The extended Pareto distribution as default loss model

by

REBECCA WIBORG

MASTER THESIS

for the degree

master of science

(Master i Modellering og dataanalyse)

The Faculty of Mathematics and Natural Sciences
University of Oslo

May 2013

Det matematisk- og naturvitenskapelige fakultet
Universitetet i Oslo

Abstract

Many families of distributions have been proposed to describe insurance losses. The process of finding the one which results in the best fit is time consuming. This thesis tries to tackle the issue of avoiding such analyses, so that the computer can handle it on its own. The approach is to introduce a flexible default loss model which results in a good fit for most historical data. The extended Pareto distribution, which comprises both heavy-tailed Pareto distributions and light-tailed Gamma distributions, is a natural choice. The true underlying distribution might not be part of the extended Pareto family, which leads to the necessity of defining a framework for maximum likelihood estimation under misspecification. In the beginning of this thesis such a framework is defined based on asymptotic theory. Then, the possibility of using the extended Pareto family as default loss model is examined. The potential reduction in error when the parametric family is further widened is also discussed.

Acknowledgements

This thesis completes my Master's degree in Modelling and Data Analysis, with the programme option Finance, Insurance and Risk, at the University of Oslo. The work which resulted in this thesis is done in the period from August 2011 to May 2013, and corresponds to 60 credits. Writing this thesis has been both interesting and challenging.

First and foremost I wish to thank my supervisor Erik Bølviken. Throughout this period Erik and I have had many interesting conversations. He has always been there to help me when I needed it, and pushed me to figure things out myself when he knew I could. For this I am sincerely grateful. Erik also provided me with the Fortran program "genlegendre1.f" which is mentioned in Appendix C.

I would also like to give a special thanks to fellow student, Finn Harald Opsjøn, for helping me with the simulations. To be able to run the simulations with the desired low level of simulation error, several measures had to be done in order to decrease the simulation time. Finn helped me a lot with this; I could not have done it without him.

I would also like to thank Mercer AS who has given me relevant work experience during the last two years. I appreciate that they have always been flexible, such that combining work and studies have been unproblematic. I would especially like to thank Tor-Eivind Høyland, Stein Erik Petersbakken and Ingrid Maudal.

During my years at the University of Oslo I have met some amazing people. I would like to take this opportunity to thank all my fellow students at the department of mathematics and my friends from back home. You have always believed in me and been great friends and supporters throughout my time as a student. I would also like to thank everyone who has taken the time to help me point out typos.

Lastly, I would like to give a sincere thanks to my parents, my two wonderful sisters and the rest of my family. My family means a lot to me, and they have always been there to support and encourage me. I would also like to give a special thanks to my dear boyfriend Christoffer, who has helped and supported me every step of the way.

Rebecca Wiborg
May 2013, Oslo

Contents

Chapter 1

Introduction

When trying to describe a set of data we have to choose a family of distributions with a density function, say $f_\theta(x)$, where θ is a vector of parameters of length p. Usually we spend time trying to find a parametric family which results in the best fit of the data. In this thesis the possibility of using the extended Pareto distribution as a default model for claim sizes in non-life insurance is studied. That is, $f_\theta(x)$ is always chosen to be the extended Pareto density without spending time studying the historical data. The extended Pareto distribution is introduced in Chapter 3, and its vector of parameters is $\theta = (\alpha, \beta, \theta)$. When the true density function, say g, is outside the parametric family, there is still asymptotic estimation theory available, this is described below.

Within the parameter space there is a vector of parameters θ_0 which minimizes the Kullback-Leibler distance. The Kullback-Leibler distance is a measure of the difference between two probability distributions which is defined in Chapter 2. In Chapter 2, it is also argued that the way θ_0 is defined implies that it is the vector of parameters which makes $f_\theta(x)$ as close to $g(x)$ as possible. Peter J. Huber proved that the maximum likelihood estimator $\hat{\theta}$ is consistent even when we do not assume that the true distribution of the data which defines the maximum likelihood estimators is part of the parametric family, see Huber [6]. Hence, $\hat{\theta} \to \theta_0$ as the number of observations n increases, also when $X_i \nsim f_\theta(x)$. There is a heuristic argument of this in Chapter 2, whereas a proof with precise mathematics and conditions is given in [6].

Let Ψ be some functional on f_θ or g describing some feature of the risk variable, e.g. the quantile or the percentile functional. Because companies are usually more interested in the error of $\Psi(\hat{\theta}) = \Psi(f_{\hat{\theta}})$ than $f_{\hat{\theta}}$ itself, error in $\Psi(\hat{\theta})$ is the main focus of this thesis. Total error is the difference between the actual and the estimated value of the functional. It is defined as

$$\Psi(\hat{\theta}) - \Psi(g) = \Psi(\hat{\theta}) - \Psi(\theta_0) + \Psi(\theta_0) - \Psi(g), \tag{1.0.1}$$

where $\Psi(\hat{\theta}) - \Psi(\theta_0)$ is random error due to estimation. The second part, $\Psi(\theta_0) - \Psi(g)$ is the bias, i.e. systematic error that occurs because the underlying distribution of the observations is outside the theoretical distribution family. In Chapter 2, the asymptotic properties of $\sqrt{n}(\hat{\theta} - \theta_0)$ are derived for a general default distribution f_θ and a general underlying distribution g. At the end of the chapter, these properties are used to derive asymptotic theory evolving total error, $\Psi(\hat{\theta}) - \Psi(g)$.

In Chapter 4 numerical methods necessary in order to access the appropriateness of the extended Pareto distribution as default loss model are given and illustrated. When the true distribution g is known, θ_0 is found by minimizing the Kullback-Leibler distance and $\hat{\theta}$ is esti-

mated. The vector of estimates $\hat{\theta}$ can be found by means of maximum likelihood with respect to data $X_1, ..., X_n$, drawn randomly from the true distribution g. It can also be found by means of the delta method which is introduced in Chapter 2. Both methods are illustrated in Chapter 4. The functionals $\Psi(g), \Psi(\theta_0)$ and $\Psi(\hat{\theta})$ are calculated by using Monte Carlo simulations. Depending on Ψ there might also exist other procedures. However, in this thesis Monte Carlo simulation is the approach used to calculate the functional Ψ on some density function g or f_θ.

Thus, if the true underlying distribution g is known, there are procedures available that make it possible to study systematic and estimation error separately. In Chapter 5, the different error terms are studied for four underlying distributions g. By studying the error terms for various underlying assumptions, it is possible to deduce how well the extended Pareto distribution works as default loss model. In Chapter 6, a fourth parameter λ is included. This results in a more flexible default distribution and reduces systematic error.

When applying Monte Carlo simulations there is a third type of error, namely Monte Carlo error. By increasing the number of Monte Carlo simulations m to a sufficient size, the simulation error is negligible. This third type of error is therefore not taken into account when the different types of error is studied.

Chapter 2

Maximum likelihood theory

As mentioned in Chapter 1, asymptotic estimation theory is still available when the true density function g is outside the parametric family. In this chapter, theory regarding the maximum likelihood estimator $\hat{\theta}$ under misspecification is derived. In Chapter 1 it is stated that θ_0 is the vector of parameters which minimizes the Kullback-Leibler distance. In Section 2.1, this measure of distance is more precisely defined, and methods for determining θ_0 are given. In Section 2.2 and Section 2.3, heuristic arguments are used to show that $\sqrt{n}(\hat{\theta}-\theta_0)$ is normally distributed. This is first shown in the one-parameter situation and then in the multi-parameter situation. The mean is always zero, but the standard deviation depends on whether the correct distribution g is a part of the parametric family or not. In Section 2.4 these results are utilized to find the mean, the standard deviation and the distribution of total error $\Psi(\hat{\theta}) - \Psi(g)$. Section 2.4 also includes a discussion of the expected value of the square of total error.

2.1 The Kullback-Leibler distance

The Kullback-Leibler distance is a non-symmetric measure of difference between two probability distributions. In the setting of this thesis it is of interest to measure the difference between the theoretical distribution f_θ and the true underlying distribution g. Thus, by minimizing the Kullback-Leibler distance, it is possible to find the set of parameters θ_0 which makes the distance between $g(x)$ and $f_{\theta_0}(x)$ as small as possible. In order to split total error into a random and systematic part, the best achievable density function $f_{\theta_0}(x)$ is required. If the underlying distribution is a part of the parametric family, f_{θ_0} is arbitrary close to the true distribution g. Then there is no systematic error, and f_{θ_0} can be regarded as the true density function.

For two continuous density distributions f_θ and g the Kullback-Leibler distance is defined as an integral or as an expectation,

$$D_{\mathrm{KL}}(f_\theta|g) = \int_{-\infty}^{\infty} g(x) \log\left(\frac{g(x)}{f_\theta(x)}\right) dx = E[\log g(X) - \log f_\theta(X)]. \tag{2.1.1}$$

The expectation is taken with respect to g, see Kullback and Leibler ([11], page 79-86). The integral $\int g(x)\log(g(x))dx$ is a constant and does not depend on θ, and the crucial quantity is $-E[\log f_\theta(X)]$. This is utilized below.

$D_{\mathrm{KL}}(f_\theta|g)$ is finite if g is absolutely continuous with respect to f_θ. That is, $g(x) = 0$ for any $x \in \mathbb{R}$ such that $f_\theta(x) = 0$, see Kullback and Leibler ([11], page 79-86). In this thesis the

choices of f_θ and g are density distributions that are defined for $x \in (0, \infty)$. Neither $f_\theta(x)$, nor $g(x)$, are ever exactly equal to zero and they approach zero for the same values of x, namely $x \to 0$ and $x \to \infty$. Hence g is absolutely continuous with respect to f_θ, and $D_{\mathrm{KL}}(f_\theta|g) < \infty$. Note that $D_{\mathrm{KL}}(f_\theta|g)$ equals zero if and only if $g(x) = f_\theta(x)$, i.e. when the true distribution is within the parametric distribution family.

The Kullback-Leibler distance can be approximated by a sum.

$$D_{\mathrm{KL}}(f_\theta|g) \approx \frac{1}{m}\sum_{i=1}^{m} \log\left(\frac{g(X_i)}{f_\theta(X_i)}\right) \tag{2.1.2}$$

where $X_1, ...X_m$ are m Monte Carlo simulations drawn independently from the true distribution g. For each i, $g(X_i)$ is a constant, and consequently θ_0 can be found by simply minimizing

$$-\frac{1}{m}\sum_{i=1}^{m} \log f_\theta(X_i). \tag{2.1.3}$$

Note that minimizing (2.1.2) is the same as maximizing the log-likelihood function of f_θ given data $X_1, ...X_m \sim g(x)$. As stated in Chapter 1, $\hat{\theta} \to \theta_0$ was the number of data m increases. Hence by choosing m sufficiently large, it is possible to find a vector of parameters $\hat{\theta}$ arbitrarily close to θ_0. Consequently, when m is close to infinity the Kullback-Leibler distance measures the difference between the best choice within the parametric family f_{θ_0} and the true distribution g. The approach where θ_0 is found by minimizing (2.1.3) is henceforth called the Monte Carlo approach.

Another approach for approximating the Kullback-Leibler distance is to utilize Gauss-Legendre quadrature. The idea is that we choose limits a and b such that the integral is approximately 0 outside (a, b). Then, for some integer N, a set of N abscissas $X = (X_1, ..., X_N)$, where $X_1 > a$ and $X_N < b$, and N weights $w = (w_1, ..., w_N)$ are found numerically. For a sufficiently large N the integral in (2.1.1) can be approximated by a weighted sum of the integrand,

$$D_{\mathrm{KL}}(f_\theta|g) \approx \int_a^b g(x) \log\left(\frac{g(x)}{f_\theta(x)}\right) dx \approx \sum_{i=1}^{N} w_i \left[g(X_i) \log\left(\frac{g(X_i)}{f_\theta(X_i)}\right)\right]. \tag{2.1.4}$$

Gauss-Legendre quadrature gives high accuracy and converges fast as N increases if the integrand is smooth, see Press et al. [15]. An integrand is considered smooth if it has derivatives of sufficiently high order, see [15]. Computationally the Gauss-Legendre quadrature is more difficult than the Monte Carlo approach. This is because the weights and abscissas have to be found numerically. Even so, the rapid convergence makes it faster and more preferable than the Monte Carlo approach.

2.2 The one-parameter situation

2.2.1 Asymptotic normality

The log-likelihood of the function $f_\theta(x) = f(x|\theta)$ is defined as $l(\theta) = \sum_{i=1}^{n} \log[f(X_i|\theta)]$. The derivative of the log-likelihood function is usually called the score function. We find the maximum likelihood estimator $\hat{\theta}$ by setting the score function equal to zero and solve for θ, i.e. $\hat{\theta}$ is defined through

$$\sum_{i=1}^{n} \frac{\partial \log[f(X_i|\hat{\theta})]}{\partial \theta} = 0.$$

When the score function is divided by $\sqrt{n}$ and Taylor expansion around θ_0 applied, then,

$$0 = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\partial \log[f(X_i|\hat{\theta})]}{\partial \theta} \approx Y + Z\sqrt{n}(\hat{\theta} - \theta_0) \tag{2.2.1}$$

where

$$Y = n^{-\frac{1}{2}} \sum_{i=1}^{n} Y_i = n^{-\frac{1}{2}} l'(\theta_0), \quad Y_i = \frac{\partial \log[f(X_i|\theta_0)]}{\partial \theta},$$

and

$$Z = n^{-1} \sum_{i=1}^{n} Z_i = n^{-1} l''(\theta_0), \quad \text{and} \quad Z_i = \frac{\partial^2 \log[f(X_i|\theta_0)]}{\partial \theta^2}.$$

From the central limit theorem, see Devore and Berk ([5], page 293), we know that Y is approximately normal for large n. By applying the law of large numbers see ([5], page 297), it can be verified that $Z \to E[Z]$, where the expectation is taken with respect to g if $X_i \sim g(x)$, and with respect to f_θ if $X_i \sim f_\theta(x)$. Hence, since $\sqrt{n}(\hat{\theta} - \theta_0) \approx -\frac{Y}{Z}$, we can conclude that $\sqrt{n}(\hat{\theta} - \theta_0)$ is normally distributed. The parameters come from the mean and variance of $Y_1, ...Y_n$ and the mean of $Z_1, ..., Z_n$.

2.2.2 Expected value

If the true distribution of $X_1, ..., X_n$ is within the theoretical distribution family, there exist a set of parameters θ_0 which makes the underlying density function of $X_1, ..., X_n$ arbitrary close to f_{θ_0}. Thus, f_{θ_0} can be considered to be the true density function. We find the expected value by noting that $\int_{-\infty}^{\infty} f_{\theta_0}(x) = 1$, consequently $\int_{-\infty}^{\infty} \frac{\partial}{\partial \theta} f_{\theta_0}(x) dx = 0$. Thus,

$$\begin{aligned} \mathrm{E}[Y_i] &= \int_{-\infty}^{\infty} \frac{\partial \log f_{\theta_0}(x)}{\partial \theta} f_{\theta_0}(x) dx = \int_{-\infty}^{\infty} \frac{\partial f_{\theta_0}(x)/\partial \theta}{f_{\theta_0}(x)} f_{\theta_0}(x) dx \\ &= \int_{-\infty}^{\infty} \frac{\partial}{\partial \theta} f_{\theta_0}(x) = 0. \end{aligned}$$

Hence, the expected value of $\sqrt{n}(\hat{\theta} - \theta_0)$ equals 0 when the true distribution is a part of the parametric family.

However, the distribution of $X_1, ..., X_n$ might well be outside the theoretical distribution family, i.e. $X_i \sim g(x)$. Recall that θ_0 is the set of parameters which minimizes the Kullback-Leibler distance, defined in (2.1.1). Since the derivative of a minimum or maximum is zero, θ_0 is the vector of parameters such that,

$$\frac{\partial}{\partial \theta} D_{\mathrm{KL}}(f_{\theta_0}|g) = \frac{\partial}{\partial \theta} \int_{-\infty}^{\infty} g(x) \log \left(\frac{g(x)}{f_{\theta_0}(x)} \right) dx = - \int_{-\infty}^{\infty} \frac{\partial}{\partial \theta} g(x) \log f_{\theta_0}(x) = 0.$$

The second equality comes from the fact that $\int_{-\infty}^{\infty} g(x) \log g(x)$ is a constant. Consequently,

$$E_g[Y_i] = \int_{-\infty}^{\infty} \frac{\partial \log f_{\theta_0}(x)}{\partial \theta} g(x) dx = -\frac{\partial}{\partial \theta} D_{\mathrm{KL}}(f_{\theta_0}|g) = 0.$$

Hence, the mean of $X_1, ..., X_n$ is still zero, and $E[\sqrt{n}(\hat{\theta} - \theta_0)] = 0$. That is, the maximum likelihood estimator is consistent also when the true underlying distribution is outside the chosen theoretical family.

2.2.3 Standard deviation

If the assumed model is correct,

$$\begin{aligned}\text{var}[Y_i] &= E[Y_i^2] \\ &= \int_{-\infty}^{\infty} \left(\frac{\partial}{\partial\theta}\log f_{\theta_0}(x)\right)^2 f_{\theta_0}(x)dx \\ &= E_{f_\theta}\left[\left(\frac{\partial}{\partial\theta}\log f(X_i|\theta_0)\right)^2\right] := I_{f_\theta}(\theta_0),\end{aligned}$$

and

$$\begin{aligned}E[Z_i] &= \int_{-\infty}^{\infty} \left(\frac{\partial^2}{\partial\theta^2}\log f_{\theta_0}(x)\right) f_{\theta_0}(x)dx \\ &= -\int_{-\infty}^{\infty} \left(\frac{\partial}{\partial\theta}\log f_{\theta_0}(x)\right)^2 f_{\theta_0}(x)dx = -I_{f_\theta}(\theta_0).\end{aligned}$$

In Appendix B the second equality is justified, i.e. it is shown that $E[Z_i] = -\text{E}[Y_i^2] = -\text{var}[Y_i]$.

$I_{f_\theta}(\theta_0)$ is usually called the Fisher information or the expected information. After inserting for $\text{var}[Y_i]$ and $E[Z_i]$, it follows that the variance of the asymptotic normal distribution is

$$\sigma^2_{f_{\theta_0}} = \frac{\text{var}[Y_i]}{(E[Z_i])^2} = \frac{1}{I_{f_\theta}(\theta_0)}.$$

Thus when $X_i \sim f_{\theta_0}(x)$, then

$$\sqrt{n}(\hat{\theta} - \theta_0) \sim \mathcal{N}(0, \sigma^2_{f_{\theta_0}}).$$

The expression for the standard deviation is different when the true model is outside the parametric family. The relationship $E[Z_i] = -\text{var}[Y_i]$ does no longer hold. All we can say is that

$$\text{var}[Y_i] = \int_{-\infty}^{\infty} \left(\frac{\partial}{\partial\theta}\log f_{\theta_0}(x)\right)^2 g(x)dx := I_g(\theta_0)$$

and

$$E[Z_i] = \int_{-\infty}^{\infty} \frac{\partial^2}{\partial\theta^2}\log f_{\theta_0}(x) g(x)dx := \lambda(\theta_0),$$

from which it follows that when $g(x) \neq f_{\theta_0}(x)$,

$$\sqrt{n}(\hat{\theta} - \theta_0) \sim \mathcal{N}(0, \sigma_g^2),$$

where

$$\sigma_g^2 = \frac{I_g(\theta_0)}{\{\lambda(\theta_0)\}^2}.$$

2.3 The multi-parameter situation

If $\theta = (\theta_1, ...\theta_p)^T$ is a vector of p unknown parameters, the log-likelihood function is $l(\theta) = \sum_{i=1}^{n} \log[f(X_i|\theta_1, ..., \theta_p)]$. The mathematics are advanced, and details are therefore not studied here. Precise mathematics and conditions are given in Huber [6]. The large sample theory also applies to the multi-parameter situation. Therefore, by a generalization of the arguments in the one-parameter situation, it is possible to prove that, for a general p, $\sqrt{n}(\hat{\theta} - \theta_0)$ is asymptotically normally distributed with expectation zero, both when the assumed model is correct and when it is false.

The expressions for the standard deviation depends, as for the one parameter situation, on whether the assumed model is correct. Since the expected value equals zero we conclude that $\text{var}[Y_i] = \text{E}[Y_i^2]$. The relationship $\text{E}[Y_i^2] = -\text{E}[Z_i]$ still holds when there is more than one parameter and the underlying distribution is a part of the parametric family. This is justified in Appendix B by a slight generalization of the argument from the one-parameter situation. The quantities are now matrices with the entry (j, l) given by

$$\text{E}[Y_i^2] = \int_{-\infty}^{\infty} \left(\frac{\partial}{\partial \theta_j} \log f_{\theta_0}(x) \right) \left(\frac{\partial}{\partial \theta_l} \log f_{\theta_0}(x) \right) f_{\theta_0}(x) dx = I_{f_\theta jl}(\theta_0),$$

and

$$\text{E}[Z_i] = \int_{-\infty}^{\infty} \left(\frac{\partial^2}{\partial \theta_j \partial \theta_l} \log f_{\theta_0}(x) \right) f_{\theta_0}(x) dx = -I_{f_\theta jl}(\theta_0).$$

Hence, when the true distribution of the data is within the theoretical distribution family, the Fisher information matrix is $I_{f_\theta}(\theta_0)$. The entry (j, l) of $I_{f_\theta}(\theta_0)$ is given by the entry (j, l) in $\text{E}[Y_i^2]$.

Thus, when the true distribution is a part of the parametric family, i.e. when $X_i \sim f_{\theta_0}(x)$,

$$\sqrt{n}(\hat{\theta} - \theta_0) \sim \mathcal{N}(0, I_{f_\theta}^{-1}(\theta_0)),$$

where $\sqrt{n}(\hat{\theta} - \theta_0)$ and 0 are column vectors of length p and $I_{f_\theta}^{-1}(\theta_0)$ is a $p \times p$ matrix.

In the situation where the true distribution is outside the parametric family, the expression for the covariance matrix is different. We need

$$I_{gjl}(\theta_0) = \int_{-\infty}^{\infty} \left(\frac{\partial}{\partial \theta_j} \log f_{\theta_0}(x) \right) \left(\frac{\partial}{\partial \theta_l} \log f_{\theta_0}(x) \right) g(x) dx \tag{2.3.1}$$

and

$$\Lambda_{jl}(\theta_0) = \int_{-\infty}^{\infty} \left(\frac{\partial^2}{\partial \theta_j \partial \theta_l} \log f_{\theta_0}(x) \right) g(x) dx. \tag{2.3.2}$$

From a result given and proved in Huber [6] and an extension of the argument from the one-parameter situation, it can be shown that the covariance matrix of the vector $\sqrt{n}(\hat{\theta} - \theta_0)$ is $\Lambda^{-1} I_g \Lambda^{-1}$, where $\Lambda = (\Lambda_{jl}(\theta_0))$ and $I_g = (I_{gjl}(\theta_0))$ for $j = 1, ..., p$ and $l = 1, ..., p$, see Huber [6].

Hence, when the true distribution is outside the parametric family,

$$\sqrt{n}(\hat{\theta} - \theta_0) \sim \mathcal{N}(0, \Lambda^{-1} I_g \Lambda^{-1}),$$

where $\sqrt{n}(\hat{\theta} - \theta_0)$ and 0 are column vectors of length p and $\Lambda^{-1} I_g \Lambda^{-1}$ is a $p \times p$ matrix.

It is easy to check that $\Lambda^{-1} I_g \Lambda^{-1}$ is reduced to σ_g^2 when $p = 1$. Equation (2.3.2) is then

$$\Lambda_{11}(\theta_0) = \int_{-\infty}^{\infty} \frac{\partial^2 \log f_{\theta_0}(x)}{\partial \theta_1^2} g(x) dx = \lambda(\theta_0)$$

and (2.3.1) is

$$I_{g11}(\theta_0) = \int_{-\infty}^{\infty} \left(\frac{\partial \log f_{\theta_0}(x)}{\partial \theta_1} \right) \left(\frac{\partial \log f_{\theta_0}(x)}{\partial \theta_1} \right) g(x) dx = I_g(\theta_0)$$

where $\lambda(\theta_0)$ and $I_g(\theta_0)$ are the expressions from the one-parameter situation. Hence, $\Lambda^{-1} I_g \Lambda^{-1} = \lambda(\theta_0)^{-1} I_g(\theta_0) \lambda(\theta_0)^{-1} = I_g(\theta_0)/\{\lambda(\theta_0)\}^2 = \sigma_g^2$, and the expression for the standard deviation from the multi-parameter situation is reduced to the one-parameter situation expression. The same happens with the expression for the standard deviation when the assumed model is correct.

Note that by replacing $g(x)$ with $f_\theta(x)$ in the expressions for $\Lambda(\theta_0)$ and $I_g(\theta_0)$, the expressions are reduced to those of $-I_{f_\theta}(\theta_0)$ and $I_{f_\theta}(\theta_0)$ respectively. Hence, the variance of $\sqrt{n}(\hat{\theta} - \theta_0)$ becomes $I_{f_\theta}^{-1}(\theta_0)$. Consequently, if g is the true distribution and it can be both within and outside the parametric family, the situation $X_i \sim f_\theta(x)$ can be regarded as a special case of the general situation where $X_i \sim g(x)$. Thus, henceforth the situations $X_i \sim g(x)$ and $X_i \sim f_\theta(x)$ are not treated separately.

The results from this section and Section 2.2 are later used to evaluate the error in $\hat{\theta}$, and thus find estimates of the differences between $\Psi(\hat{\theta})$, $\Psi(\theta_0)$, and $\Psi(g)$. The method when the asymptotic properties of $\hat{\theta}$ are used to evaluate error is called the delta method. By drawing random, independent normal data the delta method produces estimates of error in $\hat{\theta}$. As mentioned in Chapter 1, maximum likelihood can also be utilized to evaluate the error in $\hat{\theta}$. When maximum likelihood is applied, consistent estimates $\hat{\theta}$ of θ_0 are found by maximizing the log-likelihood function of f_θ, with respect to the data $X_1, ..., X_n$, where the data is drawn from the underlying distribution g. More on this subject in Chapter 4.

2.4 Extension to risk functionals

As mentioned in Chapter 1, what is usually studied is not $\hat{\theta}$ or $f_{\hat{\theta}}(x)$ itself, but some functional $\Psi(\hat{\theta}) = \Psi(f_{\hat{\theta}})$ describing some feature of the risk variable. Asymptotic properties of total error $\Psi(\hat{\theta}) - \Psi(g)$ are derived in Section 2.4.1 using the results from Section 2.3. In Section 2.4.2 the square of expected total error is decomposed into three terms, which are then examined separately.

In this section the general situation, where the true distribution g can be both within and outside the parametric family is considered. Instead of first deriving the results for the one-parameter situation and then extend it to the multi-parameter situation, as was done to find the asymptotic properties of $\sqrt{n}(\hat{\theta} - \theta_0)$, the results are given for the multi-parameter situation directly.

2.4.1 Asymptotic properties

As considered in Chapter 1 total error can be divided into a random part and a constant part, i.e. estimation error and systematic error,

$$\Psi(\hat{\theta}) - \Psi(g) = \{\Psi(\hat{\theta}) - \Psi(\theta_0)\} + \{\Psi(\theta_0) - \Psi(g)\}.$$

By applying Taylor expansion in p dimensions on $\Psi(\hat{\theta})$ around the vector θ_0, $\Psi(\hat{\theta})$ can be rewritten as,

$$\Psi(\hat{\theta}) \approx \Psi(\theta_0) + \frac{\partial \Psi(\hat{\theta})}{\partial \theta_1}|_{\hat{\theta}=\theta_0}(\hat{\theta}_1 - \theta_{01}) + ... + \frac{\partial \Psi(\hat{\theta})}{\partial \theta_p}|_{\hat{\theta}=\theta_0}(\hat{\theta}_p - \theta_{0p}).$$

Hence, estimation error can be approximated,

$$\Psi(\hat{\theta}) - \Psi(\theta_0) \approx \nabla\Psi(\theta_0)(\hat{\theta} - \theta_0),$$

where

$$\nabla\Psi(\theta_0) = \left\{ \frac{\partial \Psi(\theta_0)}{\partial \theta_1}, ..., \frac{\partial \Psi(\theta_0)}{\partial \theta_p} \right\} \tag{2.4.1}$$

is the gradient of $\Psi(\hat{\theta})$ evaluated at $\hat{\theta} = \theta_0$. Each term in the vector $\nabla\Psi(\theta_0)$ is a constant. $(\hat{\theta} - \theta_0)$ is also a vector of length p,

$$(\hat{\theta} - \theta_0) = \{(\hat{\theta}_1 - \theta_{01}), ..., (\hat{\theta}_p - \theta_{0p})\}^{\mathrm{T}}.$$

Thus total error can be written as

$$\begin{aligned} \Psi(\hat{\theta}) - \Psi(g) &\approx \nabla\Psi(\theta_0)(\hat{\theta} - \theta_0) + \{\Psi(\theta_0) - \Psi(g)\} \\ &=: a\sqrt{n}(\hat{\theta} - \theta_0) + b, \end{aligned}$$

where $a = n^{-1/2}\nabla\Psi(\theta_0)$ and $b = \Psi(\theta_0) - \Psi(g)$ is systematic error.

