Charles University in Prague

Faculty of Social Sciences

Institute of Economic Studies

RIGOROUS THESIS

Heavy Tails and Market Risk Measures: the Case of the Czech Stock Market

Author: Mgr. Radek Bulva

Consultant: MRes PhDr. Jan Zápal

Academic Year: 2011/2012

Declaration

I hereby declare that I compiled this thesis independently, using only the listed literature and resources.

In Prague on September 11, 2011

Radek Bulva

Acknowledgements

I would like to express my gratitude to my consultant Jan Zápal for his patience and invaluable remarks and comments which led to a substantial improvement of the thesis. However, the usual caveat applies.

Throughout the thesis statistical packages S-PLUS and gretl were used for computations, estimations and graph plotting.

ABSTRACT

One of the stylized facts about the behaviour of financial returns is that they tend to exhibit more probability mass in the tails of the distribution than would be suggested by the normal distribution. This phenomenon is called heavy tails. The first part of this thesis focuses on examining the tails of a distribution of returns on Czech stock market index PX. Parametric and semi-parametric approaches to estimation of the tail index, a measure of heaviness of tails, are applied and compared. The results indicate that the tails behave in a way one would expect from an emerging market stock index.

In the second part of the thesis, implications for two quantile-based market risk measures, Value at Risk and Expected Shortfall, are investigated. The main conclusion is that heavy-tailed alternatives should be preferred to the normal distribution in order to avoid serious underestimation of risks embedded in the underlying process.

JEL classification: C13, C14, C16, G15;

Keywords: Heavy Tails, Parametric and Semi-parametric Estimation, Statistics of Extremes, Extreme Value Theory, Market Risk, Value at Risk, Expected Shortfall.

ABSTRAKT

Těžké chvosty jsou jedním z mnoha dobře zdokumentovaných stylizovaných faktů o chování výnosů finančních aktiv. V první části této práce se zabýváme metodami odhadu parametru chvostu rozdělení výnosů hlavního českého akciového indexu PX a za tímto účelem zkoumáme řadu parametrických a semi-parametrických postupů. Výsledky naznačují, že chování chvostů výnosů indexu PX je v souladu s empirickými výsledky dostupnými v literatuře.
Ve druhé části práce se zaměřujeme na dvě měřítka tržního rizika, Value at Risk a Expected Shortfall. Spolu s tradičními metodami odhadu založenými na normálním rozdělení diskutujeme i metody založené na výsledcích první části práce, které berou v potaz odlišné chování chvostů. Porovnání výsledků jednotlivých metod nás vede k závěru, že modelování finančních výnosů pomocí normálního rozdělení může vést k závažnému podcenění rizik.

JEL klasifikace: C13, C14, C16, G15;

Klíčová slova: těžké chvosty, parametrické a semi-parametrické metody odhadu, teorie extrémních hodnot, tržní riziko, Value at Risk, Expected Shortfall.

TABLE OF CONTENTS

Introduction 1

1. Preliminaries 6

   1.1 Data Description 6

   1.2 The Symmetry of Tails 12

   1.3 Measuring the Fatness of Tails 14

2. Statistics of Extremes 16

   2.1 Mandelbrot's Method of Sample Moments 17

   2.2 Parametric Approaches to Tail Estimation 21

      2.2.1 The Block Maxima Method 21

      2.2.2 The Peaks over Threshold Method 30

   2.3 Semi-parametric Estimators 39

      2.3.1 The Modified Hill Estimator 46

   2.4 Literature Survey 51

   2.5 Summary 54

3. Applications to Risk Management 57

   3.1 Value at Risk 58

   3.2 Expected Shortfall 64

Conclusion 68

References 71

Appendix A 77

Appendix B 79

LIST OF FIGURES

Figure 1 Daily returns on PX index in period from 1995/01/05 to 2010/12/30...............7
Figure 2 Histogram of daily PX returns in period from 1995/01/05 to 2010/12/30. 9
Figure 3 The exponential QQ-plots of the left and the right tail of PX returns, respectively. 10
Figure 4 The mean excess plot of negative returns on PX index..................................11
Figure 5 The mean excess plot of returns on PX index. 11
Figure 6 Cumulative mean of PX returns. 19
Figure 7 Cumulative variance of PX returns...19
Figure 8 Cumulative skewness of PX returns...20
Figure 9 Cumulative kurtosis of PX returns. 20
Figure 10 GEVD CDFs H(ξ, μ, σ) for Fréchet, Weibull and Gumbel cases. 24
Figure 11 GEVD PDFs h(ξ, μ, σ) for Fréchet, Weibull and Gumbel cases...................24
Figure 12 The Frechét PDFs with different shape parameters......................................27
Figure 13 The QQ-plots of residuals of GEVD fitted to yearly, semesterly, quarterly and monthly maxima of negative daily PX returns (left tail). 31
Figure 14 The QQ-plots of residuals of GEVD fitted to yearly, semesterly, quarterly and monthly maxima of daily PX returns (right tail)..31
Figure 15 The scatter plots of residuals of GEVD fitted to yearly, semesterly, quarterly and monthly maxima of negative daily PX returns (left tail)..32
Figure 16 The scatter plots of residuals of GEVD fitted to yearly, semesterly, quarterly and monthly maxima of daily PX returns (right tail)..32
Figure 17 CDFs G(ξ, β) of the three types of the GPD...35
Figure 18 PDFs g(ξ, β) of the three types of the GPD. 35
Figure 19 The mean excess plot of the left and the right tail of PX returns. 36
Figure 20 The ML estimates of shape parameter ξ of the left tail of PX returns based on fitting the GPD as a function of a number of exceedances. 37
Figure 21 The ML estimates of shape parameter ξ of the right tail of PX returns based on fitting the GPD as a function of a number of exceedances. 37
Figure 22 The scatter plot and the QQ-plot of residuals of GPD fitted to the left tail and the right tail of PX returns above the optimal threshold of 0.019..................................40

Figure 23 The Hill estimates of ξ of the left tail of PX returns with 95% confidence interval as a function of k. 49
Figure 24 The Hill estimates of ξ of the right tail of PX returns with 95% confidence interval as a function of k. 49
Figure 25 The Hill estimates of ξ of the left tail of PX returns as a function of k. 50
Figure 26 The Hill estimates of ξ of the right tail of PX returns as a function of k. 50
Figure 27 GPD fit to the left tail of PX returns with estimates and confidence intervals of 1-day VaR and ES on 99% confidence level. 67
Figure 28 GPD fit to the right tail of PX returns with estimates and confidence intervals of 1-day VaR and ES on 99% confidence level. 67
Figure 29 Correlogram of PX returns. 77
Figure 30 Correlogram of squared PX returns. 78
Figure 31 Cumulative mean of a random sample drawn from Student-t distribution with 3 degrees of freedom. 79
Figure 32 Cumulative variance of a random sample drawn from Student-t distribution with 3 degrees of freedom. 79
Figure 33 Cumulative skewness of a random sample drawn from Student-t distribution with 3 degrees of freedom. 80
Figure 34 Cumulative kurtosis of a random sample drawn from Student-t distribution with 3 degrees of freedom. 80

LIST OF TABLES

Table 1 Percentiles of standardized PX returns (left tail). 12
Table 2 Percentiles of standardized PX returns (right tail). 12
Table 3 ML estimates of GEVD parameters of the left tail of PX returns...................... 29
Table 4 ML estimates of GEVD parameters of the right tail of PX returns. 29
Table 5 ML estimates of GPD parameters of the left tail of PX returns......................... 39
Table 6 ML estimates of GPD parameters of the right tail of PX returns. 39
Table 7 The extreme value index estimates (with asymptotic std. errors) of the left tail of PX returns based on respective fractions of tail observations. 45
Table 8 The extreme value index estimates (with asymptotic std. errors) of the right tail of PX returns based on respective fractions of tail observations. 45
Table 9 The modified Hill estimates based on OLS and WLS method for the left and the right tail of PX returns.. 51
Table 10 Comparison of the Peaks over Threshold and the Modified Hill estimates..... 54
Table 11 Values of the standardized test statistic for Hill estimates based on different numbers of tail observations. 56
Table 12 1-day VaR estimates with confidence levels 95% and 99% for both tails of PX returns based on different methods. 63
Table 13 1-day ES estimates with confidence levels 95% and 99% for both tails of PX returns based on different methods. 66
Table 14 Autocorrelation and partial autocorrelation function of PX returns. 77
Table 15 Autocorrelation and partial autocorrelation function of squared PX returns... 78

List of Abbreviations

CDF	Cumulative Distribution Function
CI	Confidence Interval
ES	Expected Shortfall
EVT	Extreme Value Theory
(G)ARCH	(Generalized) Auto-Regressive Conditional Heteroskedasticity
GEVD	Generalized Extreme Value Distribution
GPD	Generalized Pareto Distribution
IID	Independent and Identically Distributed
MDA	Maximum Domain of Attraction
ML	Maximum Likelihood (estimation)
MSE	Mean Squared Error
OLS	Ordinary Least Squares
(P)ACF	(Partial) Auto-Correlation Function
PDF	Probability Density Function
PSE	Prague Stock Exchange
SE	Standard Error
VaR	Value at Risk
WLS	Weighted Least Squares

INTRODUCTION

Since its very beginnings, major developments in financial economics have relied on the normal distribution. It is the legacy of Bachelier (1900) who spotted analogy between the diffusion of heat through substance and bond market price's movements and formulated an equivalent of Brownian motion to describe the latter. This proved to be an important stepping stone for modern financial theories developed several decades later.

It started with the pioneering work of Markowitz (1952) who argued that an investor, when contemplating an investment into some asset, takes into consideration the trade-off between its expected return and riskiness. These two quantities are then approximated by the first two statistical moments of the distribution of returns, the mean and the variance. It is the normal distribution which is uniquely determined by these two parameters, making it a natural candidate for modelling the underlying distribution.

The Modern Portfolio Theory of Markowitz was followed by the Capital Asset Pricing Model due to Sharpe (1964) and Lintner (1965). It distinguished between systemic and idiosyncratic risk and introduced the correlation of returns on an asset with overall market returns as the measure of its riskiness. In this framework, the normal distribution is assumed to govern random innovations of returns.

Then came the option pricing model of Black and Scholes (1973). At its heart lies a simple formula for pricing more sophisticated financial instruments whose value is derived from and depends on price trajectory of certain underlying asset. Returns on the underlying asset are assumed to follow a geometric Brownian motion.

And finally, in late 1980s J.P. Morgan introduced Value at Risk, a risk measure, as a response to an urgent need for new risk management tools. By definition, Value at Risk is equivalent to a quantile of the distribution of returns. As an empirical distribution cannot be extended beyond its maximum (or minimum), a theoretical distribution is used to approximate the behaviour of returns. Due to its convenient properties, such as easy conversion across different values of parameters, the normal distribution is frequently used.

