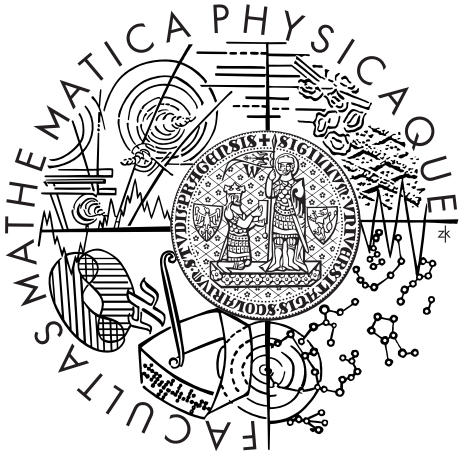
Univerzita Karlova v Praze
Matematicko-fyzikální fakulta

DIPLOMOVÁ PRÁCE



Barbora Zuzáková

Modelování velkých škod

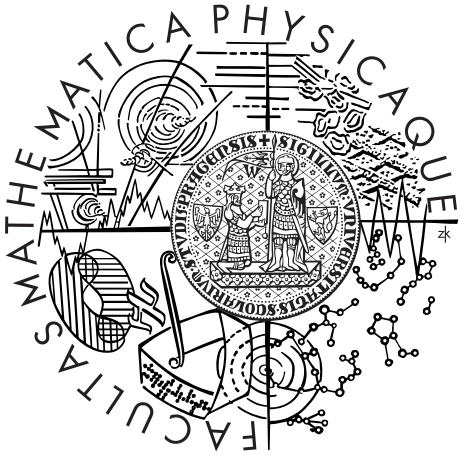
Katedra pravděpodobnosti a matematické statistiky

Vedoucí diplomové práce: RNDr. Michal Pešta, Ph.D.
Studijní program: Matematika
Studijní obor: Finanční a pojistná matematika

Praha 2013

Charles University in Prague
Faculty of Mathematics and Physics

MASTER THESIS



Barbora Zuzáková

Large claims modeling

Department of Probability and Mathematical Statistics

Supervisor: RNDr. Michal Pešta, Ph.D.
Study programme: Mathematics
Specialization: Actuarial and Financial Mathematics

Prague 2013

Acknowledgment: I would like to express my gratitude to all those who supported me to complete this thesis. Mainly, I want to thank my supervisor RNDr. Michal Pešta, Ph.D. who was abundantly helpful and offered invaluable assistance, support and guidance. I am also grateful to my friend Mgr. Daniel Krýcha, MSc. for grammar revision that substantially improved the thesis.

I declare that I carried out this master thesis independently, and only with the cited sources, literature and other professional sources.

I understand that my work relates to the rights and obligations under the Act No. 121/2000 Coll., the Copyright Act, as amended, in particular the fact that the Charles University in Prague has the right to conclude a license agreement on the use of this work as a school work pursuant to Section 60 paragraph 1 of the Copyright Act.

In Prague November 11, 2013 signature of the author

Název práce: Modelování velkých škod

Autor: Barbora Zuzáková

Katedra: Katedra pravděpodobnosti a matematické statistiky

Vedoucí diplomové práce: RNDr. Michal Pešta, Ph.D.

Abstrakt: Tato diplomová práce se zabývá statistickým modelováním vysokých škod pojistného portfolia, a to metodami vycházejícími z teorie extrémních hodnot. Zaměřuje se především na prahové modely, tedy modely, které zkoumají pouze škody překračující určitou vysokou mez. Ve srovnání s klasickým přístupem teorie extrémních hodnot, který aproximuje chování maximální výše škody, bývají prahové metody v posledních letech upřednostňovány. Veškeré teoretické přístupy a závěry jsou aplikovány na databázi léčebných výloh zdravotních pojišťoven zaznamenaných v letech 1997, 1998 a 1999, která je spravována Společností aktuárů. Naším záměrem je prokázat, že přístup založený na teorii extrémních hodnot poskytuje mnohem lepší přiblížení distribuce vysokých škod nežli modelování pomocí klasických parametrických distribučních funkcí, a umožňuje tak preciznější odhad jak vysokých kvantilů, tak pravděpodobné maximální ztráty, které jsou stěžejními ukazateli životního i neživotního pojištění.

Klíčová slova: prahové modely, obecné Paretovo rozdělení, vysoké škody.

Title: Large claims modeling

Author: Barbora Zuzáková

Department: Department of Probability and Mathematical Statistics

Supervisor: RNDr. Michal Pešta, Ph.D.

Abstract: This thesis discusses a statistical modeling approach based on the extreme value theory to describe the behaviour of large claims of an insurance portfolio. We focus on threshold models which analyze exceedances of a high threshold. This approach has gained in popularity in recent years, as compared with the much older methods based directly on the extreme value distributions. The method is illustated using the group medical claims database recorded over the periods 1997, 1998 and 1999 maintained by the Society of Actuaries. We aim to demonstrate that the proposed model outperforms classical parametric distributions and thus enables to estimate high quantiles or the probable maximum loss more precisely.

Keywords: threshold models, generalized Pareto distribution, large claims.

Contents

1. Introduction

The greatest part of the indemnities paid by the insurance companies is represented by a few large claims rising from an insurance portfolio. Therefore, it is of prime interest to calculate their precise estimates. Moreover, frequency and size of the extreme events often creates the base of pricing models for reinsurance agreements as excess-of-loss reinsurance contract which guarantees that the reinsurer reimburses all expenditure associated with a claim as soon as it exceeds a specific threshold. Portfolio managers might be further interested in estimation of high quantiles or the probable maximum loss.

The thesis discusses a universal procedure for description of statistical behaviour of the extreme claims. To this end, we present a basic features of extreme value theory, in particular we pay special attention to threshold models that are intended to identify distribution of exceedances over a high threshold. From a statistical perspective, the threshold can be loosely defined such that the population tail can be well approximated by an extreme value model, obtaining a balance between the bias due to the asymptotic tail approximation and parameter estimation uncertainty due to the sparsity of threshold excess data. We present traditional approaches for threshold selection including mean residual life plot, threshold choice plot, L-moments plot and dispersion index plot. Once the optimal threshold is identified, an asymptotic extreme value model can be fit. Maximum likelihood has emerged as a flexible and powerful modeling tool in such applications, but its performance with small samples has been shown to be poor relative to an alternative fitting procedure based on probability weighted moments. To incorporate an extra information provided by the probability weighted moments model in a likelihood-based analysis, we utilize a penalized maximum likelihood estimator that retains the modeling flexibility and large sample optimality of the maximum likelihood estimator, but improves on its small-sample properties. In this thesis we apply all three methods on different samples and provide their comparison. As the next step, the fitted model has to be verified using a graphical or statistical assessment. The present thesis considers the medical insurance claims from the SOA Group Medical Insurance Large Claims Database. A model is fitted to the amounts of the 1997, 1998 and 1999 group claims.

The thesis is organized as follows. In Chapter 2 we provide an overview of the classical extreme value theory, mainly we focus on formulation of so-called block maxima and state the theorem which is considered to be the cornerstone of the theory. In Chapter 3 we present threshold models that analyze statistical behaviour of claim exceedances over a high threshold. After a general theoretical background we introduce methods for threshold selection followed by several inference approaches accompanied by their comparison. The rest of the chapter is devoted to model validation, where we present the graphical as well as the statistical verification. The database description together with a brief descriptive statistics can be found in Chapter 4. Chapter 5 is devoted to data analysis, in other words we demonstrate the approach based on the extreme value theory to model behaviour of large claims exceeding a high threshold. The aim of the chap-

ter is to find the optimal threshold level and estimate parameters of the extremal distribution for particular claim years and prove the adequacy of the fitted models. In Chapter 6 we suggest two applications of the model that can be broadly applied in insurance strategy. Finally, conclusions are stated in Chaper 7.

2. Classical extreme value theory

This chapter is dedicated to provide a mathematical formulation of problems that involve application of extreme value theory. The first part deals with the model formulation and general statements about so called block maxima which can be considered as the cornerstone of the theory. In the next chapter we will focus on the different approach to extremal behavior, known as threshold models.

2.1 Asymptotic models

Let $X_1, \ldots, X_n$ be a sequence of independent random variables with common distribution F. In practice, the serie usually represents values of a process measured on a time scale (for instance, daily measurements of rainfall) or values of a random process recorded over a predefined time period (for instance, year observations of insurance claims). The theory focuses on the statistical behaviour of block maxima defined as:

$$M_n = \max\{X_1, \ldots, X_n\}.$$

It can be easily shown that the exact distribution function of M_n has the form of F^n for all values of n. However, it is not possible to model the distribution function according this formula immediately, since F is an unknown function. One could argue that F could be estimated from observed data and subsequently used for the distribution function of the block maxima. Nevertheless, this is not an appropriate approach. Even small discrepancies in the estimation of F can cause significant discrepancies for F^n. Inaccuracy of standard statistical techniques will be demonstrated in Section 5.3.

The extreme value theory assumes the function F to be unknown and seeks for an appropriate distribution families to model F^n directly. The *Extremal Types Theorem* defines the range of possible limit distributions for normalized block maxima. For convenience we state a modified version of the theorem, the exact formulation and an outline of the proof can be found in Coles (2001).

Theorem 1. *If there exist sequence of constants* $\{a_n > 0\}$ *and* $\{b_n\}$ *such that*

$$\mathbb{P}\left[\frac{M_n - b_n}{a_n} \leq z\right] \longrightarrow G(z), \qquad n \longrightarrow \infty \tag{2.1}$$

for a non-degenerate distribution function G*, then* G *is a member of the generalized extreme value family*

$$G(z) = \exp\left\{-\left(1 + \xi\left(\frac{z-\mu}{\sigma}\right)\right)_+^{-1/\xi}\right\}, \tag{2.2}$$

where (μ, σ, ξ) *are the location, scale and shape parameters respectively,* $\sigma > 0$ *and* $z_+ = max\,(z, 0)$.

Remark 2. The fact that the normalizing constants are unknown in practice is irrelevant. Statement (2.1) can be equivalently rewritten as:

$$\mathbb{P}\left[M_n \leq z\right] \longrightarrow G\left(\frac{z-b_n}{a_n}\right) = G^*(z), \qquad n \longrightarrow \infty$$

where G^* is just another member of the generalized extreme value family. Thus the estimation of the parameters of the functions G^* and G involves the same procedure.

The theorem implies that the normalized block maxima converge in distribution to a variable having the distribution function G, commonly termed as the *Generalized Extreme Value* (GEV) distribution. Notable feature of the previous statement is that G is the only possible limit regardless of the original distribution function F.

The generalized extreme value family includes three classes of distribution known as the Gumbel, Fréchet and negative Weibull families respectively. Each type can be obtained by a particular choice of the shape parameter ξ. The Fréchet and negative Weibull classes correspond respectively to the case when $\xi > 0$ and $\xi < 0$. The Gumbel class is defined by continuity when $\xi \to 0$. It follows that in practice these three types give quite different representations of extreme value behaviour corresponding to distinct forms of tail behaviour for the distribution function F of the original data. Consider the upper end-point z_+ of the limit distribution G, i.e., z_+ is the smallest value of z such that $G(z) = 1$. Then for the Fréchet and Gumbel distribution z_+ is infinite, whereas in the case of Weibul distribution it is finite.

Remainder of this section is devoted to the formulation of extreme quantiles estimates. Assume a serie of independent identically distributed random variables $X_1, X_2, \ldots$. Further let us split the sequence into blocks of length n, for some large n, and generate a serie of maxima corresponding to each block, to which the generalized extreme value distribution can be fitted. Then by inverting Equation (2.1) we arrive to the following expression for extreme quantiles:

$$q_p = \begin{cases} \mu - \frac{\sigma}{\xi}\left(1 - (-\log(1-p))^{-\xi}\right), & \text{for } \xi \neq 0, \\ \mu - \sigma \log\left(-\log(1-p)\right), & \text{for } \xi = 0, \end{cases} \tag{2.3}$$

where $G(q_p) = 1 - p$. q_p from the previous equation is commonly termed as the *return level* associated with the *return period* $1/p$. The quantile q_p can be interpreted as a value that is exceeded by the maximum in any relative period with probability p. Quantile analysis enables to express stochastic models on the scale of observed values, therefore it provides illustrative interpretation. In particular, if q_p is plotted against $y_p = -\log(1-p)$ on a logarithmic scale (q_p is plotted against $\log y_p$), the plot is linear in the case $\xi = 0$, convex with asymptotic limit for $\xi < 0$ and concave without any finite bound for $\xi > 0$. The graph is called a *return level plot* and indicates whether the fitted distribution resembles rather Gambel, Fréchet or negative Weibul.

2.2 Inference procedure

Behaviour of block maxima can be described by distribution in the spirit of Theorem 1. Its application involves blocking the data into sequences of equal length, determining the maximum for each block and fitting the generalized extreme value distribution to such specified values. It has been shown that determining the block size can be the crucial issue. Large blocks generate few maxima, which leads to large variance in parameters estimation and subsequently the accuracy of fitted distribution. On the other hand if the blocks are too small, the set of block maxima includes values that are rather usual than extreme. Then the limit distribution stated in Theorem 1 is likely to be poor. Thus the choice of the length for block is the trade-off between bias and variance. Although some statistical methods can be derived, in practice the length corresponds to a reasonable time unit, for instance annual maxima are often applied.

Remark 3. Any extreme value analysis suffers from limited amount of data for model estimation. Extremes are scarce, which involves large variation of model estimates. Modeling only block maxima is thus inaccurate if other data on extremes are available. This issue has motivated the search of description of extremal behaviour incorporating extra information. There are basically two well-known general approaches. One is base on exceedances of a high threshold (discussed in Chapter 3) and the other one is based on the behaviour of the r largest order statistics within a block, for small values of r. There are numerous published application of the r largest order statistic model, for instance see Coles (2001).

3. Threshold models

The threshold models are very effective methods of modeling extreme values. A threshold value, or generally set of threshold values, is used to distinguish ranges of values where the behaviour predicted by the model varies in some important way. Some of the commonly used graphical diagnostics and related statistics are described in Section 3.2.

The major benefit of threshold models is avoiding the procedure of blocking which can be very inefficient and inaccurate if one block happens to contain more extreme events than another one. The drawback with the threshold models is that once the threshold value has been identified it is treated as fixed, thus the associated uncertainty is ignored in further inference. Moreover, there are often more than one suitable threshold value stemming from various estimation procedures, thus different tail behaviour will be ignored as well when fixing the threshold. As suggested in Scarrott and MacDonald (2012) an informal approach to overcoming these problems is to evaluate the sensitivity of the inferences (e.g. parameters or quantiles) to different threshold choices.