Suppose $y = cx + d$, where $x \sim \mathcal{N}(\mu, \Sigma)$ is a p-dimensional multivariate normally distributed random vector, c and d are constants vector of length p. Then,

$$y \sim \mathcal{N}(c\mu + d, c^T\Sigma c),$$

see Patel and Read ([13], page 290). Thus, since it was shown in Section 2.3 that

$$\sqrt{n}(\hat{\theta} - \theta_0) \sim \mathcal{N}(0, \Lambda^{-1} I_g \Lambda^{-1}),$$

it follows that

$$a\sqrt{n}(\hat{\theta} - \theta_0) \sim \mathcal{N}(0, a^T\Lambda^{-1} I_g \Lambda^{-1} a),$$

and, since b is a constant

$$\Psi(\hat{\theta}) - \Psi(g) \sim \mathcal{N}(b, a^T\Lambda^{-1} I_g \Lambda^{-1} a).$$

This can be rewritten as

$$\Psi(\hat{\theta}) - \Psi(g) \sim \mathcal{N}\left(b, \frac{\tau}{n}\right), \tag{2.4.2}$$

where

$$b = \Psi(\theta_0) - \Psi(g) \quad \text{and} \quad \tau = \{(\nabla\Psi(\theta_0))^T\Lambda^{-1} I_g \Lambda^{-1} \nabla\Psi(\theta_0)\}$$

and where $\nabla\Psi(\theta_0)$ was defined in (2.4.1).

Clearly,

$$\text{var}[\Psi(\hat{\theta}) - \Psi(g)] = \frac{\tau}{n} \to 0 \quad \text{as} \quad n \to \infty.$$

Thus, total error is normally distributed with expected value equal to systematic error. The expected value is different from zero if the true distribution g is outside the parametric family. The standard deviation of total error converges to zero when $n \to \infty$. That is, when $n \to \infty$, total error equals systematic error with probability 1.

Note that since random error equals total error when $b = 0$,

$$\Psi(\hat{\theta}) - \Psi(\theta_0) \sim \mathcal{N}\left(0, \frac{\tau}{n}\right). \tag{2.4.3}$$

Some might argue that systematic error depends on n in the sense that it is possible to choose a better parametric distribution family, i.e a f_θ closer to g, when the number of observations n increases. In this thesis however, a default distribution family is chosen a priori. Therefore the last objection is ignored.

Finding $\nabla\Psi(\theta_0)$

It is often difficult to find the gradient $\nabla\Psi(\theta_0)$ explicitly, but numerical approximations are available. If $\nabla\Psi(\theta_0)$ is as defined in (2.4.1), the jth term can be approximated by

$$\nabla\Psi(\theta_0)_j \approx \frac{\Psi(\theta_0 + h\text{e}_j) - \Psi(\theta_0 - h\text{e}_j)}{2h},$$

where $e_j = (0, .., 0, 1, 0, .., 0)$ with 1 at entry j. The constant h is some small number. In the setting of this thesis the estimate of $\nabla\Psi(\theta_0)$ is very unstable due to how random extended Pareto variables are drawn. Therefore, a great amount of Monte Carlo simulations are needed. Consequently, a procedure where estimates of the total error are found by drawing random independent normal data based on (2.4.2) is not carried out or further discussed.

2.4.2 Expected total error

Expected total error is of great interest as a measure of how far the estimated reserve is from the true value. In later sections, $\text{E}[\{\Psi(\hat{\theta}) - \Psi(g)\}^2]^{1/2}$ is used to indicate the appropriateness of the extended Pareto distribution as parametric distribution family. The expectation of the square of total error can be written as

$$\text{E}[\{\Psi(\hat{\theta}) - \Psi(g)\}^2] = \text{A1} + \text{A2} + \text{A3}, \tag{2.4.4}$$

where

$$\begin{aligned}
\text{A1} &= \text{E}[\{\Psi(\hat{\theta}) - \Psi(\theta_0)\}^2] \\
\text{A2} &= \{\Psi(\theta_0) - \Psi(g)\}^2 \text{ and} \\
\text{A3} &= 2\{\text{E}[\Psi(\hat{\theta})] - \Psi(\theta_0)\}\{\Psi(\theta_0) - \Psi(g)\}.
\end{aligned}$$

It would yield a neat interpretation if the cross term A3 could be ignored. Whether that is appropriate will now be examined. This section contains some theoretical results, the technical details are however beyond the scope of this thesis and therefore omitted.

The first term, $\text{A1} = \text{E}[\{\Psi(\hat{\theta}) - \Psi(\theta_0)\}^2]$, is the expectation of the square of estimation error. In Section 2.4.1 it was argued that $\Psi(\hat{\theta}) - \Psi(\theta_0) \sim \mathcal{N}\left(0, \frac{\tau}{n}\right)$, where τ is a constant also defined in Section 2.4.1. Thus, since for a random variable X, $\text{E}[X^2] = \text{var}[X] + \text{E}[X]^2$, it follows that

$$\text{E}[\{\Psi(\hat{\theta}) - \Psi(\theta_0)\}^2] = \frac{\tau}{n}.$$

Hence, term A1 is of order n^{-1}.

The second and third term in (2.4.4) can be interpreted as skewness terms. If the true distribution is within the parametric family, $\{\Psi(\theta_0) - \Psi(g)\}$ equals zero and the two last terms in (2.4.4) are zero. Hence, when the true distribution is a part of the parametric family total error equals random error. However, when the true distribution g is outside the parametric family, the two last terms in (2.4.4) can not be neglected.

The second term, $\text{A2} = \{\Psi(\theta_0) - \Psi(g)\}^2$, is a constant. It is different from zero when g is outside the parametric family, but might be of less importance compared to the two other terms when n is small.

The third term, $\text{A3} = 2\{\text{E}[\Psi(\hat{\theta})] - \Psi(\theta_0)\}\{\Psi(\theta_0) - \Psi(g)\}$, is of order n^{-1}, but this conclusion needs some additional arguments. These arguments build on the decomposition (2.2.1) from Section 2.2, and an extension to the multi-parameter situation includes advanced mathematics beyond the scope of this thesis. Thus, only the one-parameter situation is considered here.

In Section 2.4.1, it is stated that

$$\Psi(\hat{\theta}) - \Psi(\theta_0) \approx \Psi'(\theta_0)(\hat{\theta} - \theta_0).$$

Thus A1 and A3 in (2.4.4) can be rewritten as

$$\text{A1} \approx \frac{\Psi'(\theta_0)^2}{\sqrt{n}}\text{E}[\sqrt{n}(\hat{\theta} - \theta_0)^2] \tag{2.4.5}$$

and

$$\text{A3} \approx 2\{\Psi(\theta_0) - \Psi(g)\}\frac{\Psi'(\theta_0)}{\sqrt{n}}\text{E}[\sqrt{n}(\hat{\theta} - \theta_0)]. \tag{2.4.6}$$

In Section 2.2, Taylor expansion is applied to $\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\partial \log[f(X_i|\hat{\theta})]}{\partial\theta}$ around θ_0. By adding one more term to the decomposition (2.2.1) we get,

$$0 \approx Y + Z\sqrt{n}(\hat{\theta} - \theta_0) + U\sqrt{n}(\hat{\theta} - \theta_0)^2,$$

where Y and Z are defined as in Section 2.2 and

$$U = \frac{1}{2n}\sum_{i=1}^{n}\frac{\partial^3 \log[f(X_i|\theta_0)]}{\partial\theta^3}.$$

In Section 2.2 it was shown that $\text{E}[Y] = 0$, hence

$$0 \approx \text{E}[Z\sqrt{n}(\hat{\theta} - \theta_0)] + \text{E}[U\sqrt{n}(\hat{\theta} - \theta_0)^2].$$

By the law of large numbers Z and U are constants in the limit, see Devore and Berk ([5], page 297). Therefore, by Slutsky's theorem, see Cramer ([4], page 255),

$$0 \approx \mathrm{E}[Z]\mathrm{E}[\sqrt{n}(\hat{\theta} - \theta_0)] + \mathrm{E}[U]\mathrm{E}[\sqrt{n}(\hat{\theta} - \theta_0)^2].$$

By replacing with the expressions for A1 and A3 from (2.4.5) and (2.4.6), the expressions become

$$0 \approx \frac{\mathrm{E}[Z]\sqrt{n}}{2\Psi'(\theta_0)\{\Psi(\theta_0) - \Psi(g)\}}\mathrm{A3} + \frac{\mathrm{E}[U]\sqrt{n}}{\Psi'(\theta_0)^2}\mathrm{A1}.$$

Thus,

$$\mathrm{A3} \approx \mathrm{C} \times \mathrm{A1},$$

where

$$\mathrm{C} = -\frac{2\mathrm{E}[U]\{\Psi(\theta_0) - \Psi(g)\}}{\mathrm{E}[Z]\Psi'(\theta_0)}$$

is a constant. Since A1 is of order n^{-1} and A3 is a constant times something of order n^{-1} , A3 is also of order n^{-1}.

Since A1 and A3 in (2.4.4) are of order n^{-1}, they are dominated by A2 when n is large. In particular A3 might be considered as a small remainder that can be neglected when n is large, such that

$$\mathrm{E}[\{\Psi(\hat{\theta}) - \Psi(g)\}^2] \approx \mathrm{E}[\{\Psi(\hat{\theta}) - \Psi(\theta_0)\}^2] + \{\Psi(\theta_0) - \Psi(g)\}^2. \tag{2.4.7}$$

The question is then how large n needs to be. From numerical studies shown in Section 4.4, it turns out that for $n \leq 100000$ the last term in (2.4.4) can not be neglected. Consequently, the approximation (2.4.7) is not applied in this thesis.

Chapter 3

The extended Pareto distribution

Due to its flexibility, the extended Pareto distribution is a natural choice when searching for a default loss model. In Chapter 4 and Chapter 5 its suitability as default loss model is studied. The extended Pareto distribution is a generalization of the ordinary Pareto distribution with density and cumulative distribution functions

$$f(x) = \frac{\alpha/\beta}{(1+x/\beta)^{1+\alpha}} \quad \text{and} \quad F(x) = 1 - \frac{1}{(1+x/\beta)^{\alpha}} \quad z > 0,$$

where $\alpha > 0$ is the shape and $\beta > 0$ is the scale parameter, see Kleiber and Kotz ([10], page 59). The distribution is heavy-tailed and is much used in property insurance. There exist explicit expressions for the mean, standard deviation, skewness and the kurtosis. However, due to the heavy tails the expressions are only finite for $\alpha > 1$, $\alpha > 2$, $\alpha > 3$ and $\alpha > 4$ respectively. The expressions for the mean and the standard deviation are given in Section 5.4.

Consider now

$$f(x) = \frac{\Gamma(\alpha+\theta)}{\Gamma(\alpha)\Gamma(\theta)} \frac{1}{\beta} \frac{(x/\beta)^{\theta-1}}{(1+\frac{x}{\beta})^{\alpha+\theta}}, \quad x > 0, \tag{3.0.1}$$

where $\alpha, \beta, \theta > 0$ and θ is added as an additional shape parameter, see Beirlant et al. [8]. When $\theta = 1$, the extended Pareto density takes the form of an ordinary Pareto density. When the scale parameter β is linked to the mean ξ through $\beta = \frac{\xi\alpha}{\theta}$, ξ and θ are fixed and α becomes infinite, an extended Pareto distributed random variable has the shape of a gamma distribution. By inserting $\beta = \frac{\xi\alpha}{\theta}$, it can be shown that

$$f(x) = \frac{\Gamma(\alpha+\theta)}{\Gamma(\alpha)\Gamma(\theta)} \frac{(x\theta/\xi\alpha)^{\theta-1}}{(1+\frac{x\theta}{\xi\alpha})^{\alpha+\theta}} \to \frac{(\theta/\xi)^{\theta}}{\Gamma(\theta)} x^{\theta-1} e^{-\theta x/\xi} \quad \text{as } \alpha \to \infty. \tag{3.0.2}$$

The density function to the right is the density function of a gamma distributed random variable with shape parameter θ and expectation ξ. When $\theta = 1$ the expression to the right in (3.0.2) is reduced to $f(x) = \xi e^{-x/\xi}$, which is the density function of an exponentially distributed random variable with expectation ξ. Thus, both the gamma distribution, the Pareto distribution and the exponential distribution are parts of the extended Pareto distribution family.

The density function of an extended Pareto random variable is decreasing over the real line when $\theta \leq 1$ and has a single maximum for $\theta > 1$. This is illustrated in Figure 3.1, where the extended Pareto density function is plotted twice. The solid line is the density function for $\theta = 0.8$ and the dotted line is the density function for $\theta = 1.5$. The other parameters are

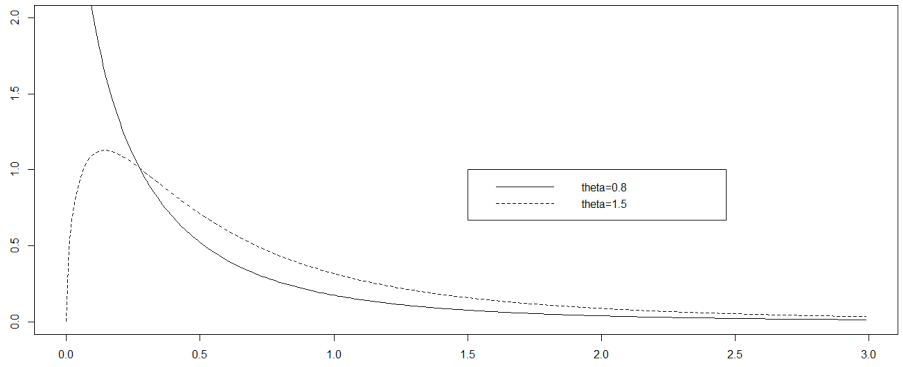

Figure 3.1: *The extended Pareto density function for* $(\alpha, \beta, \theta) = (2.5, 1, 0.8)$ *(solid line) and* $(\alpha, \beta, \theta) = (2.5, 1, 1.5)$ *(dotted line).*

fixed, $\alpha = 2.5$ and $\beta = 1$.

The mean and standard deviation in the extended Pareto distribution are

$$\mathrm{E}(X) = \frac{\theta\beta}{\alpha - 1} = \xi \quad \text{and} \quad \mathrm{sd}(X) = \xi\sqrt{\frac{\alpha + \theta - 1}{\theta(\alpha - 2)}}.$$

The skewness coefficient and kurtosis are

$$\mathrm{skew}(X) = 2\left(\frac{\alpha + 2\theta - 1}{\alpha - 3}\right)\sqrt{\frac{\alpha - 2}{\theta(\alpha + \theta - 1)}}$$

and

$$\mathrm{kurt}(X) = \frac{3(\alpha - 2)(\alpha^2\theta + 2\alpha^2 + \alpha\theta^2 + 4\alpha\theta - 4\alpha + 5\theta^2 - 5\theta + 2)}{(\alpha - 4)(\alpha - 3)\theta(\alpha + \theta - 1)}.$$

As for the ordinary Pareto distribution these expressions are only finite when $\alpha > 1$, $\alpha > 2$, $\alpha > 3$ and $\alpha > 4$ respectively. The expressions are verified in the Appendix B.

When $\theta = 1$, i.e. when we are dealing with the ordinary Pareto distribution, random variables can be drawn by using inversion. Let U^* be a randomly drawn uniformly distributed variable on the interval $[0, 1]$. The inverse Pareto cumulative distribution function is

$$Z = \beta(U^{-\frac{1}{\alpha}} - 1). \tag{3.0.3}$$

Hence, Pareto distributed random variables Z can be found by replacing U^* with U in (3.0.3), see Iyengar [7]. However, when $\theta \neq 1$ the cumulative distribution is complicated and simulation with inversion becomes difficult. Instead an extended Pareto distributed random variable can be simulated by first simulating two independent standard gamma distributed random variables, G_θ and G_α, with shape parameters θ and α. The standard gamma distribution has mean equal to one and is defined in Chapter 5. Thereafter, it can be utilized that if

$$Z = \beta\frac{\theta G_\theta}{\alpha G_\alpha},$$

Z is an extended Pareto distributed random variable with parameters (α, β, θ), see Goovaerts et al. [12].

In Chapter 4, estimates of the parameters $(\hat{\alpha}, \hat{\beta}, \hat{\theta})$ are found by optimizing the log-likelihood function given by

$$\begin{aligned} l(\alpha, \beta, \theta) = \quad & n[\log\{\Gamma(\alpha+\theta)\} - \log\{\Gamma(\alpha)\} - \log\{\Gamma(\theta)\} - \theta \log\{\beta\}] + \\ & (\theta - 1) \sum_{i=1}^{n} \log\{y_i\} - (\alpha + \theta) \sum_{i=1}^{n} \log\{1 + y_i/\beta\}, \end{aligned}$$

where $y = (y_1, ..., y_n)$ is the data drawn from the true distribution g. Henceforth, the extended Pareto distribution family is used as parametric family.

Chapter 4

Numerical methods

Consider as in Chapter 1, the decomposition

$$\Psi(\hat{\theta}) - \Psi(g) = \Psi(\hat{\theta}) - \Psi(\theta_0) + \Psi(\theta_0) - \Psi(g),$$

where $\Psi(\hat{\theta}) - \Psi(\theta_0)$ is random error due to estimation and $\Psi(\theta_0) - \Psi(g)$ is systematic error. In order to analyse these error terms separately, it is necessary to derive procedures for estimating and determining parameters and functionals. In this chapter the numerical methods are described and illustrated. The log-normal distribution with parameters $(\mu = 1, \sigma = 0)$ is used as underlying distribution when the procedures are illustrated. It is not a part of the extended Pareto distribution family, and there are both estimation and systematic error. The log-normal distribution is one of four underlying distributions that are examined in Chapter 5. The sizes of claims are often denoted by Z, and this notation is henceforth used. The R-scripts used to conduct the simulations are given in Appendix C.

4.1 Finding θ_0

4.1.1 The Monte Carlo approach

The Kullback-Leibler distance is defined in Section 2.1, and is a measure of the difference between two probability distributions. In Section 2.1 it is argued that θ_0, the vector of parameters which minimizes the Kullback-Leibler distance, is the vector which minimizes (2.1.3),

$$-\frac{1}{m}\sum_{i=1}^{m} \log f_\theta(Z_i).$$

By increasing the number of Monte Carlo simulations m, it is possible to find a vector of parameters θ arbitrarily close to θ_0. When $f_\theta(z)$ is the extended Pareto density, the Monte Carlo approach returns the values $(\alpha_0, \beta_0, \theta_0)$. A drawback with Monte Carlo simulations is that the simulation time might be high, because the number of simulations m needed often is large. As an illustration, $(\alpha_0, \beta_0, \theta_0)$ are found using $m = 10^5$ Monte Carlo simulations. The true distribution $g \sim$ log-normal(0,1) and $f_\theta \sim$ extended Pareto. The output is given in Table 4.1.

4.1.2 Gauss-Legendre quadrature

In Section 2.1, Gauss-Legendre quadrature is introduced as a second method for approximating the Kullback-Leibler distance. By means of Gauss-Legendre quadrature the parameters

Table 4.1: *Parameter values found by means of the Monte Carlo approach.*

α_0	β_0	θ_0
2.43	1.00	2.45

Table 4.2: *Parametr values found by means of Gauss-Legendre quadrature.*

α_0	β_0	θ_0
2.44	1.00	2.44

$(\alpha_0, \beta_0, \theta_0)$ stabilized fast, because N, the number of abscissas and weights, is small. In order for Gauss-Legendre quadrature to function, the limits a and b have to be chosen such that the probability mass in equation (2.1.4) is approximately 0 outside (a, b). That is, the integrand is approximately zero outside (a, b). For distributions where $(b - a)$ is large, the convergence is relatively slow, e.g. when $g \sim$log-normal$(0, 1)$ where $a = 0$ and $b \approx 100$. The problem arose for $g \sim$log-normal(μ, σ) and $g \sim$log-gamma(ξ, α) in this thesis. This complication is solved by applying a transformation.

If $H(z) = \log\{g(z)/f_\theta(z)\}$, $g(z) \sim$ log-normal(μ, σ) and $Z = e^{\mu+\sigma y}$, (2.1.4) can be rewritten as

$$\int H(z)g(z)dz = \int H(e^{\mu+\sigma y})\phi(y)dy, \tag{4.1.1}$$

where $\phi(y)$ is the Gaussian density with parameters (μ, σ). When $g(z) \sim$ log-gamma(ξ, α), $Z = e^Y - 1$ where $Y = \xi Y_0$ and Y_0 is standard gamma, the same $H(z)$ as above is used and (2.1.4) can be rewritten as

$$\int H(z)g(z)dz = \int H(e^Y - 1)f_{\mathrm{g}}(y)dy, \tag{4.1.2}$$

where $f_{\mathrm{g}}(y)$ is the density of the gamma distribution with expectation ξ and shape parameter α. The standard gamma distribution is defined in Section 5.3.

The limits (a, b) where the latter integrands in (4.1.1) and (4.1.2), are approximately zero above and beneath, depend on the parameters (μ, σ) from the log-normal distribution and (ξ, α) from the log-gamma distribution. However, when $g \sim$log-normal and when $g \sim$log-gamma, $(b - a)$ is smaller when the transformation is used. For the parameter sets (μ, α) and (ξ, α) used in this thesis, the parameters $(\alpha_0, \beta_0, \theta_0)$ stabilize for $N < 30$, when the transformation is applied.

If $g \sim$log-normal$(0, 1)$, the latter integrand in (4.1.1) is approximately zero above $a = -6$ and beneath $b = 6$, and the parameters $(\alpha_0, \beta_0, \theta_0)$ stabilized with $N \approx 25$. The parameter values found by means of Gauss-Legendre quadrature with transformation (4.1.1) are given in Table 4.2. The parameters $(\alpha_0, \beta_0, \theta_0)$ are approximately the same as what was found by means of the Monte Carlo approach. Hence, both approaches work, but the procedure with Gauss-Legendre quadrature converges faster.

Henceforth Gauss-Legendre quadrature with the transformation is used to find the vector of parameters θ_0, for all versions of the log-normal and log-gamma. Clearly, if there is a density distribution with a small value of $(b-a)$ Gauss-Legendre quadrature without the transformation can be applied. However, in this thesis the log-normal and the log-gamma are the only two underlying distributions considered where the true distribution g outside the parametric family; more on this subject in Chapter 5.

4.2 Evaluating the error in $\hat{\theta}$

4.2.1 The delta method

The delta method introduced in Section 2.2, utilizes that $\sqrt{n}(\hat{\theta}-\theta_0)$ is approximately normal. Then, random, independent normal data are drawn and used to evaluate the error in $\hat{\theta}$. In Section 2.3, it is argued that the expected value is zero and that the covariance matrix is $\Lambda^{-1}I_g\Lambda^{-1}$. The terms in Λ and I_g are given in (2.3.2) and (2.3.1) respectively. When $f_\theta \sim$ extended Pareto, Λ and I_g are 3×3 matrices. A numeric illustration is given in Section 4.3.

4.2.2 Maximum likelihood

When estimating the parameters, maximum likelihood is the most common approach. In Chapter 1, it is stated that the maximum likelihood estimator $\hat{\theta}$ converges, also when the assumed distribution is different from the underlying distribution, see Huber [6]. Consequently, $\hat{\theta}$ can be estimated by simulating n claim sizes from the true distribution. These data are used to maximize the extended Pareto log-likelihood function. When $g \sim$ log-normal(0,1) this approach leads to a lot less variability when calculating $\Psi(\hat{\theta})$ compared to when the delta method is applied. A numeric illustration is given in Section 4.3.

4.3 Calculating the reserve

Insurance companies are usually more interested in analysing properties and error of some functional Ψ of g and f_θ, than g and f_θ themselves. An example of such a functional is the reserve. In this thesis the reserve is defined as the amount of money necessary such that the liabilities are covered with probability $(1-\epsilon)\%$, for some solvency criterion ϵ. That is, the reserve is q_ϵ, the upper ϵ-percentile of the portfolio liability.

It is well established that the Poisson distribution with some parameters is a good approximation to the probability distribution of the numbers of claims N. The parameters may themselves be random, or they are constants. It is not the objective of this thesis to study the distribution and the assumptions associated with the number of claims. Therefore, the Poisson distribution's parameters $J\mu T$ are always assumed constant. J is the number of individuals in the portfolio, μ is the claim intensity and T is the observation period.

When simulating the reserve, several steps are needed. Algorithm 4.1 and Algorithm 4.2, which are given below, are two procedures that can be used to find estimates of the reserve. Algorithm 4.1 is the direct or intuitive procedure. However, long simulation time is a problem, and loops have to be avoided. Algorithm 4.2 is designed with the objective of avoiding loops, and is therefore the algorithm used in the simulation programs. The R-scripts for the programs are given in Appendix C. The density function h of the claim size Z varies for

different simulations, e.g. when calculating $\Psi(f_{\theta_0})$, $h = f_{\theta_0}$. If there is re-insurance, line 4 in Algorithm 4.1 and line 6 in Algorithm 4.2 are changed; more on this subject in Section 4.6.

Algorithm 4.1: computing the reserve (intuitive procedure)

0	Given vector of parameters θ distribution function $h(z)$ and parameters $J\mu T$.
1	Repeat m times.
2	$N \sim \text{Poisson}(\lambda = J\mu T)$
3	Simulate $Z = (Z_1, ..., Z_N) \sim h(z)$
4	$X = \sum_{i=1}^{N} Z_i$
5	Return $X_1, ..., X_m$
6	Sort $X_1, ..., X_m$ such that $X_{(1)} \geq ... \geq X_{(m)}$
7	$q_\epsilon = X_{(m\epsilon)}$ The reserve is the upper ϵ-percentile

Algorithm 4.2: computing the reserve (simulation version)

0	Given vector of parameters θ distribution function $h(z)$ and parameters $J\mu T$.
1	Simulate $N = (N_1, ..., N_m) \sim \text{Poisson}(\lambda = J\mu T)$
2	maxN=max(N) The largest value of the $1 \times m$ matrix N.
3	A= T[matrix(rep(c(1:maxN),m),maxN,m)] A matrix with m rows, each row containing the numbers 1:maxN.
4	B=matrix(rep(N,maxN),m,maxN) A matrix with one row per simulation. Row i containing the number $N[i]$ repeated maxN times.
5	I=t(B-A≥0) A maxN$\times m$ identity matrix. Each simulation have one column. For simulation i the first $N[i]$ entries are 1, the rest are 0.
6	z=matrix(h(maxN*m, z), maxN,m) maxN$\times m$ matrix, each element containing a random simulation from the $h(z)$ distribution.
7	Z=z*I Matrix with one simulation per column. For a simulation i the first $N[i]$ elements contain simulations of $h(z)$, the rest are 0.
8	X=apply(Z,2,sum) $1 \times m$ vector with values of $X_j = \sum_{i=1}^{N_j} Z_i + \sum_{N_j}^{\text{maxN}} 0$, $j = 1, ..., m$
9	return $X_1, ..., X_m$
10	sort $X_1, ..., X_m$ such that $X_{(1)} \geq ... \geq X_{(m)}$
11	$q_\epsilon = X_{(m\epsilon)}$ The reserve is the upper ϵ-percentile.

The functionals $\Psi(\theta_0)$ and $\Psi(g)$ are constants, while $\Psi(\hat{\theta})$ varies with the number of data n. As an example $\Psi(\theta_0)$ and $\Psi(g)$ are calculated. As before, $g \sim$ log-normal(0,1). The parameters in the Poisson distribution $J\mu T$ is set to 300. The output is

$$\Psi(g) = 613.3 \quad \text{and} \quad \Psi(\theta_0) = 678.7.$$

Table 4.3: *The expected value and the standrd deviation of the estimated reserve, calculated by means of the delta method and maximum likelihood.*

	$\Psi(g) = 613.3$		$\Psi(\theta_0) = 678.7$	
	The delta method		Maximum likelihood	
	mean	sd	mean	sd
$n = 50$	-	-	967.4	1078.2
$n = 100$	-	-	747.5	282.6
$n = 1000$	1603.6	6443.2	676.7	49.7
$n = 10000$	690.0	108.8	680.1	16.4
$n = 100000$	681.0	33.5	678.4	7.9

Clearly, $\Psi(g) < \Psi(\theta_0)$, and we can conclude that when g ∼log-normal(0,1), the extended Pareto distribution tends to overestimate the reserve. Systematic error is $|613.7 - 681.3| = 65.4$, and can not be neglected.