The stubborn reliance on the normal distribution is all the more surprising, given that soon after the emergence of the modern financial theory it was realised by many researchers that the distribution of returns is not normal and they set out to examine in more detail its statistical properties. These efforts were further accelerated by the advent of personal computers powerful enough to process vast datasets of high-frequency data.

Cont (2001) summarized the results of this still on-going research into eleven stylized facts about behaviour of financial returns. These are:

- Absence of (linear) autocorrelations;
- Heavy tails;
- Gain/loss asymmetry;
- Aggregational Gaussianity;
- Intermittency;
- Volatility clustering;
- Conditional heavy tails;
- Slow decay of autocorrelation in absolute terms;
- Leverage effect;
- Volume/volatility correlation;
- Asymmetry in time scales.

Throughout the years much effort has been put into finding a theoretical distribution or process which would be able to account for all the stylized facts. The classical models are stable distributions (Mandelbrot 1963), the mixed diffusion jump process (Press 1967), the Student-t distribution (Blattberg and Gonedes 1974), ARCH process (Engle 1982) and discrete mixtures of normal distributions (Kon 1984). As examples of more modern ones, normal inverse Gaussian distributions (Barndorff-Nielsen 1997), exponentially truncated stable distributions (Cont et al. 1997) and hyperbolic distributions (Eberlein et al. 1998) can be mentioned.

Cont (2001) points out that for a parametric model to be able to reproduce all the required properties it has to have at least four parameters: a location parameter, a scale parameter, a parameter describing the decay of the tails and an asymmetry parameter allowing each tail to behave differently. These requirements are met by most of the modern distributions like normal inverse Gaussian distribution, generalized hyperbolic distribution or exponentially truncated stable distribution, but their use is somewhat constrained by their complexity and difficult tractability, both analytical and computational.

Nonetheless, it is not always necessary to have a universal model perfectly describing the entire distribution of returns. Sometimes, it is more reasonable to focus on just one or a few properties of returns which are central to the purpose of enquiry and pick a less complex model that is capable of reliably reproducing these features.

The focus of this thesis will be on the tendency of financial returns to accumulate more probability mass in the tail regions than would be suggested by the normal distribution. Notions like heavy tails, thick tails, fat tails or long tails all refer to this phenomenon and will be used interchangeably in the remainder of the thesis. Investigating the tails of a distribution equals in essence to inferring on the extreme realizations of the underlying process and chances of their occurrence. This is mainly the dominion of risk managers but as Mandelbrot and Hudson (2008) point out even the ordinary small investor may be more preoccupied with the chance that a sudden price swing erases her life-time savings rather than with minor differences in expected returns.

The interest in the tails of a distribution is not exclusive to finance. It is shared especially with natural sciences. In hydrology, estimation of extreme flood discharges is of paramount importance. The 100-year discharge, i.e., level which is exceeded once in every 100 years on average, is a standard benchmark. Since estimates are usually based on discharges for a period shorter than 100 years, the need for proper modelling of tails is obvious. In meteorology, data on wind speed, ozone concentration and rainfall intensity go generally unnoticed until they break some specific threshold. Also the global warming and climate change debate is characteristic for examining extreme observations over time.

Non-life insurance industry is a typical example of thinking in terms of extremes. Surely, catastrophes like earthquakes, hurricanes or air plane crashes result in very large claims. But more ordinary events like car accidents and fires do so as well. Evaluating these risks in terms of expected (average) events, rather than extreme ones, could be deceptive.

The pioneer of heavy tails in finance is Mandelbrot (1963) who firstly spotted this phenomenon in a time series of cotton prices. Nevertheless, for a long time extreme events were frequently considered as outliers rather than observations possessing essential information, presumably in order to maintain the comfort provided by the easily tractable normal distribution. The situation started to change after the October 1987 market crash which wrought havoc among financial practitioners and forced mainstream academicians to contemplate ways of modelling extreme events instead of eliminating them as outliers.

Since then, much research has been undertaken in this area, using a wide range of methods on many different financial assets. The truth is that the majority of researchers focused on mature capital markets and currencies and, due to their short history, the emerging markets have been largely ignored. The honourable exceptions are Jondeau and Rockinger (2003), LeBaron and Samanta (2004) and Horák and Šmíd (2009). Therefore, the first aim of this thesis is to contribute with a comprehensive description of the tail behaviour of the Czech stock market index PX to this overlooked branch of research.

The second aim of the thesis is to illustrate the impact of heavy tails on estimation of two quantile-based market risk measures, Value at Risk and Expected Shortfall. In general, quantile-based risk measures are very sensitive to the choice of theoretical distribution which is used for modelling the underlying process. Therefore, using a thin-tailed distribution, such as the normal distribution, may lead to underestimation of risk. A couple of heavy-tailed alternatives will be discussed.

The thesis is structured as follows. Section 1 deals with preliminaries. Apart from data analysis and tail index definition, it also touches upon the methodological question of tail symmetry. In Section 2 the underlying theoretical results about behaviour of extreme events are presented and parametric and semi-parametric methods of tail index estimation are applied to PX returns. These results are then used in Section 3 for estimation of Value at Risk and Expected Shortfall based on a couple of different distributional assumptions.

The thesis at its present form is an updated version of master thesis successfully defended in September 2010. No major amendments were made. The biggest change is enlargement of dataset which now runs from January 1995 to December 2010 and contains about 3% more observations. The impact on estimated values is very limited and does not alter any of the previous conclusions.

Following remarks in opponent's report, some formal deficiencies have been corrected: reference to statistical packages has been added; statistically significant estimates in tables have been marked; referencing was improved; typos and mistakes have been removed. Also some arguments have been re-formulated and extensive explanations added where deemed necessary.

Apart from these changes, whole passages have been re-organized and re-written to facilitate understanding and support the flow of argument. Subsections 1.4 and 1.5 from previous version, dealing with QQ-plots and mean excess plots, were scrapped altogether. Their theoretical parts were replaced with relevant reference while the rest of their contents was condensed and incorporated into subsection 1.1.

Occasionally, a new reference was added based on author's additional reading on the topic.

1. PRELIMINARIES

How can we detect the presence of heavy tails in a dataset? How do they demonstrate themselves? What does it exactly mean to say that the tails of a distribution are heavy (or fat or thick)? How can we measure this quality? Should be both tails supposed to be identical? These are just some of the elementary question to be addressed in this section which is meant to lay foundations and prepare ground for subsequent, more profound inference on the behaviour of tails.

1.1 DATA DESCRIPTION

The data underlying this thesis are daily observations of returns on PX index, the major stock market index of the Prague Stock Exchange (PSE). PX is a price index of blue-chip shares with weights based on market capitalization.

The PSE itself was established in late 1992 and the first trading transactions took place on its floor in April 1993. The predecessor to PX, called PX50, was introduced, with initial value set to 1,000 points, a year later on April 5, 1994 but regular 5-days-a-week quotations were available since no sooner than September 19, 1994. Without any effect on the continuity of index values, PX50 was succeeded by PX on March 20, 2006.

Somewhat truncated dataset spanning 16 years, starting on January 5, 1995 and ending on December 30, 2010, will be used in following sections. The reason behind cutting off earlier and later observations is essentially a formal one – by this modification it is ensured that exactly the same data will be used in all estimation procedures[1] while the impact on results is negligible.

The dataset contains 3,998 observations of daily returns. Here and throughout the following sections, by returns we mean log-returns

$$r_{t+1} = \log\left(\frac{P_{t+1}}{P_t}\right) = \log P_{t+1} - \log P_t \,, \tag{1.1}$$

where P_t is the index value at time t. The time series is plotted in Figure 1.

[1] In subsection 2.2.1 the Block Maxima method of tail estimation will be used. It relies on temporal blocks of data and in case of larger blocks, e.g., annual or semi-annual, an issue with incomplete blocks might arise. Excluding them altogether may lead to a loss of valuable information, while using them entails the risk of distorting results.

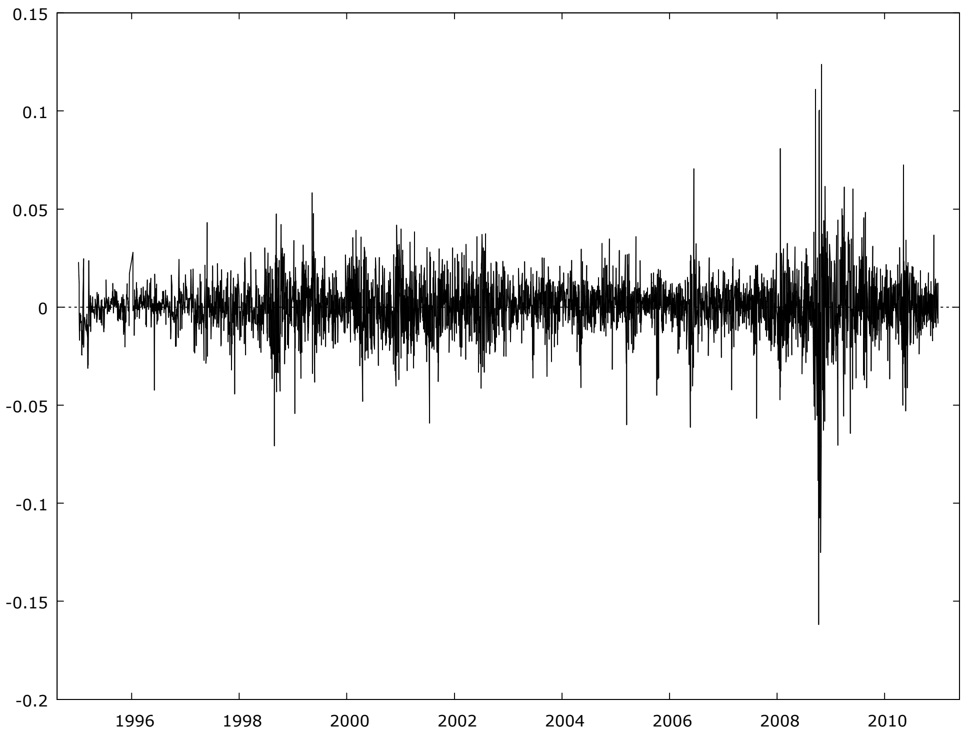

Figure 1 Daily returns on PX index in period from 1995/01/05 to 2010/12/30. Source: author's calculations.

The returns have a positive mean of 0.0197% and a high standard deviation of 1.4482%. The distribution is slightly skewed (-0.4420) and exhibits excess kurtosis (11.9680) over the normal distribution. There is a moderate serial correlation in the data – the first-order autocorrelation coefficient is statistically significant and positive at 0.0878 with some of the higher-order lags significant as well. Strong serial correlation of squared returns (with both the first- and the second-order coefficients above 0.35) then suggests (G)ARCH effects and volatility clustering. The autocorrelation and partial autocorrelation functions of simple and squared returns are included in Appendix A.