Direct comparison of the goodness of the fit for different thresholds is complicated due to the varying sample sizes. Recently, various extreme value mixture models have been developed to overcome this problem. These mixture models typically approximate the entire distribution function, thus the sample size for each threshold is invariant. For more information about these models see Scarrott and MacDonald (2012).

3.1 Asymptotic models

Let $X_1, X_2, \ldots$ be a sequence of independent and identically distributed random variables with common distribution function F. Moreover let u be some high threshold value. Then we regard as extreme events those random variables X_i that exceed the threshold u. The stochastic behaviour of extreme event X_i can be described by the following formula:

$$\mathbb{P}\left[X_i > y \mid X_i > u\right] = \frac{1 - F(y)}{1 - F(u)}, \qquad y > u.$$

Apparently if the distribution function F was known, the distribution of threshold exceedances would satisfy the previous formula. In practice this is not the case and some approximation needs to be used. The following statement extends Theorem 1 and provides us the sought approach for description of the threshold behaviour.

Theorem 4. *Let $X_1, X_2, \ldots$ be a sequence of independent random variables with common distribution function F, and let*

$$M_n = max\{X_1, \ldots, X_n\}.$$

Denote an arbitrary term in the X_i sequence by X, and suppose that F satisfies Theorem 1, so that for large n,

$$\mathbb{P}\left[M_n \leq z\right] \longrightarrow G(z), \qquad n \longrightarrow \infty$$

for a non-degenerate distribution function G given as:

$$G(z) = \exp\left\{-\left(1+\xi\left(\frac{z-\mu}{\sigma}\right)\right)_+^{-1/\xi}\right\},$$

for some $\mu, \sigma > 0$ and ξ. Then, for large enough u, the distribution function of X, conditional on $X > u$, can be approximated as:

$$\mathbb{P}\left[X \leq y \mid X > u\right] \longrightarrow H(y), \qquad u \longrightarrow \infty, \tag{3.1}$$

where

$$H(y) = 1 - \left(1+\xi\,\frac{y-u}{\sigma_u}\right)_+^{-1/\xi}, \qquad y > u, \tag{3.2}$$

where $z_+ = max\,(z, 0)$ and we adopt the conversion of using $\sigma_u = \sigma + \xi(u-\mu)$ to denote the scale parameter corresponding to excess of the threshold u.

Remark 5. The previous theorem can be rephrased in the spirit of behaviour of $X-u$, i.e., if the assumptions of the statement are satisfied, then, for large enough u, the distribution function of $X-u$, conditional on $X > u$, can be approximated as:

$$\mathbb{P}\left[X-u \leq y \mid X > u\right] \longrightarrow H^{(u)}(y), \qquad u \longrightarrow \infty, \tag{3.3}$$

where

$$H^{(u)}(y) = 1 - \left(1+\xi\,\frac{y}{\sigma_u}\right)_+^{-1/\xi}, \qquad y > 0, \tag{3.4}$$

where again $z_+ = \max\,(z, 0)$ and $\sigma_u = \sigma + \xi(u-\mu)$.

The family of distributions given by Equation(3.2) is known as *Generalized Pareto Distribution* (GPD). The previous theorem put in relation the statistical behaviour of block maxima and threshold excesses, i.e., if block maxima can be approximated by distribution function G, then threshold excesses have a corresponding approximate distribution within the generalized Pareto family. Notice that the parameters of the generalized Pareto family are uniquely determined by those associated with corresponding generalized extreme value distribution. Mainly, the shape parameter ξ is invariant to the threshold selection and block size n.

The shape parameter ξ is dominant in determining the qualitative behaviour of the generalized Pareto distribution. Similarly as it is for the generalized extreme value family, the generalized Pareto family includes three classes of distribution linked to a particular choice of parameter ξ: exponential, Pareto and beta. The Pareto and beta classes correspond respectively to the case when $\xi > 0$ and $\xi < 0$. The exponential distribution is defined by continuity when $\xi \to 0$.

Similarly as demonstrated in Chapter 2, the return level associated with the return period $1/p$ can be derived by inverting Equation (3.4). Thus, the expression for the extreme quantiles is given as:

$$q_p = \begin{cases} -\frac{\sigma_u}{\xi}\left(1 - p^{-\xi}\right), & \text{for } \xi \neq 0, \\ -\sigma_u \log p, & \text{for } \xi = 0, \end{cases} \tag{3.5}$$

where $H^{(u)}(q_p) = 1 - p$. We recall that the quantile q_p can be interpreted as a value that is exceeded by the maximum in any relative period with probability p.

3.2 Graphical threshold selection

The classical fixed threshold modeling approach uses graphical diagnostics, essentially assessing aspects of the model fit, to make an a priori threshold choice. An advantage of this approach is that it requires a graphical inspection of the data, comprehending their features and assessing the model fit. A key drawback is that they can require substantial expertise and can be rather subjective, as can be seen in the Chapter 5.

Consider a sequence of raw data $x_1, \ldots, x_n$ as realizations of independent and identically distributed random variables $X_1, \ldots, X_n$. Further let us define a high threshold u, for which the exceedances $\{x_i;\ x_i > u\}$ are identified as extreme events. In further text we refer to these excesses as y_j, i.e., $y_j = x_i$ for those indices $i \in \{1, \ldots, n\}$ for which $x_i > u$. In the spirit of Theorem 4, sequence $y_1, \ldots, y_k$ can be regarded as independent realizations of a random variable Y whose distribution function can be approximated by a member of the generalized Pareto family defined by Formula (3.2).

Unlike the method of block maxima which takes the highest event within a block as extremal and does not consider whether the event is extremal comparing the rest of observations, the threshold method identifies an event as extremal if it exceeds a high threshold. Moreover this threshold is estimated on the basis an entire data set. Nevertheless the issue of threshold selection and choice of bloc size is analogous, thus the trade-off between bias and variance arises again. If the threshold is too high only few observations are generated and the estimated limit distribution is more likely to be poor, i.e., high variance of estimated parameters is involved. On the other hand if the threshold is too low the sequence of exceedances contains values that are rather usual than extreme and as the consequence the estimated limit distribution is likely to violate the asymptotic framework of the model. In practical application as low a threshold as possible is accepted on condition that such an adopted threshold value provides reasonable approximation for the limit distribution.

In the following subsection we will discuss several concrete procedures for the threshold selection. A brief overview of the methods stated above can be found in Ribatet (2011).

3.2.1 Mean residual life plot

The first method introduced by Davison and Smith (2012), the *mean residual life plot*, is based on the theoretical mean of the generalized Pareto family. In more detail, let Y be a random variable with distribution function $GPD\,(u_0, \sigma_{u_0}, \xi)$ and let u_1 be a threshold such that $u_1 > u_0$. Then random variable $Y - u_1 | Y > u_1$ is also generalized Pareto distribution with updated parameters $u_1, \sigma_{u_1} = \sigma_{u_0} + \xi u_1$ and $\xi_1 = \xi$. Assume an arbitrary $u_1 > u_0$ and $y > 0$. The claim then stems from the following inference:

$$\begin{aligned}
\mathbb{P}\left(Y - u_1 > y | Y > u_1\right) &= \frac{1 - H^{(u_0)}(y + u_1)}{1 - H^{(u_0)}(u_1)} \\
&= \frac{\left(1 + \xi \, \frac{y+u_1}{\sigma_{u_0}}\right)_+^{-1/\xi}}{\left(1 + \xi \, \frac{u_1}{\sigma_{u_0}}\right)_+^{-1/\xi}} \\
&= \left(1 + \xi \, \frac{y}{\sigma_{u_0} + \xi u_1}\right)_+^{-1/\xi}.
\end{aligned}$$

If a random variable Y has a generalized Pareto distribution with parameters μ, σ and ξ, then

$$\mathbb{E}\left[Y\right] = \mu + \frac{\sigma}{1 - \xi}, \qquad \text{for } \xi < 1.$$

When $\xi \geq 1$, the mean is infinite. In practice, if Y represents excess over a threshold u_0, and if the approximation by a generalized Pareto distribution is sufficiently accurate, we have:

$$\mathbb{E}\left[Y - u_0 | Y > u_0\right] = \frac{\sigma_{u_0}}{1 - \xi}.$$

Now, if the generalized Pareto distribution is valid as a model for excesses of the threshold u_0, it has to be valid for all thresholds $u_1 > u_0$, provided a change of scale parameter as derived above. Thus, we have:

$$\mathbb{E}\left[Y - u_1 | Y > u_1\right] = \frac{\sigma_{u_1}}{1 - \xi} = \frac{\sigma_{u_0} + \xi u_1}{1 - \xi}. \tag{3.6}$$

We observe that for $u > u_0$ expression $\mathbb{E}\left[Y - u | Y > u_0\right]$ is a linear function of variable u. Furthermore, $\mathbb{E}\left[Y - u | Y > u\right]$ is obviously the mean of the excesses of the threshold u, which can easily be estimated using the empirical mean. By virtue of Equation (3.6), these estimated are expected to change linearly with rising values of threshold u, for which the generalized Pareto model is valid. The mean residual life plot consists of points

$$\left\{ \left(u, \frac{1}{k} \sum_{i=1}^{k} (y_i - u) \right) : u \leq x_{\max} \right\}, \tag{3.7}$$

where k denotes the number of observations that exceed threshold u, and $x_{\max}$ is the maximum of all observations x_i, $i = 1 \ldots, n$. Confidence intervals can be added to the plot since the empirical mean can be approximated by normal distribution (in the spirit of the Central Limit Theorem). However, this approximation is rather poor for high values of thresholds as the are less excesses. Moreover, by construction, the plot always converges to the point $(x_{\max}, 0)$.

3.2.2 Threshold choice plot

The *threshold choice plot* provides a complementary technique to the mean residual life plot for identification of a proper threshold. The method is based on fitting the generalized Pareto distribution at a range of thresholds, and to observe stability of parameter estimates. The argument is as follows. Let Y be a random variable with distribution function $GPD\ (u_0, \sigma_{u_0}, \xi)$ and let u_1 be a higher threshold. Then random variable $Y|Y > u_1$ has also generalized Pareto distribution with updated parameters $u_1, \sigma_{u_1} = \sigma_{u_0} + \xi(u_1 - u_0)$ and $\xi_1 = \xi$. Assume an arbitrary $u_1 > u_0$ and $y > u_1$. The derivation proceeds as follows:

$$\begin{aligned}
\mathbb{P}\left(Y > y | Y > u_1\right) &= \frac{1 - H(y)}{1 - H(u_1)} \\
&= \frac{\left(1 + \xi\ \frac{y - u_0}{\sigma_{u_0}}\right)_+^{-1/\xi}}{\left(1 + \xi\ \frac{u_1 - u_0}{\sigma_{u_0}}\right)_+^{-1/\xi}} \\
&= \left(1 + \xi\ \frac{y - u_1}{\sigma_{u_0} + \xi(u_1 - u_0)}\right)_+^{-1/\xi}.
\end{aligned}$$

Hence the scale parameter depends on value of a threshold u (unless $\xi = 0$). This difficulty can be remedied by defining the scale parameter as

$$\sigma^* = \sigma_u - \xi u = \sigma_{u_0} - \xi u_0. \tag{3.8}$$

With this new parametrization, σ^* is independent of u as it results from the last equality. Thus, estimates of σ^* and ξ are expected to be consistent for all $u > u_0$ if u_0 is suitable threshold for the asymptotic approximation. The threshold choice plots represents the points

$$\{(u, \sigma^*) : u \leq x_{\max}\} \qquad \text{and} \qquad \{(u, \xi) : u \leq x_{\max}\}, \tag{3.9}$$

where $x_{\max}$ is the maximum of the observations x_i, $i = 1 \ldots, n$.

3.2.3 L-moments plot

In statistics, a probability density or an observed data set are often summarized by its moments or cumulants. It is also common, when fitting a parametric distribution, to estimate the parameters by equating the sample moments with those of the fitted distribution. However, the moments-based method is not always satisfactory. Particularly, when the sample is too small, the numerical values of sample moments can be markedly different from those of the probability distribution from which the sample was drawn. Hosking (1990) described in detailed the alternative approach based on quantities called *L-moments*. These are analogous to the conventional moments but can be estimated by linear combinations of order statistics (thus "L" in "L-moments" emphasized that the quantities are linear functions). They have the advantage of being more robust to the presence of outliers in the data and experience also shows that they are more accurate in

small samples than even the maximum likelihood estimates. Moreover, they are able to characterize wider range of distribution functions.

In order to explain how the L-moments-based method is utilized, when selecting the threshold value, it is worthwhile to provide their exact definition.

Definition 6. *Let X be a real-valued random variable with cumulative distribution function F, and let $X_{1:n} \leq X_{2:n} \leq \ldots \leq X_{n:n}$ be the order statistics of a random sample of size n drawn from the distribution of X. L-moments of X are quantities defined as:*

$$\lambda_r \equiv r^{-1} \sum_{k=0}^{r-1} (-1)^k \binom{r-1}{k} \mathbb{E}X_{r-k:r}, \qquad r = 1, 2, \ldots, n.$$

Remark 7. L-moment λ_2 is a measure of the scale or dispersion of the random variable X (see Hosking (1990)). Higher L-moments λ_r,$r \geq 3$ are for convenience often standardized, so that they are independent of the units of measurement of X. Let us, therefore, introduce the *L-moment ratios* of X defined as quantities:

$$\tau_r = \frac{\lambda_r}{\lambda_2}, \qquad r = 3, 4, \ldots, n.$$

The L-moments $\lambda_1, \ldots, \lambda_r$ together with the L-moment ratios are useful tools for describing distribution. The L-moments can be regarded as an analogy to (conventional) central moments and the L-moment ratios are analogous to moment ratios. In particular, $\lambda_1, \lambda_2, \tau_3$ and τ_4 can be regarded as measures of location, scale, skewness and kurtosis respectively.