In Section 4.2, two possible methods for evaluating error in the vector of parameters $\hat{\theta} = (\hat{\alpha}, \hat{\beta}, \hat{\theta})$ are described, i.e. the delta method and maximum likelihood. Here $\Psi(\hat{\theta})$ is calculated with both methods. In order to reflect that there are different amounts of available data in different non-life insurance branches, n is varied between 50 and 100000. The other assumptions are kept constant with g ∼log-normal(0,1) and $J\mu T = 300$. For each n, the simulations are repeated M times, and the mean and the standard deviation of $\Psi(\hat{\theta})$ are calculated. The output is given in Table 4.3.

When $g \sim$ log-normal(0,1), the terms in the covariance matrix $\Lambda^{-1} I_g \Lambda^{-1}$ are large, especially when $n \leq 100$. This again results in large variability in the parameter estimate $(\hat{\alpha}, \hat{\beta}, \hat{\theta})$ found by means of the delta method. Some simulations returns negative values of $\hat{\alpha}$. Then $\Psi(\hat{\theta})$ can not be estimated. Thus, there is no value for $\Psi(\hat{\theta})$ when $n \leq 100$ and the delta method is applied. As expected $\Psi(\hat{\theta})$ converge to $\Psi(\theta_0)$ when $n \to \infty$ with both methods. However, from Table 4.3 it is apparent that the convergence is much faster when maximum likelihood is applied. The standard deviations are also a lot smaller. Consequently, maximum likelihood is used to estimate $\Psi(\hat{\theta})$ in Chapter 5.

4.4 Expected total error; a numerical illustration

In Section 2.4.2 the expected value of the square of total error $\mathrm{E}[\{\Psi(\hat{\theta}) - \Psi(g)\}^2]$, is decomposed into three terms,

$$\begin{aligned} \mathrm{A1} &= \mathrm{E}[\{\Psi(\hat{\theta}) - \Psi(\theta_0)\}^2] \\ \mathrm{A2} &= \{\Psi(\theta_0) - \Psi(g)\}^2 \text{ and} \\ \mathrm{A3} &= 2\{\mathrm{E}[\Psi(\hat{\theta})] - \Psi(\theta_0)\}\{\Psi(\theta_0) - \Psi(g)\}. \end{aligned}$$

The three terms are analysed separately. The first and the third term are of order n^{-1}, while the second is of order 0. One question asked is whether it is possible to look at A3 as a small remainder term that can be neglected. If so, $\mathrm{E}[\{\Psi(\hat{\theta}) - \Psi(g)\}^2]$ can be approximately by,

$$\mathrm{E}[\{\Psi(\hat{\theta}) - \Psi(g)\}^2] \approx \mathrm{E}[\{\Psi(\hat{\theta}) - \Psi(\theta_0)\}^2] + \{\Psi(\theta_0) - \Psi(g)\}^2.$$

Table 4.4: *The terms A1, A2 and A3 from* (2.4.4) *calculated for different values of* n.

	A1	A2	A3
$n = 50$	1714190	4277	43255
$n = 100$	114994	4277	12351
$n = 1000$	2692	4277	-625
$n = 10000$	300	4277	-93
$n = 100000$	60	4277	2

To study the terms and get an idea of how they vary with n, simulations are performed. Given values of $\Psi(\theta_0)$ and $\Psi(g)$, the same program that is used to estimate $\Psi(\hat{\theta})$, is used to simulated the three terms, A1, A2 and A3. The assumptions and parameters are the same as in Section 4.3. The results from the simulation are given in Table 4.4.

A1 is the dominating term when n is small, but the value decreases fast when n becomes larger. The absolute value of A3 decreases with n as well. For $n = 10^5$, A1 and A3 are still not negligible. As mentioned in Section 2.4.2, this result indicates that A3 can not be neglected when $n < 10^5$. In non-life insurance both large and small datasets occur. Datasets with more than 10^5 observations are rare and situations with $n > 10^5$ are therefore not studied. Hence, the third term A3 in (2.4.4) is not neglected in this thesis.

4.5 Quantities used to analyse a default distribution

It is of interest to study how systematic and estimation error depend on different assumptions, and how large they are compared to each other for different values of n. To analyse this, expected total, expected estimation and systematic error defined as

$$\begin{aligned}\mathcal{E}_{\text{Tot}} &= [\text{E}(\{\Psi(\hat{\theta}) - \Psi(g)\}^2)]^{1/2},\\ \mathcal{E}_{\text{Ran}} &= [\text{E}(\{\Psi(\hat{\theta}) - \Psi(\theta_0)\}^2)]^{1/2} \text{ and}\\ \mathcal{E}_{\text{Sys}} &= [\{\Psi(\theta_0) - \Psi(g)\}^2]^{1/2} = |\Psi(\theta_0) - \Psi(g)|\end{aligned} \tag{4.5.1}$$

respectively, are compared. Different specifications of the underlying density function g, different choices for amount of data n and inclusion of reinsurance are considered in Chapter 5. When $\mathcal{E}_{\text{Tot}}$ and $\mathcal{E}_{\text{Ran}}$ are defined as above, simulations which leads to large divergence from $\Psi(g)$ or $\Psi(\theta_0)$ boosts the result. Then, the difference between stable and unstable estimates of the functional are seen more clearly. The value of the error terms alone might not give a good understanding of the actual importance of systematic, estimation or total error. When $\Psi(g)$ is large, a larger value of $\mathcal{E}_{\text{Sys}}$ might for instance be acceptable. Therefore, in order to get a more nuanced impression of the importance of the error terms, the ratios

$$\text{R}_{\text{Tot}} = \frac{\mathcal{E}_{\text{Tot}}}{\Psi(g)}, \quad \text{R}_{\text{Ran}} = \frac{\mathcal{E}_{\text{Ran}}}{\Psi(g)} \quad \text{and} \quad \text{R}_{\text{Sys}} = \frac{\mathcal{E}_{\text{Sys}}}{\Psi(g)} \tag{4.5.2}$$

are also computed and analysed. Another useful quantity when analysing the appropriateness of the extended Pareto distribution as default loss model, is the expected estimated value of the functional, defined as

$$\mathcal{E}_{\hat{\theta}} = \text{E}[\Psi(\hat{\theta})]. \tag{4.5.3}$$

The error terms $\mathcal{E}_{\text{Tot}}, \mathcal{E}_{\text{Ran}}$ and $\mathcal{E}_{\text{Sys}}$, the ratios R_{Tot}, R_{Ran} and R_{Sys} and the expected value of the estimated reserve $\mathcal{E}_{\hat{\theta}}$ are in Chapter 5 used to analyse the suitability of the extended Pareto distribution as default loss model in non-life insurance. Algorithm 4.3 gives the procedure used to find these quantities.

Algorithm 4.3: computing error terms, ratios and the expected estimated reserve.

0	Given $J\mu T$, n, $\Psi(g)$ and $\Psi(\theta_0)$	
1	Repeat M times	
2	Simulate $Z = (Z_1, ..., Z_n) \sim g(z)$	
3	Compute $\hat{\theta}$ given Z	$\hat{\theta}$ is found by maximizing the log-likelihood function of $f_\theta(z)$ with respect Z.
4	Compute $\Psi(\hat{\theta}) = X_{(m\epsilon)}$	See algorithm 4.1 or 4.2, $h = f_{\hat{\theta}}$
5	Return:	
	$\mathcal{E}_{\text{Tot}} = \sqrt{\frac{1}{M}\sum_{i=1}^{M}[\Psi(g) - \Psi_i(\hat{\theta})]^2}, \mathcal{E}_{\text{Ran}} = \sqrt{\frac{1}{M}\sum_{i=1}^{M}[\Psi_i(\hat{\theta}) - \Psi(\theta_0)]^2},$	
	$\mathcal{E}_{\text{Sys}} = \lvert\Psi(g) - \Psi(\theta_0)\rvert$, $\text{R}_{\text{Tot}} = \mathcal{E}_{\text{Tot}}/\Psi(g)$, $\text{R}_{\text{Ran}} = \mathcal{E}_{\text{Ran}}/\Psi(g)$,	
	$\text{R}_{\text{Sys}} = \mathcal{E}_{\text{Sys}}/\Psi(g)$ and $\mathcal{E}_{\hat{\theta}} = \frac{1}{M}\sum_{i=1}^{M}\Psi_i(\hat{\theta})$	

In Table 4.5, the error terms $\mathcal{E}_{\text{Tot}}, \mathcal{E}_{\text{Ran}}$ and $\mathcal{E}_{\text{Sys}}$, the ratios R_{Tot}, R_{Ran} and R_{Sys} and the expected value of the estimated reserve $\mathcal{E}_{\hat{\theta}}$ are computed. The assumptions are the same as those applied earlier in this chapter, i.e. $g \sim$log-normal(0,1) and $J\mu T = 300$.

From Section 4.3, it is already clear that the reserve is overestimated. From Table 4.5, we see that the amount of overestimation increases when n decreases. Estimation error is large and dominates systematic error when n is small. However, when n increases estimation error converges to zero. Total error is also decreasing with n. However, due to systematic error, it converges to 65 and not to 0. In Chapter 5, several analysis similar to this one are preformed. The possibility of using the extended Pareto distributio as parametric family is studied more precisely.

Note that when $n = 10000$, $\mathcal{E}_{\text{Tot}} < \mathcal{E}_{\text{Sys}}$, which is a Monte Carlo error. The number of Monte Carlo simulations m are increased in Chapter 5. Therefore such errors are less likely to appear there.

Table 4.5: *The error terms, the ratios and the expected value of the estimated reserve, calculated for different values of n when g ∼log-normal(0,1).*

	$\Psi(g) = 613$			$\Psi(\theta_0) = 679$			
	$\mathcal{E}_{\text{Tot}}$	$\mathcal{E}_{\text{Ran}}$	$\mathcal{E}_{\text{Sys}}$	R_{Tot}	R_{Ran}	R_{Sys}	$\mathcal{E}_{\hat{\theta}}$
$n = 50$	4165	4155	65	6.79	6.77	0.11	967
$n = 100$	274	251	65	0.45	0.41	0.11	748
$n = 1000$	83	55	65	0.14	0.09	0.11	677
$n = 10000$	67	18	65	0.10	0.03	0.11	680
$n = 100000$	66	7	65	0.11	0.01	0.11	678

4.6 Reinsurance

Reinsurance is insurance that is purchased by an insurance company, the cedent, from one or more other insurance companies, the reinsurers, as a means for risk management. The reinsurer is paid some proportion of the premium and takes on some of the risk from the cedent. There are different reinsurance agreements. The reinsurer might pay some portion c of the total claims incurred. Then, if the total amount is Z the cedent only pays $Z^{\text{ce}} = (1-c)Z$. In this thesis, a reinsurance agreement with an upper bound b is considered. Then, the cedent pays everything up to some bound b. If the claim exceeds b, the cedent pays b. That is, a reinsurance agreement defined as

$$Z^{\text{ce}} = \begin{cases} Z & \text{if } Z < b, \\ b & \text{if } Z \geq b. \end{cases} \tag{4.6.1}$$

When the reinsurance agreement is added to the model in this thesis, the parameter estimator $\hat{\theta}$ is not effected. The reinsurance agreement is included when the claim sizes Z are drawn. Therefore $\hat{\theta}$ is estimated as before using maximum likelihood and the original, uncensored data. The vector θ_0 is also found with the same procedure as before. The reinsurance agreement with an upper bound b, is included in algorithms 1a and 1b by simply changing lines 3 and 6 respectively. That is, line 3 in Algorithm 4.1 is changed to

$$Z = (Z_1, ..., Z_n) \sim \min(h(z), b),$$

and line 6 in Algorithm 4.1 becomes

$$z = \text{matrix}(\min(h(\text{maxN} * m, z), b), \text{maxN}, m).$$

In the next section, it is studied how systematic, estimation and total error are influenced when an upper bound b is included and varied.

It is possible to let the parameter estimator $\hat{\theta}$ and the vector θ_0 be influenced by the reinsurance agreement. Then, a new density function $f_\theta^{\text{ce}}(z^{\text{ce}})$ and its associating log-likelihood function have to be calculated. The parameter estimator $\hat{\theta}$ is found by optimizing this new log-likelihood function. The vector θ_0 is found by minimizing the Kullback-Leibler distance $D_{\text{KL}}(f_\theta^{\text{ce}}|g)$. In order to find the new density function, the cumulative density function $F_\theta(z)$ is needed. When f_θ ∼extended Pareto, there is no explicit expression for $F_\theta(z)$. It can be estimated by means of numerical integration. This has however not been attempted in this thesis.

Chapter 5

Different underlying models

In this chapter, four different models are used as the underlying density model g. In Section 5.2 $g \sim$ log-normal, in Section 5.3 $g \sim$ gamma, in Section 5.4 $g \sim$ Pareto distributed and in Section 5.5 $g \sim$ log-gamma. In all four sections, three different sets of parameters defining the density distribution are applied, e.g. (μ, σ) in the log-normal distribution. These parameter values are chosen with the objective of making the results comparable. A discussion of how the parameter sets are found is given in Section 5.1. In order to make the simulations run faster, the parameters in the Poisson distribution $J\mu T$ is reduced from 300 to 50 in this chapter.

5.1 Finding the parameter sets

The mean value and the standard devaiation are natural choices of quantities to match, when trying to obtain a basis for comparison between the distributions. For the Pareto distribution the standard deviation is always larger than the mean. For the first parameter set, parameter set 1, $\text{sd}[Z] > \text{E}[Z]$. Thus, it is unproblematic to achieve correspondence between all four distributions in this situation. For parameter set 2, $\text{sd}[Z] > \text{E}[Z]$, but $\text{sd}[Z] \approx \text{E}[Z]$, which leads to very large values of the parameters that define the Pareto distribution. For parameter set 3, $\text{sd}[Z] < \text{E}[Z]$. Consequently, only the mean value of the Pareto distribution match the others when parameter set 2 and when parameter set 3 are considered. For log-normal, gamma and log-gamma distribution both the mean and the standard deviations correspond for all three parameter sets.

For the log-normal, gamma and Pareto distribution there exists explicit expressions for the means and for the standard deviations. For the log-gamma distribution finding explicit expressions for these quantities is advanced and is not attempted in this thesis. Thus, in order for the means and the standard deviations to be equal, some combination of the parameters (ξ, α) in the log-gamma distribution are chosen. Then, the mean and standard deviation for the log-gamma distribution are calculated numerically. This is done by drawing m independent log-gamma distributed random variables $Z_1, ..., Z_m$ given (ξ, α). Then, the sample mean and the sample standard deviation given by

$$\bar{Z} = \frac{1}{m}\sum_{i=1}^{m} Z_i \quad \text{and} \quad S = \left[\frac{1}{m-1}\sum_{i=1}^{m}(Z_i - \bar{Z})^2\right]^{\frac{1}{2}},$$

are calculated, see Devore and Berk ([5], page 25 and 311). In the next sections explicit expressions for the means and the standard deviations for the log-normal, gamma and Pareto distribution are given in terms of the parameters, e.g. in terms of (μ, σ) for the log-normal

Table 5.1: *Parametervalues used in the simulations for the four underlying distributions.*

	Log-normal (μ, σ)	gamma (ξ, α)	Log-gamma (ξ, α)	Pareto (α, β)
Set 1	(0.22, 1.13)	(2.37, 0.38)	(1.00,3.00)	(3.21, 5.26)
E[Z]	2.37	2.37	2.37	2.37
sd[Z]	3.84	3.84	3.84	3.84
c_v	1.61	1.61	1.61	1.61
Set 2	(1.12, 0.85)	(4.41, 0.94)	(1.50,7.00)	(4.00, 13.23)
E[Z]	4.41	4.41	4.41	4.41
sd[Z]	4.55	4.55	4.55	6.23
c_v	1.03	1.03	1.03	1.41
Set 3	(0.41, 0.78)	(2.05, 1.19)	(1.00,5.00)	(5.00, 8.21)
E[Z]	2.05	2.05	2.05	2.05
sd[Z]	1.88	1.88	1.88	2.65
c_v	0.91	0.91	0.91	1.30

distribution. Parameter values in the other distributions are calculated by connecting the explicit expressions for the expected values and standard deviations to the values $\bar{Z}$ and S and solve for the parameters.

In Table 5.1 the three parameter sets that are studied for each of the four distributions are given. The expected values and standard deviations given these parameter sets are also listed. To make it easier to compare the results from the different sets, the coefficient of variation defined as

$$c_v = \frac{\text{sd}[Z]}{\text{E}[Z]}$$

is computed, see Brown ([2], page 155). The coefficient of variation is a normalized measure of dispersion.

5.2 The log-normal distribution

The density function of a log-normal random variable with parameters (μ, σ) is

$$f(z; \mu, \sigma) = \frac{1}{z\sigma\sqrt{2\pi}} e^{-\frac{(\log z - \mu)^2}{2\sigma^2}}, \quad z > 0.$$

The expectation and standard deviation are

$$\text{E}[Z] = e^{\mu + \frac{\sigma^2}{2}} \quad \text{and} \quad \text{Sd}[Z] = \text{E}[Z]\sqrt{e^{\sigma^2} - 1},$$

see Kleiber and Kotz ([10], page 107-112). The log-normal distribution is outside the extended Pareto family. That is, no matter how the parameters (α, β, θ) are changed, $f_\theta(x)$ is not equivalent to the log-normal density. Thus, there is both systematic and estimation error present when the extended Pareto distribution is used as default loss model and the true distribution of the data is the log-normal distribution.

In Table 5.2 the error terms, the ratios and the expected value of the estimated reserve defined

Table 5.2: *The error terms, the ratios and the expected value of the estimated reserve, calculated for different values of n when g ∼log-normal(μ, σ).*

	n	$\mathcal{E}_{\text{Tot}}$	$\mathcal{E}_{\text{Ran}}$	$\mathcal{E}_{\text{Sys}}$	R_{Tot}	R_{Ran}	R_{Sys}	$\mathcal{E}_{\hat{\theta}}$
Set 1	50	977	956	79	4.65	4.55	0.38	511
$\Psi(g) = 210$	100	365	338	79	1.74	1.61	0.38	368
$\Psi(\theta_0) = 289$	1000	99	50	79	0.47	0.24	0.38	296
	10000	81	15	79	0.38	0.07	0.38	289
	100000	79	5	79	0.38	0.02	0.38	289
Set 2	50	196	190	20	0.58	0.56	0.06	402
$\Psi(g) = 337$	100	102	96	20	0.30	0.28	0.06	378
$\Psi(\theta_0) = 357$	1000	29	20	20	0.09	0.06	0.06	358
	10000	20	6	20	0.06	0.02	0.06	357
	100000	19	2	20	0.06	0.01	0.06	357
Set 3	50	56	55	4	0.37	0.36	0.03	169
$\Psi(g) = 153$	100	30	29	4	0.20	0.19	0.03	163
$\Psi(\theta_0) = 157$	1000	9	7	4	0.06	0.04	0.03	158
	10000	5	2	4	0.03	0.01	0.03	158
	100000	5	1	4	0.03	0.01	0.03	157

in (4.5.1), (4.5.2) and (4.5.3) respectively, are given for the three different parameter sets from Table 5.1. The error terms and ratios address the importance and the relative importance of the different types of error. The expected value of the estimated reserve $\mathcal{E}_{\hat{\theta}}$, illustrates in which manner $\mathcal{E}_{\hat{\theta}}$ converges to $\Psi(\theta_0)$. In Figure 5.1, the underlying density function $g(z)$ and the best achievable density function $f_{\theta_0}(z)$ are plotted for the three parameter sets from Table 5.1. In the upper row, $z \in (0, 10]$ and in the lower row $z \in [100, 200]$. Thus a small part of the tails are visible. In Figure 5.2, $f_{\theta_0}(z)$ is plotted together with five realisations of the estimated density function $f_{\hat{\theta}}(z)$ when parameter set 1 is considered. In the upper row, $z \in (0, 10]$ and in the lower row $z \in [100, 200]$. To the left $n = 50$ data from the true distribution are used to find the vector of parameters estimates $\hat{\theta}$. By drawing five sets of historical data, five realisations of the vector $\hat{\theta}$ and five density functions $f_{\hat{\theta}}(z)$ are estimated. In the middle $\hat{\theta}$ is estimated five times when $n = 1000$ and to the right when $n = 100000$. The figure illustrates the variation in the density function $f_{\hat{\theta}}(z)$ when n is small compared to when it is large. The figure will vary between simulation and does therefore merely illustrate trends. Corresponding figures for parameter sets 2 and 3 are given in Appendix A.

From Table 5.2, it is apparent that for all three parameter sets, $\Psi(g) < \Psi(\theta_0)$. Thus, the extended Pareto distribution tends to overestimate the reserve. From Table 5.2 it is also clear that systematic error is largest when the log-normal density is characterised by parameter set 1. That is, when the variation in the data compared to the mean is largest. From Figure 5.1, we see that for $z \in (0, 10]$ the differences between the density functions are small. By studying the lower row of the figure, it is evident that the best achievable density function $f_{\theta_0}(z)$ has heavier tails than the log-normal density $g(z)$. The divergence is largest when parameter set 1 is considered and is barely noticeable when parameter set 3 is considered. Thus, it seems likely that the heavy tails of the fitted extended Pareto density is the main cause of overestimation.

As would be expected, random error decreases when c_v decreases and when n increases. When c_v is small, the variability in the data compared to the mean is small, and the data is concentrated in a smaller range. The number of data needed to fit a good model is therefore

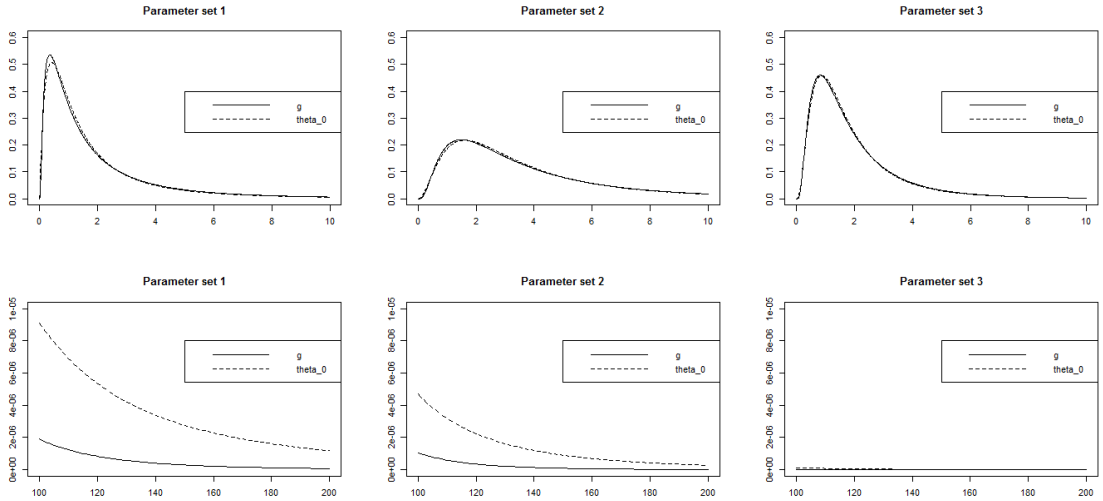

Figure 5.1: *The density functions $g(z)$ and $f_{\theta_0}(z)$ for the three parameter sets from Table 5.1, $g \sim$log-normal. Upper row: $z \in (0, 10]$. Lower row: $z \in [100, 200]$*

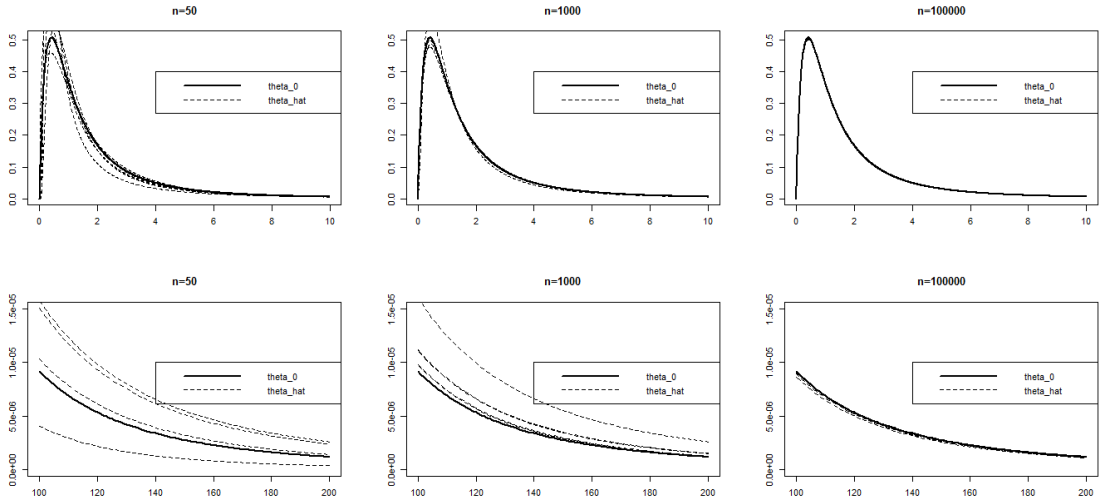

Figure 5.2: *Density functions $f_{\theta_0}(z)$ and $f_{\hat{\theta}}(z)$ for parameter set 1 from Table 5.1, $g \sim$log-normal. Upper row: $z \in (0, 10]$. Lower row: $z \in [100, 200]$.*

smaller when c_v is small. Thus, the importance of random error is smaller and converges faster to zero in parameter sets 3 and 2 compared to parameter set 1. In (2.4.3) in Section 2.4, it is stated that random error $\{\Psi(\hat{\theta}) - \Psi(\theta_0)\}$ is normally distributed with mean 0 and a standard deviation of order $n^{-1/2}$. Consequently, estimation error converges to zero as n increases.

From Table 5.2 we see that $\mathcal{E}_{\hat{\theta}} \geq \Psi(\theta_0)$ and that $\mathcal{E}_{\hat{\theta}} \to \Psi(\theta_0)$ as n increases. This tendency is also visible by studying Figure 5.2. The figure suggests that $f_{\hat{\theta}}(z)$ varies more between simulations when $n = 50$. The density functions are more stable when $n = 1000$, and when $n = 100000$ the estimated density function is almost identical to the best approximation $f_{\theta_0}(z)$ for all five simulations. The deviations from $f_{\theta_0}(z)$ seems to be especially large in the tail. From the lower row of the figure wee see that for these five simulations, the tail of $f_{\hat{\theta}}(z)$ is sometimes less heavy, but generally heavier than the tail of $f_{\theta_0}(z)$. Thus, as seen in Table 5.2, for limited choices of n the estimated reserve is on average larger than $\Psi(\theta_0)$. As n increases $f_{\hat{\theta}}(z)$ converges to $f_{\theta_0}(z)$ for both large and small values of z. The corresponding figures for parameter sets 2 and 3 are given in Appendix A, illustrates the same tendencies. Hence, small values of n, leads to even large amounts of overestimation of the reserve on average. Total error decreases when n increases and when c_v decreases. However, due to the bias, $\mathcal{E}_{\text{Tot}} \to \mathcal{E}_{\text{Sys}}$ as $n \to \infty$, and not to zero.

Inclusion of an upper bound b

The extended Pareto distribution tends to overestimate the reserve. As seen in Figure 5.1 and Figure 5.2, the main source of overestimation is the heavy tails of the fitted extended Pareto distribution. A reinsurance agreement like the one defined in equation (4.6.1) in Section 4.6, reduces the importance of this incorrection. Then, every claim Z larger than the upper bound b is censored and set equal to b. The divergence from the true reserve $\Psi(g)$ is most severe when parameter set 1 is applied. Therefore, including the reinsurance agreement, has greatest impact on the outcome when parameter set 1 is considered. In Table 5.3, the output is given when parameter set 1 is considered and for the upper bound $b = 100, 50, 25$ and 10 and for $b = \infty$, i.e. when there is no reinsurance. Note that $\mathrm{E}[Z] = 2.37$ and $\mathrm{sd}[Z] = 3.84$. Thus, only extreme values of Z are censored.