At first glance, the dependence structure in returns might raise worries as the theoretical concepts and estimation methods used in the next section are built on IID assumptions. However, as will be shown later, it is not the entire dataset which is used in estimation; instead, either the extremal observations from blocks of data or observations exceeding some high threshold are used. Danielsson and de Vries (2000) point out that the empirical properties of extreme observations surpassing some high threshold are not the same as the properties of the entire returns process and provide evidence that dependency for these extreme observations is much reduced compared to the entire process. Jondeau and Rockinger (2003) then discuss the relative resistance of block maxima to dependency in the return process.

Focusing now on tail observations, both the minimal (-16.1856%) and the maximal (12.3641%) return over the entire period were products of the recent financial crisis and occurred in the same month on October 10 and 29, 2008, respectively. Considering minima and maxima recorded for each calendar year separately, a couple of interesting insights comes up.

Firstly, there are great discrepancies between the yearly minima and maxima. The highest minimal daily return (i.e., the most severe daily loss in that year, but the softest among the minima of all the years) and the lowest maximal daily return (i.e., the beefiest gain in that year, but the slimmest among maxima of all the years) were recorded in 1995 and reached -3.1260% and 2.4586%, respectively.

Secondly, year 2008 was the only one which saw absolute values of extreme returns breaching the 7.5% threshold, i.e., in any other year than 2008 there was not a single daily gain larger than 7.5%, nor a single daily loss more severe than -7.5%. What is more, in 2008 there were eleven such unprecedented events (six positive and five negative) with nine of them occurring in one month (October) and four of them in just one trading week (October 13-17).

This observation hints at the tendency of extremes to cluster together backed, for example, by Longin (1996) who reports that in 28 years (out of 106 years in his sample of daily returns of an index of the most traded stocks on the New York Stock Exchange) the yearly minima and maxima occurred in the same week. This is not the case of PX returns though; there is not a single such week in the sample and it happened only three times[2] that the yearly extremes coincided in the same month. It should be added that, even if it was the case, this temporal clustering does not impede the earlier suggestion of reduced dependency in extreme observations, since in estimation we focus on one side of the distribution only.

Even such a brief and basic description of the data at hand is more than sufficient to draw suspicion of heavy tails. The first obvious clue is the excess kurtosis of the distribution which indicates that there are more observations in the central region around the mean of the distribution than would be implied by the normal distribution. This feature is nicely illustrated by histogram of the PX returns (Figure 2). The shape of the histogram is typical for financial returns and reveals that, apart from the bulk of the data in the centre, the empirical distribution significantly deviates from the Gaussian

[2] Apart from year 2008, this holds also for years 2002 and 2010.

curve in the tail regions in the sense that the frequency of large observations is higher than normal.

Another clue is the volatility clustering which is apparent in Figure 1: steadier periods of weaker fluctuations are intermitted by stormy periods of relatively high volatility and vice versa. Volatility clustering and its impact on heaviness of tails were investigated by, among others, Ghose and Kroner (1995). Using GARCH models, they found evidence supporting their hypothesis that the fat tails in financial data are indeed, although not exclusively, caused by temporal volatility clustering. However, as Cont (2001) points out, even after correction for volatility clustering, the residual time series usually still exhibits heavy tails (so-called conditional heavy tails).

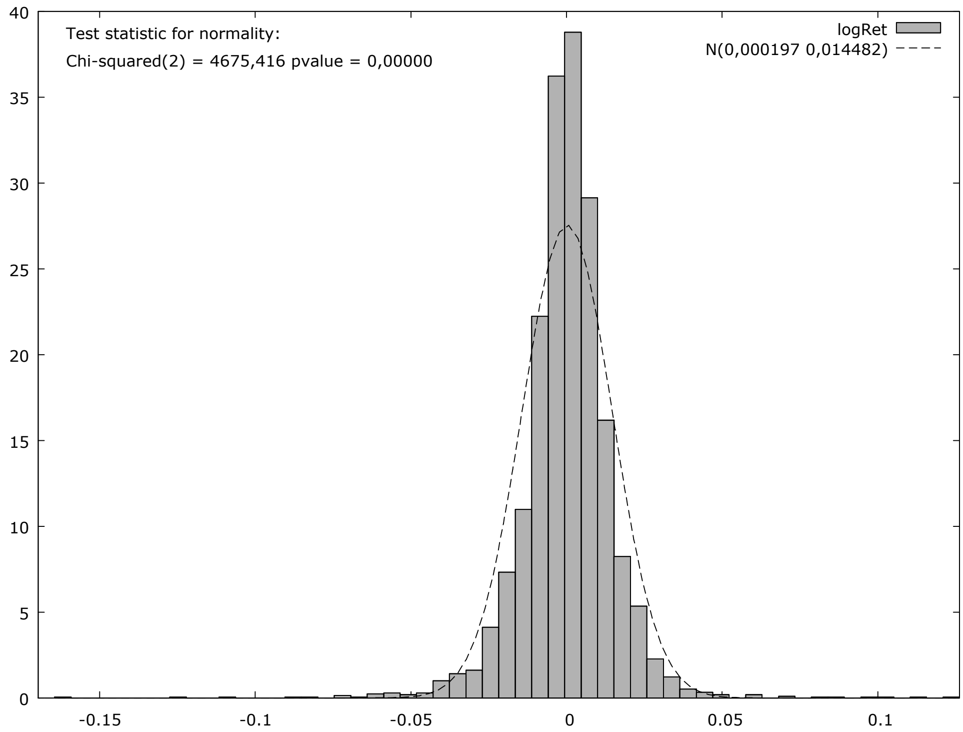

Figure 2 Histogram of daily PX returns in period from 1995/01/05 to 2010/12/30. Source: author's calculations.

More advanced statistical tools, such as the mean excess plot and Quantile-Quantile plot (shortly QQ-plot), may be used in search for evidence of fat tails. The two plots have very convenient properties for exponential distribution. Since the normal distribution has exponentially decaying tails, these properties can be exploited to investigate the nature of the tails of any underlying empirical distribution. We refer to Beirlant et al. (2004) for the theoretical background and derivation of results which will be used in this context.

For exponential tails, QQ-plot should deliver linear pattern. Figure 3 shows the exponential QQ-plots for absolute values of observations exceeding arbitrarily chosen threshold of 0.01, taken from each tail of PX returns. Up to a certain point the plots are indeed linear but then start to bend rightwards, resulting in visibly concave pattern which indicates heavier than exponential tails as values of the extreme empirical quantiles tend to be larger than their theoretical counterparts. Interestingly, the QQ-plot of the right tail seems to straighten up for the most extreme quantiles.

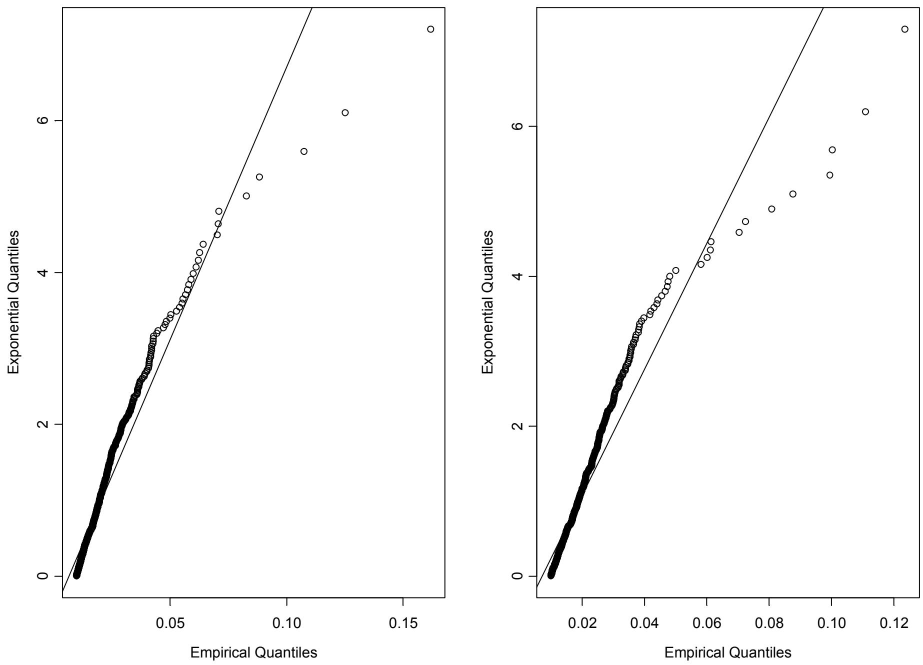

Figure 3 The exponential QQ-plots of the left and the right tail of PX returns, respectively. Source: author's calculations.

The theoretical mean excess function for the exponential distribution is constant and does not vary with the threshold above which the excesses are computed. When the empirical distribution has thicker tail than the exponential distribution, the empirical mean excess plot ultimately increases, while for thinner tail it ultimately decreases.

Figure 4 and Figure 5 show the mean excess plots for negative and positive returns on PX index. Apparently, from a certain threshold on, both plots have positive slope, indicating heavy tails. Similarly to its QQ-plot, there is a twist at the end of the mean excess plot of the right tail, hinting at relative loss of probability mass in its most extreme region. Zivot and Wang (2006) report similar behaviour of the right tail of S&P 500 returns.

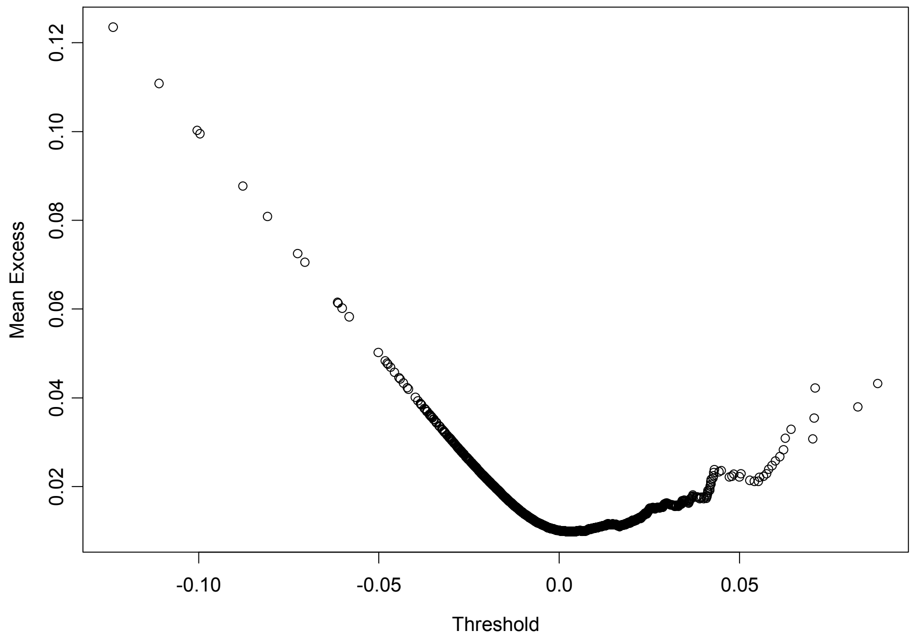

Figure 4 The mean excess plot of negative returns on PX index. Source: author's calculations.

