamma(400)$ which can be written as 399!). The largest number which can be handled in C++ is about 1.7×10^{308} . The number 400! is much larger than that and hence it is necessary to overcome this problem. The final result will be a "simple" number, since the huge numbers mentioned occur both in nominators and denominators, but it is practically not feasible to get the answer directly because of the problems mentioned above.

One method which can be used to solve the problem is the utilization of the natural logarithm of gamma function. Each factorial can be written as a gamma function (according to the identity $\Gamma(n) = (n-1)!$) and each gamma function can be written as an exponent of the gamma natural logarithm. Using this strategy we can rewrite (A.2) as

$$\begin{aligned}
 P(W = w) = & \exp \left(-\lambda(\epsilon/\mu_G + 1 - \epsilon) + w \log(\lambda(1 - \epsilon)) \right) \\
 & \times \sum_{h=1}^{\infty} \sum_{j=0}^w \left[\exp \left(h \log(\epsilon\lambda/\mu_G) + h\alpha_L \log(\beta_L) + \log(\Gamma(h\alpha_L + j)) \right. \right. \\
 & \quad \left. \left. - \log(\Gamma(h+1)) - \log(\Gamma(h\alpha_L)) - \log(\Gamma(j+1)) \right. \right. \\
 & \quad \left. \left. - (h\alpha_L + j) \log(\beta_L + 1) - j \log(\lambda(1 - \epsilon)) - \log(\Gamma(w - j + 1)) \right) \right] \\
 & + \exp \left(-\lambda(\epsilon/\mu_G + 1 - \epsilon) + w \log(\lambda(1 - \epsilon)) - \log(\Gamma(w + 1)) \right). \quad (\text{A.3})
 \end{aligned}$$

Now, instead of calculating large factorials separately, we calculate natural logarithms of gamma functions, which are certainly not that extreme.

Having the summands calculation method, we switch back to the summation index h . We know that h represents the number of groups, hence it is clearly not necessary to perform summation up to infinity. The typical assumptions for the parameters of the model are $\mu_G = 20$, $\lambda = 400$ and $\epsilon = 0.03$, which gives us $E[H] = \epsilon\lambda/\mu_G = 0.6$. In that case the probability that a realization h will be large is very small. Hence we can easily conclude that the summation of the h series will

not have to involve too many terms. However, it remains to find a specific limit for the summation index h . To obtain such limits, we return to the general appearance of (A.1). We see that $P(W = w)$ there is multiplied by $ST_R(w)$. Both factors are a function of w (and actually of many other parameters), but the most important fact is that $P(W = w) \in (0, 1)$, while $ST_R(w)$ can be quite large (all depends on the underlying parameters and the precision we would like to achieve). When $ST_R(w)$ is large (let say of order 10^{10}), the accuracy of $P(W = w)$ has to be at least of the same (reversed) order (of order 10^{-10}). This can be achieved by adding summands one by one, increasing the values of h . Numerical analysis showed that such convergence is quite stable. This means that the decimal places stabilize one by one and we can stop the summation as soon as the difference between two steps is less than the required precision (less than 10^{-10}). Note that in case the needed precision is not reached, we can get large errors already for a single summand. As an example we compare $P(W = w) \times ST_R(w) = 0.123456789 \times 10^{10} = 1234567890$ and $P(W = w) \times ST_R(w) = 0.123879889 \times 10^{10} = 1238798890$. The absolute difference for the single w is 4231000, which is definitely a very significant error. In most cases the situations explained above are not so extreme, but the total summation can be very inaccurate if the precision is not controlled at all.