In Table 5.3 the new values of the error terms, the ratios and the expected value of the estimated reserves are given. As would be expected both systematic and estimation error are decreasing with b. When $b = 100$ the actual reserve $\Psi(g)$ is approximately the same as it is without reinsurance, while $\Psi(\theta_0)$ is considerably reduced. Hence, virtually no claims are above 100 when the claims are log-normally distributed, while with the extended Pareto distribution some claims are above 100, and are censored when $b = 100$. Consequently, $\Psi(\theta_0)$ is reduced, $\Psi(g)$ is unchanged, and the difference between them is smaller. As b decreases the heavy tails of the extended Pareto distribution are less important. When $b = 10$, there is hardly any systematic error left. From Figure 5.1 we see that the density functions $g(z)$ and $f_{\theta_0}(z)$ are almost identical for $z \in (0, 10]$. Consequently the difference between $\Psi(\theta_0)$ and $\Psi(g)$ is insignificant when $b = 10$.

From Table 5.3 it is apparent that random error is decreasing with b. As discussed in Section 4.6, the estimator $\hat{\theta}$ is not effected by the inclusion of the reinsurance agreement. Consequently, the decline in random error is not due to the vector $\hat{\theta}$ converging to θ_0. From Figure 5.2, it seams likely that the estimated density function has heavier tails than $f_{\theta_0}(z)$ on average. Hence, extreme values of Z are more frequently drawn when $f_{\hat{\theta}}(z)$ is the density function. An inclusion of an upper limit b reduces estimation error, because the importance

Table 5.3: *The error terms, the ratios and the expected value of the estimated reserve, calculated for different values of n when $g \sim$log-normal$(1.12, 0.85)$ and a reinsurance limit b is added.*

	n	$\mathcal{E}_{\text{Tot}}$	$\mathcal{E}_{\text{Ran}}$	$\mathcal{E}_{\text{Sys}}$	R_{Tot}	R_{Ran}	R_{Sys}	$\mathcal{E}_{\hat{\theta}}$
$b = \infty$	50	977	956	79	4.65	4.55	0.38	511
$\Psi(g) = 210$	100	365	338	79	1.74	1.61	0.38	368
$\Psi(\theta_0) = 289$	1000	99	50	79	0.47	0.24	0.38	296
	10000	81	15	79	0.38	0.07	0.38	289
	100000	79	5	79	0.38	0.02	0.38	289
$b = 100$	50	88	80	34	0.42	0.40	0.17	246
$\Psi(g) = 209$	100	67	57	34	0.32	0.29	0.17	244
$\Psi(\theta_0) = 243$	1000	39	18	34	0.19	0.09	0.17	243
	10000	35	6	34	0.17	0.03	0.17	243
	100000	34	2	34	0.17	0.01	0.17	243
$b = 50$	50	58	56	16	0.29	0.28	0.08	216
$\Psi(g) = 201$	100	43	40	16	0.21	0.20	0.08	217
$\Psi(\theta_0) = 217$	1000	20	13	16	0.10	0.06	0.08	217
	10000	16	4	16	0.08	0.02	0.08	217
	100000	16	1	16	0.08	0.01	0.08	217
$b = 25$	50	38	38	5	0.20	0.20	0.03	192
$\Psi(g) = 189$	100	28	27	5	0.15	0.14	0.03	193
$\Psi(\theta_0) = 194$	1000	10	9	5	0.05	0.05	0.03	194
	10000	5	3	5	0.03	0.01	0.03	194
	100000	5	1	5	0.02	0.01	0.03	194
$b = 10$	50	23	23	2	0.14	0.14	0.01	160
$\Psi(g) = 163$	100	16	16	2	0.10	0.10	0.01	161
$\Psi(\theta_0) = 161$	1000	5	5	2	0.03	0.03	0.01	161
	10000	2	2	2	0.02	0.01	0.01	161
	100000	2	1	2	0.01	0.00	0.01	161

of the extreme values is reduced.

For completeness, tables similar to Table 5.3 for parameter set 2 and parameter set 3 and for $b = (15, 30, 60, 120)$ and $b = (7, 15, 25, 50)$ respectively, are given in Appendix A. The tendencies are the same. That is, including an upper limit b reduces both systematic and random error and the amount of overestimation is considerably reduces as b decreases.

Note that for $b = 50$ and $b = 25$, $\mathcal{E}_{\hat{\theta}}$ is closer to $\Psi(g)$ when n is small compared to when n is large or infinite. The estimates are however very unstable when $n = 50$ and $\mathcal{E}_{\text{Ran}}$ is still larger when n is small. Thus, the average estimated reserve might be closer to the actual value when n is small, but there is large variation and error in the estimate. The reason for $\mathcal{E}_{\hat{\theta}}$ being closer to $\Psi(g)$ than $\Psi(\theta_0)$ when $b = 50$ or 25 and n is small, is visible in Figure 5.2, and the corresponding figures for parameter sets 2 and 3 given in appendix A. The variation in $f_{\hat{\theta}}(z)$ is large when n is small, and thus the variation in $\Psi(\hat{\theta})$ is large. When $n = 50$, some density functions $f_{\hat{\theta}}(z)$ to the left and many to the right of $f_{\theta_0}(z)$ are simulated. Consequently, many large and some small claim sizes Z are drawn. This results in many large and some small estimates of $\Psi(\hat{\theta})$. When b decreases, more and more of the large draws of Z are censored, and hence the average decreases. Because nothing happens to the abnormally small values of Z, a small enough b produces estimates of $\mathcal{E}_{\hat{\theta}}$ below $\Psi(\theta_0)$ when n is small. Thus, since

$\Psi(g) < \Psi(\theta_0)$ when $b = 50$ and when $b = 25$, $\mathcal{E}_{\hat{\theta}}$ is closer to $\Psi(g)$ for small values of n.

Hence, when g ∼log-normal, the extended Pareto distribution tends to overestimate the reserve. This is mainly due to its overestimation of the tail of the density distribution. Total error decreases, but does not disappear as $n \to \infty$. The problem is smaller when c_v is small, i.e. when the variation in the data compared to the mean is small. The problem is also smaller when an upper limit b is added. Systematic error can be ignored when the upper bound $b \approx 25$. Estimation error is also strongly reduced when an upper limit is added.

5.3 The gamma distribution

The density function of a gamma distributed stochastic variable is often defined as

$$f(z) = \frac{\lambda^\alpha}{\Gamma(\alpha)} z^{\alpha-1} e^{-\lambda z}, \quad z \geq 0, \tag{5.3.1}$$

where $\lambda > 0$ is the rate parameter and $\alpha > 0$ is the shape parameter, see Rice ([14], page 53). The mean and standard deviation are given by

$$\mathrm{E}[Z] = \frac{\alpha}{\lambda} \quad \text{and} \quad \mathrm{Sd}[Z] = \frac{\sqrt{\alpha}}{\lambda},$$

see Rice ([14], page 157). A gamma distributed random variable Z can also be defined in terms of a standard gamma distributed random variable Z_0, i.e. a gamma distribution with mean equal to one. In Appendix B it is shown that the density function of a standard gamma distributed random variable Z_0 is

$$f_0(z_0) = \frac{\alpha^\alpha}{\Gamma(\alpha)} z_0^{\alpha-1} e^{-\alpha z_0}, \quad z_0 > 0, \tag{5.3.2}$$

where α is the same shape parameter as in (5.3.1). Then if $Z = \xi Z_0$, Z is gamma distributed with shape parameter α and expectation and standard deviation given by

$$\mathrm{E}[Z] = \xi \quad \text{and} \quad \mathrm{Sd}[Z] = \frac{\xi}{\sqrt{\alpha}}.$$

This is also verified in Appendix B. In Chapter 3 it is stated that an extended Pareto distributed random variable Z, is gamma distributed in the limit. Thus, the gamma distribution is a part of the extended Pareto distribution family, $\Psi(\theta_0) = \Psi(g)$, and there is no systematic error if the underlying model is gamma distributed. No systematic error is the same as stating that when n is close to infinity, there is no error and there is a perfect fit. However, n is generally limited. How well the extended Pareto distribution fits a set of gamma distributed data for limited choices of n, reveals something about the nature of the extended Pareto distribution. Therefore, it is still of interest to study how estimation error changes for different choices of n and c_v.

It is of interest to plot $f_{\hat{\theta}}(z)$ together with $f_{\theta_0}(z) = g(z)$ as done in Figure 5.2 when the underlying distribution is gamma distributed as well. However, as stated above $\alpha \to \infty$, when the extended Pareto distribution converges to the Gamma distribution. Consequently, $\hat{\alpha}$ tend to be very large. A problem occurs because the extended Pareto density, defined in (3.0.1), contains the quantity $\Gamma(\alpha)$. For large values of $\hat{\alpha}$, $\Gamma(\hat{\alpha})$ is to large for a normal computer to handle. A plot with $f_{\hat{\theta}}(z)$ and $g(z)$ is therefore not given in this section.

Table 5.4: *The error terms, the ratios and the expected value of the estimated reserve, calculated for different values of n when g ∼gamma(ξ, α).*

	n	$\mathcal{E}_{\text{Ran}}$	R_{Ran}	$\mathcal{E}_{\hat{\theta}}$
Set 1	50	860	4.23	250
$\Psi(g) = 203$	100	58	0.29	213
	1000	11	0.05	205
	10000	3	0.02	204
	100000	1	0.01	204
Set 2	50	248	0.74	349
$\Psi(g) = 335$	100	38	0.11	338
	1000	11	0.03	336
	10000	4	0.01	335
	100000	2	0.01	335
Set 3	50	31	0.21	156
$\Psi(g) = 152$	100	15	0.10	154
	1000	5	0.03	152
	10000	1	0.01	152
	100000	1	0.00	152

In Table 5.4, $\mathcal{E}_{\text{Ran}}, \text{R}_{\text{Ran}}$ and $\mathcal{E}_{\hat{\theta}}$ are given for the three parameters sets from Table 5.1. The numbers given in the table, are calculated on an remote computer with a large memory which handles large numbers, such as $\Gamma(\hat{\alpha})$. From Table 5.4 we see that the estimated reserve is larger than or equal to the true reserve, i.e. $\Psi(g) \leq \mathcal{E}_{\hat{\theta}}$, for all three parameter sets and for all values of n. Thus, the extended Pareto distribution has a tendency of overestimating the reserve for limited choices of n.

As expected, random error decreases when c_v decreases and when n increases. The amount of overestimation is larger when the variation in the observations is large compared to the mean. Since there is no systematic error, both random and total error converges to zero when n increases. Note that for all three parameter sets, $\mathcal{E}_{\text{Ran}}$ and R_{Ran} decreases a lot from $n = 50$ to $n = 100$. As stated above, the gamma distribution appears in the limit. It seams as if more than $n = 50$ historical data are needed before it converges.

Inclusion of an upper bound b

As for g ∼log-normal, the impact of including a reinsurance agreement with an upper bound b is considered. In Table 5.5, new values of $\mathcal{E}_{\text{Ran}}, \text{R}_{\text{Ran}}$ and $\mathcal{E}_{\hat{\theta}}$ are given when parameter set 1 is considered and for $b = \infty$, i.e. no reinsurance, and $b = 100, 50, 25$ and 10. From the table it is evident that the inclusion of a reinsurance agreement has a positive effect on estimation error. For completeness, tables similar to Table 5.5 when parameter set 2 and parameter set 3 are considered are given in Appendix A. For parameter set 2, $b = (15, 30, 60, 120)$ and for parameter set 3, $b = (7, 15, 25, 50)$ are considered.

As in Section 5.2, $b = 100$ and $b = \infty$ lead to the same value of the true reserve $\Psi(g)$. Even $b = 50$, has no influence on the value of $\Psi(g)$ in this section. Thus, when Z ∼gamma claim sizes substantially greater than 50 are virtually never experienced. When Z has the density of the fitted extended Pareto distribution however, this is far more likely to happen.

Table 5.5: *The error terms, the ratios and the expected value of the estimated reserve, calculated for different values of n when $g \sim gamma(2.37, 0.38)$ and a reinsurance limit b is added.*

	n	$\mathcal{E}_{\text{Ran}}$	R_{Ran}	$\mathcal{E}_{\hat{\theta}}$
$b = \infty$	50	860	4.23	250
$\Psi(g) = 203$	100	58	0.29	213
	1000	11	0.05	205
	10000	3	0.02	204
	100000	1	0.01	204
$b = 100$	50	54	0.27	212
$\Psi(g) = 203$	100	37	0.18	209
	1000	11	0.05	205
	10000	3	0.02	204
	100000	1	0.01	203
$b = 50$	50	47	0.23	208
$\Psi(g) = 203$	100	34	0.16	206
	1000	11	0.05	204
	10000	3	0.02	204
	100000	2	0.01	203
$b = 25$	50	41	0.20	198
$\Psi(g) = 199$	100	29	0.15	199
	1000	10	0.05	200
	10000	3	0.02	199
	100000	1	0.01	199
$b = 10$	50	27	0.16	164
$\Psi(g) = 169$	100	19	0.11	166
	1000	6	0.04	168
	10000	2	0.01	169
	100000	1	0.00	169

Hence, the fitted extended Pareto distribution tends to estimate too heavy tails. An inclusion of an upper limit b removes this source of error, and does therefore estimate a reserve $\Psi(\hat{\theta})$ closer to the actual reserve $\Psi(g)$.

Thus, when $g \sim$gamma with some parameters (ξ, α), the extended Pareto distribution tends to overestimate the reserve when n is limited. The problem decreases with c_v and goes to 0 as $n \to \infty$. Including an upper bound b reduces error, because extreme values of Z are censored. Consequently, a smaller amount of historical data is needed to arrive at the desired low level of estimation error.

5.4 The Pareto distribution

The density function of a Pareto distributed random variable is

$$f(z) = \frac{\alpha/\beta}{(1 + z/\beta)^{1+\alpha}} \quad z > 0,$$

as defined in Chapter 3, see Kleiber and Kotz ([10], page 59). As stated in Chapter 3, the density function is the same as the density function of an extended Pareto distributed random variable with $\theta = 1$ inserted. Hence, the Pareto distribution is part of the extended Pareto

Table 5.6: *The error terms, the ratios and the expected value of the estimated reserve, calculated for different values of n when g ∼Pareto(α, β).*

	n	$\mathcal{E}_{\text{Ran}}$	R_{Ran}	$\mathcal{E}_{\hat{\theta}}$
Set 1	50	1062	5.01	315
$\Psi(g) = 212$	100	117	0.55	237
	1000	20	0.09	214
	10000	6	0.03	212
	100000	2	0.01	212
Set 2	50	1209	3.27	469
$\Psi(g) = 370$	100	128	0.35	395
	1000	26	0.07	372
	10000	8	0.02	370
	100000	3	0.01	370
Set 3	50	144	0.87	191
$\Psi(g) = 166$	100	43	0.26	174
	1000	9	0.06	167
	10000	3	0.02	167
	100000	1	0.01	166

family, and the situation is similar to the situation when the underlying model g is gamma distributed. That is, there is no systematic error, and random error is the interesting part. The expectation and standard deviation of a Pareto distributed random variable are

$$\text{E}[Z] = \frac{\beta}{\alpha - 1} = \xi \quad \text{and} \quad \text{Sd}[Z] = \xi\sqrt{\frac{\alpha}{\alpha - 2}}.$$

In Appendix B, the expressions for the expected value and standard deviation for the extended Pareto distribution are verified. To justify the corresponding quantities for the Pareto distribution, the same argument can be used, but then $\theta = 1$ has to be inserted.

In Table 5.6, $\mathcal{E}_{\text{Ran}}, \text{R}_{\text{Ran}}$ and $\mathcal{E}_{\hat{\theta}}$ are given for the three parameter sets from Table 5.1. In Figure 5.3, $f_{\theta_0}(z) = g(z)$ is plotted against five realisations of $f_{\hat{\theta}}(z)$ when parameter set 1 is considered, as in Figure 5.2. The corresponding figures for parameter set 2 and parameter set 3 are given in Appendix A.

From Table 5.6, we see that the average estimated reserve $\mathcal{E}_{\hat{\theta}}$ is larger than the true value $\Psi(g)$, for all parameter sets when n is small. Hence, for limited choices of n the estimated reserve is on average overestimated. The amount of overestimation is decreasing with c_v and estimation error converges to 0 as $n \to \infty$. This is also visible from Figure 5.3 and the corresponding figures for parameter set 2 and parameter set 3 given in Appendix A. The difference between $f_{\theta_0}(z)$ and $f_{\hat{\theta}}(z)$ is large and varies a lot between simulations when $n = 50$. The figure suggests that the main reason for the overestimation, is that the fitted extended Pareto distribution tends to overestimate the tails when n is small. That is, the probability for large and extreme values of Z, are too high with the fitted extended Pareto density. When $n = 100000$, $\hat{\theta}$ converges to θ_0, and there is almost no difference between $f_{\theta_0}(z)$ and $f_{\hat{\theta}}(z)$.

Inclusion of an upper bound b

In Table 5.7, an upper bound b is added when parameter set 1 is considered. Corresponding tables when parameter set 2 and parameter set 3 are considered are given in Appendix A.

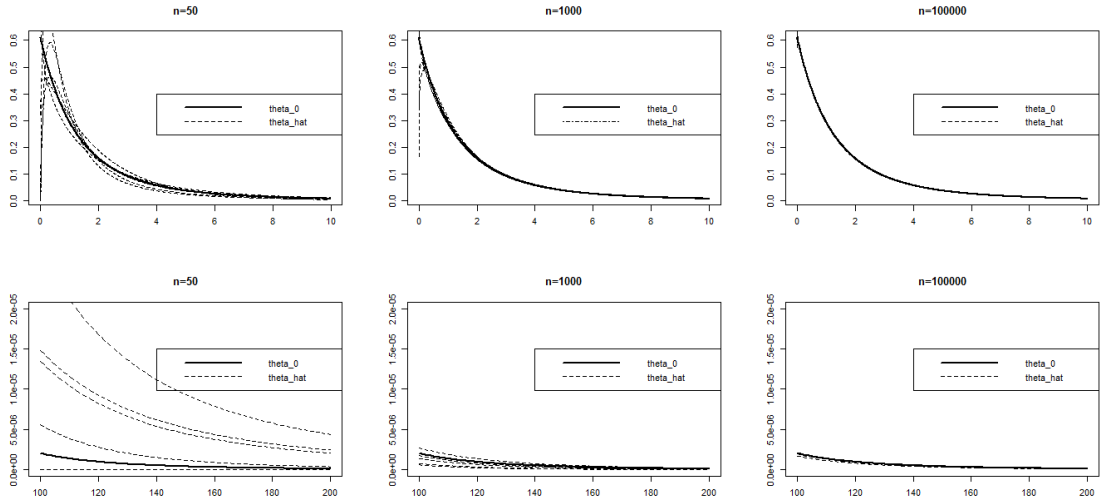

Figure 5.3: *Density functions $f_{\theta_0}(z)$ and $f_{\hat{\theta}}(z)$ for parameter set 1 from Table 5.1, $g \sim$Pareto. Upper row: $z \in (0, 10]$. Lower row: $z \in [100, 200]$.*

Table 5.7: *The error terms, the ratios and the expected value of the estimated reserve, calculated for different values of n when $g \sim$Pareto$(3.21, 5.26)$ and a reinsurance limit b is added.*

	n	$\mathcal{E}_{\text{Ran}}$	R_{Ran}	$\mathcal{E}_{\hat{\theta}}$
$b = \infty$	50	1062	5.01	315
$\Psi(g) = 212$	100	117	0.55	237
	1000	20	0.09	214
	10000	6	0.03	212
	100000	2	0.01	212
$b = 100$	50	68	0.32	216
$\Psi(g) = 210$	100	48	0.23	211
	1000	16	0.08	210
	10000	5	0.02	210
	100000	2	0.01	210
$b = 50$	50	50	0.25	201
$\Psi(g) = 202$	100	36	0.18	201
	1000	12	0.06	201
	10000	4	0.02	202
	100000	1	0.01	202
$b = 25$	50	38	0.20	188
$\Psi(g) = 191$	100	27	0.14	189
	1000	9	0.05	190
	10000	3	0.01	191
	100000	1	0.01	191
$b = 10$	50	24	0.15	164
$\Psi(g) = 166$	100	17	0.10	165
	1000	5	0.03	166
	10000	2	0.01	166
	100000	1	0.00	166

Table 5.8: *The error terms, the ratios and the expected value of the estimated reserve, calculated for different values of n when g ∼log-gamma(ξ, α).*

	n	$\mathcal{E}_{\text{Tot}}$	$\mathcal{E}_{\text{Ran}}$	$\mathcal{E}_{\text{Sys}}$	R_{Tot}	R_{Ran}	R_{Sys}	$\mathcal{E}_{\hat{\theta}}$
Set 1	50	296	295	7	1.38	1.38	0.03	290
$\Psi(g) = 214$	100	127	125	7	0.60	0.58	0.03	251
$\Psi(\theta_0) = 221$	1000	25	24	7	0.12	0.11	0.03	223
	10000	10	7	7	0.04	0.03	0.03	221
	100000	7	2	7	0.03	0.01	0.03	221
Set 2	50	122	120	7	0.36	0.35	0.02	372
$\Psi(g) = 342$	100	77	75	7	0.22	0.22	0.02	363
$\Psi(\theta_0) = 349$	1000	21	19	7	0.06	0.06	0.02	351
	10000	9	6	7	0.03	0.02	0.02	349
	100000	7	2	7	0.02	0.01	0.02	348
Set 3	50	40	40	1	0.26	0.26	0.01	163
$\Psi(g) = 154$	100	25	25	1	0.16	0.16	0.01	159
$\Psi(\theta_0) = 155$	1000	6	6	1	0.04	0.04	0.01	155
	10000	2	2	1	0.01	0.01	0.01	155
	100000	1	1	1	0.01	0.00	0.01	155

The upper bounds are $b = (15, 30, 60, 120)$ for parameter set 2 and $b = (7, 15, 25, 50)$ for parameter set 3. As expected, random error decreases when b is added. From Table 5.7 it is apparent that when $b = 100$, the true reserve $\Psi(g)$ is almost unchanged , while $\mathcal{E}_{\hat{\theta}}$ is considerably reduced. This suggests, as seen in Figure 5.3, that the estimated density $f_{\hat{\theta}}(z)$ has too heavy tails. Consequently, the incorrection is largely reduced when an upper bound b is included. From the upper row in Figure 5.3, we see that for $z \in (0, 10]$ and $n = 50$ there is still divergence of $f_{\hat{\theta}}(z)$ from $f_{\theta_0}(z)$. Hence, including an upper bound b and decreasing it to $b = 10$ has a strong positive effect on estimation error. However, small values of n are still not unproblematic.

Thus, the extended Pareto distribution tends to overestimate the reserve when g ∼Pareto and n is small. The error decreases as c_v increases and goes to zero as $n \to \infty$. Estimation error is strongly reduced when an upper bound b is added because extreme values of Z drawn from the fitted distribution are censored.

5.5 The log-gamma distribution

If Z is log-gamma distributed, then $Y = \log(1 + Z)$ is gamma distributed, and $Z = e^Y - 1$. The density function of a log-gamma distributed random variable is

$$f(z) = \frac{1}{\xi(1+z)} \frac{\alpha^\alpha}{\Gamma(\alpha)} \{\log(1+z)/\xi\}^{\alpha-1} e^{-\alpha \log(1+z)/\xi}, \quad z > 0. \tag{5.5.1}$$

The density is derived in Appendix B. As mentioned, the derivation of the expected value and standard deviation are not conducted. The log-gamma distribution is not a part of the extended Pareto distribution family, hence there is both systematic and estimation error.

In Table 5.8 the quantities defined in (4.5.1), (4.5.2) and (4.5.3) are given for the three parameter sets given in Table 5.1. In Figure 5.4, the density functions $g(z)$ and $f_{\theta_0}(z)$ are plotted for the three parameter sets from Table 5.1. In Figure 5.5, $f_{\theta_0}(z)$ is plotted together

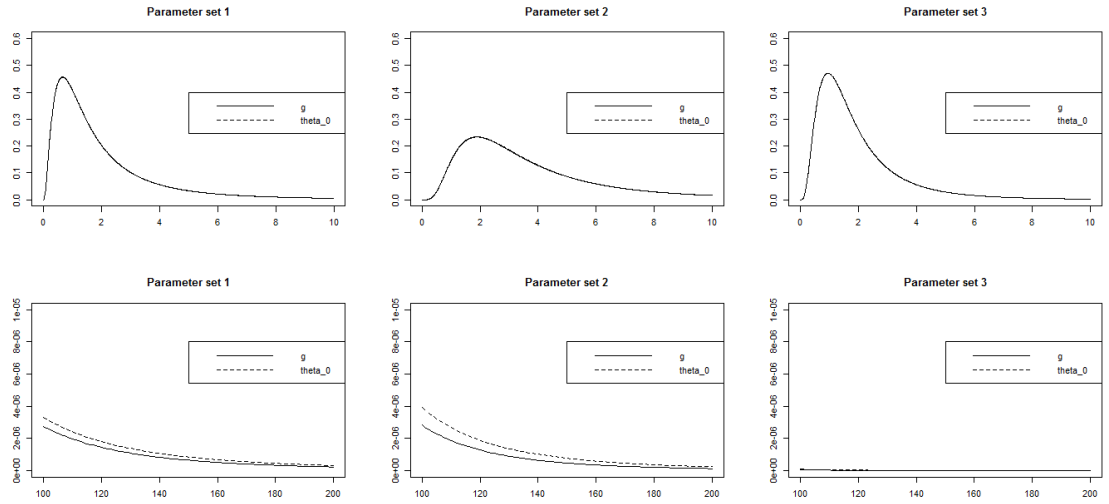

Figure 5.4: *The density functions $g(z)$ and $f_{\theta_0}(z)$ for the three parameter sets from Table 5.1, $g \sim$log-gamma. Upper row: $z \in (0, 10]$. Lower row: $z \in [100, 200]$*

with five realisations of $f_{\hat{\theta}}(z)$ for parameter set 1 and for $n = 50, 1000$ and 100000. Corresponding figures for parameter set 2 and parameter set 3 are given in Appendix A. In all figures $z \in (0, 10]$ in the upper row and $z \in [100, 200]$ in the lower row. Hence, some of the tails are visible in the lower rows.

From Table 5.8 we see that $\Psi(g) < \Psi(\theta_0)$ for all values of (ξ, α). Hence, the extended Pareto distribution tends to overestimate the reserve when $g \sim$log-gamma as well. From Figure 5.4, it is apparent that the density functions $g(z)$ and $f_{\theta_0}(z)$ are almost identical for $z \leq 10$, and that the extended Pareto distribution estimates too heavy tails. The deviation is however much smaller compared to when $g \sim$log-normal, as seen in Figure 5.1. Thus, the log-gamma distribution is not part of the extended Pareto distribution family, but the divergence is much smaller compared to the situation where the underlying distribution is the log-normal distribution. This can also be seen by comparing Table 5.2 and Table 5.8. Systematic error is substantially reduced compared to the situation when $g \sim$log-normal. From Table 5.8, we see that systematic error is negligible for all three parameter sets. However, random error can not be ignored when n is small, but it decreases when c_v decreases and converges to zero as $n \to \infty$. Consequently total error converges to systematic error, and since there is barely any systematic error, total error is negligible in the limit.

In Figure 5.5, the convergence of $f_{\hat{\theta}}(z)$ to $f_{\theta_0}(z)$, in the situation where parameter set 1 is considered, is plotted. As expected the variation in $f_{\hat{\theta}}(z)$ between simulations is large when $n = 50$ and barely visible when $n = 100000$. The figure suggests that small values of n increases the amount of overestimation, because small values of n leads to estimated density functions with tails that on average are heavier than the tails of the best achievable density function $f_{\theta_0}(z)$. Since the tails of $f_{\theta_0}(z)$ are too heavy compared to $g(z)$, small values of n increases the amount of overestimation further. The corresponding figures for parameter set 2 and parameter set 3 given in Appendix A shows the same tendencies.