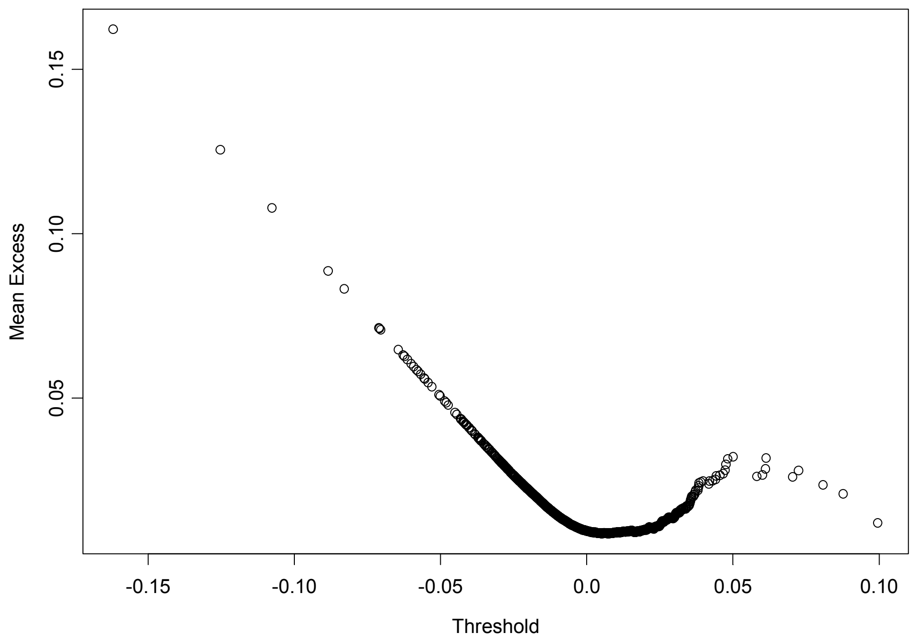

Figure 5 The mean excess plot of returns on PX index. Source: author's calculations.

Finally, Table 1 and Table 2 reports percentiles of the standardized PX returns. The time series was centred by subtracting the mean and divided by the standard deviation to have unit variance so that it would be easily comparable to the theoretical quantiles of the standard normal distribution. In the light of preceding paragraphs, it is hardly surprising that the 1st and 99th percentiles in absolute values are larger than the corresponding normal quantiles. On the other hand, there are not enough observations for the 5th and 95th percentiles to be compatible with the normal distribution. Notice also that while the mass in the left tail is thicker than normal already for 2.5% of most extreme observations, the right tail is still thinner than normal at this point.

	1%	2.5%	5%	10%
PX returns	-2.8663	-2.0132	-1.5218	-1.0421
Normal	-2.3263	-1.9600	-1.6449	-1.2816

Table 1 Percentiles of standardized PX returns (left tail). Source: author's calculations.

	99%	97.5%	95%	90%
PX returns	2.4357	1.8373	1.4471	1.0327
Normal	2.3263	1.9600	1.6449	1.2816

Table 2 Percentiles of standardized PX returns (right tail). Source: author's calculations.

1.2 The Symmetry of Tails

The Efficient Market Hypothesis maintains that prices reflect all publicly available information and, therefore, the returns should follow the so-called random walk. As a consequence, one would naturally expect both tails of the distribution of returns to be perfectly symmetric. Yet, the empirical evidence often contradicts the classical theories and their assumptions. There are plenty of hints in the preceding subsection that the tails of PX returns do not follow identical patterns, starting with the histogram through the QQ-plots through the mean excess plots to the percentiles.

It is a wide-spread and generally accepted notion that the left tail should be heavier than the right one. This belief is commonly explained by the predisposition of markets to occasionally inflate price bubbles. While the bubble is typically inflated gradually by a series of moderate gains over a longer time period, it usually takes no more than a couple of sudden and exceptionally strong negative returns to correct for it.

More elaborate explanation was offered by Campbell and Hentschel (1992). Their starting point is a simple premise that returns are moved by news. Apart from generating volatility, news tends to cluster together in the sense that news is often followed by additional news. A piece of good news increases stock prices, but some of this increase is offset by the increase in risk premium requested by investors for higher volatility. On the other hand, the impact of a piece of bad news is a decrease in stock prices which is further intensified by the increase in the risk premium. As a result, clustering of news might lead to the left tail of returns being heavier than the right one.

Whatever the theoretical justification, the empirical evidence on this issue is ambiguous. Horák and Šmíd (2009) found that left tails of stock returns are significantly heavier than the right ones, but only in case of three from their sample of 22 world-wide stocks. Their inference was based on the asymptotic normality of the Hill estimator and therefore rested on certain distributional assumptions. Instead, LeBaron and Samanta (2004) used Monte Carlo approach and bootstrapping methods on their sample of 8 emerging markets' and 12 developed markets' stock indices. Their results were mixed but in most cases there was no statistically significant asymmetry. Nevertheless, when present, tail asymmetry was more likely to be found in emerging markets, especially in Asia and South America. And finally, Jondeau and Rockinger (2003), using the block maxima method on the sample of 20 stock indices including mature as well as emerging markets, reported they were not able to reject the hypothesis that the tail indices of the left and the right tail were equal for any country in their sample.

The discussion about the asymmetry of tails is important also from the methodological point of view. Basically, there are two alternatives. If symmetry is assumed, one can take the absolute values of centred observations and treat the data as coming from a single tail.[3] This approach has the advantage of effectively doubling the scarce empirical material available for estimation. The obvious disadvantage is a possibly serious distortion of results in case the tails differ significantly.

Alternatively, avoiding the assumption of symmetry and investigating the tails separately allows for more precise modelling and comparison of the behaviour of the underlying distribution of either tail. Eventually, statistical tests may be performed to decide upon symmetry. The snag here is that by using less data the results tend to give

[3] It should be noted here that, apart from Mandelbrot's method of cumulative sample moments, all methods of estimation used later focus only on one tail of the distribution.

more blurry picture about the tails in terms of higher standard errors. Besides, in certain applications one is interested in behaviour of only one tail (e.g., downside risks).

Although the empirical evidence of asymmetry is shabby and depends on statistical tools used, we feel that our dataset is large enough to provide sufficient material for investigating either tail separately. It has to be stressed that from here on we will work with absolute values of observations from the left tail, i.e., losses will be positive numbers. The reason for this is merely a formal one as the statistical methods which will be used throughout the following sections are constructed in a way that they assume observations from the right tail of the distribution. We will return to the question of symmetry of distribution of PX returns at the end of Section 2.

1.3 Measuring the Fatness of Tails

In our exposition so far, we have treated heavy tails simply as tails with more probability mass in their furthermost regions than would be suggested by the normal distribution. Such a characterization is clearly insufficient and calls for a more rigorous definition.

Let X be a random variable. Suppose that CDF $F(x)$ varies regularly at infinity with tail index α

$$\lim_{t\to\infty}\frac{1-F(tx)}{1-F(t)}=x^{-\alpha}, \tag{1.2}$$

where $\alpha>0$ and $x>0$. Then the unconditional distribution of X is heavy-tailed and a parametric form for the tail shape of $F(x)$ can be obtained by taking a second order expansion of $F(x)$ as $x\to\infty$. The only non-trivial possibility under mild assumptions is

$$F(x)\simeq 1-ax^{-\alpha}\left[1+bx^{-\beta}\right] \tag{1.3}$$

for x large and where $\alpha,\beta>0$. The crucial coefficient is the Tail Index α, which determines the fatness of the tail. Coefficient a is the scale parameter and embodies the dependency effect through the so-called extremal index. Coefficients β and b are the second order counterparts of α and a. For more details and proofs see Danielsson and de Vries (1997).

Alternatively, the tail index α can be defined as the order of the highest absolute moment which is finite. Consequently, the tail index is sometimes called the maximal moment exponent. The higher the tail index, the thinner the tail; for a Gaussian or exponential tail $\alpha = +\infty$ and therefore all moments are finite, while for a power-law distribution with exponent k, the tail index is equal to k (Cont 2001).

In some applications, the tail index is substituted by the Extreme Value Index ξ. It is defined simply as an inverse to the tail index, i.e., $\xi = \alpha^{-1}$. To add to the confusion, in the framework of the Generalized Extreme Value distribution and the Generalized Pareto distribution, introduced in Section 2, ξ is referred to as the Shape Parameter.

2. STATISTICS OF EXTREMES

The first academic forays into the area of extreme events can be traced back to early 20^{th} century. In its initial stages, the statistics of extremes was preoccupied with theoretical results on limiting behaviour of sample maxima, due to Fisher and Tippett (1928), which laid the foundations of so-called Extreme Value Theory (EVT). In 1970s EVT was substantially enriched with theoretical results on limiting behaviour of extreme observations exceeding a high threshold, elaborated by Balkema and de Haan (1974) and Pickands (1975).

From mid-1970s, a parallel line of research on extreme events started to emerge thanks to the ground-breaking works of Hill (1975) and again Pickands (1975) and the attention started to gradually shift towards semi-parametric estimators. Their simplicity and convenient asymptotic properties were very appealing, although many challenges remained to be addressed.

In the following decades, researchers focused mainly on accommodating and mitigating the most severe drawbacks of the various estimation methods and on investigating the validity of these results for non-IID observations. Among others, a second-order reduced-bias estimator (weighted Hill estimator) and the mixed moment estimator were developed. For a summary of these developments see Gomes et al. (2008).

Nowadays, the statistics of extremes is firmly established and has its place in many fields of research. The best evidence of its importance is an increasing number of self-contained publications focusing solely on this branch of statistics: Embrechts et al. (1997) has already become a classic reference in the field of finance and insurance; Beirlant et al. (2004) offer a broader spectrum of applications in their exposition; and de Haan and Ferreira (2006) contributed with an extensive introductory text.

In this section, we will focus on only a handful of estimators which are most commonly used in finance and related areas. Whether by chance or by design, these estimators correspond to the key developments in the field and therefore provide an excellent overview.

2.1 MANDELBROT'S METHOD OF SAMPLE MOMENTS

Introduced by Mandelbrot (1963), the method of sample moments cannot be counted among the proper tail index estimation methods. The reason is simple: it does not provide exact numerical estimates. Instead, and more fittingly, it can be thought of as a sort of educated guess in what regions the tail index could lie.

The method is based on the alternative definition of tail index as the maximal moment exponent, implying that heavy-tailed distributions do not possess a complete set of statistical moments, i.e., for $\alpha < +\infty$ and $k \geq \alpha$

$$E\left(X^k\right) \approx \pm\infty. \tag{2.1}$$

In words, sample moments of order higher than the theoretical value of the tail index are infinite. Therefore, by cumulatively computing sample moment as a function of sample size, only two outcomes are possible. If the theoretical moment is finite, the sample moment will eventually settle down to a region defined around its theoretical limit and will fluctuate around that value. In the opposite case, when the true value is infinite, the sample moment will either diverge or exhibit erratic behaviour and large fluctuations.