A.2 Situation when $C \sim \text{Gamma}$ and $L \sim \text{IG}$

In this section we discuss the application of (4.8). All the suggestions and recommendations which were presented in the previous section in order to get the final result, are applicable here as well. The difficulties with the application of (4.8) are similar to the ones in case of (A.2). These are summation to infinity and huge numbers. The problem of infinite summation was already discussed in the previous section, therefore, the main issue here is the calculation of the large factorials, which (as in the previous section) will be solved by utilization of the natural logarithm of the gamma function. Each factorial can be written as a gamma function (according to the identity $\Gamma(n) = (n - 1)!$) and each gamma function can be written as an exponent of the gamma natural logarithm. Using this strategy we can rewrite (4.8) as

$$\begin{aligned}
 P(W = w) &= \exp(w \log(\lambda(1 - \epsilon)) - \lambda(\epsilon/\mu_G + 1 - \epsilon)) \\
 &\times \left(\exp(-\log(\Gamma(w + 1))) + \sum_{h=1}^{\infty} \sum_{j=0}^w \sum_{k=0}^{j-1} \exp\left(h \log(\epsilon\lambda/\mu_G) \right. \right. \\
 &+ k \log(\beta_L) + j \log(h\alpha_L) + \log(\Gamma(j + k)) \\
 &- (h\alpha_L/\beta_L)\{(1 + 2\beta_L)^{1/2} - 1\} - (j + k) \log(1 + 2\beta_L) / 2 \\
 &\left. \left. - \log(\Gamma(h + 1)) - k \log(2h\alpha_L) - \log(\Gamma(j + 1)) - \log(\Gamma(j - k)) \right) \right)
 \end{aligned}$$

$$\left. - \log(\Gamma(k+1)) - j \log(\lambda(1-\epsilon)) - \log(\Gamma(w-j+1)) \right)$$

which solves the 'huge numbers' problem.

A.3 Situation when $C \sim IG$ and $L \sim Gamma$

In this section we discuss the calculation difficulties when the model assumptions are $C \sim IG$ and $L \sim Gamma$. Most of the problems arise from the calculation of $P(W = w)$. However, this part was already discussed in Section A1. One more tricky point is hidden in the expression involving $e^{x_1} \Phi(-x_2)$ (see (4.13)) and it again concerns products of very small and very large numbers. For modern computers $e^{x_1} = \infty$ if $x_1 > \log(1.7 \times 10^{308})$. The situation with $\Phi(-x_2)$ is similar. If x_2 is large (say 100), most of the numerical algorithms give $\Phi(-x_2) = 0$. The problem occurs while multiplying these two numbers when $1/\exp(x_1)$ is close to $\Phi(-x_2)$. If $x_1 = 3010$ and x_2 such that $\Phi(-x_2) \approx e^{-3000}$ it is clear that $e^{x_1} \Phi(-x_2)$ is a quite large number. But it is practically impossible to get this number accurately by applying direct multiplication. However, we write it as

$$e^{x_1} \Phi(-x_2) = \frac{e^{x_1}}{\sqrt{2\pi}} \int_{x_2}^{\infty} e^{-u^2/2} du = \frac{1}{\sqrt{2\pi}} \int_{x_2}^{\infty} e^{-(u^2-2x_1)/2} du. \quad (\text{A.4})$$

Hence, instead of calculating $e^{x_1} \Phi(-x_2)$, we calculate (A.4) by using standard numerical algorithms (i.e. one of the quadrature rules). This solves the multiplication problem.

Bibliography

- W. Albers. Stop-loss premium under dependence. *Insurance: Mathematics and Economics*, 24:173–185, 1999.
- W. Albers and W. C. M. Kallenberg. Estimation effects on stop-loss premiums under dependence. Technical Report 1837, University of Twente, Enschede, May 2007.
- W. Albers, W. C. M. Kallenberg, and V. Lukocius. A flexible model for actuarial risks under dependence. *Scandinavian Actuarial Journal*, February 2008.
- S. Bouma. Risk management in the insurance industry and solvency ii. Technical report, Capgemini’s Compliance and Risk Management Centre of Excellence, 2006.
- Y. P. Chaubey, J. Garrido, and S. Trudeau. On the computation of aggregate claims distributions: some new approximations. *Insurance: Mathematics and Economics*, 23(3):215–230, 1998.
- W. Cody and K. Hillstrom. Chebyshev approximations for the natural logarithm of the gamma function. *Mathematics of Computation*, 21:198–203, 1967.
- J. Dhaene, M. Denuit, M.J. Goovaerts, R. Kaas, and D. Vyncke. The concept of comonotonicity in actuarial science and finance: theory. *Insurance: Mathematics and Economics*, 31:3–33, 2002a.
- J. Dhaene, M. Denuit, M.J. Goovaerts, R. Kaas, and D. Vyncke. The concept of comonotonicity in actuarial science and finance: applications. *Insurance: Mathematics and Economics*, 31:133–161, 2002b.
- J. Dhaene, S. Vanduffel, M. J. Goovaerts, R. Kaas, Q. Tang, and D. Vyncke. Risk measures and comonotonicity: A review. *Stochastic Models*, 22:573 – 606, 2006.
- P. Embrechts, F. Lindskog, and A.J. McNeil. *Modelling dependence with copulas and applications to risk management*. In *Handbook of heavy tailed distributions in finance*, chapter 8. Amsterdam: Elsevier/North-Holland, 2003.