As derived in Hosking (1990), the L-kurtosis and L-skewness for generalized Pareto distribution are given by formulas:

$$\tau_3 = \frac{1+\xi}{3-\xi},$$
$$\tau_4 = \frac{(1+\xi)(2+\xi)}{(3-\xi)(4-\xi)},$$

thus parameter τ_4 can be expressed in terms of τ_3 as:

$$\tau_4 = \tau_3 \, \frac{1+5\tau_3}{5+\tau_3}. \tag{3.10}$$

The previous equation creates the headstone for the *L-moment plot* which is represented by points defined by:

$$\left\{(\hat{\tau}_{3,u}, \hat{\tau}_{4,u}) : u \leq x_{\max}\right\}, \tag{3.11}$$

where $\hat{\tau}_{3,u}$ and $\hat{\tau}_{4,u}$ are estimations of the L-kurtosis and L-skewness based on excess over threshold u and $x_{\max}$ is the maximum of the observations x_i, $i = 1 \ldots, n$. In practical applications, the theoretical curve defined by Equation (3.10) is traced as a guideline and the choice of the accurate threshold value is assessed according to position of the estimated points with respect to this line.

3.2.4 Dispersion index plot

The last graphical technique for threshold selection presented in this thesis, the *dispersion index plot*, is particularly useful when analysing time series. According to the Extreme value theory, the occurrence of excesses over a high threshold in a fixed period, generally this is a year, must be distributed as Poisson process. As for a random variable having Poisson distribution, the ratio of the variance and the mean is equal to 1. Cunnane (1979) introduced a dispersion index statistics defined as:

$$DI = \frac{s^2}{\lambda}, \tag{3.12}$$

where s^2 is the intensity of the a Poisson process and λ the mean number of events in a block, both corresponding to a high threshold value. Equation (3.12) creates the basis for the dispersion index plot which can be represented by points defined as:

$$\{(u, DI) : u \leq x_{\max}\}, \tag{3.13}$$

where $x_{\max}$ is the maximum of the observations x_i, $i = 1 \ldots, n$. The plot enables to test if the ratio DI differs from 1 for various levels of threshold. In practice, confidence levels for DI are appended to the graph (DI can be considered to have χ^2 distribution with $M - 1$ degree of freedom, M being the total number of fixed periods).

3.3 Inference procedure

By almost universal consent, a starting point for modeling the extreme values of a process is based on distributional models derived from asymptotic theory. Of less common agreement is the method of inference by which such asymptotic behaviour is approximated. In this section, we aim to get the reader acquainted with the most commonly used ones: the maximum likelihood, the probability weighted moments and penalized maximum likelihood. In order to learn about remaining approaches we refer to the following papers: Pickands (1975) introduces the whole problematics of fitting the generalized Pareto distribution, Hosking and Wallis (1987) provides an overview of the *moments based method*, Juaréz and Schucany (2004) broadly discuss so called *minimum density power divergence estimator*, Peng and Welsh (2001) deal with the *median* approach, and the *maximum goodness-of-fit estimator* is summarized by Luceno (2006). Recently, Zvang (2007) discussed the *likelihood moment* method.

As a general framework for extreme value modeling, the maximum likelihood based estimators have many advantages. As Coles and Dixon (1999) pointed out, likelihood functions can be constructed for complex modeling situations enabling, for example, non-stationarity, covariate effects and regression modeling. This combination of asymptotic optimality, ready-solved inference properties and modeling flexibility represents a comprehensive package for alternative methods to compete against. Moreover, approximate inference, such as confidence intervals, are straightforward. An argument that is addressed against the maximum likelihood method is its small sample properties. Hosking, Wallis, and Wood (1985)

showed that dealing with small sample, probability weighted moments estimator performs better in terms of bias and mean square error. Despite many advantages of maximum likelihood, this fact remains a source criticism, since it is common in practice to make inference with very few data. We aim to utilize both of the methods in the computational part and provide their comparison. In light of Coles and Dixon (1999) who suggested a modification of the maximum likelihood estimator using a penalty function for fitting generalized extreme value distribution, we demonstrate the same approach for inference for the generalized Pareto distribution. As they showed, the penalized maximum likelihood estimator retains all of the advantages of the maximum likelihood and additionally improves small sample properties which are comparable with those of the probability weighted moments estimator.

3.3.1 Maximum likelihood

Let us recall that the values $y_1, \ldots, y_k$ are the k extreme events excessing threshold u, i.e., $y_j = x_i$ for those indices $i \in \{1, \ldots, n\}$ for which $x_i > u$. For convenience, we denote the excesses of threshold u as z_j, i.e., $z_j = y_j - u$ for $j = 1, \ldots, k$. Considering distribution function of the form given by Formula (3.4), the likelihood function can be then derived from the following relation (for more detailed description see a standard asymptotic likelihood theory summarized by Cox and Hinkley (1974), for example):

$$L(\sigma_u, \xi) = \prod_{j=1}^{k} \frac{\mathrm{d}H^{(u)}(z_j; \sigma_u, \xi)}{\mathrm{d}z_j} \tag{3.14}$$

Thus for $\xi \neq 0$ the log-likelihood function can be expressed as:

$$l(\sigma_u, \xi) = -k \log \sigma_u - \left(1 + \frac{1}{\xi}\right) \sum_{j=1}^{k} \log \left(1 + \xi \frac{z_j}{\sigma_u}\right), \tag{3.15}$$

provided $(1 + \xi z_j/\sigma_u) > 0$ for $i = j, \ldots, k$, otherwise, $l(\sigma_u, \xi) = -\infty$. In the case $\xi = 0$, the log-likelihood has the following form:

$$l(\sigma_u) = -k \log \sigma_u - \frac{1}{\sigma_u} \sum_{j=1}^{k} z_j. \tag{3.16}$$

The log-likelihood function defined above can be made arbitrarily large by taking $\xi < -1$ and ratio σ_u/ξ close enough to $\max \{z_j; j = 1, \ldots, k\}$. Thus the maximum likelihood estimators are taken to be the values ξ and σ_u, which yield a local maximum of the function. The analytical maximization of the log-likelihood is not possible, thus to find the local maximum requires using of numerical methods. In the practical application, we adopted a procedure based on a quasi-Newton method (also known as a variable metric algorithm). Quasi-Newton methods are based on Newton's method to find the stationary point of a function, where the gradient is 0. Newton's method assumes that the function can be locally approximated as quadratic in the region around the optimum, and uses the first and

second derivatives to find the stationary point. In higher dimensions, Newton's method uses the gradient and the Hessian matrix of second derivatives of the function to be minimized. In quasi-Newton methods the Hessian matrix does not need to be computed. The Hessian is updated by analyzing successive gradient vectors instead.

Remark 8. A potential complication with the use of maximum likelihood methods concerns the regularity conditions that are required for the usual asymptotic properties. Such conditions are not satisfied because the end-points of the generalized Pareto distribution are functions of the parameter values: $-\sigma_u/\xi$ is an upper end-point of the distribution when $\xi > 0$, or a lower end-point when $\xi < 0$, respectively. This issue was studied by Smith (1985) who proved that maximum likelihood estimators are regular (in the sense of having the usual asymptotic properties) whenever $\xi > -0.5$, for $-1 < \xi < -0.5$ the estimators are generally obtainable, but do not have the standard asymptotic properties, and the estimators are unlikely to be obtainable whenever $\xi < 1$.

3.3.2 Probability weighted moments

The probability weighted moments of a continuous random variable X with distribution function F are the quantities

$$M_{p,r,s} = \mathbb{E}\left[X^p (F(X))^r (1 - F(X))^s\right],$$

where p, r and s are real parameters. As suggested in Hosking and Wallis (1987), for the generalized Pareto distribution it is convenient to set the parameters as follows: $p = 1$ and $r = 0$. Then the explicit formula for the probability weighted moments exists and has the form:

$$\alpha_s \equiv M_{1,0,s} = \frac{\sigma_u}{(s+1)(s+1-\xi)},$$

which exists provided that $\xi < 1$. Consequently, the previous equation for $s = 0, 1$ enables us to derive the analytical expressions for parameters σ_u and ξ, those are then given by:

$$\sigma_u = \frac{2\alpha_0\alpha_1}{\alpha_0 - 2\alpha_1}, \tag{3.17}$$

$$\xi = 2 - \frac{\alpha_0}{\alpha_0 - 2\alpha_1}. \tag{3.18}$$

The probability weighted moments estimator $\hat{\sigma}_u$ and $\hat{\xi}$ are obtained by replacing α_0 and α_1 in Equation (3.18) by estimators based on the ordered sample $z_{1:k}, \ldots, z_{k:k}$. The statistic

$$a_s = \frac{1}{k} \sum_{j=1}^{k} \frac{(k-j)(k-j-1)\cdots(k-j-s+1)}{(k-1)(k-1)\cdots(k-s)} \; z_{j:k} \tag{3.19}$$

is an unbiased estimator of α_s. Probability weighted moment α_s can be estimated as:

$$\tilde{a}_s = \frac{1}{k}\sum_{j=1}^{k}(1-p_{j:k})^r z_{j:k}, \tag{3.20}$$

where $p_{j:k}$ is a so called plotting position, i.e., it is an empirical estimate of the considered distribution function based on the sample $z_1, \ldots, z_k$. Reasonable choice for the plotting position is $p_{j:k} = (j+\gamma)/(n+\delta)$, where γ and δ are suitable constants. For detailed discussion on deriving of the unbiased and consistent estimators we refer to Landwehr, Matalas, and Wallis (1979). Whichever variant is used, the estimators of α_r, σ_u, and ξ are asymptotically equivalent. In the computational part, we aim to use the biased estimator given by Formula (3.20) with parameters $\gamma = -0.35$ and $\delta = 0$. These values were recommended by Landwehr, Matalas, and Wallis (1979) for the Wakeby distribution, of which the generalized Pareto distribution is a special case.

3.3.3 Penalized maximum likelihood

Penalized maximum likelihood is a simple method how to incorporate into an inference information that is supplementary to that provided by data. Many of improvements have been suggested so far, in order to balance the appropriate likelihood function. The most commonly applied one is non-parametric smoothing, that penalizes roughness of the likelihood function. The presented modification is taken from Coles and Dixon (1999); they use a penalty function to provide the likelihood function with the information that the value of ξ is smaller than 1, and that values close to 1 are less likely than smaller values. The penalty function has the explicit form:

$$P(\xi) = \begin{cases} 1 & \text{if } \xi \leq 0 \\ \exp\left\{-\lambda\left(\frac{1}{1-\xi}-1\right)^{\alpha}\right\} & \text{if } 0 < \xi < 1 \\ 0 & \text{if } \xi \geq 1 \end{cases} \tag{3.21}$$

for a range of non-negative values of parameters α and λ. With declining value of parameter α the penalty function raises for values of ξ which are large, but less than 1, while the penalty function descends for values of ξ close to 0. The shape of the penalty function corresponding to different values of λ is straightforward; declining values of λ lead to more severe relative penalty for ξ close to 1. The penalty function for various values of α and λ is shown in Figure 3.1. After numerous experimentation, Coles and Dixon (1999) found that the combination $\alpha = \lambda = 1$ leads to a reasonable performance across a range of values for ξ and sample sizes. Thus our results are reported with respect to this choice.

The corresponding penalized likelihood function is then:

$$L_p(\sigma_u, \xi) = L(\sigma_u, \xi) \cdot P(\xi), \tag{3.22}$$

where L is the likelihood function defined by Formula (3.14). Apparently, the values of σ_u and ξ which maximize the penalized likelihood function given by Formula (3.22) are the maximum penalized likelihood estimator.

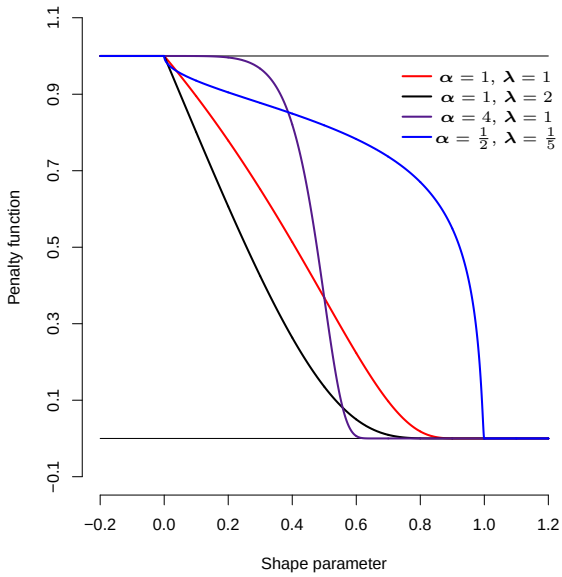


Figure 3.1: Penalty function for various values of α and λ plotted against shape parameter ξ.

Remark 9. The performance of different estimators is often judged in terms of the accuracy of estimation of extreme quantiles, since these quantities are the principal requirement of an extreme value analysis. In practice, estimates of quantiles g_p are obtained by substituting estimates of σ_u and ξ into Formula (3.5), i.e.,

$$\hat{q}_p = \begin{cases} -\frac{\hat{\sigma}_u}{\hat{\xi}}\left(1 - p^{-\hat{\xi}}\right), & \text{for } \hat{\xi} \neq 0, \\ -\hat{\sigma}_u \log p, & \text{for } \hat{\xi} = 0. \end{cases} \tag{3.23}$$

3.3.4 Comparison of methods

Coles and Dixon (1999) carried out multiple simulations in order to assess the accuracy of the maximum likelihood method and the method of probability weighted moments. They simulated thousands of samples of different sizes from the generalized extreme value distribution with various shape parameter keeping the other parameter (μ and σ) fixed. Although in their computations they used the method of block maxima for fitting the generalized extreme value distribution, Hosking and Wallis (1987) showed that the same conclusions can made when applying the threshold methods.

The authors investigated the frequency of convergence failure of the Newton -Raphson iteration algorithm used for maximization of the likelihood function. Hosking and Wallis (1987) concluded that the vast majority of failures of the algorithms were caused by the non-existence of a local maximum of the likelihood

function rather than by failure of the algorithm itself. Failure of the algorithm to converge occurred exclusively for samples for which the other estimation methods gave large negative estimates of ξ.

Both studies showed that for small sample sizes ($n \leq 15$), in terms of bias and mean square error, maximum likelihood appeared to be a poor competitor to the probability weighted moments estimator, especially for the estimate of a particularly extreme quantile q_p, when ξ is positive. This poor relative performance of the maximum likelihood estimator was explained by examining the sampling behaviour of the original parameters of the considered distribution (see Coles and Dixon (1999)). It was revealed that main difference between the methods arose in the distribution of the shape parameter estimate, which, in the case of maximum likelihood, was positively skewed. Moreover, the expression for quantiles is a non-linear function of ξ, thus a small positive bias for ξ translated to a substantial bias for q_p. For more detailed discussion on bias and mean square error of parameter estimates, we refer to Hosking and Wallis (1987).