Naturally, the simplicity of this procedure brings about certain considerable limitations. Firstly, it does not produce a precise estimate of the tail index value. The best possible outcome is an interval bounded by two integers. On the other hand, in certain circumstances even such a limited insight is sufficient, for instance, when the purpose of enquiry is to check whether the variance of a sample is well defined.

Secondly, as emphasised by Cont (2001), sample moments of higher orders tend to be unstable. For instance, the standard deviation of kurtosis, i.e., the fourth sample moment, involves the eighth moment of the distribution which reaches enormous numerical values. Thus, despite being finite, kurtosis is likely to behave erratically.

And thirdly, by working with sample moments of the distribution we implicitly assume symmetric behaviour of both tails. Drawbacks of such assumption have been thoroughly discussed in subsection 1.2.

Figure 6 through Figure 9 plot the first four sample moments of PX returns as cumulative functions of number of observations.[4] As a point of reference, similar plots of cumulative sample moments of a random series drawn from Student-t distribution

[4] Although actually being functions of number of observations, they are plotted against time axis to facilitate intuitive identification of breaking points.

with 3 degrees of freedom (DF) are included in Appendix B. The convenient property of the Student-t distribution is that its statistical moments are finite up to its number of DF. The number of DF was chosen based on empirical evidence summarized in subsection 2.4 and should provide a feasible benchmark. In all four figures the impact of recent financial turmoil starting in late 2008 is apparent. While there is only a minor dip in the mean, jumps in the second and the third moment are more pronounced, although their relative significance pales in the face of the hike experienced by the fourth moment. All these moves are in line with intuition: fluctuations naturally increase variance while their unprecedented magnitude in this case led to the sharp spike in kurtosis; on the other hand, decreasing mean and leaning of the distribution leftwards documented by skewness give us a clue about the general direction of these wild market moves, i.e., downwards.

It might be very tempting and convenient to discard the crisis time observations and cut off the sample in September 2008. In that case we would have all four moments more or less converging and thus indicating that the value of tail index could be expected to lie somewhere above 4. But it is exactly these extreme times when the fat tails are formed by putting on their weight and disregarding them would be misleading, especially when they are our primary focus.

So, what to make of these figures? The first moment is apparently stable and there is nothing to suggest divergence or erratic oscillation. By contrast, this is definitely not the case of the fourth moment with its gigantic leap upwards. Therefore, we may consider it a safe bet to state that the tail index will lie between 1 and 4. Regarding the second and the third moment, there are also some moderate hikes and slumps. Curiously, at first glance the behaviour of cumulative skewness looks more stable than that of cumulative variance but this is mostly due to larger scale of Y-axis in case of skewness caused by wilder oscillations at the beginning of the sample. However, when compared to the patterns of the second and the third moment of benchmark distribution in Appendix B, no palpable qualitative difference is apparent and thus we may conclude that the tail index estimates produced in the following subsections are likely to lie between 3 and 4.

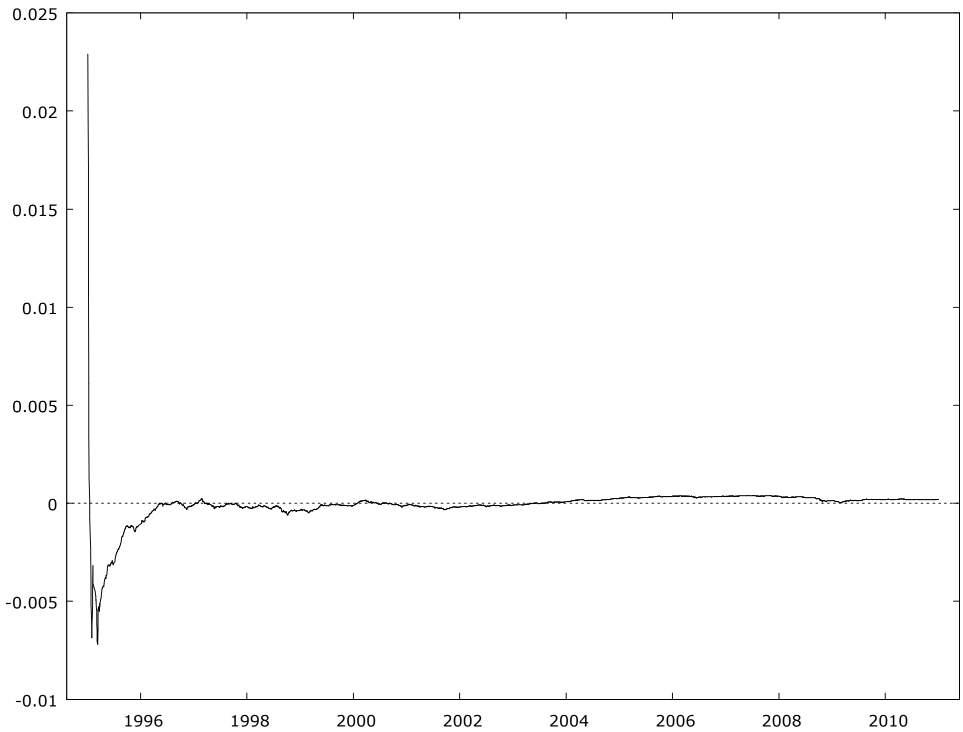

Figure 6 Cumulative mean of PX returns. Source: author's calculations.

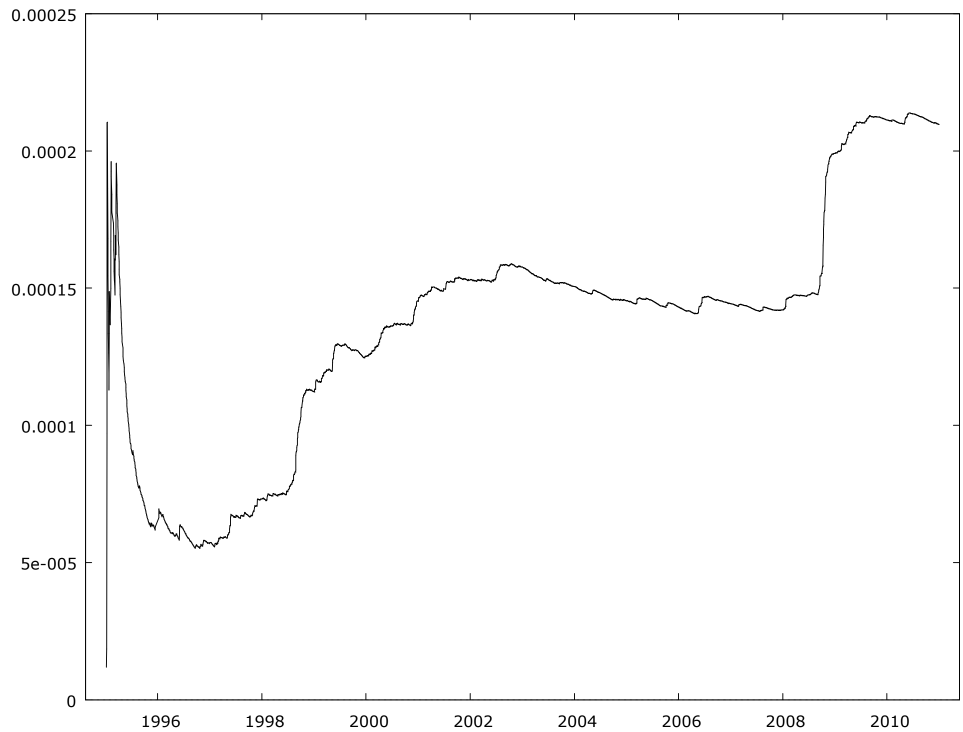

Figure 7 Cumulative variance of PX returns. Source: author's calculations.

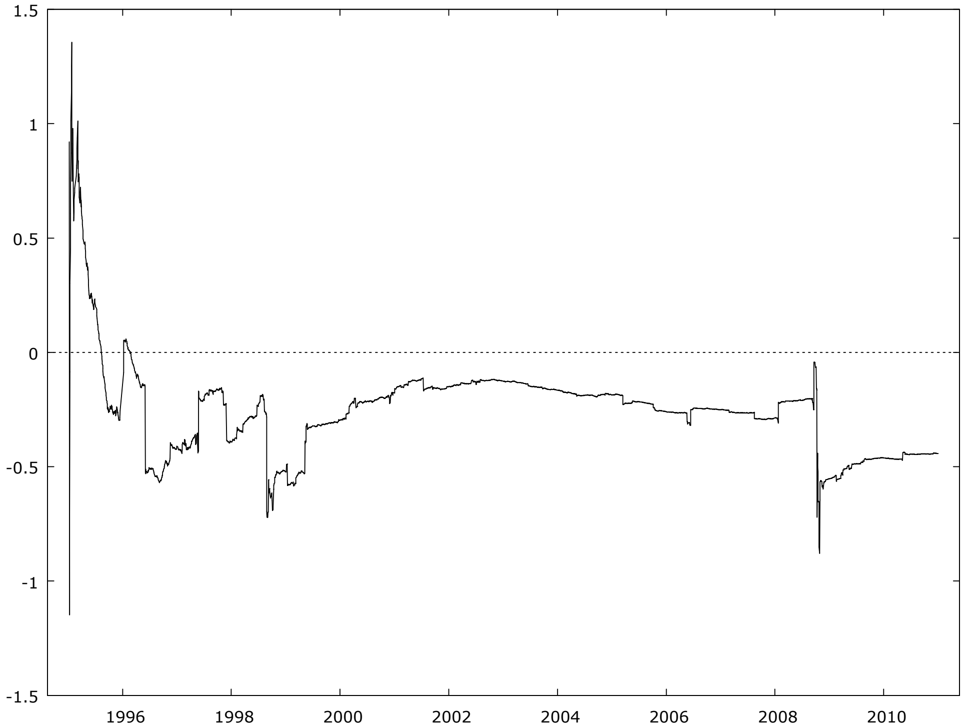

Figure 8 Cumulative skewness of PX returns. Source: author's calculations.

Figure 9 Cumulative kurtosis of PX returns. Source: author's calculations.

2.2 PARAMETRIC APPROACHES TO TAIL ESTIMATION

As suggested by their name, the parametric methods of tail estimation are based on fitting a theoretical statistical distribution to certain type of extreme events. Two methods will be introduced, each with its respective theoretical background. The main difference between them lies in the type of extreme events they are concerned with. While the Block Maxima method draws on information contained in extreme observations from temporal blocks of data, the Peaks over Threshold method relies on information embedded in tail observations exceeding a predetermined high threshold.

2.2.1 THE BLOCK MAXIMA METHOD

The Block Maxima method has its theoretical cornerstone in the work of Fisher and Tippett (1928) complemented by Gnedenko (1943). Their findings are summarized in the so-called Fisher-Tippett Theorem, also known as the Extreme Value Theorem or the First Theorem in Extreme Value Theory. In a sense, it is an analogue to the classical Central Limit Theorem (CLT), which says that, regardless of underlying distribution, the limiting distribution of the sample mean is the normal distribution. In a similar fashion, the Fisher-Tippet Theorem states that the limiting distribution of the sample maximum is an extreme value distribution.