- Hart, W. Cheney, C. Lawson, Maehly, C. Mesztenyi, J. Rice, Thacher, and Witzgall. *Computer Approximations*. Wiley and sons, New York, 1968.
- R. M. Hayne. The common shock model for correlated insurance losses. *Variance*, 1:40–52, 2007.
- H. Joe. *Multivariate Models and Dependence Concepts*. London: Chapman and Hall, 1997.
- R. Kaas, M. Goovaerts, J. Dhaene, and M. Denuit. *Modern Actuarial Risk Theory*. Kluwer, Boston, 2001.
- R. J. A. Laeven. *Essays on risk measures and stochastic dependence*. PhD thesis, University of Amsterdam, 2005.
- F. Lindskog. Modeling dependence with copulas and applications to risk management. Master's thesis, Department of Mathematics, ETH Zurich, 2000.
- V. Lukocius, W. Albers, and W. C. M. Kallenberg. Accuracy of approximations in actuarial overdispersion models. Technical Report 1842, University of Twente, Enschede, June 2007.
- R. B. Nelsen. *An introduction to copulas. Lecture Notes in Statistics*. New York: Springer Verlag, 1999.
- R. Reijnen. Nauwkeurigheid van benaderingen stoploss herverzekeringspremie. Master's thesis, University of Twente, Enschede, 2003.
- R. Reijnen, W. Albers, and W. C. M. Kallenberg. Approximations for stop-loss reinsurance premiums. *Insurance: Mathematics and Economics*, 36:273–250, 2005.
- K.K. Saha and S.R. Paul. Bias-corrected maximum likelihood estimator of the negative binomial dispersion parameters. *Biometrics*, (61):179–185, 2005.
- H. L. Seal. Approximation to risk theory's $f(x, t)$ by means of the gamma distribution. *ASTIN Bulletin*, 9:213–218, 1977.
- B. L. Shea. Chi-squared and incomplete gamma integral. *Applied Statistics*, 37 (3):466–473, 1988. Algorithm AS239.
- G. Willmot. Mixed compound poisson distributions. *ASTIN Bulletin*, 16:59–79, 1986.

Summary

The present thesis considers an advanced actuarial modeling technique which introduces dependence as a natural extension of existing independence models. The dependence structure considered here, together with classical ways of modeling insurance, is briefly introduced in Chapter 1. The basic assumption of our modeling technique is the possibility of special claims where at once a whole group has damage. Therefore, we distinguish between simple (or 'individual') claims and special (or 'group') claims. The (so called) dependence model has gone through several development steps. Starting with the independence model (Model 1) in three steps (called Model 2, 3 and 4) dependence is introduced in a more and more flexible and general way. The final Model 4 is the main target of investigation of the thesis. Chapter 1 briefly discusses all these development steps, introduces some risk measures, discusses the need of approximations and the influence of the model and of the approximation and estimation step on the risk measures considered. In the present research we mostly concentrate on the distribution of the aggregate sum of claims S , which usually is far from normally distributed, i.e. not symmetric and having a heavy tail. Therefore the risk measures which concern the tail behavior of the distributions involved, are of main interest. The net Stop-Loss premium (SLP), the variance of the Stop-Loss contract and the Value at Risk (VaR) are chosen as representative risk measures.