Coles and Dixon (1999) emphasize that the relative poor estimation of ξ by the maximum likelihood method can be attributed to the different assumptions. In particular, the probability weighted moments estimator a priori assumes that $\xi < 1$. Thus, the estimated parameter space of $(-\infty, \infty)$ is reduced to $(\infty, 1)$. A consequence of this fact is a reduction in the sampling variation of the estimate of ξ. Moreover, increasing value of ξ leads to higher negative bias in the estimate of ξ. Taking into account the non-linearity in Equation (3.5) as a function of ξ, underestimation of ξ is penalized much less heavily, in terms of mean square error, than overestimation. Otherwise specified, the method of the probability weighted moments provides a certain possibility of trade-off between bias and variance in the estimate of ξ. As a consequence, the distribution of extreme quantiles q_p, which are computed on the basis of the probability weighted moments estimates, avoids the heavy upper tail of the distribution provided by the maximum likelihood estimates. Obviously, the restriction $\xi < 1$ viewed as prior information should be available also to the likelihood-based analysis when comparing the methods fairly. One possibility is to adopt a suitable penalty function in order to incorporate the required structural information about ξ. Thus, we arrive to the method of penalized maximum likelihood estimator presented in Subsection 3.3.3.

3.4 Model checking

In this section we introduce a statistical approach as well as a graphical approach to check the validity of an extrapolation based on the fitting generalized Pareto distribution.

3.4.1 Graphical verification of the fitted model

Although the graphical approach is not a sufficient justification, it can provide a reasonable prerequisite. We discuss four methods of goodness-of-fit, namely probability plots, quantile plots, return level plots and density plots.

A *probability plot* is a comparison of the empirical distribution and the fitted distribution. Assume $z_{1:k} \leq z_{2:k} \leq \cdots \leq z_{k:k}$ to be the ordered sample of the k extreme excesses over a high threshold $z_1, \ldots, z_k$, i.e. $z_j = x_j - u$ for those indices $i \in \{1, \ldots, n\}$ for which $x_i > u$ and $x_1, \ldots, x_n$ is sequence of the original data. Then the empirical distribution function evaluated at $z_{j:k}$ can be computed as:

$$\tilde{H}^{(u)}\left(z_{j:k}\right) = \frac{1}{k}\sum_{i=1}^{k} \mathbb{I}(z_{j:i} \leq z_{j:k}) = \frac{j}{k}. \tag{3.24}$$

By substituting of estimated parameters (obtained by application of arbitrary estimating procedure presented in Section 3.3 into Equation (3.4), the corresponding model-based estimates of the distribution function evaluated at $z_{j:k}$ are:

$$\hat{H}^{(u)}\left(z_{j:k}\right) = 1 - \left(1 + \hat{\xi}\,\frac{z_{j:k}}{\hat{\sigma}_u}\right)^{-1/\hat{\xi}}. \tag{3.25}$$

If the model performs well, then $\tilde{H}^{(u)}\left(z_{j:k}\right)$ is approximately equal to $\hat{H}^{(u)}\left(z_{j:k}\right)$ for each j, thus a probability plot, consisting of the points

$$\left\{\left(\tilde{H}^{(u)}\left(z_{j:k}\right), \hat{H}^{(u)}\left(z_{j:k}\right)\right); j = 1, \ldots, k\right\},$$

should lie close to the unit diagonal. Any substantial departures from the guideline indicate some failure in the model. A disadvantage of the probability plot is that both estimates of distribution function, $\tilde{H}^{(u)}\left(z_{j:k}\right)$ as well as $\hat{H}^{(u)}\left(z_{j:k}\right)$, approach 1 as $z_{j:k}$ increases. It is the accuracy of the model for large values that is of the greatest interest, in other words, the probability plot provides the least information in the region of most concern.

A *quantile plot* represents an alternative that overcomes the deficiency described in the last paragraph. The quantile plot consists of the points

$$\left\{\left((\hat{H}^{(u)})^{-1}\left(\frac{j}{k}\right), z_{j:k}\right); j = 1, \ldots, k\right\},$$

where $(\hat{H}^{(u)})^{-1}$ denotes the estimate of the quantile, i.e.,

$$(\hat{H}^{(u)})^{-1}\left(\frac{j}{k}\right) = -\frac{\hat{\sigma}_u}{\hat{\xi}}\left(1 - \left(1 - \frac{j}{k}\right)^{-\hat{\xi}}\right).$$

Similarly as in the previous case, any departures from the unit diagonal indicate model failure.

A *return level plot* is particularly convenient for interpreting extreme value models. The tail of the distribution is compressed, therefore the return level estimates for long return periods are displayed. The plot consists of the locus of points

$$\{(\log y_p, \hat{q}_p); 0 < p < 1\},$$

where $\hat{q}_p$ is the estimate of quantile q_p given by Formula (3.23) and $y_p = -\log(1-p)$ is the return level function (see Section 2.1). If the model is suitable for data, the model-based curve and empirical estimates should be in reasonable agreement. Any substantial or systematic departures from guideline suggests an inadequacy of the model. Moreover, the guideline is a straightforward indicator of the type of the generalized Pareto distribution. If the guideline is linear ($\xi = 0$) the exponential distribution is the most likely distribution for the excesses over a high threshold, convex guideline suggest that the excesses have beta distribution and concave guideline corresponds to Pareto distribution.

All methods of goodness-of-fit presented so far are derived from comparison of model-based and empirical estimates of the distribution function. The last graphical check is an equivalent diagnostic, based on the density function though. A *density plot* is a comparison of the probability density function of a fitted model with a histogram of the data. This is generally less informative than the previous plots though. The reason is that the histogram can vary substantially with the choice of grouping intervals.

3.4.2 Statistical verification of the fitted model

In order to check the validity of an extrapolation based on the fitting generalized Pareto distribution, empirical distribution function statistics is employed. This method compares the hypothesized distribution function $\hat{H}^{(u)}$ and the empirical distribution function $\tilde{H}^{(u)}$. We present several test statistics including Kolmogorov-Smirnov (D), Cramer-von Mises (W^2), and also Anderson-Darling statistics (A^2).

Assume $z_{1:k} \leq z_{2:k} \leq \cdots \leq z_{k:k}$ to be the ordered sample of the k extreme excesses over a high threshold $z_1, \ldots, z_k$, as it was defined above. Then the hypothesized distribution function $\hat{H}^{(u)}$ and the empirical distribution function $\tilde{H}^{(u)}$ have the form given by Formulas (3.24) and (3.25). The *Kolmogorov-Smirnov statistic* measures the maximum deviation of the empirical distribution function and the fitted distribution function, it is defined as:

$$\begin{aligned} D^+ &= \max_{1\leq j\leq k}\left\{\frac{j}{k} - \hat{H}^{(u)}\left(z_{j:k}\right)\right\} \\ D^- &= \max_{1\leq j\leq k}\left\{\hat{H}^{(u)}\left(z_{j:k}\right) - \frac{j-1}{k}\right\} \\ D &= \max\left\{D^+, D^-\right\}. \end{aligned} \tag{3.26}$$

The hypothesis regarding the distributional form is rejected if the test statistic, D, is greater than the critical value obtained from a table. There are several variations of these tables in the literature distinguishing by different scalings for the test statistic and critical regions. These alternatives are provided by software programs that perform the test.

The *Cramér–von Mises test* is based on a statistic of the type:

$$W^2 = \frac{1}{12k} + \sum_{j=1}^{k} \left(\hat{H}^{(u)}\left(z_{j:k}\right) - \frac{2j-1}{2k} \right)^2. \tag{3.27}$$

As it was in the case of Kolmogorov-Smirnov statistic, the Cramer-von Mises statistic measures in some way the discrepancy between the empirical distribution function and the theoretical function F. Also in this case, the critical values do not depend on the specific distribution being tested, thus they can be obtained from a table.

The last test statistic used in the computational part in oder to assess goodness of the fit is the *Anderson-Darling statistic*. This test gives more weight to the tails than the Kolmogorov-Smirnov test. It has the following form:

$$A^2 = -\frac{1}{k} \sum_{j=1}^{k} (2j-1) \left(\log \hat{H}^{(u)}\left(z_{j:k}\right) + \log \hat{H}^{(u)}\left(z_{k+1-j:k}\right) \right) - k. \tag{3.28}$$

Some of the tests, for example Kolmogorov-Smirnov and Cramer-von Mises, are distribution free in the sense that the critical values do not depend on the specific distribution being tested. The Anderson-Darling test makes use of the specific distribution in calculating critical values. This has the advantage of allowing a more sensitive test and the disadvantage that critical values must be calculated for each distribution. Currently, tables of critical values are available for several distribution families, including generalized Pareto distribution.

It has been proven that the Cramer-von Mises test statistic and the Anderson-Darling statistics performs better than the Kolmogorov-Smirnov statistics, and the Anderson-Darling test gives more weight to the tails of the distribution than the Cramer-von Mises. Historically, the Kolmogorov-Smirnov test statistic has been the most used empirical distribution function statistic, but it tends to be the least powerful, overall. In practice, it is often recommended to use the Anderson-Darling test at the first place and the Cramer-von Mises test as the second choice. For more detailed discussion on advantages and disadvantages of the test statistics see Stephens (1974).

4. Data description

In the last decades the Society of Actuaries (SOA) has been leading a series of projects exploring and analysing medical insurance large claims. As a part of the research, a number of insurers were requested to collect diverse information about the claims of each claimant. SOA has released dozens of reports and papers discussing medical insurance large claims. The research by Grazier and G'Sell (2004), initiated in October 1998, provides analysis of factors affecting claims incidence rates and distribution of claim sizes. The study considers claims from years 1997, 1998 and 1999. Although the published monograph is rather descriptive, it creates a relevant source of information to our work. Namely we processed the same database as the authors did, therefore in later data description we will refer to their article. Likewise our work the paper by Cebrián, Deniut, and Lambert (2003) focuses on a statistical modeling of large claims based on the extreme value theory with particular emphasis on the threshold methods of fitting generalized Pareto distribution. A significant part of the article is devoted to a practical implementation of the method carried out on 1991 and 1992 group medical claims database maintained by the SOA. The study considered censored data, i.e., claimants with paid charges exceeding $25,000.

We consider the SOA Group Medical Insurance Large Claims Database (available online at web pages of the Society of actuaries (2004)), which records all the claim information over the period 1997, 1998 and 1999. Since each of the records for specific year contains sufficiently high number of observations we will carry out our practical implementation on 1997 data. Those related to years 1998 and 1999 will serve as an assessment for predicting the future claim sizes based on the extreme value theory and classical forecasting based on the fitting of standard distributions, respectively. Each record of row of the file represents one particular claim and comprises of 27 fields, which can be divided into three subgroups. First 5 columns provide a general information about a claimant such as his/her date of birth or sex. Other 12 columns quantify various types of medical charges and expenses. In our computational experiments we consider only the total paid charges (column 17). The last 10 fields summarize details connected to the diagnosis. For more detailed description of the data we refer to Grazier and G'Sell (2004) or Claim Database Documentation attached to the thesis.

Detailed descriptive statistic is presented only for the claim year 1997, though we provide a brief comment on behavior of the claims in the course of the whole period at the end of the paragraph. The data set consists of approximately 1.2 million observations of $2 billion in total paid charges. The size of the claims range from $0.01 to $1 225 908.30 with the sample mean equal to $1 613.58. Comparing the quantiles of the distribution, which are stated in Table 4.1 together with overall descriptive statistics, we can conclude that the observations are mainly concentrated in lower values.

Figure 4.1 displays box plots for the claims and the logarithmic claims. The red

mean	1613.58	minimum	0.01
1st quantile	101.00	maximum	1225908.30
median	293.70	standard deviation	7892.97
3rd quantile	993.60	skewness	34.82

Table 4.1: Descriptive statistics of the claims (1997).

line in both pictures express the mean. The left box plot in Figure 4.1 clearly shows that the mass is concentrated in lower values. Note that the claims are censored on the vertical axis in order to distinguish at a glance different values of quantiles, the median and the mean. The right graph demonstrates shifting of the mass when applying the logarithmic transformation. We observe that the data are markedly right-skewed (the skewness coefficient equals to 34.82), even on the log-scale. This indicates the long-tailed behavior of the underlying data.

In order to compare the goodness of prediction of future large claims based on standard statistical methods and the method based on the fitting of generalized Pareto distribution, we are interested in finding a distribution which traces behaviour of the claims the best. With respect to the number of observations, we carried out the computation on a random sample of 10 000. Figure 4.2 displays the empirical density and the empirical distribution of the logarithmic claims approximated by the fitted normal distribution. Considering both upper graphs we can conclude that the estimated normal distribution describes well the logarithmic claim amounts although the empirical data appear to be slightly skewed. The quantile-quantile plot in Figure 4.2 provides us with another evidence of goodness

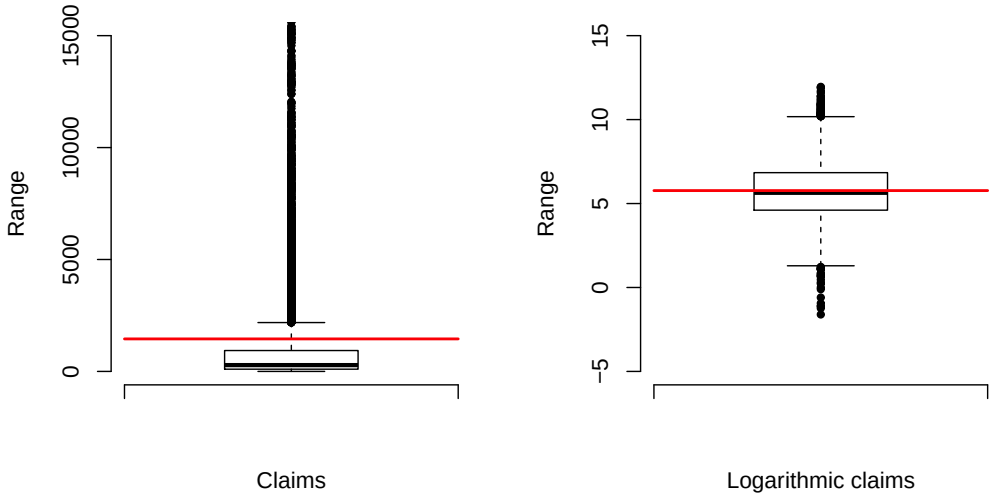


Figure 4.1: Box plots of the claims 1997 (left) and the logarithmic claims °1997 (right).