To better illustrate the analogy between CLT and the Fisher-Tippett Theorem, CLT can be re-formulated in the following way:

(Re-formulated) Central Limit Theorem: *Let* $X_1, X_2, X_3, ...$ *be a sequence of IID random variables drawn from an unknown distribution with mean* μ *and variance* $0 < \sigma^2 < \infty$ *and let* $S_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ *be a sequence of partial means. If we normalize this sequence of partial sums with coefficients* $a_n = \frac{\sigma}{\sqrt{n}}$ *and* $b_n = \mu$, *then the resulting statistic*

$$Z_n = \frac{S_n - b_n}{a_n} \xrightarrow{d} N(0,1) . \qquad (2.2)$$

In words, CLT says that the limiting distribution of the normalized sequence of partial means is the standard normal distribution. But what happens if one is not interested in the behaviour of averages of the underlying statistical process but rather in its extremes? Let $M_n = \max(X_1,...,X_n)$ be the maximum of a subsample of n observations. If the variables $X_1, X_2, X_3,...$ are IID with a distribution function F, then the exact distribution of the statistic M_n can be immediately written as a function of the underlying distribution F and the subsample size n

$$F_{M_n}(x) = (F(x))^n . \tag{2.3}$$

Obviously, the limiting distribution of M_n as $n \to \infty$ is either nil (for x smaller than the maximum of the sample) or equal to one (for x larger than the maximum of the sample). This result is not very exciting since the exact limiting distribution is degenerate. Moreover, the distribution of the parent variable is typically unknown and, consequently, so is the exact distribution of the extremes. For theoretical purposes as well as for practical ones, it is the asymptotic behaviour of the extremes which is of much more interest. That is the area where the Fisher-Tippett Theorem provides guidance.

Fisher-Tippett (and Gnedenko) Theorem: *Let $X_1, X_2, X_3,...$ be a sequence of IID random variables*[5] *and let $M_n = \max(X_1,...,X_n)$ be the maximum of the first n terms. If there are constants $\sigma_n > 0$ and μ_n and some non-degenerate distribution function H such that*

$$Z_n = \frac{M_n - \mu_n}{\sigma_n} \xrightarrow{d} H \tag{2.4}$$

then H belongs to one of the three standard extreme value distributions:

[5] The Fisher-Tippett Theorem applies to IID observations, but the GEVD may be shown (c.f. Embrechts et al. 1997) to be the correct limiting distribution for maxima computed from stationary time series in general, including stationary GARCH-type processes.

$$Fréchet:\ \phi_\alpha(z)=\begin{cases}0, & z\le 0\\ \exp\left(e^{-z^{-\alpha}}\right), & z>0\end{cases}$$

$$Weibull:\psi_\alpha(z)=\begin{cases}\exp\left(-\left(-z^{\alpha}\right)\right), & z\le 0\\ 1, & z>0\end{cases}$$

$$Gumbel:\Lambda(z)=\exp\left(-e^{-z}\right),\ z\in\mathbb{R}\ ,$$

where $\alpha>0$.

In words, the Fisher-Tippett Theorem says that the limiting distribution of the standardized sequence of sample maxima is one of the extreme value distributions. The parallel between Fisher-Tippett and CLT is nearly perfect.

The Fréchet, Weibull and Gumbel distributions can be re-written in terms of a single-parameter (ξ) family:

$$H_\xi(z)=\begin{cases}\exp\left(-\left(1+\xi z\right)^{-\frac{1}{\xi}}\right), & if\ \xi\neq 0\\ \exp\left(-e^{-z}\right), & if\ \xi=0\end{cases} \tag{2.5}$$

where x is such that $1+\xi x>0$. The distribution function $H_\xi(z)$ is called the Generalized Extreme Value Distribution (GEVD). This representation is obtained from the Fréchet distribution by setting $\xi=\alpha^{-1}$, from the Weibull distribution by setting $\xi=-\alpha^{-1}$ and by interpreting the Gumbel distribution as the limit case for $\xi=0$. Figure 10 and Figure 11 show examples of PDFs and CDFs of the respective extreme value distributions.

The Fisher-Tippett Theorem implies that to infer on the distribution of extreme events, knowledge of the exact parametric form of the marginal distribution F is not necessary. The value of ξ depends solely on the tail behaviour of F.

A distribution F is said to belong to the Maximum Domain of Attraction (MDA) of an extreme value distribution H_ξ, denoted by $F\in D_M\left(H_\xi\right)$, if $M_n=\max\left(X_1,...,X_n\right)$ satisfies Equation (2.4), where X_i's are random variables with distribution F.

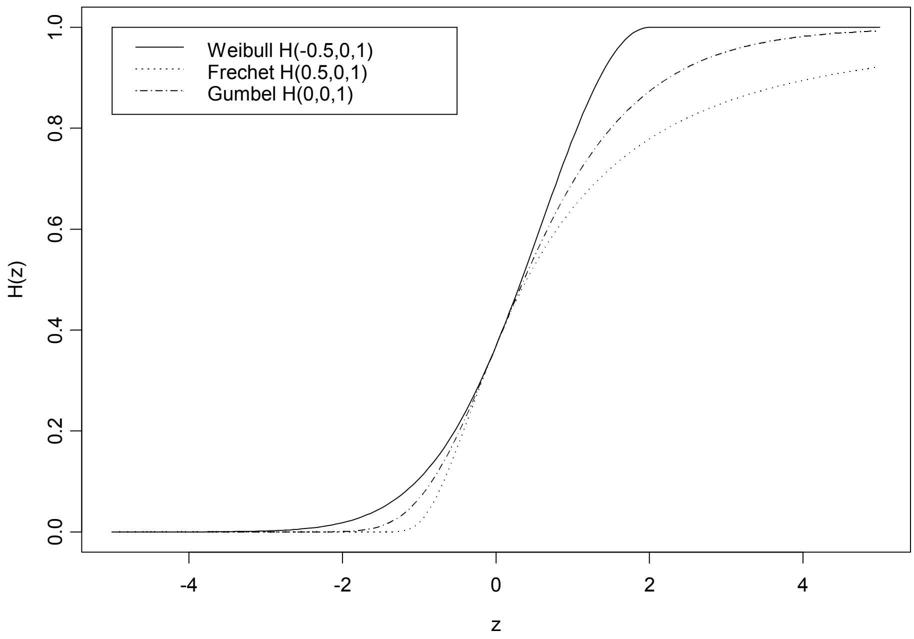

Figure 10 GEVD CDFs H(ξ, μ, σ) for Fréchet, Weibull and Gumbel cases. Source: Zivot and Wang (2006, p. 146).

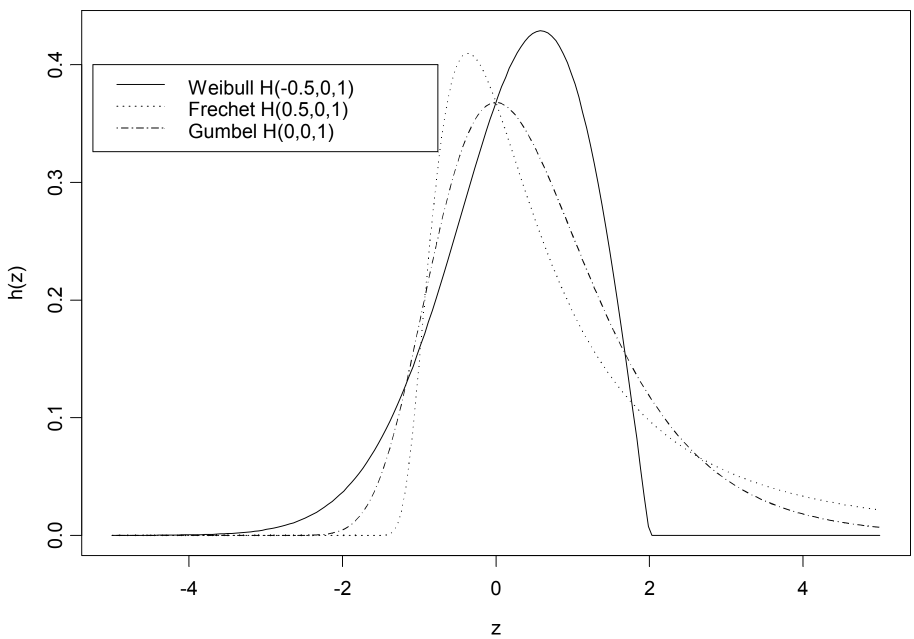

Figure 11 GEVD PDFs h(ξ, μ, σ) for Fréchet, Weibull and Gumbel cases. Source: Zivot, Wang (2006, p. 146).

Distributions with finite support, such as beta and uniform distributions, have $\xi < 0$ and belong to MDA of the Weibull distribution. Distributions with tails governed by power-law with exponent α fall into the Fréchet class and $\xi = \alpha^{-1} > 0$. Pareto, Burr and log-gamma distributions are examples of distributions from MDA of the Fréchet distribution. These are the heavy-tailed distributions. The Gumbel distribution can be regarded as a transitional limiting form between the Fréchet and the Weibull distributions. For small values of ξ, the latter distributions are very close to the Weibull. For instance, normal, log-normal and exponential distributions belong to MDA of the Gumbel distribution (Longin 1996).

To be formally correct, the necessary and sufficient first order condition for $F \in D_M\left(H_\xi\right)$ has to be mentioned:

$$F \in D_M\left(H_\xi\right) \Leftrightarrow \lim_{t\to\infty} \frac{U(tx)-U(t)}{a(t)} = \begin{cases} \dfrac{x^\xi - 1}{\xi} & if \quad \xi \neq 0 \\ \ln x & if \quad \xi = 0 \end{cases} \tag{2.6}$$

for every $x > 0$ and some positive measurable function a, with U standing for a quantile type function associated to F by

$$U(t) = \inf\left\{x : F(x) \geq 1 - \frac{1}{t}\right\}. \tag{2.7}$$

For heavy-tailed models, i.e., distributions $F \in D_M\left(H_\xi\right)$ with $\xi > 0$, $a(t)$ in Equation (2.6) may be picked such that $a(t) = \xi U(t)$. Then for every $x > 0$

$$F \in D_M\left(H_\xi\right) \Leftrightarrow \lim_{t\to\infty} \frac{U(tx)}{U(t)} = x^\xi, \tag{2.8}$$

i.e., U is of regular variation with index ξ (Gomes et al. 2008).

The Fisher-Tippett Theorem has been extended to time series. The same result stands if the variables are correlated (the sum of squared correlation coefficients remaining finite). Various processes based on the normal distribution, e.g., auto-regressive processes with normal disturbances, discrete mixtures of normal distributions and mixed diffusion jump processes, have thin tails so that they lead to the Gumbel distribution for the extremes. The maxima of ARCH processes follow the Fréchet distribution (see Longin (1996) and references therein).