The form of dependence introduced in the thesis is quite new and differs considerably from structures available in literature. Therefore, the structure and the assumptions of the model have to be motivated. This aspect is covered in Chapter 2, which is devoted to an illustrative example of the proposed dependence structure. Chapter 3 describes Model 4 in more detail, presents its main characteristics, discusses the impact of the model parameters (see Section 3.3) and describes the region of interest for the values of the parameters involved (see Table 3.7).

Calculation aspects are covered in Chapter 4. There we present several calculation methods for the chosen risk measures. These are convolutions, approximations and simulations. The major calculation tool which is used while analyzing dependence and estimation effects is the approximation, being the most transparent one. The methods of convolutions and simulations are used in testing the accuracy of the approximations, which is the main topic of Chapter 5. Here we

carefully analyze the chosen approximations, introduce the retention level for SLP (and the variance of the Stop-Loss contract) and the representative probabilities for VaR . All decisions about the accuracy of the approximations are based on the criterion chosen, which is defined in (5.1) and (5.2). As a result, we present a rule of thumb for Model 4. The general recommendation is to use the Gamma-IG approximation for the SLP and the variance of the Stop-Loss contract and the IG approximation for the VaR whenever the aggregated sum S is modeled by Model 4. This rule is valid in the given region of parameters of interest, which is presented in Table 3.7. It is restricted to the distributions under consideration, which are presented in Table 3.2. The schematic overview of the rule of thumb for Model 4 is presented in Table 5.3.

In Chapter 6 we present an explicit example of the dependence effect. The example concerns a concrete line of the insurance business, the so-called workers compensation insurance. The underlying data of this branch of insurance perfectly fit our model (dependence structure is present) and, moreover, it was suggested by the users committee of our STW-project as a potential source of problems due to dependence. However, the needed structure is not explicitly present in the available data. Therefore, it only serves as a basis for the construction of the data set needed. Hence, real data are not used in this chapter, but compared to Chapter 2 (where all data were simulated), we go a step further here, since real data are the starting point. This contributes to the reality aspect of the example. Another goal of the chapter is to discuss the structure of the data which is needed for the implementation of Model 4. It turns out that the dependence effect can indeed be very large and that even small dependencies cannot be ignored.

As usual in stochastic modeling, all the parameters which are involved in the model considered, have to be estimated. Replacing the unknown model parameters by their estimates (which are obtained from observed data) will result in estimation errors. Just as with ignoring the dependence effect, it is too optimistic to act as if the estimation errors are negligible, unless the number of observations is large. This topic, the effect of the estimation step, is exactly the issue which is addressed in Chapter 7. We start with discussing the observations and defining the estimators of the model parameters, the summary of which is presented in (7.1). The influence of the estimators on the insurance products (measures) considered certainly depends on their behavior as a function of the parameters which are going to be estimated, as well as on the accuracy of the estimators. Therefore, we consider two points: how fluctuate the quantities considered (VaR and SLP) and how accurate are the estimators. The first aspect is covered in Section 7.2 where we give some impression about the order of magnitude of the fluctuation of the risk measures by considering their partial derivatives w.r.t. the model parameters. The second aspect is covered in Section 7.3. There we discuss the asymptotic behavior of the estimators which later is used in the analysis of the estimation effect and construction of the confidence bounds for the risk measures considered. The estimation error is dominated by the part of the parameters related to the special claims, because by their nature we do not have

many observations of them. Although restricted to a few parameters, it is seen that the estimation error may be quite substantial.

Chapter 8 deals with the real data obtained through one of the members of the users committee of our STW-project. The main goal is to perform a data analysis, as an illustration of the model implementation: in particular finding ways to determine dependencies within the data and illustration of the dependence and estimation effects. As soon as we assume the same distribution for the simple and special claims sizes, the number of claims becomes the major aspect through which dependence is introduced. Therefore, the significant challenge is to recognize and distinguish these two types of claims.

Our method is based on simple linear regression in the context of a QQ-plot e.g. for daily claims. Since days with special claims have much larger numbers of claims, the numbers of claims of these days will appear above the one-sided 95%-confidence interval associated with the fitted linear regression line in the QQ-plot. In a few iteration steps the special days are detected. As such special days also contain simple claims, the number of simple claims has to be subtracted to obtain the number of special claims. To test the method we performed a brief simulation using data from two different normal distributions. The results of the testing procedure were satisfactory to apply the method.