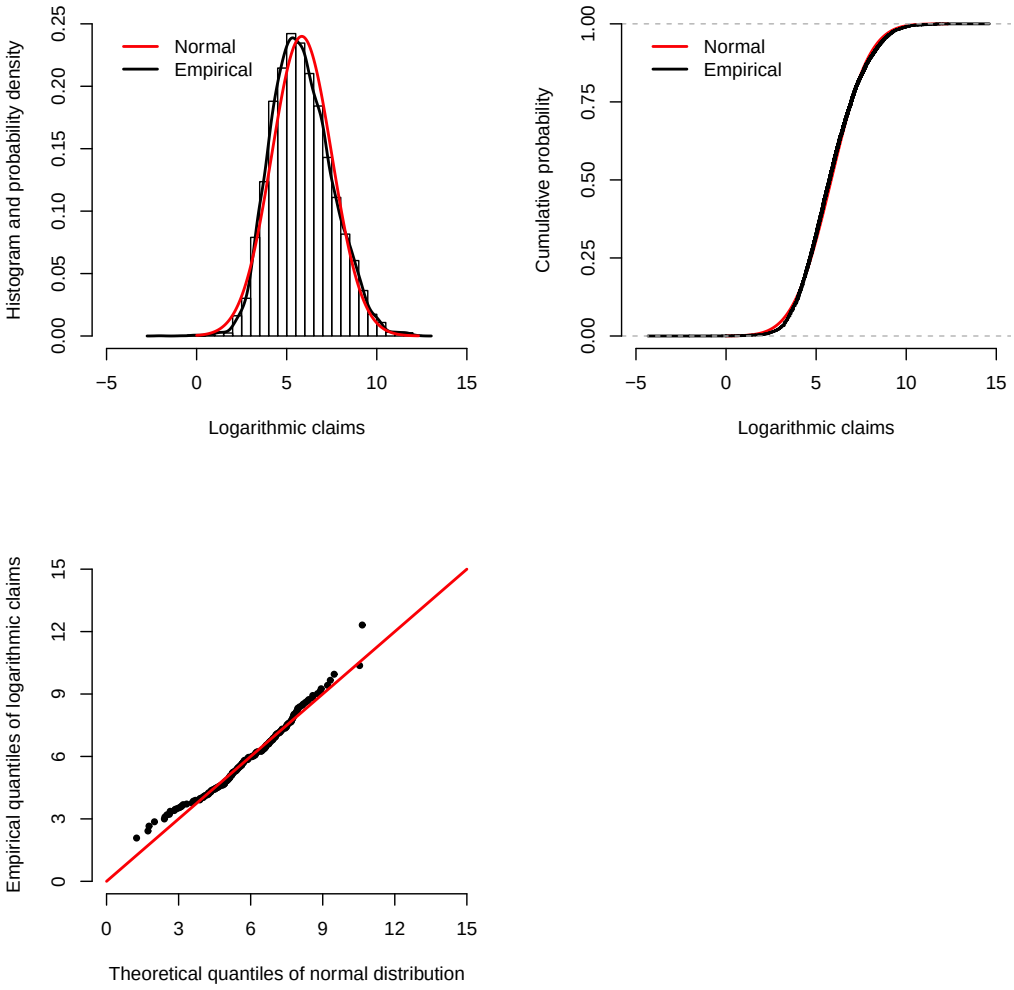


Figure 4.2: The empirical density of the logarithmic claims (1997) approximated by normal distribution (top left), the empirical distribution of the logarithmic claims (1997) approximated by normal distribution (top right), the quantile-quantile plot comparing the empirical quantiles of the logarithmic claims (1997) and the theoretical quantiles of normal distribution (bottom left).

of the fit. The graph displays theoretical quantiles of the fitted normal distribution plotted against empirical quantiles of logarithmic values of the claims. The axis of the quadrant represented by the red line approximates well the points, thus we can conclude that the empirical quantiles can be treated as quantiles of normal distribution.

The accuracy of the fitted distribution was verified by the Shapiro-Wilk test on the confidence level 0.05. We run the test ten times on different random samples consisting of 100 observations. Based on the p-values (the confidence level was exceeded in nine cases out of ten) we have not rejected the null hypothesis and thus we can conclude that logarithmic claims are normally distributed or the

claims are log-normally distributed, respectively [3].

	$\hat{\mu}$	$\hat{\sigma}$
1997	5.820	1.666
1998	5.835	1.687
1999	5.800	1.708

Table 4.2: Mean and standard deviation of the fitted normal distribution to the logarithmic data (claim years 1997, 1998 and 1999).

Analysis of the remaining claim years leads to similar conclusions as drawn above, i.e., we cannot reject the hypothesis of the normal distribution of the logarithmic claims for any of the sample. The parameters of the estimated distributions can be found in Table 4.2, where $\hat{\mu}$ and $\hat{\sigma}$ denote the parameters of the fitted normal distribution. One can observe that the sample mean and standard deviation do not differ significantly regarding the claim year.

[3] When considering several hypotheses in the same test the problem of multiplicity arises. Application of the Holm–Bonferroni method is one of the many ways to address this issue. It modifies the rejection criteria in order to control the overall probability of witnessing one or more type I errors at a given level.

5. Data analysis

Lognormal, log-gamma, gamma, as well as other parametric models have been often applied to model claim sizes in both life and non-life insurance (see for instance Klugman, Panjer, and Willmot (1998)). However, if the main interest is in the tail behaviour of loss distribution, it is essential to have a special model for largest claims. Distribution providing a good overall fit can be inadequate at modeling tails. Extreme value theory together with generalized Pareto distribution focus on the tails and are supported by strong theoretical arguments introduced in Chapters 2 and 3. The following chapter is devoted to reexamination of SOA Group Medical Insurance Large Claims Database using the threshold approach. To check the improvement achieved by using the generalized Pareto distribution instead of a classical parametric model, i.e., the traditional approach of fitting standard distributions based on the analysis performed in Chapter 4, we provide a brief discussion on these two methods as a summary of this chapter.

5.1 Application of graphical threshold selection

The first section focuses on identifying the optimal threshold level in order to construct the further inference. We attempt to apply graphical methods presented in Section 3.2 for each claim year 1997, 1998 and 1999. We provide a detailed analysis and description of respective plots for the claim year 1997 together with discussion on the optimal threshold choice plot. The results based on the same procedure for the remaining claim years 1998 and 1999 are summarized in Table 5.2 at the end of this section.

The mean residual life plot displays empirical estimates of the sample mean excesses plotted against a range of the thresholds, along with confidence interval estimates. Confidence intervals can be added to the plot based on the approximate normal distribution of sample mean stemming from the Central limit theorem. The threshold is chosen to be the lowest level where all the higher threshold based sample mean excesses are consistent with a straight line. Coles (2001) acknowledges that the interpretation of the mean residual life plot is not always simple in practice. The subsequent analysis shows that his attitude is reasonable.

Figure 5.1 shows the mean residual life plot with approximate 95% confidence intervals for the claim year 1997. In our analysis we proceeded censored data, i.e., claim ammounts exceeding \$100 000 over the related period. The left censorship is not a problem since we are interested in the extreme behavior, thus there is no truncation due to benefit maxima. The lower bound of \$100 000 for the threshold selection seems to be reasonable considering descriptive statistics of the claim database (see Table 4.1) and shape of the mean residual life plot for lower threshold than 100 000. The increasing variance of the mean residual life for high thresholds leads to wide confidence intervals, which must be taken into account when assessing the threshold choice. Once the confidence intervals are taken into account, the graph appears to ascend from $u = 100\ 000$ to $u \approx 400\ 000$, within it is approximately linear. Upon $u \approx 700\ 000$ it decays slightly. It is tempting

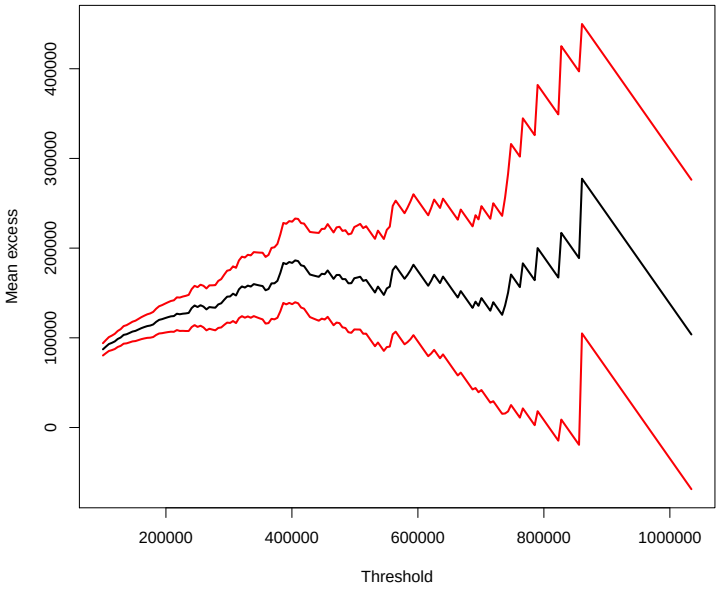


Figure 5.1: The threshold selection using the mean residual life plot (claim year 1997).

to conclude that there is no stability until $u \approx 700\ 000$, after which there is an evidence of approximate linearity. This conclusion suggests to take the threshold equal to 700 000. However, there are just 10 observations above this level, too few to make meaningful inference. Moreover, the estimated mean residual life for large values of u is unreliable due to the limited sample of data on which the estimate and confidence interval are based. Hence, it is preferable to assume that there is some evidence of linearity above $u = 400\ 000$ and work with lower threshold level, which provides us with 50 observations.

As already stated, the threshold is selected to be the lowest level where all the higher threshold based sample mean excesses are consistent with a straight line, i.e., for any higher $u > u_0$ the mean residual life is linear in u with gradient $\xi/(1-\xi)$ and intercept $\sigma_{u_0}/(1-\xi)$. In further consideration we were particularly interested in estimation of the shape parameter ξ based on fitting a linear model to sample mean excesses above threshold choices stemming from the previous discussion (400 000, 600 000 and 700 000). A simple linear regression estimates of parameters for three threshold values provide three fitted mean residual life straight lines in Figure 5.2. In the computation we applied the weighted least squares method for fitting the linear model, where the weights were chosen as to correspond to distribution of the claim mass. One way to achieve the desired effect is to assign to each mean excess the normalized number of observations in the original sample that exceed considered value. Table 5.1 displays obtained estimates of the shape parameter for three threshold choices.

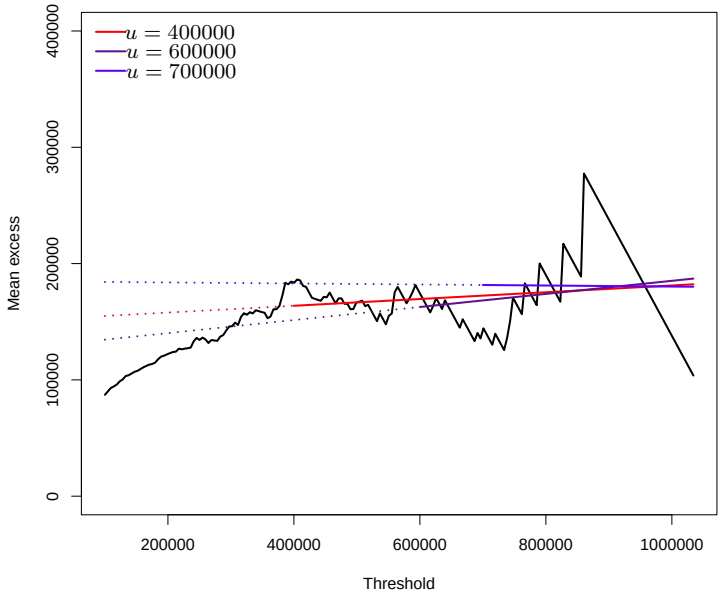


Figure 5.2: Mean residual life plot (claim year 1997) supplemented by the mean residual life estimates for thresholds $u = 400\,000$, $600\,000$ and $700\,000$.

u	ξ
400000	0.0283
600000	0.0532
700000	−0.0044

Table 5.1: Estimates of the shape parameter ξ based on the linear regression model for the mean residual life (claim year 1997).

The threshold choice plot displays empirical estimates of the modified scale parameter and shape both plotted against a range of the thresholds, along with confidence interval estimates. Confidence intervals can be added to the plot based on the so called delta method, as suggested in Coles (2001), or using Fisher information, as described in Ribatet (2011). The threshold is chosen to be the lowest level where all the higher threshold based modified scale and shape parameters are constant. However, in practice decisions are not so clear-cut and threshold selection is not always simple.

Figure 5.3 shows the threshold choice plot with approximate 95% confidence intervals for the claim year 1997. In order to reduce time consumption of the computation, we again considered claim amount exceeding \$100 000 over the related period. As in the case of the mean residual life plot, we observe the increasing variance of the modified scale as well as the shape parameter for high thresholds leading to wide confidence intervals, which must be taken into consideration

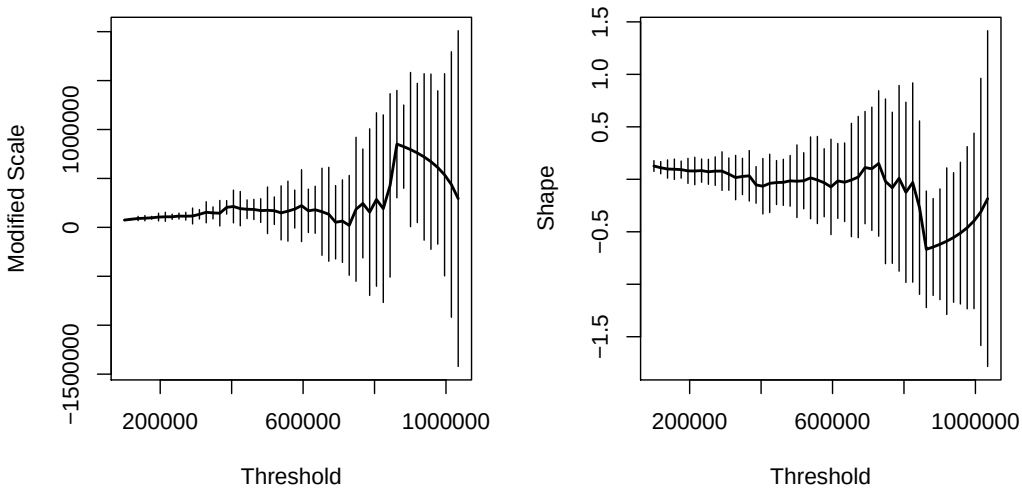


Figure 5.3: The threshold selection using the threshold choice plot (claim year 1997).

when assessing the threshold choice. Both graphs appear to fluctuate around a horizontal line from $u = 100\ 000$ to $u \approx 800\ 000$. The change in pattern for the threshold of $u \approx 800\ 000$ can be characterized by a sharp jump, whereupon the curve of modified scale decays significanty, or the curve of shape increases respectively. This observation suggests that there is no stability until u reaches 850 000. Moreover, exceeding this level there is no evidence of approximate constancy for either modified scale or shape parameter. On the other hand, there are just 3 observations above this level, too few to make any further inference. Moreover, the accuracy of the related estimates is disputed due to the limited sample of data. Given these circumstances, it is probably better to choose a lower. In order to have a reasonably wide sample of data, a threshold lower then 600 000 is preferred. However it is difficult to make any decision about the exact value since the modified scale curve and the scale curve can be both well approximated by a horizontal line. In this case, we adopt conclusion arising from mean residual life plot or L-moments plot.