The GEVD characterizes the limiting distribution of the standardized maxima. It turns out that the GEVD is invariant to location and scale transformations such that for location and scale parameters μ and $\sigma > 0$

$$H_{\xi}(z) = H_{\xi}\left(\frac{x-\mu}{\sigma}\right) = H_{\xi,\mu,\sigma}(x) \tag{2.9}$$

(Zivot and Wang 2006).

The Fisher-Tippett Theorem may then be re-interpreted as follows. For n large enough

$$\Pr(Z_n < z) = \Pr\left(\frac{M_n - \mu_n}{\sigma_n} < z\right) \approx H_{\xi}(z). \tag{2.10}$$

Letting $x = \sigma_n z + \mu_n$ then yields

$$\Pr(M_n < x) \approx H_{\xi,\mu,\sigma}\left(\frac{x-\mu_n}{\sigma_n}\right) = H_{\xi,\mu_n,\sigma_n}(x), \tag{2.11}$$

which is used in practical applications to make inferences about the maximum M_n.

The GEVD is characterised by three parameters: a shape parameter $\xi = \alpha^{-1}$, a location parameter μ_n and a scale parameter σ_n). All three parameters have their interpretation. The location parameter μ_n indicates where extremes are located on average. The scale parameter σ_n indicates the extent to which extreme realizations are dispersed. The shape parameter ξ describes the fatness of the tail; the higher the parameter, the more probability mass there is in the tail. Figure 12 illustrates this point.

The shape parameter ξ is an intrinsic parameter of the underlying process and does not depend on the number of observations n the maximum is selected from. As n increases, the value of the tail index should settle down around a particular value. At the same time the absolute value of the location parameter μ should rise, since extremes selected over longer periods are inevitably larger. There is little to say about the behaviour of the scale parameter σ ex ante as the distribution of extreme observations may contract or expand (Longin 1996).

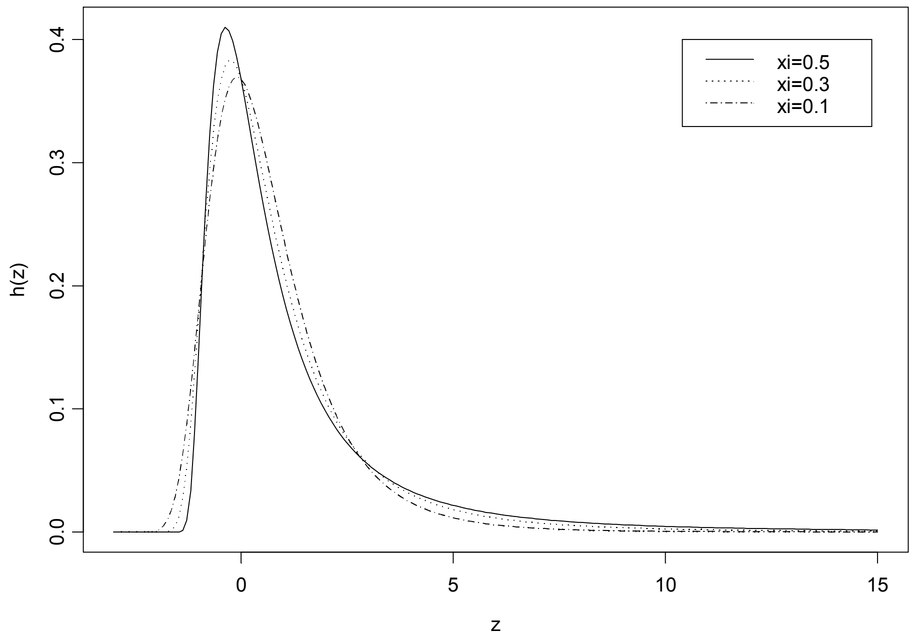

Figure 12 The Frechét PDFs with different shape parameters. Source: author's calculations.

The parameters of the GEVD can be estimated by the parametric Maximum Likelihood (ML) estimation method. Let $x_1,...,x_N$ be a sample of IID realizations of size N with unknown CDF F and let M_N denote the sample maximum. For inference on M_N using Equation (2.11), the parameters ξ, σ_N and μ_N have to be estimated. Since there is only one value of M_N, it is not possible to form a likelihood function for ξ, σ_N and μ_N. However, if one is interested in the maximum M_n over a large finite subsample or block of size $n < N$, then a sub-sampling method can be used to form the likelihood function for the parameters ξ, σ_n and μ_n of the GEVD for M_n.

The practical procedure is as follows. The sample is divided into m non-overlapping blocks of essentially equal size $n = N/m$

$$\left[x_1,...,x_n \left| x_{n+1},...,x_{2n} \right| ... \left| x_{mn-n+1},...,x_{mn} \right. \right]$$

and M_n^j is defined as the maximum value of x_i in block $j = 1,...,m$. The likelihood function for the parameters ξ, σ_n and μ_n of the GEVD is then constructed from the sample of block maxima $\left\{ M_n^1,...,M_n^m \right\}$. It is assumed that the block size n is sufficiently large so that the Fisher-Tippett Theorem holds.

The log-likelihood function, assuming IID observations from a GEVD with $\xi \neq 0$, can be written as

$$l(\mu,\sigma,\xi) = -m\ln(\sigma) - \left(1+\frac{1}{\xi}\right)\sum_{j=1}^{m}\ln\left[1+\xi\left(\frac{M_n^j-\mu}{\sigma}\right)\right] - \sum_{j=1}^{m}\left[1+\xi\left(\frac{M_n^j-\mu}{\sigma}\right)\right]^{-1/\xi} \tag{2.12}$$

such that

$$1+\xi\left(\frac{M_n^j-\mu}{\sigma}\right) > 0\ . \tag{2.13}$$

If $\xi = 0$, the log-likelihood function is

$$l(\mu,\sigma) = -m\ln(\sigma) - \sum_{j=1}^{m}\left(\frac{M_n^j-\mu}{\sigma}\right) - \sum_{j=1}^{m}\exp\left[-\left(\frac{M_n^j-\mu}{\sigma}\right)\right]. \tag{2.14}$$

For $\xi > -0.5$, the ML estimates of μ, σ and ξ are consistent and asymptotically normal with asymptotic variance given by the inverse of the observed information matrix.

The finite sample properties of ML estimates depend on the number of blocks m and the block size n. The bias of the estimate decreases with increasing block size n and the variance of the estimate decreases with increasing number of blocks m. Obviously, given a sample of observations, one cannot increase the size of blocks (and thus reduce the bias) and at the same time increase the number of blocks (and thus reduce the uncertainty). For details on ML estimation see Zivot and Wang (2006) and references therein.

In Table 3 and Table 4 ML estimates of the parameters for the left and the right tail, respectively, for four different values of n are reported. Overall, there is not much to be drawn from these results in terms of the key parameter ξ as the estimates are hindered by high standard errors. In case of the left tail, only one estimate is statistically significant. The right tail estimates ended up better, with three of them being significant. The trouble is that for both tails it is the estimates based on smaller blocks that are significant – unfortunately, these estimates also carry the most severe bias whose build-up with decreasing size of blocks is well-documented by the results. Despite their statistical significance, it is unlikely that the tail index would reach values exceeding four.

This is consistent with findings of Longin (1996) who performed similar analysis on much larger sample containing 106 years of U.S. stock market daily data. Based on results of a goodness-of-fit test, he concludes that while the actual distribution of extremes converges relatively quickly to GEVD as the size of blocks increases, sufficiently reliable estimates are achieved only for extremes selected over periods longer than a semester.

Looking at the other two parameters, the estimates of location parameter μ are in line with the intuition that it should increase with the size of blocks. The estimates of scale parameter σ are then relatively stable for both tails with the exception of monthly blocks where higher dispersion of extremes is evident. All estimates of μ and σ are significant.

Size of block	Year	Semester	Quarter	Month
No. of blocks	16	32	64	192
Xi	0.2800	0.2065	0.1552	0.1485**
SE	0.1990	0.1476	0.0930	0.0530
Alpha	3.5714	4.8426	6.4433	6.7340
Sigma	0.0125**	0.0129**	0.0126**	0.0099**
SE	0.0029	0.0021	0.0014	0.0006
Mu	0.0459**	0.0343**	0.0257**	0.0168**
SE	0.0035	0.0026	0.0018	0.0008

Table 3 ML estimates of GEVD parameters of the left tail of PX returns. Source: author's calculations. (* and ** denote estimates significant at 5% and 1% level, respectively)

Size of block	Year	Semester	Quarter	Month
No. of blocks	16	32	64	192
Xi	0.4330	0.3519*	0.2011*	0.1658**
SE	0.3372	0.1655	0.0836	0.0510
Alpha	2.3095	2.8417	4.9975	6.0314
Sigma	0.0118**	0.0104**	0.0109**	0.0086**
SE	0.0034	0.0018	0.0012	0.0005
Mu	0.0351**	0.0293**	0.0248**	0.0175**
SE	0.0037	0.0021	0.0015	0.0007

Table 4 ML estimates of GEVD parameters of the right tail of PX returns. Source: author's calculations. (* and ** denote estimates significant at 5% and 1% level, respectively)

To check the fit of GEVD to underlying data, two plots of residuals, a scatter plot and a QQ-plot using the exponential distribution as the reference distribution, may be used. If the estimated model is correct, crude residuals, defined as

$$W_i = \left(1 + \hat{\xi}\,\frac{M_n^{(i)} - \hat{\mu}}{\hat{\sigma}}\right)^{-\frac{1}{\hat{\xi}}}, \tag{2.15}$$

will be IID unit exponentially distributed random variables (Zivot and Wang 2006). In such case, the scatter plot will not reveal any significant non-modelled trend in the data and the QQ-plot will follow linear pattern.

Figure 13 through Figure 16 show QQ-plots and scatter plots of the residuals of fitted GEVDs in Table 3 and Table 4. QQ-plots of the left tail look better than their right tail counterparts which exhibit substantial departures from the linear trend at some points. The scatter plots then do not reveal any clear-cut evidence of non-modelled pattern but to say they are evenly distributed would be an over-statement.

2.2.2 The Peaks over Threshold Method

In contrast with the Block Maxima method, the Peaks over Threshold (POT) method makes a better use of chronically limited empirical material on extreme events. While the Block Maxima method neglects, regardless of their magnitude, all observations in a block other than the maximal one, the POT method gathers information from all observations surpassing certain predetermined high threshold in the entire sample, an approach shared with semi-parametric estimators which will be introduced later.

The theoretical background underpinning this method is an extension of the Fisher-Tippett Theorem elaborated by Pickands (1975) and Balkema and de Haan (1974). Just like the mean statistic converges in distribution to the normal distribution and the maximum statistic converges in distribution to one of the extreme value distributions, the exceedances converge in distribution to the Generalized Pareto Distribution (GPD), provided the threshold is set high enough. This is formalized in the following theorem, also known as the Second Theorem in Extreme Value Theory.