We finish the chapter with the illustration of the dependence and estimation effects on the basis of the estimated parameters. The results are in line with the main message of the thesis: the dependence should not be ignored. The same holds for the estimation effect.

The last chapter contains the review of the thesis together with recommendations for future development. It also contains several discussions of the methods and assumptions which are used in the thesis.

Samenvatting

Dit proefschrift bestudeert een geavanceerde actuariële modelleertechniek waarin afhankelijkheid als een natuurlijke uitbreiding van bestaande onafhankelijke modellen wordt geïntroduceerd. De afhankelijkheidsstructuur waar het hier om gaat, wordt kort geïntroduceerd in Hoofdstuk 1, samen met klassieke manieren om verzekeringen te modelleren. De basis aanname van onze modelleertechniek is de mogelijkheid van het vóórkomen van speciale claims waarbij een hele groep tegelijkertijd schade heeft. Daarom zullen we enkelvoudige (of ‘individuele’) claims en speciale (of ‘groep’) claims onderscheiden. Het voorgestelde afhankelijkheidsmodel heeft verschillende ontwikkelingsstappen doorlopen. Beginnend met het onafhankelijke model (Model 1) wordt in drie stappen (die Model 2,3 en 4 worden genoemd) afhankelijkheid ingevoerd op een steeds meer flexibele en algemene manier. Het uiteindelijke Model 4 is het hoofdonderwerp van het onderzoek in dit proefschrift. Hoofdstuk 1 bespreekt kort al deze ontwikkelingsstappen, introduceert enkele risicomaten, bespreekt de noodzaak van het gebruik van benaderingen en de invloeden van het model en van de benaderings- en schattingsstap op de beschouwde risicomaten. In het huidige onderzoek concentreren we ons vooral op de verdeling van de totale som S van de claims. Deze is meestal verre van normaal, bijvoorbeeld niet-symmetrisch en met een zware staart. Het gevolg is dat met name die risicomaten van belang zijn die betrekking hebben op de staart van de betrokken verdelingen. De netto Stop-Loss premie (SLP), de variantie van het Stop-Loss contract en de Value at Risk (Var) worden daarom gekozen als representatieve risicomaten.

De vorm van afhankelijkheid die in het proefschrift wordt geïntroduceerd, is volstrekt nieuw en verschilt aanzienlijk van structuren die in de literatuur beschikbaar zijn. Daarom dienen de structuur en de veronderstellingen van het model terdege te worden gemotiveerd. Dit aspect komt aan de orde in Hoofdstuk 2, dat gewijd is aan een illustratief voorbeeld van de voorgestelde afhankelijkheidsstructuur. Vervolgens beschrijft Hoofdstuk 3 Model 4 in meer detail, presenteert de belangrijkste kenmerken ervan, bespreekt de invloed van de model parameters (zie Paragraaf 3.3) en beschrijft het gebied waarin de parameterwaarden liggen die in de praktijk van belang zijn (zie Tabel 3.7).

Aspecten met betrekking tot de berekeningen worden behandeld in Hoofdstuk 4. Daar presenteren we verscheidene berekeningsmethoden voor de gekozen

risicomaten. Dit zijn convoluties, benaderingen en simulaties. De belangrijkste berekeningsmethode die gebruikt wordt bij het analyseren van afhankelijkheid en schattingseffecten, is die van het gebruik van benaderingen, gezien het feit dat deze het meest transparant is. De methodes van convolutie en simulatie worden gebruikt bij het testen van de nauwkeurigheid van de benaderingen, het belangrijkste onderwerp van Hoofdstuk 5. Hier analyseren we zorgvuldig de gekozen benaderingen, voeren we retentieniveaus voor de SLP (en de variantie van het Stop-Loss contract) in en representatieve kansen voor VaR . Alle beslissingen omtrent de nauwkeurigheid van de benaderingen worden gebaseerd op het gekozen criterium, dat gedefinieerd wordt in (5.1) en (5.2). Als gevolg hiervan kunnen we een vuistregel voor Model 4 formuleren. De algemene aanbeveling is het gebruiken van de Gamma-IG benadering voor de SLP en de variantie van het Stop-Loss contract en van de IG benadering voor de VaR , steeds onder de aanname dat de cumulatieve som S wordt gemodelleerd volgens Model 4. Deze regel is geldig in het eerder aangegeven parametergebied, zoals gepresenteerd in Tabel 3.7. Verder is deze regel ook beperkt tot de beschouwde verdelingen, zoals weergegeven in Tabel 3.2. Het schematisch overzicht van de vuistregel voor Model 4 wordt gepresenteerd in Tabel 5.3.