The L-moments plot displays empirical estimates of L-skewness plotted against empirical estimates of L-kurtosis, along with the theoretical curve defined by Equation (3.10) which is traced as a guideline. Decision about choice of the optimal threshold value is made based on position of estimated points with respect to this theoretical line. Unfortunately, it has been shown by Ribatet (2011) that the graphic has often poor performance on real data.

Figure 5.4 depicts the L-moments plot for the claims year 1997. The red line is associated with the theoretical curve which reflects relation between L-skewness and L-kurtosis. The black line corresponds to their empirical estimates. As

in the previous graphical illustrations we use lower bound for threshold limit equal to \$100 000 since we have already shown that a lower threshold value is rather unlikely. The empirical estimates of L-skewness and L-kurtosis for this lowest admissible threshold are represented by the inner endpoint of the black curve in the graph. As the threshold limit rises, the points corresponding to pairs of empirical estimates of L-moments scroll along the guideline towards the beginning of the coordinate system. The theoretical relation between L-skewness and L-kurtosis given by Equation (3.10) seems to be accurate for sample of data generated by a threshold up to 500 000. Once the threshold reaches 500 000, the estimated points lie on the outer curve and converge to a point on the horizontal axis as the threshold approaches the maximal value of the sample data. Thus the particular shape of L-moments plot is tempting to select a threshold value that does not exceed 500 000.

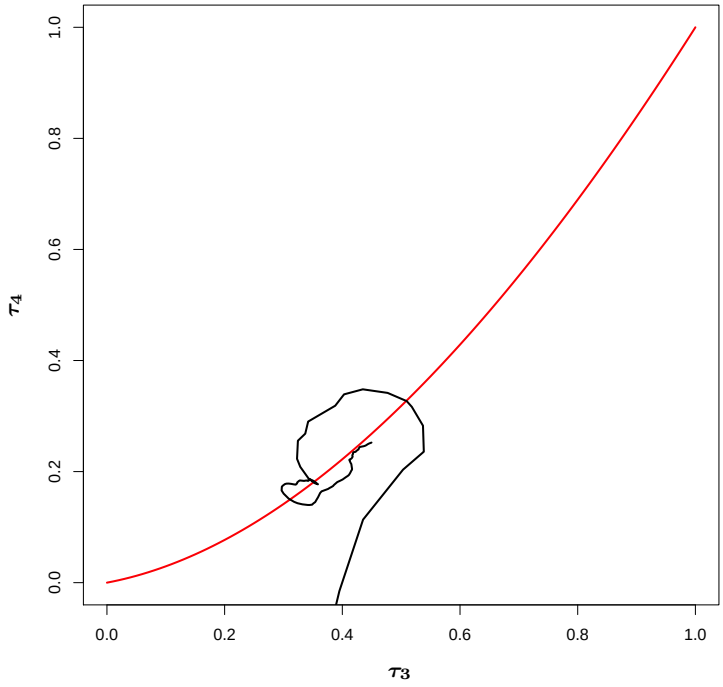


Figure 5.4: The threshold selection using the L-moments plot (claim year 1997).

Remark 10. In order to assess the optimal threshold level based on dispersion index plot, one needs to have two time series at disposal. The first one tracks the absolute values of the examined process, the other one records moments on a time scale when the events occur. Since the second information was not available for our claim database, utilization of this method was unfeasible.

We carried out a similar procedure for determining the optimal threshold also for the remaining claim years 1998 and 1999. The analysis did not result in

any surprising conclusions that should be commented with special emphasis here. Conversely, analogous discussion on the graphical methods for threshold selection could be provided. However, instead of length interpretation of the graphical results, we only state valid threshold choices together with the estimates of the shape parameter ξ based on the linear regression model for the mean residual life in Table 5.2. [3]

1997		1998		1999	
u	ξ	u	ξ	u	ξ
400000	0.0283	300000	0.1708	300000	0.1259
600000	0.0532	600000	0.0939	700000	−0.3350
700000	−0.0044	700000	−0.1267	1000000	−1.2795

Table 5.2: Estimates of the shape parameter ξ based on the linear regression model for the mean residual life (claim years 1997, 1998 and 1999).

To make conclusion about the threshold choice that will be used in further analysis, we take into account all results obtained from the previous graphical analysis. In the spirit of the mean residual life plot the threshold of either 400 000 or 700 000 should be adopted. Whereas the value of 700 000 provides better asymptotic linearity of the mean residual life corresponding to higher thresholds, the value of 400 000 generates large sample of data that is convenient for further distribution estimation. As suggested in paragraph discussing the threshold choice plot, a threshold lower than 600 000 is preferred in order to derive a meaningful inference for generalized Pareto distribution. The L-moments plot is tempting to select a threshold lower than 500 000 as the relation between empirical L-skewness and L-kurtosis is rather valid. Considering all these arguments we suggest to use the threshold value equal to 400 000 in further analysis.

5.2 Fitting the generalized Pareto distribution

The previous section was devoted to selecting the optimal threshold level in order to identify the extremal events that are supposed to have the generalized Pareto distribution. Considering the threshold choices from the previous part we sought for adequate shape and scale parameters of the generalized Pareto family for approximating the tail behavior of our data set. We utilize all three approaches for fitting distribution, the maximum likelihood, the probability weighted moments and the penalized maximum likelihood, that are presented in Section 3.3 pointing out their differences reflecting in estimated parameters and their confidence intervals. Let us recall that the analytical maximization of the log-likelihood is

[3] The figures which served as a basis for the threshold selection, can be found on a compact disc attached to the thesis.

not possible, thus to find the local maximum requires using of numerical methods. In the implementation, we exploit a procedure based on a quasi-Newton method. In the case of the penalized maximum likelihood we adopted a proposal of Coles and Dixon (1999) to set the parameters α and λ both equal to 1. Finally, the constants γ and δ of the plotting position in the probability weighted moments method was set as follows: $\gamma = -0.35$ and $\delta = 0$.

Once the parameters are estimated we check the model using statistical tests (see Subsection 3.4.2) accompanied by graphical verification (see Subsection 3.4.1). We carried out the computation for each claim year separately. The whole procedure is presented in detail for the claim year 1997 whereas the results for the claim years 1998 and 1999 are briefly summarized at the end of the section. Although we suggested to carry out further computation with the lowest value of threshold (for the claim year 1997 it was 400 000), we present the results obtained for all threshold levels that were identified as reasonable choices in the previous section.

Table 5.3 and Table 5.4 report the estimates and confidence intervals of scale and shape parameters obtained for different values of threshold levels applying the maximum likelihood, the penalized maximum likelihood and the probability weighted moments methods. First let us compare the maximum likelihood and probability weighted moments estimator. We observe difference in the scale and shape estimates for all choices of threshold level. For thresholds equal to 400 000 and 600 000 the maximum likelihood based scale parameter is smaller than the probability weighted moments based whereas for threshold equal to 700 000 it is greater. In terms of the shape parameter, it is the other way around, the maximum likelihood based shape parameter is bigger than the probability weighted moments based for threshold choices 400 000 and 600 000 and smaller for threshold equal to 700 000. However the sign of the shape parameter remains consistent with varying the estimation procedure. The exceedances over thresholds equal to 400 000 and 600 000 have beta distribution while those over 700 000 can be approximated by Pareto distribution. Note that the statement holds true when applying the penalized maximum likelihood estimator, except the case of exceedences over 700 000 which seem to be exponentially distributed. We revise that for small sample sizes (n = 15), in terms of bias and mean square error, the probability weighted moments estimator performs better than the maximum likelihood estimator. The relatively better estimation of ξ by the probability weighted moments method can be attributed to the a priori assumption that $\xi < 1$. The difference becomes even more significant for the estimate of a particularly extreme quantile q_p, when ξ is positive. Threshold 400 000 generates 50 observations thus the maximum likelihood estimates can be considered sufficiently accurate. For higher threshold only few observation are available (for 600 000 there are 16 exceedances, for 700 000 there are only 10). Therefore in these cases one should rely on the probability weighted moments estimator.

The motivation for using the penalized maximum likelihood method is to incorporate the extra information that $\xi < 1$ to the likelihood-based analysis in order to assess fairly the maximum likelihood and probability weighted moments estima-

	Threshold u		
	400000	600000	700000
`mle`	183913	174424	145182
	(113914, 253911)	(150547, 198301)	(14828, 275535)
`pmle`	183913	174424	145182
	(113914, 253911)	(57283, 291565)	(121910, 168454)
`pwm`	214691	207588	116705
	(124034, 305348)	(52591, 362586)	(3560, 229851)

Table 5.3: Estimates and confidence intervals of the scale parameter σ_u for different choices of the threshold (claim year 1997) applying the maximum likelihood method (`mle`), the penalized maximum likelihood method (`pmle`) and the probability weighted moments method (`pwm`).

	Threshold u		
	400000	600000	700000
`mle`	-0.0572	-0.0607	0.0493
	(-0.3186, 0.2041)	(-0.3918, 0.2704)	(-0.6011, 0.6996)
`pmle`	-0.0572	-0.0607	0.0000
	(-0.3186, 0.2041)	(-0.5210, 0.3995)	(-0.0859, 0.0859)
`pwm`	-0.1674	-0.1901	0.1961
	(-0.5076, 0.1729)	(-0.7991, 0.4188)	(-0.5512, 0.9434)

Table 5.4: Estimates and confidence intervals of the shape parameter ξ for different choices of the threshold (claim year 1997) applying the maximum likelihood method (`mle`), the penalized maximum likelihood method (`pmle`) and the probability weighted moments method (`pwm`).

tor. However, as it follows from Formula (3.21), for negative value of parameter ξ the penalty function equals to 1 and the estimates based on the maximum likelihood method and the penalized maximum likelihood method are equal. Therefore setting the threshold level as 400 000 or 600 000 leads to the same estimates of the scale and shape parameters (see Table 5.3 and Table 5.4). Note that the confidence intervals vary due to different standard errors of estimated parameters. Both methods should generate different estimates for threshold 700 000 where ξ is positive. However we observe that the scale parameter remains consistent and the estimates of the shape parameter do not differ significantly. These results can be attributed to the fact that the value of the penalty function is close to 1 for positive values of ξ close enough to 0. From the behaviour of the penalty function, i.e. for all selected threshold levels its values are equal or very close to 1, it is not surprising that the estimates based on the penalized maximum likelihood method correspond rather those based on the maximum likelihood method than those based on the probability weighted moments method.

Once the parameters of the generalized Pareto distribution are specified, goodness of the fit has to be checked. Table 5.5 contains p-values of statistical tests presented earlier in Subsection 3.4.2, namely the Anderson-Darling test, Kolmogorov-Smirnov test and Cramer-von Mises test. We stated that the Anderson-Darling statistics performs better than the remaining two, particularly when assessing the tail distribution, thus we make conclusion predominantly on the basis of results of the Anderson-Darling test. We observe that p-values of the test statistic exceed considerably the confidence level of 0.05 for all threshold choices and all estimating approaches. The same conclusion can be drawn when considering the Kolmogorov-Smirnov test or Cramer-von Mises test. Hence we statistically confirmed suitability of the fitted generalized Pareto distribution for all choices of threshold level.

	Threshold u		
	400000	600000	700000
`mle`	0.7799	0.7801	0.8581
	0.8183, 0.1871	0.7316, 0.8657	0.9445, 0.5794
`pmle`	0.7799	0.7801	0.8750
	0.8183, 0.9870	0.7316, 0.3842	0.9596, 0.5794
`pwm`	0.9733	0.8522	0.9117
	0.9671, 0.6521	0.9513, 0.9516	0.8953, 0.5794

Table 5.5: P-values of Anderson-Darling (top), Kolmogorov-Smirnov (bottom left) and Cramer-von Mises (bottom right) tests (claim year 1997).