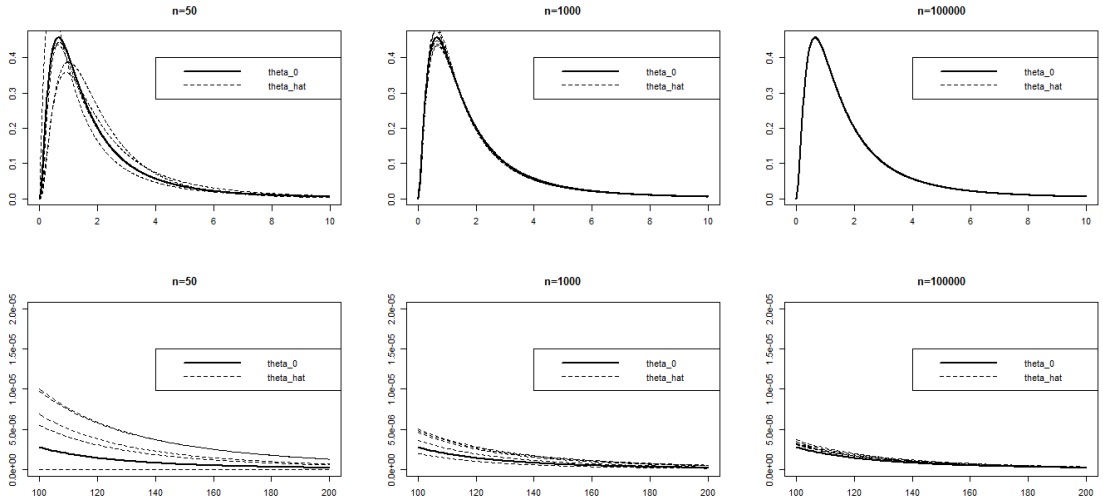

Figure 5.5: *Density functions $f_{\theta_0}(z)$ and $f_{\hat{\theta}}(z)$ for parameter set 1 from Table 5.1, $g \sim$log-gamma. Upper row: $z \in (0, 10]$. Lower row: $z \in [100, 200]$.*

Table 5.9: *The error terms, the ratios and the expected value of the estimated reserve, calculated for different values of n when $g \sim$log-gamma$(1, 3)$ and a reinsurance limit b is added.*

	n	$\mathcal{E}_{\text{Tot}}$	$\mathcal{E}_{\text{Ran}}$	$\mathcal{E}_{\text{Sys}}$	R_{Tot}	R_{Ran}	R_{Sys}	$\mathcal{E}_{\hat{\theta}}$
$b = \infty$	50	296	295	7	1.38	1.38	0.03	290
$\Psi(g) = 214$	100	127	125	7	0.60	0.58	0.03	251
$\Psi(\theta_0) = 221$	1000	25	24	7	0.12	0.11	0.03	223
	10000	10	7	7	0.04	0.03	0.03	221
	100000	7	2	7	0.03	0.01	0.03	221
$b = 100$	50	66	66	4	0.32	0.31	0.02	218
$\Psi(g) = 210$	100	47	47	4	0.22	0.22	0.02	214
$\Psi(\theta_0) = 214$	1000	16	15	4	0.08	0.07	0.02	213
	10000	6	5	4	0.03	0.02	0.02	214
	100000	4	2	4	0.02	0.01	0.02	214
$b = 50$	50	48	48	1	0.24	0.24	0.01	201
$\Psi(g) = 198$	100	35	35	1	0.18	0.18	0.01	199
$\Psi(\theta_0) = 199$	1000	11	11	1	0.06	0.06	0.01	199
	10000	4	4	1	0.02	0.02	0.01	199
	100000	2	1	1	0.01	0.01	0.01	199
$b = 25$	50	35	35	1	0.19	0.19	0.00	184
$\Psi(g) = 185$	100	26	26	1	0.14	0.14	0.00	185
$\Psi(\theta_0) = 186$	1000	8	8	1	0.04	0.04	0.00	186
	10000	3	3	1	0.01	0.01	0.00	186
	100000	1	1	1	0.01	0.01	0.00	186
$b = 10$	50	22	22	0	0.13	0.13	0.00	162
$\Psi(g) = 163$	100	16	16	0	0.10	0.10	0.00	163
$\Psi(\theta_0) = 163$	1000	5	5	0	0.03	0.03	0.00	163
	10000	2	2	0	0.01	0.01	0.00	163
	100000	1	1	0	0.00	0.00	0.00	163

Inclusion of an upper bound b

Figure 5.4 and Figure 5.5 suggest that the main source of error are the tails. Thus, as in previous sections including an upper bound b i desirable. Table 5.9 gives the output when parameter set 1 is considered and an upper bound b is added as before. Corresponding tables when parameter set 2 and parameter set 3 are considered are giden in Appendix A. As expected, systematic error decreases even further when b decreases. As seen in the previous sections, estimation error decrease when the upper bound is introduced as well. When b decreases, the amount of data necessary before total error can be neglected is reduced.

Thus, when g ∼log-gamma, systematic error is present, but unimportant. The reserve is still overestimated on average. The amount of overestimation decreases with n. Total error is decreasing with c_v, since a decline in c_v has a positive effect on both systematic and estimation error. Including an upper limit b reduces systematic error further and has a large positive effect on estimation error.

5.6 Summary

The possibility of using the extended Pareto distribution as default loss model has now been examined. The main general trend is that it is a conservative procedure, since for each of the four underlying distributions, the extended Pareto distribution tends to overestimate the reserve. The amount of overestimation is increasing with c_v. Thus, larger variability results in larger errors. Due to the decline in estimation error, the difference between the average estimated reserve and the true reserve is reduced when the amount of data n increases. Too heavy tails seams to be the main reason for the overestimation. Consequently, the incorrection is largely reduced when a reinsurance agreement, with an upper bound b is added. When g ∼gamma, Pareto or log-gamma there is no or an insignificant amount of systematic error, whereas when $g \sim$ log-normal, the bias can not be ignored. However, for a small enough b systematic error is unimportant in every situation considered in this chapter. When parameter set 1 is applied, $b \approx 25$ is sufficient. This is not unrealistic value b because the expected value of a claim is more than 10 times smaller, i.e. $\mathrm{E}[Z] = 2.37$. Estimation error is negligible for a sufficiently large n and it decreases with c_v. It is also substantially decreased when an upper bound b is added. In Chapter 6, including a fourth parameter λ is discussed. This might improve the situation because it reduces systematic error.

Chapter 6

A fourth parameter

When the extended Pareto family is the default loss model and the true distribution is for example log-normal, systematic error is not inconsiderable. Thus, it would make sense to try to "widen" the parametric family by including additional parameters, alternatively choose a different parametric family. Including one or more parameters results in a more flexible parametric family and reduces systematic error. A different parametric family might be a better approximation to a wider range of historical data. The log-normal or log-gamma distribution may for instance become parts of the parametric family. Estimation error might increase when more parameters are added. For a different parametric family, it is unknown what happens to systematic and estimation error. In this chapter, adding a fourth parameter by transforming the data is examined.

6.1 The simple power transformation

In [1], Box and Cox discuss how a transformation of the observed data can improve the approximation to the assumptions regarding the data which are made a priori. In the setting of this thesis, the objective is to transform the observed data such that the density function of the transformed data is a part of the extended Pareto family. The original observations are denoted by $(y_1, ..., y_n)$ and the transformed observations by $(z_1, ..., z_n)$. The transformation applied in this thesis is

$$z = \frac{(1+y)^\lambda - 1}{\lambda}, \quad y > 0, \tag{6.1.1}$$

which is a slight modification of the simple power transformation, see Box and Cox [1]. Thus, $z \in (0, \infty)$ and $z \to \log(1+y)$ as $\lambda \to 0$. The transformed observations are strictly larger than zero, i.e. $z > 0$ since $y > 0$. When $\lambda = 1$, the transformed data equals the original data.

As seen in Section 5.2, the log-normal distribution is not a part of the extended Pareto distribution family. It is desirable to be able to include the log-normal distribution in the parametric family after the transformation. When the transformation is applied and $\lambda = 0$, log-normally distributed data $y_1, ..., y_n$ are transformed into $z_1, ...z_n$ which are normally distributed. In Chapter 3 it is stated that the gamma distribution is a part of the parametric family in the limit. When the shape parameter α in a standard gamma distribution defined in (5.3.2) grows, the distribution converges to a Gaussian distribution, see Kim and Sung [9]. The standard deviation equals $1/\sqrt{\alpha}$ and is small, since α is large. The standard deviation can be made larger by multiplying Z_0 by ξ. That is, look at the distribution of $Z = \xi Z_0$ when α is large. The standard deviation is then $\text{sd}[Z] = \xi/\sqrt{\alpha}$, which can be made infinitely large

by increasing ξ. However, note that since the mean also increases by the same amount, the coefficient of variation c_v is fixed and equals $c_v = 1/\sqrt{\alpha}$. Thus, when the transformation is applied, the log-normal distribution is a part of the parametric family for $\lambda = 0$, but not for any parameter set (μ, σ). That is, the coefficient of variation c_v has to be small. A transformation which includes the log-normal distribution for any choice of (μ, σ) is something which is of interest to develop further.

A drawback with the transformation is that another parameter λ has to be estimated. In [3], Carroll and Ruppert argues that the variance of the parameter estimators can be much larger when the transformation parameters λ is unknown and has to be estimated, compared to when it is known. In the setting of this thesis λ is not known and it has to be estimated together with the other parameters, even if this increases estimation error. The question is then how much the variances in the parameter estimates $(\hat{\alpha}, \hat{\beta}, \hat{\theta}, \hat{\lambda})$ increase, and how much slower the convergence is. Hence, the decision whether to use the transformation or not, is a trade between less systematic or less estimation error. It is of interest to analyse how much estimation error increase compared to the reduction in systematic error.

6.2 Calculating the reserve

The transformed observations might have a density function closer to the parametric family, then the density function of the original data. Correspondingly, the density function of the original observations might be closer to the parametric family when the transformation parameter λ is added. That is, instead of transforming the observations, the parametric family is widened because a fourth parameter is added. In Appendix B, it is shown that the new density function, i.e. the density function of $Y = (Z\lambda + 1)^{1/\lambda} - 1$, where $Z \sim$extended Pareto(α, β, θ), is

$$f_Y(y) = \frac{\Gamma(\alpha+\theta)}{\Gamma(\alpha)\Gamma(\theta)} \frac{(\lambda\beta)^{\alpha+1}}{\beta} \frac{(1+y)^{\lambda-1}\{(1+y)^\lambda - 1\}^{\theta-1}}{\{\lambda\beta + (1+y)^\lambda - 1\}^{\alpha+\theta}}, \quad y > 0 \tag{6.2.1}$$

where $\alpha, \beta, \theta > 0$, $\lambda \in \mathbb{R}$ and $y = (y_1, ..., y_n)$ are the original observations. It is also shown that the new log-likelihood function is

$$\begin{aligned} l(\alpha, \beta, \theta, \lambda) =& n[\log\Gamma(\alpha+\theta) - \log\Gamma(\alpha) - \log\Gamma(\theta) + (\alpha+1)\log(\lambda\beta) - \log\beta] \\ &+ (\lambda-1)\sum_{i=1}^{n}\log(1+y_i) + (\theta-1)\sum_{i=1}^{n}\log\{(1+y_i)^\lambda - 1\} \\ &- (\alpha+\theta)\sum_{i=1}^{n}\log\{\lambda\beta + (1+y_i)^\lambda - 1\}. \end{aligned} \tag{6.2.2}$$

Thus, the parametric family is widen, since a fourth parameter is added . Consequently, more distributions are within the parametric family. Note that when $\lambda = 1$, the density function (6.2.1) equals the extended Pareto density function defined in (3.0.1) in Chapter 3. The new density and log-likelihood functions are well-defined. Hence, it should be unproblematic to do the same kind of simulations as in Chapter 4 and in Chapter 5. By simply replacing the density and log-likelihood functions from Chapter 3 with the new defined above in the simulation programs, it should be possible to find an exact value of $\Psi(\theta_0)$ and an estimated value of $\Psi(\hat{\theta})$. The problem is, as stated above, slow convergence of the parameter estimators θ_0 and $\hat{\theta}$.

Table 6.1: *The value of θ_0 from three different simulations for each parameter set when the Monte Carlo approach is applied.*

Parameter set	Simulation	α_0	β_0	θ_0	λ_0
1	1	67.95	34.50	1.99	0.11
	2	9.77	5.20	1.98	0.29
	3	32.47	16.62	1.98	0.14
2	1	26.88	13.42	3.85	0.33
	2	21.41	10.95	3.82	0.35
	3	28.66	14.10	3.88	0.32
2	1	216.57	58.13	3.95	0.12
	2	247.72	67.40	3.90	0.12
	3	6.01	2.06	3.87	0.70

Table 6.2: *The lower, mean and upper value of $\Psi(\theta_0)$ from 5 simulations when the Monte Carlo approach is used to calculate θ_0.*

Parameter set	Lower	Mean	Upper
1	233.13	234.05	234.56
2	339.28	342.39	351.93
3	154.66	155.34	156.60

6.2.1 Finding $\Psi(\theta_0)$

In Chapter 5, Gauss-Legendre quadrature is utilized to find θ_0 and $\Psi(\theta_0)$. However, when $\theta = (\alpha, \beta, \theta, \lambda)$ it turns out that the program used to find θ_0 in Chapter 4 and Chapter 5 can not be applied due to slow convergence of the parameter estimator θ_0. One solution could be to transform the integrand. The transformation discussed in Section 4.1.2 does not improve the speed of the convergence sufficiently. Thus, the integrand has to be transformed in a different manner. Alternatively, the program itself has to be changed. Due to time limitation, neither of these alternatives have been attempted.

In Section 4.1.1, it is stated that also Monte Carlo simulations can be used to find the vector of parameters θ_0 which minimize the Kullback-Leibler distance. By increasing the number of Monte Carlo simulations, it is possible to find a set of parameters arbitrarily close to θ_0. The question is then how large m needs to be. In Table 6.1, θ_0 is given for three different simulations. In Table 6.2, the lower, upper and mean values of $\Psi(\theta_0)$, from five simulations are given. The underlying distribution g is log-normally distributed, and the three parameter sets from Table 5.1 are considered in both tables. The number of Monte Carlo simulations are $m = 10^8$. Clearly, the vector of parameters $\theta_0 = (\alpha_0, \beta_0, \theta_0, \lambda_0)$ is very unstable. The corresponding reserve $\Psi(\theta_0)$ is still not perfectly stable, but it is surprisingly stable compare to θ_0. Hence, the vector of parameters converges slow when a fourth parameters is added, even so there is in only a small variance in the resulting reserve $\Psi(\theta_0)$. Due to large simulation time, raising m further has not been attempted.

Table 6.3: *The error terms, the ratios and the expected value of the estimated reserve, calculated for different values of n when $g \sim$log-normal(μ, σ), and transformation* (6.1.1) *is applied.*

	n	$\mathcal{E}_{\text{Tot}}$	$\mathcal{E}_{\text{Ran}}$	$\mathcal{E}_{\text{Sys}}$	R_{Tot}	R_{Ran}	R_{Sys}	$\mathcal{E}_{\hat{\theta}}$
Set 1	50	-	-	14	-	-	0.11	-
$\Psi(g) = 210$	100	-	-	14	-	-	0.11	-
$\Psi(\theta_0) = 234$	1000	57	45	14	0.27	0.21	0.11	248
	10000	33	15	14	0.16	0.07	0.11	241
	100000	31	10	14	0.15	0.05	0.11	240
	1000000	31	10	14	0.15	0.05	0.11	240
Set 2	50	91704	91704	5	271	271	0.01	1660
$\Psi(g) = 337$	100	95	94	5	0.28	0.28	0.01	366
$\Psi(\theta_0) = 342$	1000	18	17	5	0.05	0.05	0.01	345
	10000	8	7	5	0.02	0.02	0.01	342
	100000	6	5	5	0.02	0.01	0.01	342
	1000000	5	4	5	0.02	0.01	0.01	342
Set 2	50	-	-	3	-	-	0.02	-
$\Psi(g) = 152$	100	40	40	3	0.26	0.26	0.02	163
$\Psi(\theta_0) = 155$	1000	7	6	3	0.05	0.04	0.02	156
	10000	3	2	3	0.02	0.01	0.02	155
	100000	3	1	3	0.02	0.01	0.02	155
	1000000	3	1	3	0.02	0.01	0.02	155

6.2.2 Estimating $\Psi(\hat{\theta})$ and the different error terms

It is of interest to analyse the change in total, systematic and estimation error when transformation (6.1.1) is applied. As seen above, a perfectly stable value of $\Psi(\theta_0)$ has not been found, and the mean value of $\Psi(\theta_0)$ from Table 6.2 is used as an estimate of $\Psi(\theta_0)$. In Table 6.3, the error terms, the ratios and the expected value of the estimated reserve defined in (4.5.1), (4.5.2) and (4.5.3) respectively, are given for the three different parameter sets from Table 5.1, when $g \sim$log-normal. To be able to examine the convergence of estimation error more carefully, the situation $n = 10^6$ is added.

As expected systematic error is substantially reduced. For parameter set 1, $\Psi(\theta_0)$ is reduced from 289 to 234, which results in R_{Sys} decreasing from 0.38 to 0.11. Thus, even though including the transformation parameter λ doesn't eliminate all systematic error, it has a strong positive effect on systematic error. It is surprising that when $n \geq 100$, estimation error is almost unchanged. There might be a slight tendency that estimation error converges slower to zero when the transformation parameter λ is added, but even so the great decrease in systematic error decreases total error considerably when $n \geq 100$.

Not a number values are produced when $n = 100$ and when $n = 50$ and parameter set 1 is considered and when $n = 50$ and parameter set 3 is considered. This is most likely due to a numerical error. Correcting this is time consuming, and has therefore not been attempted. For parameter set 2 estimation error is substantially increased when $n = 50$. Thus, including a fourth parameter might not be preferable when n is very limited.

Hence, when $n \geq 100$ including the transformation has a positive effect on total error, when $g \sim$log-normal(μ, σ). In order to know with certainty whether the inclusion of λ has a posi-

tive effect on total error in general, more analyses have to be carried out. We need to know what happens to total error for other underlying distributions. For g ∼gamma, Pareto or log-gamma, there are no or an insignificant amount of systematic error. Including λ might increase estimation error more than it did when g ∼log-normal, which will result in an increased total error. Hence, the results above suggests that including a fourth parameter λ might have a positive effect on total error. Investigating the effect of adding one or more parameters, is something which is of interest to develop further.

Chapter 7

Concluding remarks

A framework based on asymptotic theory for maximum likelihood estimation under misspecifications has been defined, and procedures for testing a specific family of distributions as default loss models are developed. The thesis has reviewed the extended Pareto family as basis for such an automatic procedure, with the main conclusion that it a conservative procedure which tends to overestimate the reserve.

Total error of some functional Ψ is throughout the thesis divided into a constant part and a random part, i.e. systematic error and estimation error,

$$\Psi(\hat{\theta}) - \Psi(g) = \Psi(\hat{\theta}) - \Psi(\theta_0) + \Psi(\theta_0) - \Psi(g).$$

Systematic error is independent of n while estimation error decreases when n increases. For every situation considered, systematic error is negligible when a reinsurance agreement with a sufficiently small upper bound b is added. Estimation error is always present when the amount of data is limited. It might however be larger when a flexible distribution, like the extended Pareto distribution, is applied an several parameters have to be estimated. In real life situation it is generally unknown how large the bias is, it might be unimportant, but it might also be larger than anything seen in this thesis. Hence, the extended Pareto family seems to work well as default loss model in many situations. The large amount of overestimation in some situations might be due to the estimator $\hat{\theta}$ being estimated based on a likelihood criterion. Other criteria, which might give less weight to extreme values of Z, is something which is of interest to examine further.

Before the extended Pareto can be introduced as default loss model, more testes should be completed and situations which leads to large errors should be characterized. In chapter 6, other ways of attacking the problem are suggested, and the possibility of adding one more parameter is briefly examined. The estimates themselves become unstable, but the reserve evaluated from them is not, at least not to the same degree. That brings hope that one can construct even more flexible models than what considered in this thesis, arriving at a truly automatic model for modelling losses.

Appendix A

Additional tables and figures

Additional figures, Chapter 5

In Figure 5.2, Figure 5.3 and Figure 5.5, the density function $f_{\theta_0}(z)$ is plotted together with five realisations of density function $f_{\hat{\theta}}(z)$ when parameter set 1 is considered and g ∼log-normal, g ∼Pareto and g ∼log-gamma respectively. The corresponding plots for parameter set 2 and parameter set 3 are given here. In each figure, $n = 50$ observations from the true distribution are used to estimate $\hat{\theta}$ in the left column. In the middle $n = 1000$ observations are used and to the right, $n = 100000$ observations are used. In Figure A.1 and in Figure A.2, g ∼log-normal and parameter set 2 and parameter set 3 are considered respectively. In Figure A.3 and in Figure A.4 g ∼Pareto and parameter set 2 and parameter set 3 are considered respectively. In Figure A.5 and in Figure A.6 g ∼log-gamma and parameter set 2 and parameter set 3 are considered respectively.

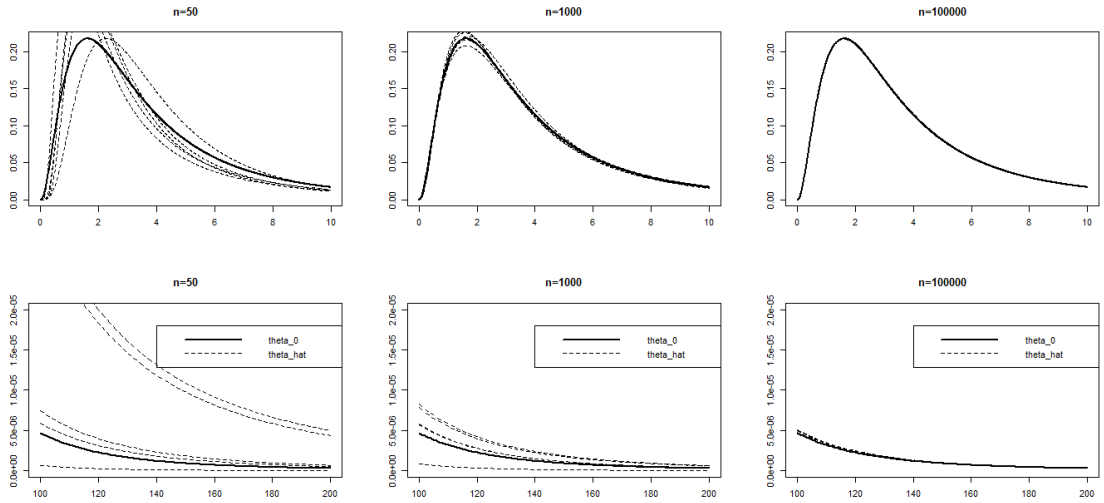

Figure A.1: *Density functions $f_{\theta_0}(z)$ and $f_{\hat{\theta}}(z)$ for parameter set 2 from Table 5.1, g ∼log-normal. Upper row: $z \in (0, 10]$. Lower row: $z \in [100, 200]$.*

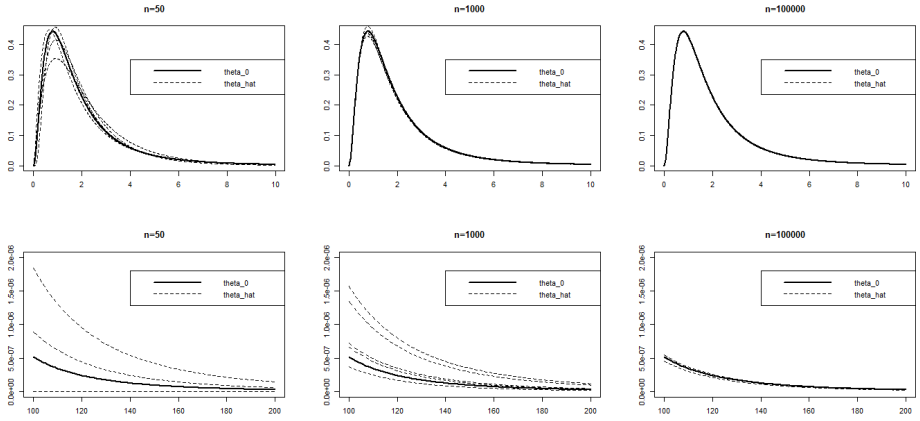

Figure A.2: *Density functions $f_{\theta_0}(z)$ and $f_{\hat{\theta}}(z)$ for parameter set 3 from Table 5.1, $g \sim$log-normal. Upper row: $z \in (0, 10]$. Lower row: $z \in [100, 200]$.*

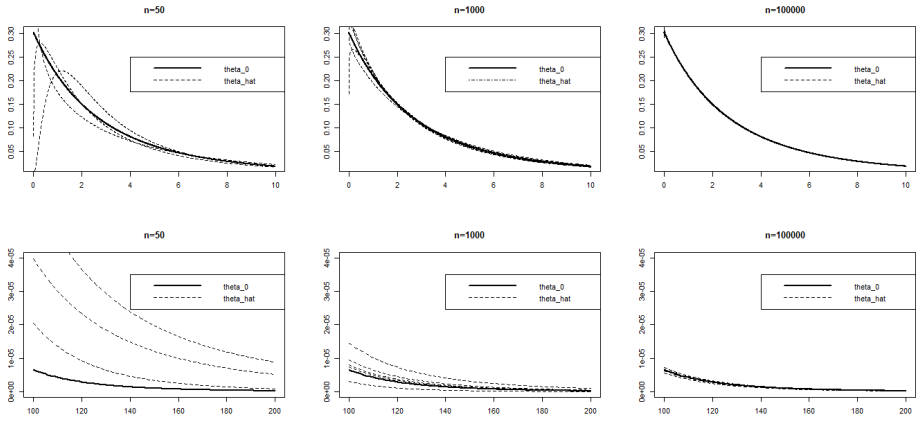

Figure A.3: *Density functions $f_{\theta_0}(z)$ and $f_{\hat{\theta}}(z)$ for parameter set 2 from Table 5.1, $g \sim$Pareto. Upper row: $z \in (0, 10]$. Lower row: $z \in [100, 200]$.*

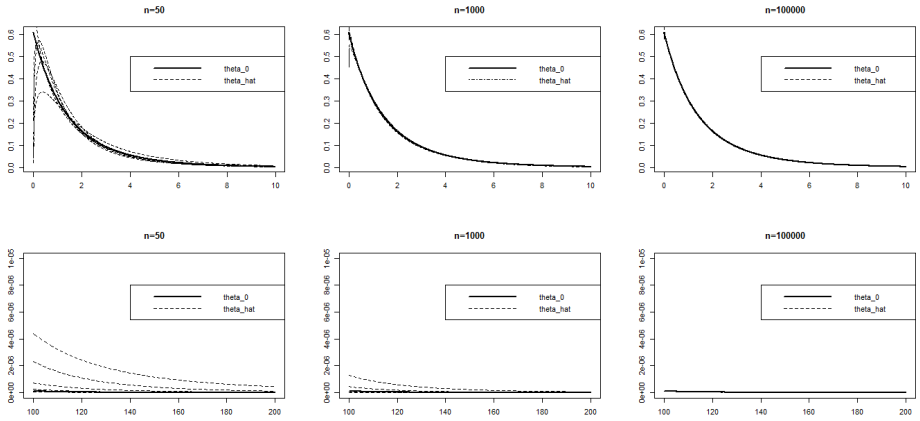

Figure A.4: *Density functions $f_{\theta_0}(z)$ and $f_{\hat{\theta}}(z)$ for parameter set 3 from Table 5.1, $g \sim$Pareto. Upper row: $z \in (0, 10]$. Lower row: $z \in [100, 200]$.*

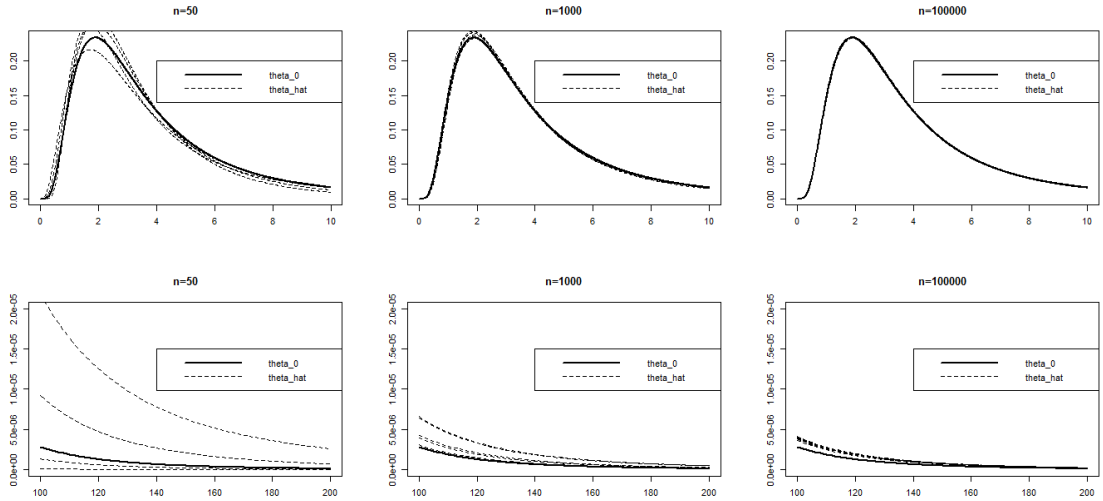

Figure A.5: *Density functions $f_{\theta_0}(z)$ and $f_{\hat{\theta}}(z)$ for parameter set 2 from Table 5.1, $g \sim$log-gamma. Upper row: $z \in (0, 10]$. Lower row: $z \in [100, 200]$.*

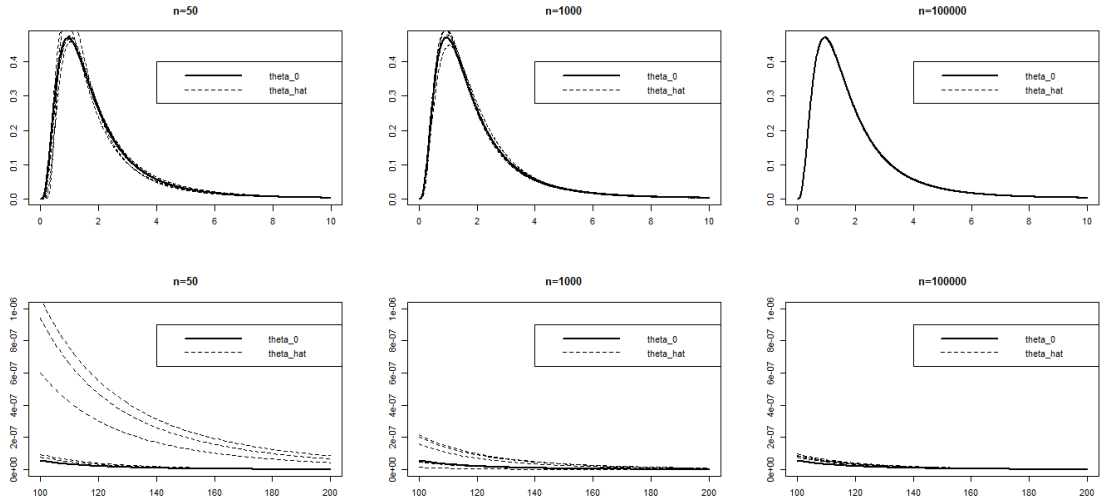

Figure A.6: *Density functions $f_{\theta_0}(z)$ and $f_{\hat{\theta}}(z)$ for parameter set 3 from Table 5.1, $g \sim$log-gamma. Upper row: $z \in (0, 10]$. Lower row: $z \in [100, 200]$.*

Additional tables, Chapter 5

In Chapter 5, an upper limit b is added for the four different choices of underlying models, but only parameter set 1 from Table 5.1 is considered. For completeness, the output when parameter set 2 and 3 are considered is given here. The upper limits considered are $b = 120, 60, 30$ and 15 for parameter set 2 and $b = 50, 25, 15$ and 7 for parameter set 3. In Table A.1 and Table A.2 $g \sim$log-normal, in Table A.3 and Table A.4 $g \sim$gamma, in Table A.5 and Table A.6 $g \sim$Pareto and in Table A.7 and Table A.8 $g \sim$log-gamma. As stated in Chapter 5, the tendencies are the same as when parameter set 1 is considered. That is, including an upper limit b improves the situation since both systematic and estimated error decreases.