In Hoofdstuk 6 bespreken we een expliciet voorbeeld van het afhankelijkheids-effect. Dit voorbeeld heeft betrekking op een concreet onderdeel van het verzekeringsbedrijf, de zogeheten werknemers compensatie verzekering. De onderliggende data bij deze verzekeringstak passen perfect in ons model (afhankelijkheidsstructuur is aanwezig). Bovendien werd het voorgesteld vanuit de gebruikerscommissie van ons STW-project, als een mogelijke bron van problemen ten gevolge van afhankelijkheid. De benodigde structuur is echter niet expliciet aanwezig in de beschikbare gegevens. Derhalve dient het slechts als basis voor de constructie van de benodigde dataset. Dus in dit hoofdstuk worden nog steeds niet ‘echte’ data gebruikt, maar in vergelijking met Hoofdstuk 2 (waarin alle gegevens werden gesimuleerd), gaan we hier een stap verder, omdat reële data wel als uitgangspunt fungeren. Dit draagt bij aan het realiteitsgehalte van het voorbeeld. Een ander doel van het hoofdstuk is om uiteen te zetten wat de structuur moet zijn van data die geschikt zijn voor de uitvoering van Model 4. Het blijkt dat het effect van afhankelijkheid inderdaad erg groot kan zijn en dat zelfs kleine afhankelijkheden niet genegeerd kunnen worden.

Zoals gebruikelijk in stochastische modellen, moeten alle parameters die in het beschouwde model voorkomen, worden geschat. Vervangen van de onbekende parameters in het model door hun schattingen (die worden verkregen uit waargenomen data), zal resulteren in schattingsfouten. Net als bij het negeren van de afhankelijkheid, is het te optimistisch om te doen alsof die schattingsfouten te verwaarlozen zijn, tenzij het aantal waarnemingen groot is. Dit onderwerp, het effect van de schattingsstap, is precies hetgeen wordt behandeld in Hoofdstuk 7. We beginnen met het bespreken van de data en het definiëren van de schatters van de modelparameters, de samenvatting waarvan wordt gepresenteerd in (7.1). De invloed van de schatters op de verzekeringsproducten (maten) die we beschouwen,

hangt zowel af van hun gedrag als functie van de te schatten parameters, als van de nauwkeurigheid van de schatters zelf. Het eerste aspect wordt in Sectie 7.2 bekeken, waar we een indruk geven van de orde van grootte van de fluctuatie van de risicomaten door hun partiële afgeleiden m.b.t. de modelparameters te beschouwen. Het tweede aspect komt in Sectie 7.3 aan de orde. Daar bestuderen we het asymptotische gedrag van de schatters en gebruiken dit vervolgens bij de analyse van het schattingseffect en tevens bij de constructie van betrouwbaarheidsintervallen voor de beschouwde risicomaten. De schattingsfout wordt gedomineerd door dat deel van de parameters dat betrekking heeft op de speciale claims, doordat daar naar de aard der zaak relatief weinig waarnemingen voor beschikbaar zijn. Hoewel het zich zo dus eigenlijk tot enkele parameters beperkt, blijkt het schattingseffect toch heel substantieel te kunnen zijn.

Het laatste hoofdstuk houdt zich bezig met echte data, die zijn verkregen via een van de leden van de gebruikerscommissie van het STW-project. Het belangrijkste doel is het uitvoeren van een data-analyse als illustratie van de implementatie van het model: in het bijzonder het vinden van manieren om afhankelijkheden in de data vast te stellen en een illustratie van effecten van zowel de afhankelijkheid als de schattingsstap. Zodra we aannemen dat de gewone en de speciale claimhoogten dezelfde verdeling volgen, wordt het aantal schadegevallen het belangrijkste aspect in het model waardoor de afhankelijkheid wordt ingevoerd. Daarom is het een duidelijke uitdaging om deze beide soorten claims te herkennen en te onderscheiden.