For completeness, we attach the graphical assessment of goodness of the fit for threshold equal to 400 000 (for theoretical description of the graphical diagnostic see Subsection 3.4.1). Figure 5.5 and Figure 5.6 depict the probability plot, the quantile-quantile plot, the density plot and the return level plot for maximum likelihood and probability weighted moments approach. Note that we do not display the results for the penalized maximum likelihood estimator since for threshold 400 000 the scale and the shape parameters do not differ from those maximum likelihood based. Same parameters generate the same distribution and the whole graphical diagnostic is thus identical. The probability plot displays the empirical distribution function evaluated at the exceedances over threshold 400 000 against the model based distribution function evaluated at the same points. The fit of the distribution is accurate when the points are situated close enough to the quadrant axis. In both figures the quadrant axis (the red guideline) can be considered as a satisfactory approximation for plotted points. The quantile-quantile plot provides similar description of the fitted model as the probability plot, the plotted points correspond to combinations of the empirical quantiles and the model based quantiles. Hence the quadrant axis again expresses the perfect fit. The quantile-quantile plots for our data set confirm that the estimated model is adequate. Similar conclusion can be made by assessing the density plot which compares the model based density and the empirical density accompanied by the histogram. The last graphical diagnostic used was the return level plot

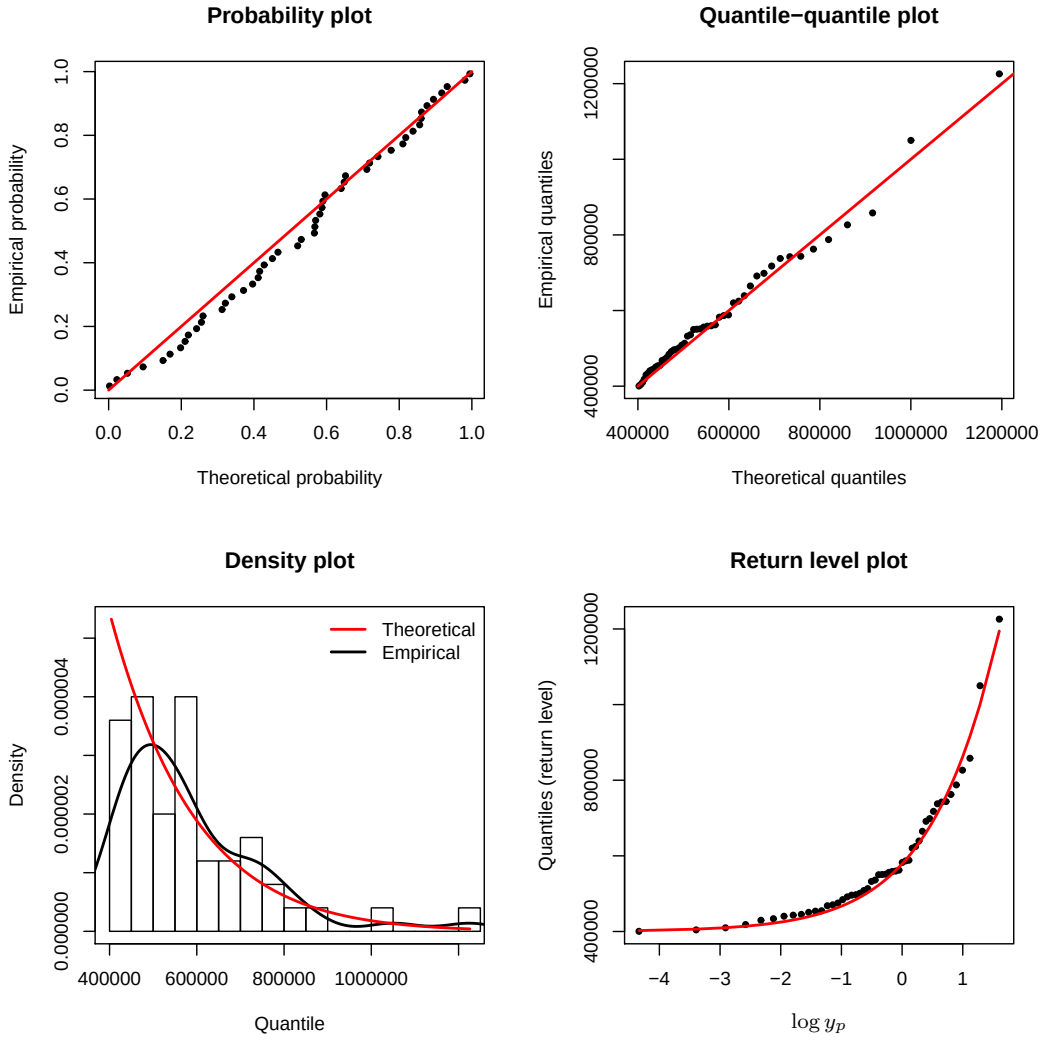


Figure 5.5: Graphical diagnostic for a fitted generalized Pareto model using the maximum likelihood model or the penalized maximum likelihood model equivalently (claim year 1997, threshold 400 000).

which depicts empirical quantiles and the model based quantiles both against $\log y_p$ where $y_p = -log(1-p)$ and p is the reciprocal value of the return level. In Figures 5.5 and 5.6 the model based quantiles are related to the red curve whereas the empirical quantiles to the black points. We can claim that the red guideline is an adequate approximation for the dependence of the empirical quantiles on $\log y_p$. Moreover, from the return level plot a particular type of the generalized Pareto distribution can be recognized. The convex guideline suggests that the excesses have beta distribution which corresponds to negative value of the shape parameter estimated for threshold 400 000.

The data analysis showed that the modeling of the tail behaviour by generalized Pareto distribution is appropriate. Although the goodness of the fit of the distribution was justified by statistical tests for all threshold choices and estimating

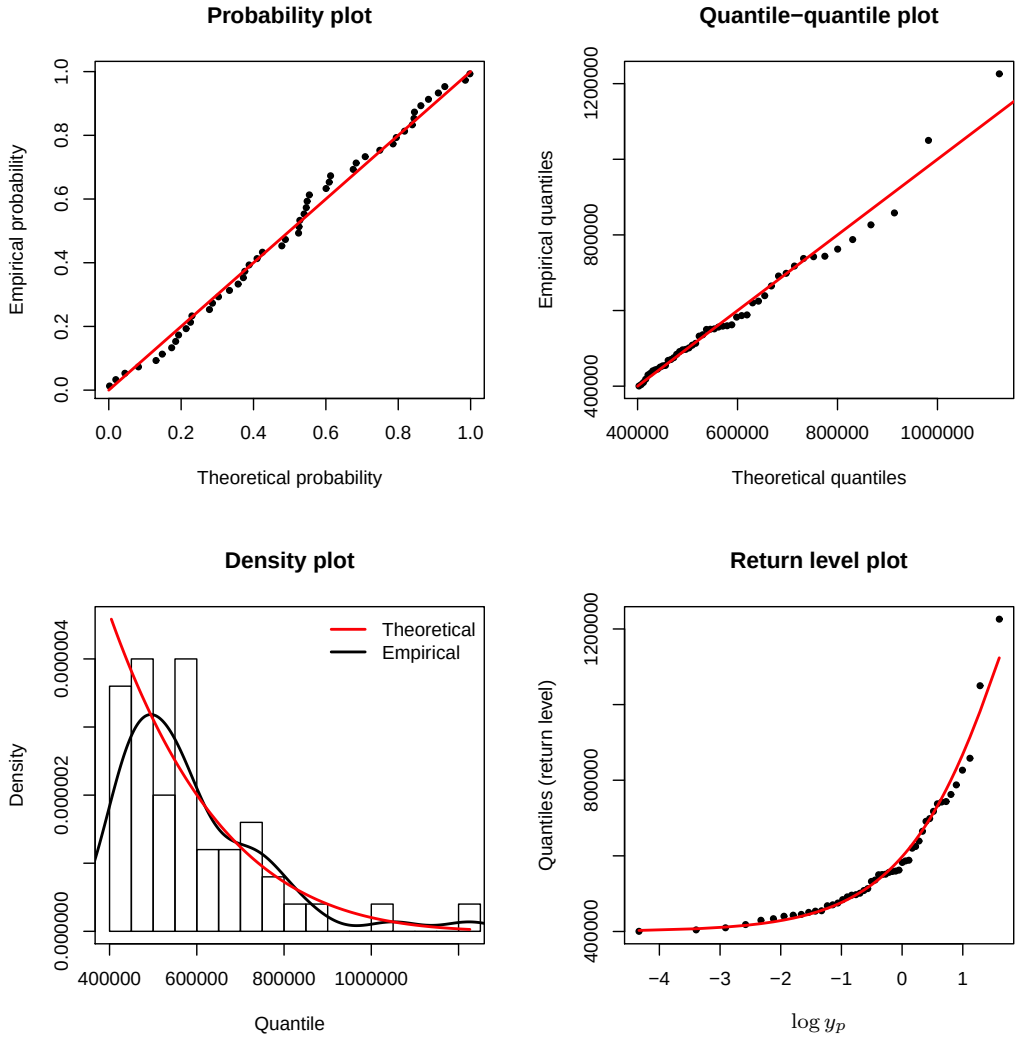


Figure 5.6: Graphical diagnostic for a fitted generalized Pareto model using the probability weighted moments model (claim year 1997, threshold 400 000).

procedure, we suggest to utilize shape and scale parameters computed on the basis of the probability weighted moments estimator for higher values of threshold level, i.e. for 600 000 and 700 000. As it was noted above, for small sample sizes the probability weighted moments estimator performs better in terms of bias and mean square error than the maximum likelihood estimator, in particular the estimate of an extreme quantile q_p can be significantly biased, especially when ξ is positive. For threshold equal to 400 000 one can decide for particular estimating procedure arbitrarily. In conclusion, if one particular choice of the generalized Pareto distribution has to be made, we recommend to adopt the threshold level of 400 000. We believe that the number of exceedances over this threshold is adequate to construct reasonable fit. Higher thresholds generate few observation and thus the estimated model can suffer from inaccuracies.

We repeated the same computational procedure for determining the parameters

for the generalized Pareto distribution also for the remaining claim years 1998 and 1999. As a starting point we adopted the threshold level suggested earlier in Section 5.1, i.e. for the claim year it is 300 000. However for the claim year 1999 the hypothesis that the exceedances over 300 000 have generalized Pareto distribution was rejected, thus we adjusted the threshold level to be in accordance with the generalized Pareto distribution. Table 5.6 summarized the estimated parameters of the distribution for all claim years and particular choices of threshold levels. We present the results obtained by applying the probability weighted moments estimator only since two higher thresholds generate few observation for each claim year, thus the probability weighted moments based estimates are considered to be more accurate. For estimating the parameters for the lowest threshold choice an arbitrary estimating procedure could be used.

u	σ_u	ξ
400000	214691	−0.1674
600000	207588	−0.1901
700000	116705	0.1961

Claim year 1997

u	σ_u	ξ
300000	106967	0.1606
600000	103532	0.3566
700000	141152	0.3299

Claim year 1998

u	σ_u	ξ
500000	192410	0.3459
700000	221663	0.4577
1000000	973970	−0.3248

Claim year 1999

Table 5.6: Parameters of the generalized Pareto distribution for the claim years 1997, 1998 and 1999 estimated using the probability weighted moments approach.

5.3 Comparison with standard parametric fit

To check the improvement achieved by using the generalized Pareto distribution instead of a classical parametric model, one can adopt the methods introduced in Section 3.4. In this section we provide a comparison based on the quantile-quantile plot and Kolmogorov-Smirnov test. First let us recall the results from Chapter 4, i.e. the logarithmic data can be approximated by normal distribution. In other words, we assume that the data are lognormally distributed with specific parameters. Further we assume that the distribution of exceedances over a high threshold can be well described by generalized Pareto distribution with specific parameters. We aim to compare goodness-of-fit of lognormal approximation and generalized Pareto approximation taking into account a tail fraction given by different choices of the threshold level.

We provide a brief summary of an analysis carried out on the claim year 1997.

	Threshold u		
	400000	600000	700000
σ_u	214691	207588	116705
ξ	-0.1674	-0.1901	0.1961

Table 5.7: Estimates of the scale and shape parameters for different choices of the threshold using the probability weighted moments approach (claim year 1997).

Computational results from previous parts are exploited in order to model the tail behaviour of the data, namely we assumed that the logarithmic data are normally distributed with mean equal to 5.82 and standard deviation 1.666 (see Table 4.2). Further we considered three threshold levels (400 000, 600 000 and 700 000) and assumed that exceedances over these thresholds have generalized Pareto distribution with parameters as stated in Table 5.7. Figure 5.7 depicts quantile-quantile plots for the fitted lognormal model and generalized Pareto model for different threshold choices. Comparing the graphs corresponding to a particular threshold level we can state that the lognormal approximation is a poor competitor to the generalized Pareto approximation. Kolmogorov-Smirnov tests confirms conclusions made from the graphical approach. As indicated in

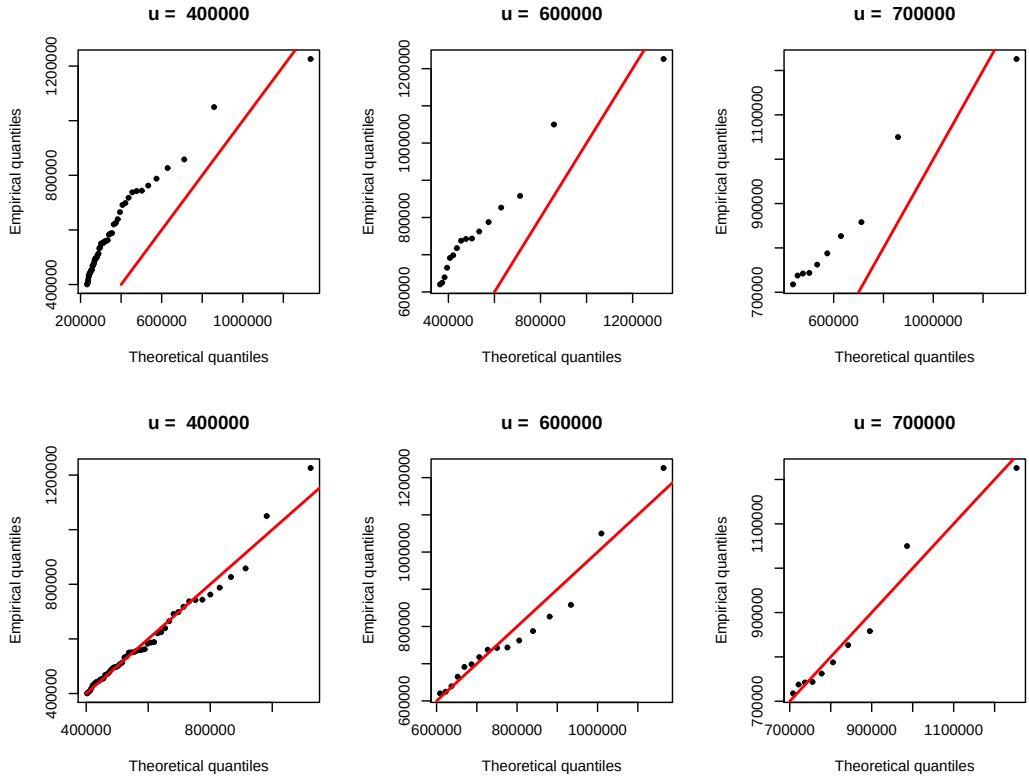


Figure 5.7: Quantile-quantile plots for a fitted lognormal model (top) and generalized Pareto model (bottom) for different threshold levels (claim year 1997).

Table 5.8 p-values of the tests assessing adequacy of the lognormal distribution were considerably close to 0 whereas p-values of testing the generalized Pareto distribution exceeded the confidence level 0.05 in all cases.

	Threshold u		
	400000	600000	700000
`lnorm`	0.0000	0.0000	0.0000
`gpd`	0.9671	0.9513	0.8953

Table 5.8: P-values of Kolmogorov-Smirnov test for a fitted lognormal model and generalized Pareto model for different threshold levels (claim year 1997).

From demonstrated results, we can conclude that no satisfactory fit for the extremal events is obtained using a classical parametric model. Contrary to the generalized Pareto distribution, the fit does not significantly improve when increasing the threshold. Therefore the results highly support the use of the generalized Pareto models instead of the traditional parametric models and demonstrate the interest of extreme value theory when the concern lies in the right tail.