Table A.1: *The error terms, the ratios and the expected value of the estimated reserve, calculated for different values of n when $g \sim$log-normal$(1.12, 0.85)$ and a reinsurance limit b is added.*

	n	$\mathcal{E}_{\text{Tot}}$	$\mathcal{E}_{\text{Ran}}$	$\mathcal{E}_{\text{Sys}}$	R_{Tot}	R_{Ran}	R_{Sys}	$\mathcal{E}_{\hat{\theta}}$
$b = \infty$	50	196	190	20	0.58	0.56	0.06	402
$\Psi(g) = 337$	100	102	96	20	0.30	0.28	0.06	378
$\Psi(\theta_0) = 357$	1000	29	20	20	0.09	0.06	0.06	358
	10000	20	6	20	0.06	0.02	0.06	357
	100000	19	2	20	0.06	0.01	0.06	357
$b = 120$	50	77	74	16	0.23	0.22	0.05	359
$\Psi(g) = 337$	100	56	53	16	0.17	0.016	0.05	356
$\Psi(\theta_0) = 353$	1000	23	17	16	0.07	0.05	0.05	353
	10000	16	5	16	0.05	0.02	0.05	353
	100000	15	2	16	0.05	0.01	0.05	353
$b = 60$	50	60	59	8	0.18	0.18	0.03	344
$\Psi(g) = 335$	100	44	43	8	0.13	0.13	0.03	344
$\Psi(\theta_0) = 343$	1000	16	13	8	0.05	0.04	0.03	343
	10000	9	4	8	0.03	0.01	0.03	343
	100000	9	2	8	0.03	0.00	0.03	343
$b = 30$	50	45	45	1	0.14	0.14	0.01	328
$\Psi(g) = 328$	100	33	33	1	0.10	0.01	0.01	328
$\Psi(\theta_0) = 329$	1000	11	11	1	0.03	0.03	0.01	329
	10000	4	3	1	0.01	0.01	0.01	329
	100000	2	1	1	0.01	0.00	0.01	329
$b = 15$	50	33	33	1	0.11	0.11	0.01	302
$\Psi(g) = 305$	100	24	24	1	0.08	0.08	0.01	302
$\Psi(\theta_0) = 304$	1000	8	8	1	0.03	0.03	0.01	303
	10000	3	2	1	0.01	0.01	0.01	304
	100000	2	1	1	0.01	0.00	0.01	304

Table A.2: *The error terms, the ratios and the expected value of the estimated reserve, calculated for different values of n when $g \sim$log-normal$(0.41, 0.78)$ and a reinsurance limit b is added.*

	n	$\mathcal{E}_{\text{Tot}}$	$\mathcal{E}_{\text{Ran}}$	$\mathcal{E}_{\text{Sys}}$	R_{Tot}	R_{Ran}	R_{Sys}	$\mathcal{E}_{\hat{\theta}}$
$b = \infty$	50	56	55	4	0.37	0.36	0.03	169
$\Psi(g) = 153$	100	30	29	4	0.20	0.19	0.03	163
$\Psi(\theta_0) = 157$	1000	9	7	4	0.06	0.04	0.03	158
	10000	5	2	4	0.03	0.01	0.03	158
	100000	5	1	4	0.03	0.01	0.03	157
$b = 50$	50	29	28	4	0.19	0.19	0.03	160
$\Psi(g) = 153$	100	21	20	4	0.13	0.13	0.03	159
$\Psi(\theta_0) = 157$	1000	7	6	4	0.05	0.04	0.03	159
	10000	5	2	4	0.03	0.02	0.03	157
	100000	4	1	4	0.03	0.00	0.03	157
$b = 25$	50	24	23	3	0.15	0.15	0.02	155
$\Psi(g) = 152$	100	17	17	3	0.11	0.11	0.02	155
$\Psi(\theta_0) = 155$	1000	6	5	3	0.04	0.04	0.02	155
	10000	3	2	3	0.02	0.01	0.02	155
	100000	3	1	3	0.02	0.00	0.02	155
$b = 15$	50	20	20	1	0.13	0.13	0.01	151
$\Psi(g) = 151$	100	14	14	1	0.10	0.10	0.01	152
$\Psi(\theta_0) = 152$	1000	5	5	1	0.03	0.03	0.01	152
	10000	2	1	1	0.01	0.01	0.01	152
	100000	1	1	1	0.01	0.00	0.01	152
$b = 7$	50	15	15	1	0.11	0.11	0.00	141
$\Psi(g) = 143$	100	11	11	1	0.07	0.07	0.00	142
$\Psi(\theta_0) = 142$	1000	3	3	1	0.02	0.02	0.00	142
	10000	1	1	1	0.01	0.01	0.00	142
	100000	1	0	1	0.01	0.00	0.00	142

Table A.3: *The error terms, the ratios and the expected value of the estimated reserve, calculated for different values of n when $g \sim$gamma$(4.41, 0.94)$ and a reinsurance limit b is added.*

	n	$\mathcal{E}_{\text{Ran}}$	R_{Ran}	$\mathcal{E}_{\hat{\theta}}$		n	$\mathcal{E}_{\text{Ran}}$	R_{Ran}	$\mathcal{E}_{\hat{\theta}}$
$b = \infty$	50	248	0.74	349	$b = 30$	50	47	0.14	334
$\Psi(g) = 335$	100	38	0.11	338	$\Psi(g) = 333$	100	33	0.10	334
	1000	11	0.03	336		1000	11	0.03	334
	10000	4	0.01	335		10000	4	0.01	333
	100000	2	0.01	335		100000	1	0.00	333
$b = 120$	50	53	0.16	340	$b = 15$	50	38	0.12	311
$\Psi(g) = 335$	100	36	0.11	338	$\Psi(g) = 315$	100	26	0.08	312
	1000	11	0.03	335		1000	9	0.03	314
	10000	4	0.01	335		10000	3	0.01	315
	100000	1	0.00	335		100000	1	0.00	315
$b = 60$	50	51	0.15	338					
$\Psi(g) = 335$	100	36	0.11	337					
	1000	11	0.03	335					
	10000	4	0.01	335					
	100000	1	0.00	335					

Table A.4: *The error terms, the ratios and the expected value of the estimated reserve, calculated for different values of n when $g \sim gamma(2.05, 1.19)$ and a reinsurance limit b is added.*

	n	$\mathcal{E}_{\text{Ran}}$	R_{Ran}	$\mathcal{E}_{\hat{\theta}}$
$b = \infty$	50	31	0.21	156
$\Psi(g) = 152$	100	15	0.10	154
	1000	5	0.03	152
	10000	1	0.01	152
	100000	1	0.00	152
$b = 50$	50	21	0.14	152
$\Psi(g) = 150$	100	14	0.10	151
	1000	4	0.030	150
	10000	1	0.01	150
	100000	1	0.00	150
$b = 25$	50	20	0.13	152
$\Psi(g) = 150$	100	14	0.10	151
	1000	4	0.03	150
	10000	1	0.01	150
	100000	1	0.00	150
$b = 15$	50	19	0.13	151
$\Psi(g) = 150$	100	14	0.09	150
	1000	4	0.03	150
	10000	1	0.01	150
	100000	1	0.00	150
$b = 7$	50	14	0.11	143
$\Psi(g) = 142$	100	11	0.08	142
	1000	3	0.02	142
	10000	1	0.01	142
	100000	0	0.00	142

Table A.5: *The error terms, the ratios and the expected value of the estimated reserve, calculated for different values of n when $g \sim Pareto(4.00, 13.23)$ and a reinsurance limit b is added.*

	n	$\mathcal{E}_{\text{Ran}}$	R_{Ran}	$\mathcal{E}_{\hat{\theta}}$		n	$\mathcal{E}_{\text{Ran}}$	R_{Ran}	$\mathcal{E}_{\hat{\theta}}$
$b = \infty$	50	1209	3.27	469	$b = 30$	50	57	0.17	332
$\Psi(g) = 370$	100	128	0.35	395	$\Psi(g) = 337$	100	40	0.11	333
	1000	26	0.07	372		1000	13	0.04	337
	10000	8	0.02	370		10000	4	0.01	337
	100000	3	0.01	370		100000	1	0.00	337
$b = 120$	50	91	0.25	371	$b = 15$	50	39	0.13	295
$\Psi(g) = 366$	100	65	0.18	368	$\Psi(g) = 298$	100	28	0.10	297
	1000	21	0.06	366		1000	9	0.03	298
	10000	7	0.02	366		10000	3	0.01	298
	100000	2	0.01	366		100000	1	0.00	298
$b = 60$	50	71	0.20	354					
$\Psi(g) = 357$	100	52	0.14	356					
	1000	17	0.05	356					
	10000	5	0.02	356 6					
	100000	2	0.01	357					

Table A.6: *The error terms, the ratios and the expected value of the estimated reserve, calculated for different values of n when $g \sim$Pareto$(5.00, 8.21)$ and a reinsurance limit b is added.*

	n	$\mathcal{E}_{\text{Ran}}$	R_{Ran}	$\mathcal{E}_{\hat{\theta}}$		n	$\mathcal{E}_{\text{Ran}}$	R_{Ran}	$\mathcal{E}_{\hat{\theta}}$
$b = \infty$	50	144	0.87	191	$b = 15$	50	26	0.16	157
$\Psi(g) = 166$	100	43	0.26	174	$\Psi(g) = 158$	100	19	0.12	157
	1000	9	0.06	167		1000	6	0.04	158
	10000	3	0.02	167		10000	2	0.01	158
	100000	1	0.01	166		100000	1	0.00	158
$b = 50$	50	37	0.23	168	$b = 7$	50	18	0.13	141
$\Psi(g) = 166$	100	27	0.16	167	$\Psi(g) = 141$	100	13	0.09	139
	1000	9	0.05	166		1000	4	0.03	141
	10000	3	0.02	166		10000	1	0.01	141
	100000	1	0.01	166		100000	0	0.00	141
$b = 25$	50	31	0.19	162					
$\Psi(g) = 163$	100	22	0.17	162					
	1000	7	0.04	163					
	10000	2	0.01	163					
	100000	1	0.00	163					

Table A.7: *The error terms, the ratios and the expected value of the estimated reserve, calculated for different values of n when $g \sim$log-gamma$(1.50, 7.00)$ and a reinsurance limit b is added.*

	n	$\mathcal{E}_{\text{Tot}}$	$\mathcal{E}_{\text{Ran}}$	$\mathcal{E}_{\text{Sys}}$	R_{Tot}	R_{Ran}	R_{Sys}	$\mathcal{E}_{\hat{\theta}}$
$b = \infty$	50	122	120	7	0.36	0.35	0.02	372
$\Psi(g) = 342$	100	77	75	7	0.22	0.22	0.02	363
$\Psi(\theta_0) = 349$	1000	21	19	7	0.06	0.06	0.02	351
	10000	9	6	7	0.03	0.02	0.02	349
	100000	7	2	7	0.02	0.01	0.02	348
$b = 120$	50	68	67	4	0.20	0.20	0.01	349
$\Psi(g) = 341$	100	50	50	4	0.15	0.15	0.01	349
$\Psi(\theta_0) = 345$	1000	17	16	4	0.05	0.05	0.01	346
	10000	7	5	4	0.02	0.02	0.01	346
	100000	5	2	4	0.02	0.01	0.01	345
$b = 60$	50	55	55	2	0.16	0.16	0.01	336
$\Psi(g) = 335$	100	41	41	2	0.12	0.12	0.01	338
$\Psi(\theta_0) = 337$	1000	13	13	2	0.04	0.04	0.01	337
	10000	5	4	2	0.01	0.01	0.01	337
	100000	3	1	2	0.01	0.00	0.01	337
$b = 30$	50	45	45	1	0.14	0.14	0.00	323
$\Psi(g) = 324$	100	32	32	1	0.10	0.10	0.00	324
$\Psi(\theta_0) = 325$	1000	10	10	1	0.03	0.03	0.00	325
	10000	3	3	1	0.01	0.01	0.00	325
	100000	1	1	1	0.00	0.00	0.00	325
$b = 15$	50	33	33	0	0.11	0.11	0.00	301
$\Psi(g) = 303$	100	23	23	0	0.08	0.08	0.00	302
$\Psi(\theta_0) = 303$	1000	7	7	0	0.02	0.02	0.00	303
	10000	2	2	0	0.01	0.01	0.00	303
	100000	1	1	0	0.00	0.00	0.00	303

Table A.8: *The error terms, the ratios and the expected value of the estimated reserve, calculated for different values of n when $g \sim$log-gamma$(1.00, 5.00)$ and a reinsurance limit b is added.*

	n	$\mathcal{E}_{\text{Tot}}$	$\mathcal{E}_{\text{Ran}}$	$\mathcal{E}_{\text{Sys}}$	R_{Tot}	R_{Ran}	R_{Sys}	$\mathcal{E}_{\hat{\theta}}$
$b = \infty$	50	40	40	1	0.26	0.26	0.01	163
$\Psi(g) = 154$	100	25	25	1	0.16	0.16	0.01	159
$\Psi(\theta_0) = 155$	1000	6	6	1	0.04	0.04	0.01	155
	10000	2	2	1	0.01	0.01	0.01	155
	100000	1	1	1	0.01	0.00	0.01	155
$b = 50$	50	27	27	1	0.17	0.17	0.00	157
$\Psi(g) = 154$	100	19	18	1	0.12	0.12	0.00	157
$\Psi(\theta_0) = 155$	1000	6	6	1	0.04	0.04	0.00	155
	10000	2	2	1	0.01	0.01	0.00	155
	100000	1	1	1	0.01	0.00	0.00	155
$b = 25$	50	23	23	0	0.15	0.15	0.00	154
$\Psi(g) = 153$	100	16	16	0	0.11	0.11	0.00	153
$\Psi(\theta_0) = 153$	1000	5	5	0	9.03	0.03	0.00	153
	10000	2	2	0	0.01	0.01	0.00	153
	100000	1	1	0	0.00	0.00	0.00	153
$b = 15$	50	20	20	0	0.13	0.13	0.00	150
$\Psi(g) = 151$	100	14	14	0	0.10	0.10	0.00	151
$\Psi(\theta_0) = 151$	1000	5	5	0	0.03	0.03	0.00	151
	10000	2	2	0	0.01	0.01	0.00	151
	100000	1	1	0	0.00	0.00	0.00	151
$b = 7$	50	15	15	0	0.10	0.10	0.00	141
$\Psi(g) = 142$	100	10	10	0	0.07	0.07	0.00	142
$\Psi(\theta_0) = 142$	1000	3	3	0	0.02	0.02	0.00	142
	10000	1	1	0	0.01	0.01	0.00	142
	100000	0	0	0	0.00	0.00	0.00	142

Appendix B

Mathematical arguments

Chapter 2

Section 2.2.3

Proof that $\mathrm{E}[Z_i] = -\mathrm{E}[Y_i^2]$, when the assumed model is correct and θ is a scalar, i.e. the one-parameter situation.

If $\mathrm{E}[Z_i] = -\mathrm{E}[Y_i^2]$, this is equivalent to

$$\int_{-\infty}^{\infty} \left(\frac{\partial^2}{\partial \theta^2} \log f_{\theta_0}(x) \right) f_{\theta_0}(x) dx = -\int_{-\infty}^{\infty} \left(\frac{\partial}{\partial \theta} \log f_{\theta_0}(x) \right)^2 f_{\theta_0}(x) dx,$$

or

$$E\left[\frac{\partial^2}{\partial \theta^2} \log f(X_i|\theta_0) \right] = -E\left[\left(\frac{\partial}{\partial \theta} \log f(X_i|\theta_0) \right)^2 \right].$$

From the argument regarding the expected value in Section 2.2.2, we know that since $\int_{-\infty}^{\infty} f_{\theta_0}(x)dx = 1$,

$$\begin{aligned} 0 &= \frac{\partial}{\partial \theta} \int_{-\infty}^{\infty} f_{\theta_0}(x) dx \\ &= \int_{-\infty}^{\infty} \frac{\partial f_{\theta_0}(x)/\partial \theta}{f_{\theta_0}(x)} f_{\theta_0}(x) dx \\ &= \int_{-\infty}^{\infty} \left[\frac{\partial}{\partial \theta} \log f_{\theta_0}(x) \right] f_{\theta_0}(x) dx. \end{aligned}$$

By differentiating once more we get

$$\begin{aligned} 0 &= \frac{\partial}{\partial \theta} \int_{-\infty}^{\infty} \left[\frac{\partial}{\partial \theta} \log f_{\theta_0}(x) \right] f_{\theta_0}(x) dx \\ &= \int_{-\infty}^{\infty} \left[\frac{\partial^2}{\partial \theta^2} \log f_{\theta_0}(x) \right] f_{\theta_0}(x) dx + \int_{-\infty}^{\infty} \left[\frac{\partial}{\partial \theta} \log f_{\theta_0}(x) \right] \frac{\partial}{\partial \theta} f_{\theta_0}(x) dx \\ &= E\left[\frac{\partial^2}{\partial \theta^2} \log f(X_i|\theta_0) \right] + \int_{-\infty}^{\infty} \left[\frac{\partial}{\partial \theta} \log f_{\theta_0}(x) \right]^2 f_{\theta_0}(x) dx \\ &= E\left[\frac{\partial^2}{\partial \theta^2} \log f(X_i|\theta_0) \right] + E\left[\left(\frac{\partial}{\partial \theta} \log f(X_i|\theta_0) \right)^2 \right] \\ &= \mathrm{E}[Z_i] + \mathrm{var}[Y_i]. \end{aligned}$$

Hence, $\mathrm{E}[Z_i] = -\mathrm{var}[Y_i] = \mathrm{E}[Y_i^2]$, when the underlying distribution is part of the parametric family.

Section 2.3

Proof that $\mathrm{E}[Z_i] = -\mathrm{E}[Y_i^2]$, when the assumed model is correct and $\theta = (\theta_1, .., \theta_p)$ is a vector of length p, i.e. the multi-parameter situation.

In the multi-parameter situation the quantities $\mathrm{E}[Y_i^2]$ and $\mathrm{E}[Z_i]$ are

$$\mathrm{E}[Y_i^2] = \int_{-\infty}^{\infty} \left(\frac{\partial}{\partial \theta_j} \log f_{\theta_0}(x) \right) \left(\frac{\partial}{\partial \theta_l} \log f_{\theta_0}(x) \right) f_{\theta_0}(x) dx$$

and

$$\mathrm{E}[Z_i] = \int_{-\infty}^{\infty} \left(\frac{\partial^2}{\partial \theta_j \partial \theta_l} \log f_{\theta_0}(x) \right) f_{\theta_0}(x) dx.$$

The arguments in the multi-parameter situation are the same as those in the one-parameter situation above,

$$0 = \frac{\partial}{\partial \theta_j} \int_{-\infty}^{\infty} f_{\theta_0}(x) dx = \int_{-\infty}^{\infty} \left[\frac{\partial}{\partial \theta_j} \log f_{\theta_0}(x) \right] f_{\theta_0}(x) dx.$$

By differentiating once more we get,

$$\begin{aligned} 0 &= \frac{\partial}{\partial \theta_l} \int_{-\infty}^{\infty} \left[\frac{\partial}{\partial \theta_j} \log f_{\theta_0}(x) \right] f_{\theta_0}(x) dx \\ &= \int_{-\infty}^{\infty} \left[\frac{\partial^2}{\partial \theta_j \partial \theta_l} \log f_{\theta_0}(x) \right] f_{\theta_0}(x) dx + \int_{-\infty}^{\infty} \left[\frac{\partial}{\partial \theta_j} \log f(x|\theta) \right] \frac{\partial}{\partial \theta_l} f_{\theta}(x) dx \\ &= \mathrm{E}[Z_i] + \mathrm{E}[Y_i^2]. \end{aligned}$$

Hence, $\mathrm{E}[Z_i] = -\mathrm{var}[Y_i] = -\mathrm{E}[Y_i^2]$ in the multi-parameter situation as well.

Chapter 3

Finding the mean, the standard deviation, the skewness and the kurtosis of the extended Pareto distribution.

Let $f_\theta(x)$ be the extended Pareto density, then,

$$\begin{aligned} E(X^i) &= \int_0^{\infty} x^i f_\theta(x) dx \\ &= \frac{\Gamma(\alpha+\theta)}{\Gamma(\alpha)\Gamma(\theta)} \int_0^{\infty} \frac{1}{\beta} \frac{(x/\beta)^{\theta-1}}{(1+\frac{x}{\beta})^{\alpha+\theta}} x^i dx \\ &= \frac{\Gamma(\alpha+\theta)}{\Gamma(\alpha)\Gamma(\theta)} \beta^i \int_0^{\infty} \frac{1}{\beta} \frac{(z/\beta)^{\theta+i-1}}{(1+x/\beta)^{\alpha+\theta}} dx. \end{aligned}$$

Define $\tilde{\alpha} = \alpha + i$ and $\tilde{\theta} = \theta - 1$. Then,

$$\begin{aligned} E(X^i) &= \frac{\Gamma(\alpha+\theta)}{\Gamma(\alpha)\Gamma(\theta)}\beta^i \int_0^\infty \frac{\Gamma(\tilde{\alpha})\Gamma(\tilde{\theta})}{\Gamma(\tilde{\alpha}+\tilde{\theta})} \left(\frac{1}{\beta} \frac{\Gamma(\tilde{\alpha}+\tilde{\theta})}{\Gamma(\tilde{\alpha})\Gamma(\tilde{\theta})} \frac{(x/\beta)^{\tilde{\theta}-1}}{(1+x/\beta)^{\tilde{\alpha}+\tilde{\theta}}} \right) dx \\ &= \frac{\Gamma(\alpha+\theta)}{\Gamma(\alpha)\Gamma(\theta)}\beta^i \frac{\Gamma(\tilde{\alpha})\Gamma(\tilde{\theta})}{\Gamma(\tilde{\alpha}+\tilde{\theta})} \\ &= \frac{\Gamma(\alpha+i)\Gamma(\theta-i)}{\Gamma(\alpha)\Gamma(\theta)}\beta^i \end{aligned}$$

since $(\tilde{\alpha} + \tilde{\theta}) = \alpha + \theta$. From this it follows that

$$\begin{aligned} E(X) &= \beta\frac{\Gamma(\alpha-1)\Gamma(\theta+1)}{\Gamma(\alpha)\Gamma(\theta)} = \beta\frac{\theta}{\alpha-1}, \\ E(X^2) &= \beta^2\frac{\Gamma(\alpha-2)\Gamma(\theta+2)}{\Gamma(\alpha)\Gamma(\theta)} = \beta^2\frac{\theta(\theta+1)}{(\alpha-1)(\alpha-2)}, \\ E(X^3) &= \beta^3\frac{\Gamma(\alpha-3)\Gamma(\theta+3)}{\Gamma(\alpha)\Gamma(\theta)} = \beta^3\frac{\theta(\theta+1)(\theta+2)}{(\alpha-1)(\alpha-2)(\alpha-3)} \quad \text{and} \\ E(X^4) &= \beta^4\frac{\Gamma(\alpha-4)\Gamma(\theta+4)}{\Gamma(\alpha)\Gamma(\theta)} = \beta^4\frac{\theta(\theta+1)(\theta+2)(\theta+3)}{(\alpha-1)(\alpha-2)(\alpha-3)(\alpha-4)}. \end{aligned}$$

Thus, the expectation is

$$E(X) = \beta\frac{\theta}{\alpha-1} = \xi.$$

The variance is defined as, $\text{var}(X) = E(X^2) - E(X)^2$, thus

$$\begin{aligned} \text{var}(X) &= \frac{\beta^2\theta(\theta+1)}{(\alpha-1)(\alpha-2)} - \left(\frac{\beta\theta}{\alpha-1}\right)^2 \\ &= \frac{\beta^2\theta^2(\theta+1)(\alpha-1)}{\theta(\alpha-1)^2(\alpha-2)} - \left(\frac{\beta\theta}{\alpha-1}\right)^2 \\ &= \xi^2\left(\frac{(\theta+1)(\alpha-1)}{\theta(\alpha-2)} - 1\right) \\ &= \xi^2\left(\frac{\alpha+\theta-1}{\theta(\alpha-2)}\right). \end{aligned}$$

Hence, the standard deviation in the extended Pareto distribution is

$$\text{sd}(X) = \xi\sqrt{\frac{\alpha+\theta-1}{\theta(\alpha-2)}} = \sigma.$$

The skewness is defined as,

$$\text{skew}(X) = \frac{\text{E}(X-\xi)^3}{\sigma^3}, \quad \text{where} \quad \xi = \text{E}(X) \quad \text{and} \quad \sigma = \text{sd}(X).$$

The numerator can be rewritten as

$$\begin{aligned} \text{E}(X-\xi)^3 &= \text{E}(X^3) - 3\xi\text{E}(X^2) + 3\xi^2\text{E}(X) - \xi^3 \\ &= \text{E}(X^3) - 3\xi(\text{E}(X^2) + \xi\text{E}(X)) - \xi^3 \\ &= \text{E}(X^3) - 3\xi\sigma^2 - \xi^3. \end{aligned}$$

Thus, it follows that

$$\text{skew}(X) = \frac{\text{E}(X^3) - 3\xi\sigma^2 - \xi^3}{\sigma^3}.$$

By plugging in the expressions for $\text{E}(X^3), \xi$ and σ which are found above, and then do some algebra to simplify, it can be shown that,

$$\text{skew}(X) = 2\left(\frac{\alpha + 2\theta - 1}{\alpha - 3}\right)\sqrt{\frac{\alpha - 2}{\theta(\alpha + \theta - 1)}}.$$

The kurtosis is defined as,

$$\text{kurt}(X) = \frac{\text{E}(X - \xi)^4}{\sigma^4} \quad \text{where} \quad \xi = \text{E}(X) \quad \text{and} \quad \sigma = \text{sd}(X).$$

The numerator can be rewritten as,

$$\begin{aligned}\text{E}(X - \xi)^4 &= \text{E}(X^4) - 4\text{E}(X^3)\xi + 6\text{E}(X^2)\xi^2 - 4\text{E}(Z)\xi^3 + \xi^4 \\ &= \text{E}(X^4) - 4\text{E}(X^3)\xi + 6\text{E}(X^2)\xi^2 - 3\text{E}(X)^4\end{aligned}$$

Hence, it follows that

$$\text{kurt}(X) = \frac{\{\text{E}(X^4) - 4\text{E}(X^3)\xi + 6\text{E}(X^2)\xi^2 - 3\text{E}(X)^4\}}{\sigma^4}.$$

By plugging in the expressions for ξ, σ and $\text{E}(X^i)$ for $i = 1, 2, 3, 4$, and carry out some calculations, it can be shown that

$$\text{kurt}(X) = \frac{3(\alpha - 2)(\alpha^2\theta + 2\alpha^2 + \alpha\theta^2 + 4\alpha\theta - 4\alpha + 5\theta^2 - 5\theta + 2)}{(\alpha - 4)(\alpha - 3)\theta(\alpha + \theta - 1)}.$$

Chapter 5

Section 5.3

Finding the density function for the standard gamma distribution.