Onze methode is gebaseerd op enkelvoudige lineaire regressie in de context van een QQ-plot bijvoorbeeld voor dagelijkse claims. Omdat dagen met speciale claims een veel hoger aantal claims tellen, zal dat aantal op dergelijke dagen uitkomen boven het eenzijdige 95%-betrouwbaarheidsinterval dat samenhangt met de gefitte regressie lijn in de QQ-plot. In een paar iteratiestappen worden de speciale claims zo gedetecteerd. Omdat zulke speciale dagen ook gewone claims bevatten, moeten deze van het totaal afgetrokken worden om de hoeveelheid speciale claims over te houden. Om de methode te testen, hebben we een kleine simulatie uitgevoerd met waarnemingen uit twee verschillende normale verdelingen. De resultaten zijn bevredigend, zodat de methode toegepast kan worden. We besluiten het hoofdstuk met het illustreren, op basis van de geschatte parameters, van de effecten van zowel de afhankelijkheid als het schatten.

Het laatste hoofdstuk biedt een overzicht van het proefschrift, tezamen met aanbevelingen voor toekomstige ontwikkelingen. Het bevat ook een discussie van de gebruikte methoden en aannamen in het proefschrift.

Acknowledgements

I would like to finish the book with the words of gratitude to those who helped and assisted me in accomplishing this project.

First of all, I would like to express my sincere appreciation to my research supervisors, Prof. Dr. Wim Albers and Dr. Wilbert Kallenberg, for their kind help, guidance, support and encouragement throughout the past four years. This book would be full of linguistic nonsenses, small mathematical errors and misprints without their careful assistance.

I am grateful to all the members of my promotion committee: Prof. Dr. W. Albers, Dr. W.C.M. Kallenberg, Prof. Dr. A.A. Stoorvogel, Prof. Dr. A. Bagchi, Prof. Dr. J.H.J. Einmahl, Prof. Dr. C.A.J. Klaassen and Prof. Dr. R.J.M.M. Does for their attentive reading of a draft of this thesis.

I would like to cordially thank the STW for the financial support of this project and all the members of the users committee: A.J.J van Dalen (Unive Verzekeringen), Drs. P.W.F. Hoogveld (Watson Wyatt Insurance Consulting B.V.), Drs. H. A. van Mourik (Achmea Schade), Dr. ir. W.J.H. Stortelder (De Nederlandsche Bank NV), B. Teeuwen (Nationale Nederlanden) and J. van der Vorst (Ernst & Young Actuarissen) for sharing important practical information related to the project and providing the data of different lines of insurance business.

I also thank all faculty members and employees of the SK and SST groups for their pleasant company. In particular, I would like to thank Thyra Kamphuis-Kuijpers for her help in administrative matters.

It is impossible to mention all the people which helped and supported me during these four years. However, I would like to point out several names which made my life easier during the hard periods. Hundreds liters of coffee were drunk with Jaroslav Krystul during these years. Special thank for his friendship and support. Sincere thank to Olga Krystul for her pleasant company and bright smile. Many thanks to Dima, Pasha, Masha, Emad and Kitam, Arianto, Zaher, Beata, Piotr, Vera, Alex, Maria and Maxim for the nice moments we have spent together. Additional thanks to Stas, Lina and my twin brother Sergej.

Finally, I reserve infinite thanks to my family and especially to my wife Olga, which was strong enough to stand and support me during the hard periods of the project. Hundreds of thanks to my daughter Arina, who gave me an infinite amount of inspiration. Heartfelt thanks to my parents, Jelena and Viktor, and

parents in law, Tatjana and Antanas, for their endless love. Separately I would like to thank my grandmother, Valentina Selkovnikova, and grandfather, Sergej Selkovnikov. They will live forever in my heart. I would like to dedicate this thesis to them.

Enschede, October 2008

Viktor Lukočius