6. Applications of the model

In this chapter, we outline the practical utilization of the generalized Pareto distribution to model the behaviour of the exceedances over high thresholds for solving some actuarial issues. We focus on two applications, point estimation of high quantiles and prediction of probable maximum loss.

6.1 Point estimation of high quantiles

The high quantiles of the distribution of the claim amounts is a measure that provides useful information for insurance companies. In other fields of statistics quantiles are usually estimated by their empirical counterparts (see, for instance, Table 4.1 in Chapte 4). However when one is interested in extremal quantiles, this approach is not longer valid since estimation based on a low number of large observations tempts to be strongly imprecise. Thus, in such cases usage of the generalized Pareto model is preferred.

Let us denote as $F^{(u)}$ the common cumulative distribution function of $X - u$, conditional on $X > u$. From Formula (3.3), we see that a potential estimator for the distribution $F^{(u)}$ is $H^{(u)}$, provided u is sufficiently large. In other words, $H^{(u)}$ approximates the conditional distribution of the losses, given that they exceed the threshold u. Thus quantile estimators derived from this function (given by Equation (3.23)) are conditional quantile estimators that indicate the scale of losses that could be experienced if the threshold u was to be exceeded. To estimate the high unconditional quantiles, we aim relate the unconditional cumulative distribution function F to the conditional cumulative distribution function $F^{(u)}$. Denoting $\bar{F}(y) = 1 - F(y)$, we obtain:

$$\bar{F}^{(u)}(y) = \mathbb{P}\left[X - u \geq y \mid X > u\right] = \frac{\bar{F}(u+y)}{\bar{F}(u)}, \qquad y > 0.$$

Applying the generalized Pareto approximation for the distribution of the exceedances, we have:

$$\bar{F}(u+y) \approx \bar{F}(u)\left(1 - H^{(u)}(y)\right), \qquad y > 0.$$

Provided we have a large enough sample, we can estimate $\bar{F}(u)$ by its empirical counterpart, i.e., $\bar{F}(u) = N_u/n$ where N_u and n are the number of claims above the threshold u and the total number of claims, respectively. To sum up, we can estimate the probability that the claim amount is larger than y, for any amount $y > u$ by

$$\hat{\bar{F}}(y) = \frac{N_u}{n}\left(1 - \hat{H}^{(u)}(y-u)\right) = \frac{N_u}{n}\left(1 + \hat{\xi}\,\frac{y-u}{\hat{\sigma}_u}\right)^{-1/\hat{\xi}}.$$

Substituting $F(q_p) = 1 - p$ to the latter equation, we obtain the following estimator of the quantile:

$$\hat{q}_p = u - \frac{\hat{\sigma}_u}{\hat{\xi}}\left(1 - \left(\frac{pn}{N_u}\right)^{-\hat{\xi}}\right). \tag{6.1}$$

Figure 6.1 displays the generalized Pareto quantiles based on the shape and scale parameters estimated for threshold 400 000 against the empirical quantiles. According to the graph these quantiles rather agree. Note that the total number of observation n exceeds 1 200 000 whereas threshold level 400 000 generates only 50 observation, thus the ratio n/N_u is considerably high and we have to set probability p proportionally low in order to obtain reasonable quantiles.

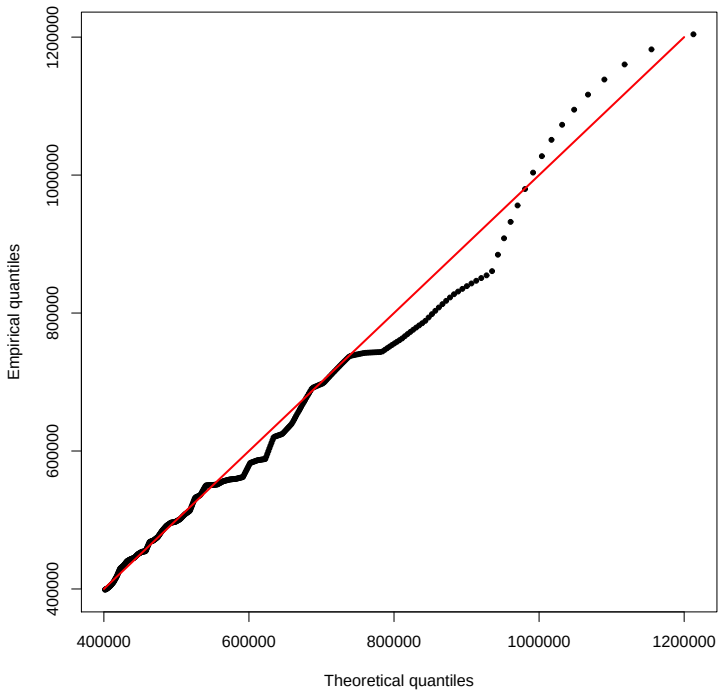


Figure 6.1: Generalized Pareto quantiles against their empirical analogues (claim year 1997, threshold 400 000).

6.2 Probable maximum loss

Another useful measure in the field of insurance is the probable maximum loss which can be interpeted as the worst loss likely to happen. Generally, the probable maximum loss can be obtained by solving equation

$$\mathbb{P}\left[M_n \leq PML_p\right] = 1 - p$$

for some small p and PML_p denotes the probable maximum loss. This means that the probable maximum loss is a high quantile of the maximum of a random sample of size n thus the solution is given as:

$$PML_p = F_{M_n}^{-1}(1-p),$$

i.e., the probable maximum loss is computed as the $1-p$ quantile of the distribution of the maximum loss. We will use an approach based on the generalized Pareto approximation of the exceedances. Under the condition that the exceedances over threshold u can be approximated by the generalized Pareto distribution, the number N_u of exceedances over that threshold is roughly Poisson (for more details we refer to Cebrián, Deniut, and Lambert (2003)). In that situation, it can be proved that the distribution of the maximum above the threshold M_{N_u} of these N_u exceedances can be approached by a generalized extreme value distribution. More precisely, consider a random variable N_u distributed according the Poisson law with mean λ, and let $X_1 - u, \ldots, X_{N_u} - u$ be a sequence of N_u independent and identically distributed random variables with common cumulative distribution function $H^{(u)}$,. Then for $M_{N_u} = \max\{X_1, \ldots, X_{N_u}\} - u$ the cummulative dstribution function is given as:

$$\mathbb{P}\left[M_{N_u} \leq x\right] \approx G(x)$$

with the following modification of location and scale parameters:

$$\begin{aligned} \mu &= \frac{\sigma_u}{\xi}\left(\lambda^{\xi} - 1\right), \\ \sigma &= \sigma_u \lambda^{\xi}. \end{aligned}$$

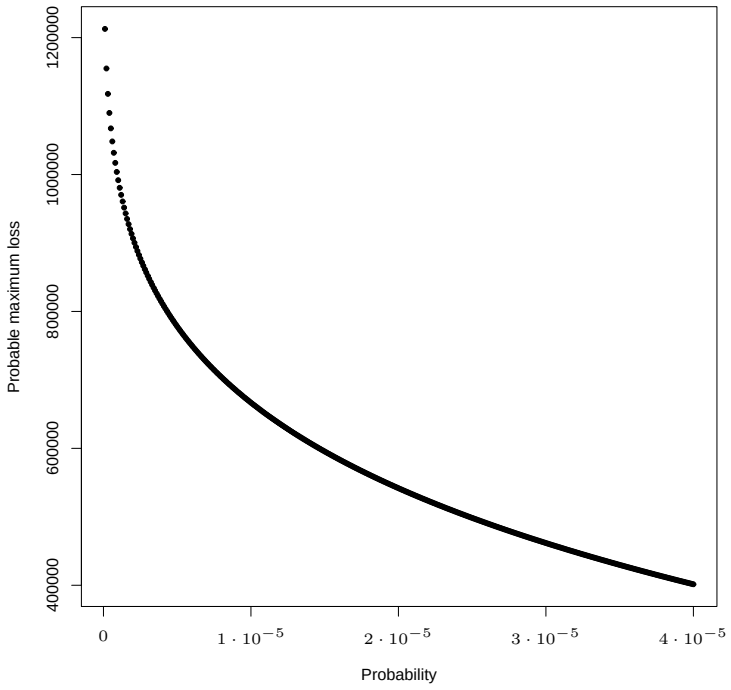


Figure 6.2: Probable maximum loss (claim year 1997).

Using this distribution and Formula (2.3), the previous probable maximum loss definition results in:

$$PML_p = u + \frac{\sigma_u}{\xi}\left(\left(-\frac{\lambda}{\log(1-p)}\right)^{\xi} - 1\right).$$

Thus when estimating the probable maximum loss for particular p we can use the following formula:

$$PML_p = u + \frac{\hat{\sigma}_u}{\hat{\xi}}\left(\left(-\frac{\hat{\lambda}}{\log(1-p)}\right)^{\hat{\xi}} - 1\right),$$

where $\hat{\lambda} = N_u/n$.

Figure 6.2 shows the probable maximum loss for different probability levels. Similarly as in the previous subsection, p has to be chosen sufficiently low to balance the estimate of λ.

7. Conclusion

The presented study has shown the usefulness of extreme value theory for analysis of an insurance portfolio. Indeed, this theory undoubtedly allows to determine distribution of the large losses, high quantiles, and probable maximum loss. However, the generalized Pareto distribution is valid only above very high thresholds, thus the estimation of the optimal threshold level seems to be a crucial issue. We introduced the graphical threshold selection including the mean residual life plot, the threshold choice plot and the L-moments plot. We found out that the graphical approach is ambiguous since each of the plots might prefer a different threshold choice. In other words, we were not able to choose a single threshold level for either of the data sample. In the end we prioritized the lowest threshold suggested by the mean residual life plot since it ensured generating a reasonable amount of exceedances available for further inference. To estimate the shape and scale parameters of the generalized Pareto distribution we applied the maximum-likelihood, penalized maximum-likelihood and probability weighted moments methods. We observed slight difference in the estimated parameters for each of the methods. However the sign of the shape parameter remained consistent with varying the estimation procedure, thus the specific shape of the generalized Pareto distribution was independent of the applied method. We emphasize that for small sample sizes, in terms of bias and mean square error, the probability weighted moments estimator performs better than the maximum likelihood estimator. The lowest threshold levels generated sufficiently enough observations to consider the maximum likelihood estimates to be accurate. For higher thresholds only few observation were available, therefore in these cases one should rely on the probability weighted moments estimator. The estimates of the shape and scale parameters based on the penalized maximum likelihood method were very close to those estimated by the maximum likelihood methods. This fact can be attributed to the specific behaviour of the penalty function, for all selected threshold levels its values were equal or very close to 1 and thus the penalized likelihood function was approximately the same as the likelihood function and their maximization led to similar estimates. In the next step of the estimation procedure we checked goodness of the fit. Although the Kolmogorov-Smirnov and Cramer-von Mises tests provided satisfactory results, our conclusions were derived mainly from p-values of the Anderson-Darling which seems to perform better than the previously mentioned ones. We found out that the generalized Pareto distribution with corresponding parameters can be considered as a valid model for all predetermined threshold levels and all estimating methods. At the end of the computational part we showed that no satisfactory fit for the extreme events can be obtained using a classical parametric model, and thus the generalized Pareto distribution performs better when modeling large claims.

Bibliography

A. C. Cebrián, M. Deniut, and P. Lambert. Generalized pareto fit to the society of actuaries' large claims database. *North American Actuarial Journal*, 7:18–36, 2003.

S. G. Coles. *An introduction to statistical modelling of extreme values*. Springer Series in Statistics, London, 2001.

S. G. Coles and M. J. Dixon. Likelihood-based inference for extreme value models. *Extremes*, 2(1):5–23, 1999.

D. R. Cox and D. V. Hinkley. *Theoretical Statistics*. Chapman and Hall, London, 1974.

C. Cunnane. Note on the poisson assumption in partial duration series model. *Water Resources Research*, 15(2):489–494, 1979.

A. C. Davison and R. L. Smith. Models for exceedances over high thresholds (with discussion). *Journal of the Royal Statistical Society*, 52:237–254, 2012.

K. L. Grazier and W. G'Sell. Group medical insurance claims database collection and analysis. Technical report, 2004.

J. R. M. Hosking. L-moments: analysis and estimation of distributions using linear combination of order statistics. *Journal of the Royal Statistical Society*, 52(1):105–125, 1990.

J. R. M. Hosking and J. R. Wallis. Parameter and quantile estimation for the generalized pareto distribution. *Technometrics*, 29:339–349, 1987.

J. R. M. Hosking, J. R. Wallis, and E. F. Wood. Estimation of the generalized extreme value distribution by the method of probability weighted moments. *Technometrics*, 27:251–261, 1985.

S. F. Juaréz and W. R. Schucany. Robust and efficient estimation for the generalized pareto distribution. *Extremes*, 7(3):237–251, 2004.

S. A. Klugman, H. H. Panjer, and G. E. Willmot. *Loss models: from data to decisions*. Wiley, New York, 1998.

J. M. Landwehr, N. C. Matalas, and J. R. Wallis. Probability weighted moments compared with some traditional techniques in estimating gumbel parameters and quantiles. *Water Resources Research*, 15:1066–1064, 1979.

A. Luceno. Fitting the generalized pareto distribution to data using maximum goodness-of-fit estimators. *Computational Statistics & Data Analysis*, 51(2): 904–917, 2006.

L. Peng and A. H. Welsh. Robust estimation of the generalized pareto distribution. *Extremes*, 4(1):53–65, 2001.

J. III Pickands. Statistical inference using extreme order statistics. *Annals of Statistics*, 3:119–131, 1975.

M. Ribatet. *A user's guide to the POT package (version 1.4)*. University of Montpellier II, 2011.

C. Scarrott and A. MacDonald. A review of extreme value threshold estimation and uncertainty quantification. *Statistical Journal*, 10(1):33–60, 2012.

R. L. Smith. Maximum likelihood estimation in a class of non-regular cases. *Biometrika*, 72:67–92, 1985.

Society of actuaries. Group medical claims database 1997, 1998 and 1999. `http://www.soa.org`, 2004. Accessed 9th October 2013.

M. A. Stephens. Edf statistics for goodness of fit and some comparisons. *Journal of the American Statistical Association*, 69(347):730–737, 1974.

J. Zvang. Likelihood moment estimation for the generalized pareto distribution. *Australian and New Zealand Journal of Statistics*, 49(1):69–77, 2007.