In Section 5.3, the density function of a gamma distributed random variable Z is defined as

$$f(z) = \frac{\lambda^\alpha}{\Gamma(\alpha)} z^{\alpha-1} e^{-\lambda z}, \quad z \geq 0. \tag{B.0.1}$$

The expected value and standard deviation are

$$\text{E}[Z] = \frac{\alpha}{\lambda} \quad \text{and} \quad \text{Sd}[Z] = \frac{\sqrt{\alpha}}{\lambda}.$$

For a standard gamma distributed random variable Z_0, $\text{E}[Z_0] = 1$, thus $\alpha = \lambda$. Hence, (B.0.1) can be rewritten as,

$$f_0(z_0) = \frac{\alpha^\alpha}{\Gamma(\alpha)} z_0^{\alpha-1} e^{-\alpha z_0}, \quad z_0 > 0,$$

which is the same as (5.3.2).

Finding the mean and standard deviation of $Z = \xi Z_0$

The standard deviation of Z_0 is

$$\mathrm{sd}[Z_0] = \frac{\sqrt{\alpha}}{\alpha} = \frac{1}{\sqrt{\alpha}}.$$

Hence, the mean and the standard deviation of $Z = \xi Z_0$ are

$$\mathrm{E}[Z] = \xi \quad \text{and} \quad \mathrm{sd}[Z] = \frac{\xi}{\sqrt{\alpha}}.$$

Section 5.5

Finding the log-gamma density function.

The log-gamma distribution is defined such that if X is log-gamma distributed, then $Y = \log(1 + X)$ is gamma distributed, i.e. $X = e^Y - 1$. A gamma distributed random variable Y can be written as $Y = \xi Y_0$, where $\xi = \mathrm{E}[Y]$ and Y_0 is a standard gamma as described above and in Section 5.3. The cumulative distribution function $F_{\text{l-g}}$ of a log-gamma distributed random variable is then

$$\begin{aligned} F_{\text{l-g}}(x) &= P(X \leq x) \\ &= P(e^{\xi Y_0} - 1 \leq x) \\ &= P\left(Y_0 \leq \frac{\log(x+1)}{\xi}\right) \\ &= F_{\text{g}}\left(\frac{\log(x+1)}{\xi}\right), \end{aligned}$$

where F_{g} is the cumulative distribution function for the standard gamma distribution.

The density distribution is found by differentiating the cumulative distribution once,

$$\begin{aligned} f_{\text{l-g}} &= \frac{d}{dx} F_{\text{l-g}}(x) \\ &= \frac{d}{dx} F_{\text{g}}\left(\frac{\log(x+1)}{\xi}\right) \\ &= \frac{1}{\xi(1+x)} f_{\text{g}}\left(\frac{\log(x+1)}{\xi}\right) \\ &= \frac{1}{\xi(1+z)} \frac{\alpha^\alpha}{\Gamma(\alpha)} \{\log(1+z)/\xi\}^{\alpha-1} e^{-\alpha \log(1+z)/\xi}, \end{aligned}$$

which is the same as (5.5.1).

Chapter 6

Section 6.2

Finding the new density function of the observations Y when transformation (6.1.1) **is applied.**

The transformations applied in Section 6 is

$$Z = \frac{(1+Y)^\lambda - 1}{\lambda} \Rightarrow Y = (Z\lambda + 1)^{1/\lambda} - 1,$$

where Y is the original observations. We assume that the transformed data $Z_1, ..., Z_n$ are extended Pareto distributed.

The cumulative distribution function of the observations is

$$F_Y(y) = F_Z\left(\frac{(1+y)^\lambda - 1}{\lambda}\right).$$

By differentiating once we get the density function,

$$f_Y(y) = \frac{\mathrm{d}}{\mathrm{d}y}F_Y(y) = \frac{\mathrm{d}}{\mathrm{d}y}F_Z\left(\frac{(1+y)^\lambda - 1}{\lambda}\right) = (1+y)^{\lambda-1} f_Z\left(\frac{(1+y)^\lambda - 1}{\lambda}\right),$$

where f_Z is the extended Pareto density function. Hence, the density function of Y is

$$f_Y(y) = (1+y)^{\lambda-1}\frac{\Gamma(\alpha+\theta)}{\Gamma(\alpha)\Gamma(\theta)}\frac{1}{\beta}\frac{\left(\frac{(1+y)^\lambda-1}{\lambda\beta}\right)^{\theta-1}}{\left(1+\frac{(1+y)^\lambda-1}{\lambda\beta}\right)^{\alpha+\theta}}, \qquad \alpha,\beta,\theta > 0, \lambda \in \mathbb{R},$$

which can be rewritten as

$$f_Y(y) = \frac{\Gamma(\alpha+\theta)}{\Gamma(\alpha)\Gamma(\theta)}\frac{(\lambda\beta)^{\alpha+1}}{\beta}\frac{(1+y)^{\lambda-1}\{(1+y)^\lambda - 1\}^{\theta-1}}{\{\lambda\beta + (1+y)^\lambda - 1\}^{\alpha+\theta}}, \qquad \alpha,\beta,\theta > 0, \lambda \in \mathbb{R}.$$

It follows that the log-likelihood function needed to find the estimates $(\hat{\alpha}, \hat{\beta}, \hat{\theta}, \hat{\lambda})$ is

$$\begin{aligned} l(\alpha,\beta,\theta,\lambda) =& n[\log\Gamma(\alpha+\theta) - \log\Gamma(\alpha) - \log\Gamma(\theta) + (\alpha+1)\log(\lambda\beta) - \log\beta] \\ &+ (\lambda-1)\sum_{i=1}^{n}\log(1+y_i) + (\theta-1)\sum_{i=1}^{n}\log\{(1+y_i)^\lambda - 1\} \\ &- (\alpha+\theta)\sum_{i=1}^{n}\log\{\lambda\beta + (1+y_i)^\lambda - 1\}. \end{aligned}$$

Appendix C

R-scripts

R-code used to find total, estimation and systematic error

The R-code below is used to find the error terms, ratios, expected value of estimated reserve defines in (4.5.1), (4.5.2) and (4.5.3) respectively. These quantities are used to analyse the possibility of using the extended Pareto distribution as default loss model. The output from the programs are used to find the values given in the tables in Chapter 4 and Chapter 5.

The R-code below is used to determine the parameters in Table 5.1 at the beginning of Chapter 5. The parameters (ξ, α) from the log-gamma distribution is input in the first program, where the parameter sets from set 1 is calculated. In the second program, also α in the Pareto distribution is an input parameter. This program calculates the parameter sets 2 and 3.

```
parset1=function(m, xi, alpha){
#Simulate m log-Gamma variables
x= exp(xi*rgamma(m, alpha)/alpha)-1
#Determine the mean and standard deviation.
mean_lg=mean(x)
sd_lg=sd(x)

#log-normal
sigma=sqrt(log((sd_lg/mean_lg)**2+1))
mu=log(mean_lg)-sigma**2/2

#Gamma
xi_g=mean_lg
alpha_g=(xi_lg/sd_lg)**2

#Pareto
a=(sd_lg/mean_lg)**2
alpha_p=2*a/(a-1)
beta_p=mean_lg*(alpha_p-1)

list(mu_ln=mu, sigma_ln=sigma, xi_g=xi_g, alpha_g=alpha_g, mean_lg=mean(x), sd_
    lg=sd(x), beta_p=beta_p,alpha_p=alpha_p)
}


parset2and3=function(m, xi, alpha, alpha_p){
#Simulate m log-Gamma variables
x= exp(xi*rgamma(m, alpha)/alpha)-1
#Determine the mean and standard deviation.
```

```
mean_lg=mean(x)
sd_lg=sd(x)

#log-normal
sigma=sqrt(log((sd_lg/mean_lg)**2+1))
mu=log(mean_lg)-sigma**2/2

#Gamma
xi_g=mean_lg
alpha_g=(xi_lg/sd_lg)**2

#Pareto
beta_p=mean_lg*(alpha_p-1)
sd_p=mean_lg*sqrt(alpha_p/(alpha_p-2))

list(mu_ln=mu, sigma_ln=sigma, xi_g=xi_g, alpha_g=alpha_g, mean_lg=mean(x), sd_
    lg=sd(x), beta_p=beta_p, sd_p=sd_p)
}
```

The program below gives the negative of the log-likelihood function for the extended Pareto distribution. It is minimized within other programs, in order to find the maximum likelihood estimates $\hat{\theta} = (\hat{\alpha}, \hat{\beta}, \hat{\theta})$.

```
paretolik=function(s,z)
{
t=exp(s)
l_1=-lgamma(t[1]+t[2])+lgamma(t[1])+lgamma(t[2])+t[2]*log(t[3])
l_2=-(t[2]-1)*mean(log(z))+(t[1]+t[2])*mean(log(1+z/t[3]))
l_1+l_2
}
```

Simulating $\Psi(g), \Psi(\theta_0)$ and $\Psi(\hat{\theta})$ when there is systematic error

The programs below are those used to estimated $\Psi(g), \Psi(\theta_0)$ and $\Psi(\hat{\theta})$ when $g \sim$log-normal or $g \sim$log-gamma. That is, when the underlying distribution is outside the parametric family. The programs are given for the situation where $g \sim$log-normal. What has to be changed when $g \sim$log-gamma is illustrated within the programs. The input parameters are those used in Chapter 4.

The program below is used to find θ_0 with the Monte Carlo approach, as discussed in Section 4.1.1.

```
minKL_lognormal=function(m=100000,mu=0, sigma=1)
{
#Generate m log-normal data
z=rlnorm(m, meanlog=mu, sdlog=sigma)
#When g-log-gamma, this is changed to z=exp(xi*rgamma(m, alpha)/alpha)-1.

#Minimize the negative log-likelihood function and return the parameters who
    minimize it.
o=optim(c(1,0.7,log(mean(z))),paretolik,z=z)
alpha0=exp(o$par[1])
theta0=exp(o$par[2])
beta0=exp(o$par[3])
#Value of the Kullback-Leibler distance.
```

```
f= (gamma(alpha0+theta0)/(gamma(alpha0)*gamma(theta0)*beta0))*((z/beta0)**(
    theta0-1))/((1+ z/beta0)**(alpha0+theta0))
I = (-1/m)*sum(log(f))
list(Imin=min(I),alpha0=alpha0,theta0=theta0,beta0=beta0)
}
```

The program below is used to find θ_0 by means of Gauss-Legendre quadrature, as discussed in Section 4.1.2. The vectors x and w are found with a program called "genlegendre1.f" and are then written into the file "genlegendre.res". The program was provided by my supervisor. It is written in Fortran, and is not given here.

```
GL_lognormal=function(mu=0, sigma=1)
{
#Load the parameters from the file#genlegendre.res".
A=matrix(scan("genlegendre.res"), byrow=T, ncol=2)
x=A[,1]
w=a[,2]
f=function(s,x){
#The negative of Equation (2.1.4) from Chapter 2.
t=exp(s)
sum(-w*dnorm(x, mu, sigma)*log((gamma(t[1]+t[2])/(gamma(t[1])*gamma(t[2])*t[3])
    )*((exp(x)/t[3])**(t[2]-1))/((1+ exp(x)/t[3])**(t[1]+t[2]))))
#when g-log-gamma the first part is changed to sum(-w*dgamma(x, scale=xi/alpha,
     shape=alpha)*"
}


#Minimize f and return the parameters who minimize it.
o=optim(c(1,2,3),f,x=x)
alpha0=exp(o$par[1])
theta0=exp(o$par[2])
beta0=exp(o$par[3])
list(alpha0=alpha0, theta0=theta0, beta0=beta0)
}
```

The program below is used to find $\Psi(\theta_0)$, $\Psi(g)$ and systematic error as discussed in Section 4.3.

```
reserve0=function(m=100000,eps=0.01, JmuT=300, mu=0, sigma=1){
#Load the parameters found with the program GL_lognormal(mu, sigma)
alpha0= GL_lognormal(mu,sigma)$alpha0
theta0= GL_lognormal(mu,sigma)$theta0
beta0= GL_lognormal(mu,sigma)$beta0
#Changed to GL_logggamma when g-log-gamma

#Procedure as in Algorithem 4.2.
N = rpois(m, JmuT)
#Matrices used to avoid loops.
maxN=max(N)
A=t(matrix(rep(c(1:maxN),m),maxN,m))
B=matrix(rep(N,maxN),m,maxN)
I=t(B-A>=0)

zlnorm=matrix(rlnorm(maxN*m, mu, sigma), maxN,m)
#Changed to zlgamma=matrix(exp(xi*rgamma(maxN*m, alpha)/alpha)-1, maxN,m) when
    g-log-gamma.
Zlnorm=zlnorm*I
Xlnorm=apply(Zlnorm,2,sum)
```

```
zexpa0= matrix(beta0*rgamma(maxN*m, theta0)/rgamma(maxN*m, alpha0), maxN,m)
Zexpa0=zexpa0*I
Xexpa0=apply(Zexpa0,2,sum)

reserve_lnorm = sort(Xlnorm)[(1-eps)*m]
reserve_expa0 = sort(Xexpa0)[(1-eps)*m]
error_sys = abs(reserve_expa0-reserve_lnorm)
list(reserve_lnorm =reserve_lnorm, reserve_expa0=reserve_expa0, error_sys=error
    _sys)
}
```

The program below is used to estimate $\Psi(\hat{\theta})$ by means of maximum likelihood as discussed in Section 4.2.2 and Section 4.3. It is also used to calculate the error terms, the ratios and the expected value of the estimated reserve defined in (4.5.1), (4.5.2) and (4.5.3) respectively. It also returns the three terms A1, A2 and A3 from decomposition (2.4.4) of total error given in Section 4.4.

```
reservetot1=function(m=10000,eps=0.01, JmuT=300,mu=0, sigma=1, n=100000, M=100)
    {
#Procedure as in Algorithem 4.2 repeated M times.
N = matrix(rpois(m*M, JmuT),m,M)
maxN=max(N)
Xexpa_hat=matrix(0,m,M)
for(j in 1:M){
z=rlnorm(n,mu,sigma)
#Changed to z=exp(xi*rgamma(m, alpha)/alpha)-1 when g-log-gamma.
o=optim(c(1,0.7,log(mean(z))),paretolik,z=z)
alpha_hat=exp(o$par[1])
theta_hat=exp(o$par[2])
beta_hat=exp(o$par[3])
Ni=N[,j]
Zexpa_hat= matrix(beta_hat*rgamma(maxN*m, theta_hat)/rgamma(maxN*m, alpha_hat),
    maxN,m)
#Changed to zlgamma=matrix(exp(xi*rgamma(maxN*m, alpha)/alpha)-1, maxN,m) when
    g-log-gamma.
A=t(matrix(rep(c(1:maxN),m),maxN,m))
B=matrix(rep(Ni,maxN),m,maxN)
Zexpa_hat=Zexpa_hat*t((B-A>=0))
Xexpa_hat[,j]=apply(Zexpa_hat,2,sum)
}
#Write these numbers manually because it saves simulation time. Loading them
    from the other program takes long. They could have been loaded in by writin:
     reserve_lnorm=reserve0(m=100000,eps,JmuT,mu,sigma)$reserve_lnorm, and
    reserve_expa0=reserve0(m=100000,eps,JmuT,mu,sigma)$reserve_expa0
reserve_lnorm = 613.3
reserve_expa0 = 678.7
reserve_expa_hat = apply(Xexpa_hat,2,sort)[(1-eps)*m,]

mean_reserve_expa_hat=mean(reserve_expa_hat)
sd_reserve_expa_hat=sd(reserve_expa_hat)

E_Tot=sqrt(mean((reserve_lnorm-reserve_expa_hat)**2))
E_Ran=sqrt(mean((reserve_expa0-reserve_expa_hat)**2))
E_Sys=abs(reserve_lnorm-reserve_expa0)

R_Tot=E_Tot/reserve_lnorm
R_Ran=E_Ran/reserve_lnorm
```

```
R_Sys=E_Sys/reserve_lnorm

E_hattheta=mean_reserve_expa_hat

A1=mean((reserve_expa_hat-reserve_expa0)**2)
A2=(reserve_expa0-reserve_lnorm)**2
A3=2*mean(reserve_expa_hat-reserve_expa0)*(reserve_expa0-reserve_lnorm)


list(mean_reserve_expa_hat=mean_reserve_expa_hat,sd_reserve_expa_hat=sd_reserve
    _expa_hat,E_Tot=E_Tot, E_Ran=E_Ran, E_Sys=E_Sys,R_Tot=R_Tot, R_Ran=R_Ran, R_
    Sys=R_Sys, E_hattheta=E_hattheta,A1=A1, A2=A2, A3=A3)
}
```

The program below is used to estimate $\Psi(\hat{\theta})$ by means of the delta method as discussed in Section 4.2.1.

```
reserve_norm=function(M=100, m=10000,eps=0.01, JmuT=300, n=100000, mu=0, sigma
    =1){
#Load the parameters found with the program GL_lognormal(mu, sigma).
alpha0= GL_lognormal(mu,sigma)$alpha0
theta0= GL_lognormal(mu,sigma)$theta0
beta0= GL_lognormal(mu,sigma)$beta0
#Changed to GL_loggamma when g-log-gamma.
parameter0=c(alpha0, theta0, beta0)

#The integrands for every term in the Ig matrix.
f_alpha2=function(x){
((digamma(alpha0+theta0) - digamma(alpha0) - log(1+x/beta0))**2)*dlnorm(x)
}

f_alphatheta=function(x){
(digamma(alpha0+theta0) - digamma(alpha0) - log(1+x/beta0))*(digamma(alpha0+
    theta0) - digamma(theta0) +log(x/beta0) -log(1+ x/beta0))*dlnorm(x)
}

f_alphabeta=function(x){
(digamma(alpha0+theta0) - digamma(alpha0) - log(1+x/beta0))*(-theta0/beta0 + (
    alpha0+theta0)*((x/(beta0**2))/(1+x/beta0)))*dlnorm(x)
}

f_theta2=function(x){
((digamma(alpha0+theta0) - digamma(theta0) +log(x/beta0) -log(1+ x/beta0))**2)*
    dlnorm(x)
}

f_thetabeta=function(x){
(digamma(alpha0+theta0) - digamma(theta0) +log(x/beta0) -(1+ x/beta0))*(-theta0
    /beta0 + (alpha0+theta0)*log((x/(beta0**2))/(1+x/beta0)))*dlnorm(x)
}

f_beta2=function(x){
((-theta0/beta0 + (alpha0+theta0)*((x/(beta0**2))/(1+x/beta0)))**2)*dlnorm(x)
}

#The integrands for every term in the lambda matrix.
f2_alpha2=function(x){
(psigamma(alpha0+theta0,1)-psigamma(alpha0,1))*dlnorm(x)
}
```

```
f2_alphatheta=function(x){
psigamma(alpha0+theta0,1)*dlnorm(x)
}
f2_theta2=function(x){
(psigamma(alpha0+theta0) -psigamma(theta0))*dlnorm(x)
}

f2_alphabeta=function(x){
((x/(beta0**2))/(1+x/beta0))*dlnorm(x)
}
f2_thetabeta=function(x){
(-1/beta0 + (x/(beta0**2))/(1+x/beta0))*dlnorm(x)
}
f2_beta2=function(x){
(theta0/(beta0**2) +((alpha0+theta0)*x/(beta0**3)) * ((x/beta0-2)/(1+x/beta0)**
    2))*dlnorm(x)
}

#Finding the Ig-matrix.
Ig = matrix(0,3,3)
Ig[1,1]= integrate(f_alpha2,0,100000)$value
Ig[2,1]= integrate(f_alphatheta,0,100000)$value
Ig[1,2]= Ig[2,1]
Ig[1,3]= integrate(f_alphabeta,0,10000)$value
Ig[3,1]= Ig[1,3]
Ig[3,3]= integrate(f_beta2,0,10000)$value
Ig[2,3]= integrate(f_thetabeta,0,10000)$value
Ig[3,2]= Ig[2,3]
Ig[2,2]= integrate(f_theta2,0,10000)$value

#Finding the Lambda-matrix given parameter0.
lambda=matrix(0,3,3)
lambda[1,1]=integrate(f2_alpha2,0,10000)$value
lambda[1,2]=integrate(f2_alphatheta,0,10000)$value
lambda[2,2]=integrate(f2_theta2,0,10000)$value
lambda[2,1]=lambda[1,2]
lambda[1,3]=integrate(f2_alphabeta,0,10000)$value
lambda[3,1]=lambda[1,3]
lambda[3,3]=integrate(f2_beta2,0,10000)$value
lambda[2,3]=integrate(f2_thetabeta,0,10000)$value
lambda[3,2]=lambda[2,3]

#The covariance matrix, formula given in Section 2.3.
cov_mat= (solve(lambda)%*%Ig%*%solve(lambda))/sqrt(n)

#Given cov_mat and parameter0, we can use the delta method to estimate the
    reserve.
N = matrix(rpois(m*M, JmuT),m,M)
maxN=max(N)
Xexpa_hat=matrix(0,m,M)
for(j in 1:M){
par_hat=rnorm(3,parameter0, cov_mat)
alpha_hat=par_hat[1]
theta_hat=par_hat[2]
beta_hat=par_hat[3]
Ni=N[,j]
A=t(matrix(rep(c(1:maxN),m),maxN,m))
B=matrix(rep(Ni,maxN),m,maxN)
I=t(B-A>=0)
```

```
Zexpa_hat= matrix(beta_hat*rgamma(maxN*m, theta_hat)/rgamma(maxN*m, alpha_hat),
    maxN,m)
#Changed to zlgamma=matrix(exp(xi*rgamma(maxN*m, alpha)/alpha)-1, maxN,m) when
    g-log-gamma.
Zexpa_hat=Zexpa_hat*I
Xexpa_hat[,j]=apply(Zexpa_hat,2,sum)
}
reserve_expa_hat = apply(Xexpa_hat,2,sort)[(1-eps)*m,]


list(mean_reserve_expa_hat=mean(reserve_expa_hat), sd_reserve_expa_hat=sd(
    reserve_expa_hat))
}
```

Simulating $\Psi(g)$ and $\Psi(\hat{\theta})$ when there is no systematic error

The program below is used to estimated $\Psi(g)$ and $\Psi(\hat{\theta})$ when $g \sim$gamma or $g \sim$Pareto. That is, when the underlying distribution is within the parametric family. The program is given for the situation where $g \sim$gamma. What has to be changed when $g \sim$Pareto is illustrated within the program.

```
reservetot2=function(m=100000,eps=0.01, JmuT=50, xi=2.37, alpha=0.38, n=100000,
    M=10000){
#Procedure as in Algorithem 4.1 repeated M times.
N = rpois(m, JmuT)
maxN=max(N)
Xexpa_hat=matrix(0,m,M)
for(j in 1:M){
z=xi*rgamma(n,shape=alpha)/alpha
#Changed to z=beta*(runif(n)**(-1/alpha)-1) when g-Pareto.
o=optim(c(1,0.7,log(mean(z))),paretolik,z=z)
alpha_hat=exp(o$par[1])
theta_hat=exp(o$par[2])
beta_hat=exp(o$par[3])
Ni=N[,j]
Zexpa_hat= matrix(beta_hat*rgamma(maxN*m, theta_hat)/rgamma(maxN*m, alpha_hat),
    maxN,m)
A=t(matrix(rep(c(1:maxN),m),maxN,m))
B=matrix(rep(N,maxN),m,maxN)
I=t(B-A>=0)
Zexpa_hat=Zexpa_hat*I
Xexpa_hat=apply(Zexpa_hat,2,sum)
}
reserve_expa_hat = sort(Xexpa_hat)[(1-eps)*m]

zgamma=matrix(xi*rgamma(maxN*m,shape=alpha)/alpha, maxN, m)
#Changed to zpareto=matrix(beta*(runif(maxN*m)**(-1/alpha)-1), maxN, m) when g-
    Pareto.
Zgamma=zgamma*I
Xgamma=apply(Zgamma,2,sum)
reserve_gamma = sort(Xgamma)[(1-eps)*m]

E_Ran = sqrt(mean((reserve_gamma-reserve_expa_hat)**2))

R_Ran=E_Ran/reserve_gamma

list(E_Ran=E_Ran, R_Ran=R_Ran,mean_reserve_expa_hat=mean(reserve_expa_hat))
}
```

Reinsurance

In Section 4.6, reinsurance is introduced. This is included by changing two lines in reserve0() for finding $\Psi(\theta_0)$ and $\Psi(g)$. In the program reservetot1() used to estimate $\Psi(\hat{\theta})$ when $g \sim$log-normal or $g \sim$log-gamma, one line has to be changed. While in the program reservetot2() used to estimate $\Psi(\hat{\theta})$ and $\Psi(g)$ when $g \sim$gamma or $g \sim$Pareto, two lines have to be changed. For a given value of b, in reserve0() we use

```
zlnorm=matrix(pmin(rlnorm(maxN*m, mu, sigma),b), maxN,m)
#Changed to zlgamma=matrix(pmin(exp(xi*rgamma(maxN*m, alpha)/alpha)-1,b), maxN,
    m) when g-log-gamma.

zexpa0= matrix(pmin(beta0*rgamma(maxN*m,theta0)/rgamma(maxN*m, alpha0),b), maxN
    ,m)
```

In reservetot1() we use

```
zexpa_hat= matrix(pmin(beta_hat*rgamma(maxN*m,theta_hat)/rgamma(maxN*m, alpha_
    hat),b), maxN,m)
```

In reservetot2() we use

```
Zexpa_hat= matrix(pmin(beta_hat*rgamma(maxN*m,theta_hat)/rgamma(maxN*m, alpha_
    hat),b), maxN,m)

zgamma=matrix(pmin(xi*rgamma(maxN*m,shape=alpha)/alpha,b), maxN, m)
#Changed to zpareto=matrix(pmin(beta*(runif(maxN*m)**(-1/alpha)-1),b), maxN, m)
     when g-Pareto.
```

R-code for the graphical representations

The programs below is used to make Figure 3.1 in Chapter 3.

```
#The extended Pareto density function.
density_expa=function(x,alpha, beta, theta){
t1=gamma(theta+alpha)/(gamma(theta)*gamma(alpha)*beta)
t2= (x/beta)**(theta-1)
t3=(1+ x/beta)**(alpha + theta)
t1*t2/t3
}

plot_density_exPa=function(alpha=2.5, beta=1, theta1=0.8, theta2=1.5){
x=0:50000/10000+0.000000001
plot(x,density_expa(x,alpha, beta, theta1), type="l", ylim=c(0,2), xlab="",
    ylab= "")
lines(x, density_expa(x,alpha, beta, theta2), lty=4)
legend(3,1, c("theta=0.8", "theta=1.5"), lty = c(1,4))
}
```

The programs below is used to make Figure 5.1 in Section 5.2, and the corresponding figure for the situation where $g \sim$log-gamma, Figure 5.4 in Section 5.5.

```
plotdensity_lognormal=function(mu1=0.22, sigma1=1.13, mu2=1.12, sigma2=0.85,
    mu3=0.41, sigma3=0.78){
#The extended Pareto density function.
densityexpa=function(x,alpha_hat, beta_hat, theta_hat){
t1=gamma(theta_hat+alpha_hat)/(gamma(theta_hat)*gamma(alpha_hat)*beta_hat)
t2= (x/beta_hat)**(theta_hat-1)
```

```
t3=(1+ x/beta_hat)**(alpha_hat + theta_hat)
t1*t2/t3
}
#Replaced by the log-gamma density when the corresponding figure in Section 5.5
     is made.

#Load the parameters from the file#genlegendre.res".
A=matrix(scan("genlegendre.res"), byrow=T, ncol=2)
x_gl=A[,1]
w=a[,2]

#The negative of equation (2.1.4) from chapter 2 for the 3 different parameter
    sets.
f1=function(s,x_gl){
t=exp(s)
sum(-w*dnorm(x_gl, mu1, sigma1)*log((gamma(t[1]+t[2])/(gamma(t[1])*gamma(t[2])*
    t[3]))
*((exp(x_gl)/t[3])**(t[2]-1))/((1+ exp(x_gl)/t[3])**(t[1]+t[2]))))
}
f2=function(s,x_gl){
t=exp(s)
sum(-w*dnorm(x_gl, mu2, sigma2)*log((gamma(t[1]+t[2])/(gamma(t[1])*gamma(t[2])*
    t[3]))
*((exp(x_gl)/t[3])**(t[2]-1))/((1+ exp(x_gl)/t[3])**(t[1]+t[2]))))
}
f3=function(s,x_gl){
t=exp(s)
sum(-w*dnorm(x_gl, mu3, sigma3)*log((gamma(t[1]+t[2])/(gamma(t[1])*gamma(t[2])*
    t[3]))
*((exp(x_gl)/t[3])**(t[2]-1))/((1+ exp(x_gl)/t[3])**(t[1]+t[2]))))
}
#When g-log-gamma the first part is changed to sum(-w*dgamma(x_gl, scale=xi/
    alpha, shape=alpha)*"

x1=seq(from=0.000001, to=10, 0.01)
x2=seq(from=100, to=200, 1)

o=optim(c(1,2,3),f1,x_gl=x_gl)
alpha01=exp(o$par[1])
theta01=exp(o$par[2])
beta01=exp(o$par[3])
o=optim(c(1,2,3),f2,x_gl=x_gl)
alpha02=exp(o$par[1])
theta02=exp(o$par[2])
beta02=exp(o$par[3])
o=optim(c(1,2,3),f3,x_gl=x_gl)
alpha03=exp(o$par[1])
theta03=exp(o$par[2])
beta03=exp(o$par[3])

par(mfrow=c(2,3))
plot(x1,densityexpa(x1,alpha01, beta01, theta01), type="l", xlab="", ylab="",
     lty=2, main="Parameter set 1", ylim=c(0,0.6))
lines(x1,dlnorm(x1, mu1, sigma1), lty=1)
legend(5,0.4, c("g", "theta_0"), lty = c(1,2))
plot(x1,densityexpa(x1,alpha02, beta02, theta02), type="l", xlab="", ylab="",
     lty=2, main="Parameter set 2", ylim=c(0,0.6))
lines(x1,dlnorm(x1, mu2, sigma2))
legend(5,0.4, c("g", "theta_0"), lty = c(1,2))
plot(x1,densityexpa(x1,alpha03, beta03, theta03), type="l", xlab="", ylab="",
     lty=2, main="Parameter set 3", ylim=c(0,0.6))
```

```
lines(x1,dlnorm(x1, mu3, sigma3))
legend(5,0.4, c("g", "theta_0"), lty = c(1,2))

plot(x2,densityexpa(x2,alpha01, beta01, theta01), type="l", ylim=c(0,10**(-5)),
     xlab="", ylab="", lty=2, main="Parameter set 1")
lines(x2,dlnorm(x2, mu1, sigma1))
legend(150, 8*10**(-6), c("g", "theta_0"), lty = c(1,2))
plot(x2,densityexpa(x2,alpha02, beta02, theta02), type="l", ylim=c(0,10**(-5)),
     xlab="", ylab="", lty=2, main="Parameter set 2")
lines(x2,dlnorm(x2, mu2, sigma2))
legend(150, 8*10**(-6), c("g", "theta_0"), lty = c(1,2))
plot(x2,densityexpa(x2,alpha03, beta03, theta03), type="l", ylim=c(0,10**(-5)),
     xlab="", ylab="", lty=2, main="Parameter set 3")
lines(x2,dlnorm(x2, mu3, sigma3))
legend(150, 8*10**(-6), c("g", "theta_0"), lty = c(1,2))
}
```

The programs below is used to make Figure 5.2 in Section 5.2, and the corresponding figure for the situation where g ∼log-gamma, Figure 5.5 in Section 5.5. It is also used to make Figure A.1, Figure A.2, Figure A.5 and Figure A.6 in Appendix A.

```
densityplot=function(mu=0.22, sigma=1.13){
#The extended Pareto density function.
density_expa=function(x,alpha_hat, beta_hat, theta_hat){
t1=gamma(theta_hat+alpha_hat)/(gamma(theta_hat)*gamma(alpha_hat)*beta_hat)
t2= (x/beta_hat)**(theta_hat-1)
t3=(1+ x/beta_hat)**(alpha_hat + theta_hat)
t1*t2/t3
}
#Replaced by the log-gamma density when the corresponding figure in Section 5.5
     is made.

#Load the parameters from the file#genlegendre.res".
A=matrix(scan("genlegendre.res"), byrow=T, ncol=2)
x_gl=A[,1]
w=a[,2]

f=function(s,x_gl){
#The negative of equation (2.1.4) from Chapter 2.
t=exp(s)
sum(-w*dnorm(x_gl, mu, sigma)*log((gamma(t[1]+t[2])/(gamma(t[1])*gamma(t[2])*t
    [3]))
*((exp(x_gl)/t[3])**(t[2]-1))/((1+ exp(x_gl)/t[3])**(t[1]+t[2]))))
}
#When g-log-gamma the first part is changed to sum(-w*dgamma(x_gl, scale=xi/
    alpha, shape=alpha)*

par(mfrow=c(2,3))
x1=seq(from=0.000001, to=10, 0.01)
x2=seq(from=100, to=200, 1)

o=optim(c(1,2,3),f,x_gl=x_gl)
alpha0=exp(o$par[1])
theta0=exp(o$par[2])
beta0=exp(o$par[3])
plot(x1, density_expa(x1,alpha0, beta0, theta0), lty=1, type="l", main="n=50",
    xlab="", ylab= "", lwd=2)
#-------------------------
n1=50
```

```
z1=rlnorm(n1,mu,sigma)
#Changed to z=exp(xi*rgamma(m, alpha)/alpha)-1 when g-log-gamma.
o1=optim(c(1,0.7,log(mean(z1))),paretolik,z=z1)
alpha_hat11=exp(o1$par[1])
theta_hat11=exp(o1$par[2])
beta_hat11=exp(o1$par[3])
lines(x1, density_expa(x1,alpha_hat11, beta_hat11, theta_hat11), lty=2)
#-----------------------------
n1=50
z1=rlnorm(n1,mu,sigma)
o1=optim(c(1,0.7,log(mean(z1))),paretolik,z=z1)
alpha_hat12=exp(o1$par[1])
theta_hat12=exp(o1$par[2])
beta_hat12=exp(o1$par[3])
lines(x1, density_expa(x1,alpha_hat12, beta_hat12, theta_hat12), lty=2)
#-----------------------------
n1=50
z1=rlnorm(n1,mu,sigma)
o1=optim(c(1,0.7,log(mean(z1))),paretolik,z=z1)
alpha_hat13=exp(o1$par[1])
theta_hat13=exp(o1$par[2])
beta_hat13=exp(o1$par[3])
lines(x1, density_expa(x1,alpha_hat13, beta_hat13, theta_hat13), lty=2)
#-----------------------------
n1=50
z1=rlnorm(n1,mu,sigma)
o1=optim(c(1,0.7,log(mean(z1))),paretolik,z=z1)
alpha_hat14=exp(o1$par[1])
theta_hat14=exp(o1$par[2])
beta_hat14=exp(o1$par[3])
lines(x1, density_expa(x1,alpha_hat14, beta_hat14, theta_hat14), lty=2)
#-----------------------------
n1=50
z1=rlnorm(n1,mu,sigma)
o1=optim(c(1,0.7,log(mean(z1))),paretolik,z=z1)
alpha_hat15=exp(o1$par[1])
theta_hat15=exp(o1$par[2])
beta_hat15=exp(o1$par[3])
lines(x1, density_expa(x1,alpha_hat15, beta_hat15, theta_hat15), lty=2)
#-----------------------------

legend(4,0.4, c("theta_0", "theta_hat"), lty = c(1,2), lwd=c(2,1))

#-----------------------------
#-----------------------------
plot(x1, density_expa(x1,alpha0, beta0, theta0), lty=1, type="l", main="n=1000"
    , xlab="␣", ylab= "␣", lwd=2)
#-----------------------------
n2=1000
z2=rlnorm(n2,mu,sigma)
o2=optim(c(1,0.7,log(mean(z2))),paretolik,z=z2)
alpha_hat21=exp(o2$par[1])
theta_hat21=exp(o2$par[2])
beta_hat21=exp(o2$par[3])
lines(x1, density_expa(x1,alpha_hat21, beta_hat21, theta_hat21), lty=2)
#-----------------------------
n2=1000
z2=rlnorm(n2,mu,sigma)
o2=optim(c(1,0.7,log(mean(z2))),paretolik,z=z2)
alpha_hat22=exp(o2$par[1])
theta_hat22=exp(o2$par[2])
```

```
beta_hat22=exp(o2$par[3])
lines(x1, density_expa(x1,alpha_hat22, beta_hat22, theta_hat22), lty=2)
#-----------------------------
n2=1000
z2=rlnorm(n2,mu,sigma)
o2=optim(c(1,0.7,log(mean(z2))),paretolik,z=z2)
alpha_hat23=exp(o2$par[1])
theta_hat23=exp(o2$par[2])
beta_hat23=exp(o2$par[3])
lines(x1, density_expa(x1,alpha_hat23, beta_hat23, theta_hat23), lty=2)
#-----------------------------
n2=1000
z2=rlnorm(n2,mu,sigma)
o2=optim(c(1,0.7,log(mean(z2))),paretolik,z=z2)
alpha_hat24=exp(o2$par[1])
theta_hat24=exp(o2$par[2])
beta_hat24=exp(o2$par[3])
lines(x1, density_expa(x1,alpha_hat24, beta_hat24, theta_hat24), lty=2)
#-----------------------------
n2=1000
z2=rlnorm(n2,mu,sigma)
o2=optim(c(1,0.7,log(mean(z2))),paretolik,z=z2)
alpha_hat25=exp(o2$par[1])
theta_hat25=exp(o2$par[2])
beta_hat25=exp(o2$par[3])
lines(x1, density_expa(x1,alpha_hat25, beta_hat25, theta_hat25), lty=2)
#-----------------------------

legend(4,0.4, c("theta_0", "theta_hat"), lty = c(1,2), lwd=c(2,1))
#-----------------------------
#-----------------------------
plot(x1, density_expa(x1,alpha0, beta0, theta0), lty=1, type="l", main="n
    =100000", xlab="␣", ylab= "␣", lwd=2)
#-----------------------------
n3=100000
z3=rlnorm(n3,mu,sigma)
o3=optim(c(1,0.7,log(mean(z3))),paretolik,z=z3)
alpha_hat31=exp(o3$par[1])
theta_hat31=exp(o3$par[2])
beta_hat31=exp(o3$par[3])
lines(x1, density_expa(x1,alpha_hat31, beta_hat31, theta_hat31), lty=2)
#-----------------------------
n3=100000
z3=rlnorm(n3,mu,sigma)
o3=optim(c(1,0.7,log(mean(z3))),paretolik,z=z3)
alpha_hat32=exp(o3$par[1])
theta_hat32=exp(o3$par[2])
beta_hat32=exp(o3$par[3])
lines(x1, density_expa(x1,alpha_hat32, beta_hat32, theta_hat32), lty=2)
#-----------------------------
n3=100000
z3=rlnorm(n3,mu,sigma)
o3=optim(c(1,0.7,log(mean(z3))),paretolik,z=z3)
alpha_hat33=exp(o3$par[1])
theta_hat33=exp(o3$par[2])
beta_hat33=exp(o3$par[3])
lines(x1, density_expa(x1,alpha_hat33, beta_hat33, theta_hat33), lty=2)
#-----------------------------
n3=100000
z3=rlnorm(n3,mu,sigma)
o3=optim(c(1,0.7,log(mean(z3))),paretolik,z=z3)
```

```
alpha_hat34=exp(o3$par[1])
theta_hat34=exp(o3$par[2])
beta_hat34=exp(o3$par[3])
lines(x1, density_expa(x1,alpha_hat34, beta_hat34, theta_hat34), lty=2)
#-----------------------------
n3=100000
z3=rlnorm(n3,mu,sigma)
o3=optim(c(1,0.7,log(mean(z3))),paretolik,z=z3)
alpha_hat35=exp(o3$par[1])
theta_hat35=exp(o3$par[2])
beta_hat35=exp(o3$par[3])
lines(x1, density_expa(x1,alpha_hat35, beta_hat35, theta_hat35), lty=2)
legend(4,0.4, c("theta_0", "theta_hat"), lty = c(1,2), lwd=c(2,1))

#-----------------------------
plot(x2, density_expa(x2,alpha0, beta0, theta0), lty=1, type="l",ylim=c(0,15*
    (10**(-6))), main="n=50", xlab="␣", ylab= "␣", lwd=2)
lines(x2, density_expa(x2,alpha_hat11, beta_hat11, theta_hat11), lty=2)
lines(x2, density_expa(x2,alpha_hat12, beta_hat12, theta_hat12), lty=2)
lines(x2, density_expa(x2,alpha_hat13, beta_hat13, theta_hat13), lty=2)
lines(x2, density_expa(x2,alpha_hat14, beta_hat14, theta_hat14), lty=2)
lines(x2, density_expa(x2,alpha_hat15, beta_hat15, theta_hat15), lty=2)
legend(140,10*(10**(-6)), c("theta_0", "theta_hat"), lty = c(1,2), lwd=c(2,1))

#-----------------------------
plot(x2, density_expa(x2,alpha0, beta0, theta0), lty=1, type="l",ylim=c(0,15*
    (10**(-6))), main="n=1000", xlab="␣", ylab= "␣", lwd=2)
lines(x2, density_expa(x2,alpha_hat21, beta_hat21, theta_hat21), lty=2)
lines(x2, density_expa(x2,alpha_hat22, beta_hat22, theta_hat22), lty=2)
lines(x2, density_expa(x2,alpha_hat23, beta_hat23, theta_hat23), lty=2)
lines(x2, density_expa(x2,alpha_hat24, beta_hat24, theta_hat24), lty=2)
lines(x2, density_expa(x2,alpha_hat25, beta_hat25, theta_hat25), lty=2)
legend(140,10*(10**(-6)), c("theta_0", "theta_hat"), lty = c(1,2), lwd=c(2,1))

#-----------------------------
plot(x2, density_expa(x2,alpha0, beta0, theta0), lty=1, type="l",ylim=c(0,15*
    (10**(-6))), main="n=100000", xlab="␣", ylab= "␣", lwd=2)
lines(x2, density_expa(x2,alpha_hat31, beta_hat31, theta_hat31), lty=2)
lines(x2, density_expa(x2,alpha_hat32, beta_hat32, theta_hat32), lty=2)
lines(x2, density_expa(x2,alpha_hat33, beta_hat33, theta_hat33), lty=2)
lines(x2, density_expa(x2,alpha_hat34, beta_hat34, theta_hat34), lty=2)
lines(x2, density_expa(x2,alpha_hat35, beta_hat35, theta_hat35), lty=2)
legend(140,10*(10**(-6)), c("theta_0", "theta_hat"), lty = c(1,2), lwd=c(2,1))
}
```

The program below is used to make Figure 5.3 in Section 5.4. It is also used to make Figure A.3 and Figure A.4 in Appendix A.

```
densityplot_pareto=function(alpha=3.21, beta=5.26, n1=50){

densitypareto=function(x, alpha, beta){
(alpha/beta)/((1+x/beta)**(1+alpha))
}

density_expa=function(x,alpha_hat, beta_hat, theta_hat){
t1=gamma(theta_hat+alpha_hat)/(gamma(theta_hat)*gamma(alpha_hat)*beta_hat)
t2= (x/beta_hat)**(theta_hat-1)
t3=(1+ x/beta_hat)**(alpha_hat + theta_hat)
t1*t2/t3
```

```
}

par(mfrow=c(2,3))
x1=seq(from=0.0001, to=10, 0.01)
x2=seq(from=100, to=200, 1)

plot(x1, densitypareto(x1, alpha, beta),lty=1, type="l", main="n=50", xlab="␣",
     ylab= "␣", lwd=2)
#-------------------------

z1=beta*(runif(n1)**(-1/alpha)-1)
o1=optim(c(1,0.7,log(mean(z1))),paretolik,z=z1)
alpha_hat11=exp(o1$par[1])
theta_hat11=exp(o1$par[2])
beta_hat11=exp(o1$par[3])
lines(x1, density_expa(x1,alpha_hat11, beta_hat11, theta_hat11), lty=2)
#-------------------------

z1=beta*(runif(n1)**(-1/alpha)-1)
o1=optim(c(1,0.7,log(mean(z1))),paretolik,z=z1)
alpha_hat12=exp(o1$par[1])
theta_hat12=exp(o1$par[2])
beta_hat12=exp(o1$par[3])
lines(x1, density_expa(x1,alpha_hat12, beta_hat12, theta_hat12), lty=2)
#-------------------------

z1=beta*(runif(n1)**(-1/alpha)-1)
o1=optim(c(1,0.7,log(mean(z1))),paretolik,z=z1)
alpha_hat13=exp(o1$par[1])
theta_hat13=exp(o1$par[2])
beta_hat13=exp(o1$par[3])
lines(x1, density_expa(x1,alpha_hat13, beta_hat13, theta_hat13), lty=2)
#-------------------------

z1=beta*(runif(n1)**(-1/alpha)-1)
o1=optim(c(1,0.7,log(mean(z1))),paretolik,z=z1)
alpha_hat14=exp(o1$par[1])
theta_hat14=exp(o1$par[2])
beta_hat14=exp(o1$par[3])
lines(x1, density_expa(x1,alpha_hat14, beta_hat14, theta_hat14), lty=2)
#-------------------------

z1=beta*(runif(n1)**(-1/alpha)-1)
o1=optim(c(1,0.7,log(mean(z1))),paretolik,z=z1)
alpha_hat15=exp(o1$par[1])
theta_hat15=exp(o1$par[2])
beta_hat15=exp(o1$par[3])
lines(x1, density_expa(x1,alpha_hat15, beta_hat15, theta_hat15), lty=2)
#-------------------------

legend(4,0.4, c("theta_0", "theta_hat"), lty = c(1,2), lwd=c(2,1))


#-------------------------
#-------------------------
plot(x1, densitypareto(x1, alpha, beta),lty=1, type="l", main="n=1000", xlab="␣
    ", ylab= "␣", lwd=2)
#-------------------------
n2=1000
z2=beta*(runif(n2)**(-1/alpha)-1)
o2=optim(c(1,0.7,log(mean(z2))),paretolik,z=z2)
```

```
alpha_hat21=exp(o2$par[1])
theta_hat21=exp(o2$par[2])
beta_hat21=exp(o2$par[3])
lines(x1, density_expa(x1,alpha_hat21, beta_hat21, theta_hat21), lty=2)
#-----------------------------
n2=1000
z2=beta*(runif(n2)**(-1/alpha)-1)
o2=optim(c(1,0.7,log(mean(z2))),paretolik,z=z2)
alpha_hat22=exp(o2$par[1])
theta_hat22=exp(o2$par[2])
beta_hat22=exp(o2$par[3])
lines(x1, density_expa(x1,alpha_hat22, beta_hat22, theta_hat22), lty=2)
#-----------------------------
n2=1000
z2=beta*(runif(n2)**(-1/alpha)-1)
o2=optim(c(1,0.7,log(mean(z2))),paretolik,z=z2)
alpha_hat23=exp(o2$par[1])
theta_hat23=exp(o2$par[2])
beta_hat23=exp(o2$par[3])
lines(x1, density_expa(x1,alpha_hat23, beta_hat23, theta_hat23), lty=2)
#-----------------------------
n2=1000
z2=beta*(runif(n2)**(-1/alpha)-1)
o2=optim(c(1,0.7,log(mean(z2))),paretolik,z=z2)
alpha_hat24=exp(o2$par[1])
theta_hat24=exp(o2$par[2])
beta_hat24=exp(o2$par[3])
lines(x1, density_expa(x1,alpha_hat24, beta_hat24, theta_hat24), lty=2)
#-----------------------------
n2=1000
z2=beta*(runif(n2)**(-1/alpha)-1)
o2=optim(c(1,0.7,log(mean(z2))),paretolik,z=z2)
alpha_hat25=exp(o2$par[1])
theta_hat25=exp(o2$par[2])
beta_hat25=exp(o2$par[3])
lines(x1, density_expa(x1,alpha_hat25, beta_hat25, theta_hat25), lty=2)

legend(4,0.4, c("theta_0", "theta_hat"), lty = c(1,4), lwd=c(2,1))
#-----------------------------
#-----------------------------
plot(x1, densitypareto(x1, alpha, beta),lty=1, type="l", main="n=100000", xlab=
    "␣", ylab= "␣", lwd=2)
#-----------------------------
n3=100000
z3=beta*(runif(n3)**(-1/alpha)-1)
o3=optim(c(1,0.7,log(mean(z3))),paretolik,z=z3)
alpha_hat31=exp(o3$par[1])
theta_hat31=exp(o3$par[2])
beta_hat31=exp(o3$par[3])
lines(x1, density_expa(x1,alpha_hat31, beta_hat31, theta_hat31), lty=2)
#-----------------------------
n3=100000
z3=beta*(runif(n3)**(-1/alpha)-1)
o3=optim(c(1,0.7,log(mean(z3))),paretolik,z=z3)
alpha_hat32=exp(o3$par[1])
theta_hat32=exp(o3$par[2])
beta_hat32=exp(o3$par[3])
lines(x1, density_expa(x1,alpha_hat32, beta_hat32, theta_hat32), lty=2)
#-----------------------------
n3=100000
z3=beta*(runif(n3)**(-1/alpha)-1)
```

```
o3=optim(c(1,0.7,log(mean(z3))),paretolik,z=z3)
alpha_hat33=exp(o3$par[1])
theta_hat33=exp(o3$par[2])
beta_hat33=exp(o3$par[3])
lines(x1, density_expa(x1,alpha_hat33, beta_hat33, theta_hat33), lty=2)
#-------------------------
n3=100000
z3=beta*(runif(n3)**(-1/alpha)-1)
o3=optim(c(1,0.7,log(mean(z3))),paretolik,z=z3)
alpha_hat34=exp(o3$par[1])
theta_hat34=exp(o3$par[2])
beta_hat34=exp(o3$par[3])
lines(x1, density_expa(x1,alpha_hat34, beta_hat34, theta_hat34), lty=2)
#-------------------------
n3=100000
z3=beta*(runif(n3)**(-1/alpha)-1)
o3=optim(c(1,0.7,log(mean(z3))),paretolik,z=z3)
alpha_hat35=exp(o3$par[1])
theta_hat35=exp(o3$par[2])
beta_hat35=exp(o3$par[3])
lines(x1, density_expa(x1,alpha_hat35, beta_hat35, theta_hat35), lty=2)
legend(4,0.4, c("theta_0", "theta_hat"), lty = c(1,2), lwd=c(2,1))

#-------------------------

plot(x2, densitypareto(x2, alpha, beta), lty=1, type="l", ylim=c(0, 2*(10**
    (-5))),main="n=50", xlab="␣", ylab= "␣",lwd=2)
lines(x2, density_expa(x2,alpha_hat11, beta_hat11, theta_hat11), lty=2)
lines(x2, density_expa(x2,alpha_hat12, beta_hat12, theta_hat12), lty=2)
lines(x2, density_expa(x2,alpha_hat13, beta_hat13, theta_hat13), lty=2)
lines(x2, density_expa(x2,alpha_hat14, beta_hat14, theta_hat14), lty=2)
lines(x2, density_expa(x2,alpha_hat15, beta_hat15, theta_hat15), lty=2)
legend(140,1.5*(10**(-5)), c("theta_0", "theta_hat"), lty = c(1,2), lwd=c(2,1))

#-------------------------
plot(x2, densitypareto(x2, alpha, beta), lty=1, type="l", ylim=c(0, 2*(10**
    (-5))),main="n=1000", xlab="␣",ylab="␣",lwd=2)
lines(x2, density_expa(x2,alpha_hat21, beta_hat21, theta_hat21), lty=2)
lines(x2, density_expa(x2,alpha_hat22, beta_hat22, theta_hat22), lty=2)
lines(x2, density_expa(x2,alpha_hat23, beta_hat23, theta_hat23), lty=2)
lines(x2, density_expa(x2,alpha_hat24, beta_hat24, theta_hat24), lty=2)
lines(x2, density_expa(x2,alpha_hat25, beta_hat25, theta_hat25), lty=2)
legend(140,1.5*(10**(-5)), c("theta_0", "theta_hat"), lty = c(1,2), lwd=c(2,1))

#-------------------------
plot(x2, densitypareto(x2, alpha, beta), lty=1, type="l", ylim=c(0, 2*(10**
    (-5))),main="n=100000", xlab="␣", ylab= "␣",lwd=2)
lines(x2, density_expa(x2,alpha_hat31, beta_hat31, theta_hat31), lty=2)
lines(x2, density_expa(x2,alpha_hat32, beta_hat32, theta_hat32), lty=2)
lines(x2, density_expa(x2,alpha_hat33, beta_hat33, theta_hat33), lty=2)
lines(x2, density_expa(x2,alpha_hat34, beta_hat34, theta_hat34), lty=2)
lines(x2, density_expa(x2,alpha_hat35, beta_hat35, theta_hat35), lty=2)
legend(140,1.5*(10**(-5)), c("theta_0", "theta_hat"), lty = c(1,2), lwd=c(2,1))
}
```

Bibliography

[1] G.E.P Box and D. Cox. An analysis of transformations. *Journal of the Royal Statitical Society, Seres B*, 26:211–219, 1964.

[2] Charles E. Brown. *Applied Multivariate Statistics in Geohydrology and Related Sciences.* Springer Berlin Heidelberg, 1998.

[3] R.J. Carrol and D. Ruppert. On prediction and the power transformation family. *Biometrika*, pages 609–615, 1981.

[4] H. Cramer. *Mathematical Methods in Statistics.* Princeton University Press, 1946.

[5] Jay L. Devore and Kenneth N. Berk. *Modern Mathematical Statistics with Applications.* Thomson Books/Cole, 2007.

[6] Peter J. Huber. The behavior of maximum likelihood estimates under nonstandard conditons. *University of California Press*, pages 221–233, 1967.

[7] G. Iyengar. Inverse transform method. University Lecture from Aalborg Universitet, 2002.

[8] J.L. Teugels J. Beirlant and P. Vynckier. *Practical analysis of extreme values.* Leuven University Press, 1996.

[9] Ju Yong Lee Sunggon Kim and Dan Keun Sung. A shifted gamma distribution model for long-range dependent internet traffic. *Communications Letters, IEEE*, pages 124–126, 2003.

[10] C. Kleiber and S. Kotz. *Statistical Size Distributions in Economics and Actuarial Sciences.* John Wiley and Sons, Inc. Hoboken, 2003.

[11] S. Kullback and R. A. Leibler. *The Annals of Mathematical Statistics*, volume 22. Institute of Mathematical Statistics, 1951. Subtitle, On Information and Sufficiency.

[12] N. De Pril M.J. Goovaerts, L. D'Hooge. On the infinite divisibility of the ratio of two gamma-distributed variables. *North-Holland Publishing Company*, pages 291–297, 1978.

[13] Jagdish K. Patel and Campbell B. Read. *Handbook of the normal distribution*, volume 40. Marcel Dekker, Inc., 1982.

[14] J. A. Rice. *Mathematical Statistics and Data Analysis.* Thomson Books/Cole, 3rd edition, 2007.

[15] William T. Vetterling William H. Press, Saul A. Teukolsky and Brian P. Flannery. *Numerical Recipes, The Art of Scientific Computing. Third Edition.* Cambidge University Press, 32 Avenue of the Americas, New York, NY 10013-2473, USA, 2007.