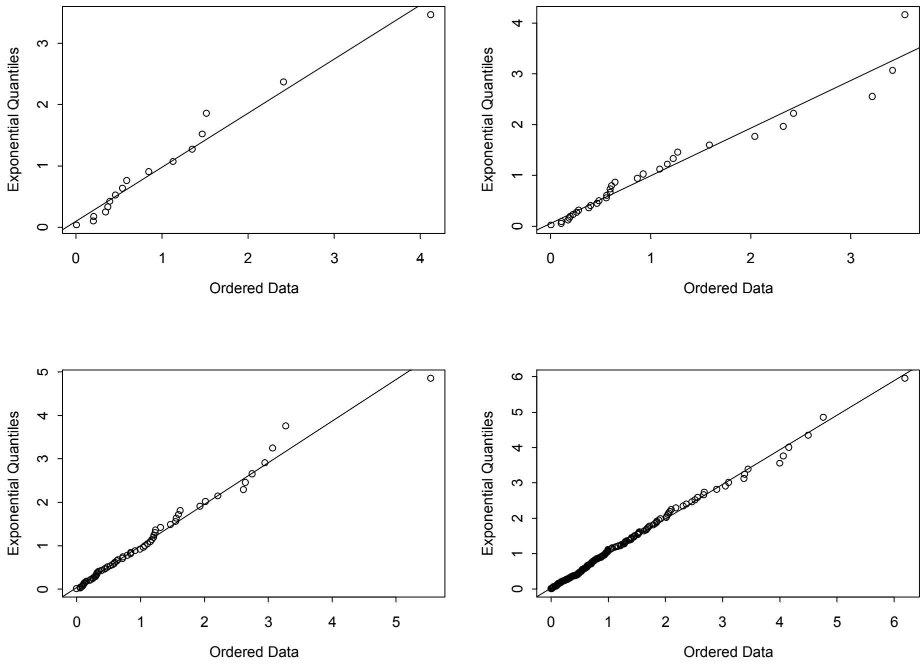

Figure 13 The QQ-plots of residuals of GEVD fitted to yearly (upper left), semesterly (upper right), quarterly (lower left) and monthly (lower right) maxima of negative daily PX returns (left tail). Source: author's calculations.

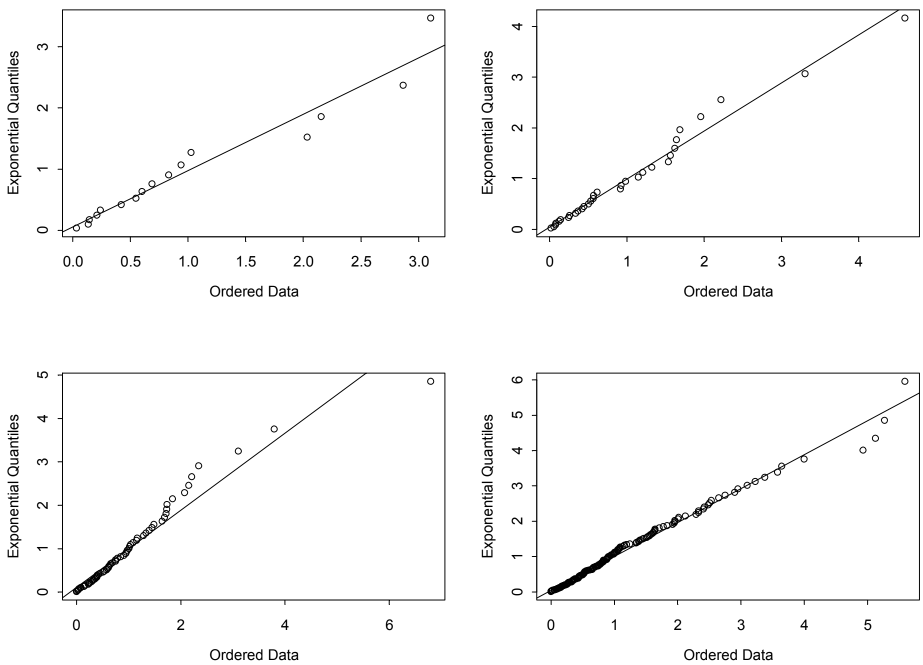

Figure 14 The QQ-plots of residuals of GEVD fitted to yearly (upper left), semesterly (upper right), quarterly (lower left) and monthly (lower right) maxima of daily PX returns (right tail). Source: author's calculations.

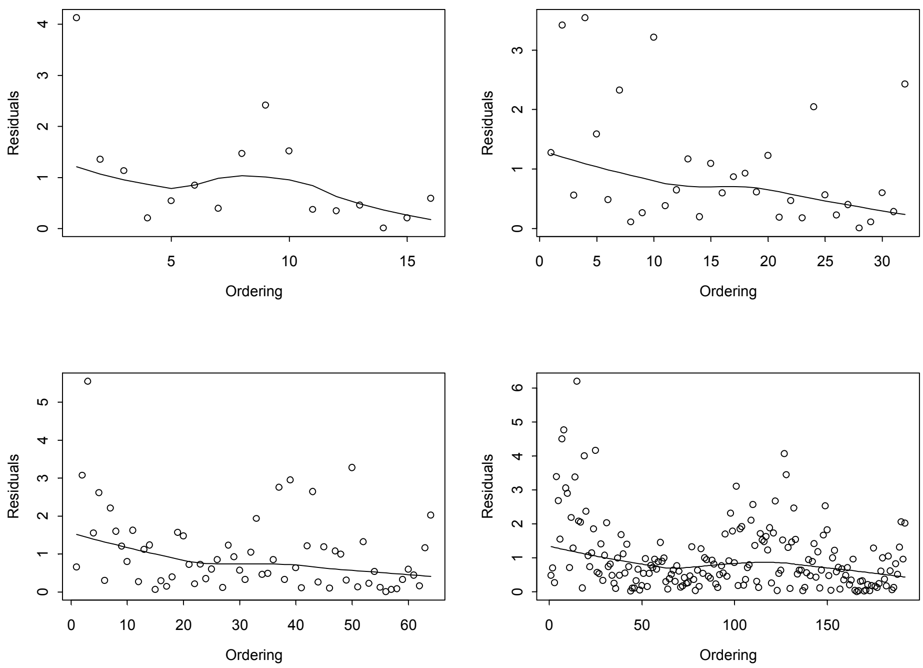

Figure 15 The scatter plots of residuals of GEVD fitted to yearly (upper left), semesterly (upper right), quarterly (lower left) and monthly (lower right) maxima of negative daily PX returns (left tail). Source: author’s calculations.

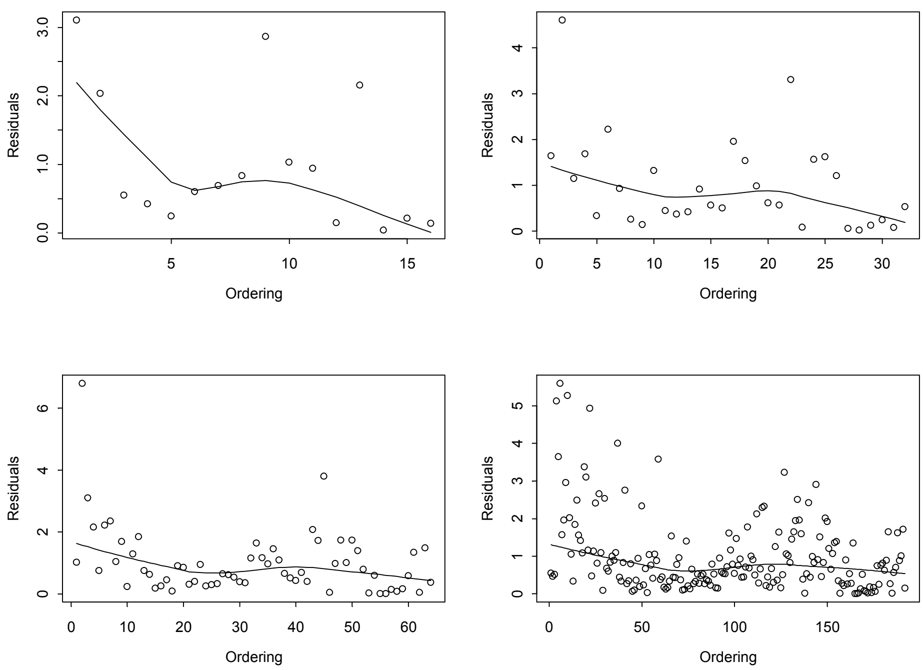

Figure 16 The scatter plots of residuals of GEVD fitted to yearly (upper left), semesterly (upper right), quarterly (lower left) and monthly (lower right) maxima of daily PX returns (right tail). Source: author’s calculations.

Pickands, Balkema and de Haan Theorem: *Let* $X_1, X_2, X_3, \ldots$ *be a sequence of IID random variables drawn from an unknown distribution* F *and let* u *be a predetermined threshold. Let* F_u *denote the conditional distribution function of the exceedance* $Y = X - u$ *given that* X *exceeds* u

$$F_u(y) = \Pr\left(X - u \le y \middle| X > u\right) = \frac{F(y+u) - F(u)}{1 - F(u)}. \qquad (2.16)$$

For the class of underlying distribution functions F *such that the CDF of the standardized values of* M_n *converges to a GEVD and for* u *large enough there exists a positive function* β_u *such that the excess distribution* $F_u(y)$ *is well approximated by the generalized Pareto distribution* $G_{\xi,\beta_u}(y)$ *where*

$$G_{\xi,\beta_u}(y) = \begin{cases} 1 - \left(1 + \dfrac{\xi}{\beta_u} y\right)^{-\frac{1}{\xi}}, & if\ \xi \ne 0 \\ 1 - \exp\left(-\dfrac{y}{\beta_u}\right), & if\ \xi = 0 \end{cases} \qquad (2.17)$$

for $y \in \left[0, (\overline{x} - u)\right]$ *if* $\xi \ge 0$ *and* $y \in \left[0, -\dfrac{\beta_u}{\xi}\right]$ *if* $\xi < 0$, *where* $\overline{x} \le +\infty$ *denotes the rightmost point of the distribution function* F.

The GPD is determined by two parameters: ξ is a shape parameter and β_u is a scaling parameter. The GPD is generalized in the sense that it subsumes three types of distributions under a common parametric form. For $\xi > 0$, GPD is a re-parameterized version of the ordinary Pareto distribution. For $\xi < 0$, GPD becomes the Lomax distribution, also known as the Pareto Type II distribution. And finally, setting $\xi = 0$ yields exponential distribution. Figure 17 and Figure 18 show examples of CDFs and PDFs of the three types of GPD.

As is apparent from the theorem, the limiting GEVD for block maxima and the limiting GPD for threshold exceedances are intimately inter-connected. For a given value of u, the parameters ξ, μ and σ of the GEVD determine the parameters ξ and β of GPD. In particular, the shape parameter ξ of the GEVD is identical to the shape parameter ξ in the GPD. Consequently, if $\xi < 0$ then F belongs to MDA of the Weibull distribution and $G_{\xi,\beta}$ is the Lomax distribution, also called Pareto Type II distribution; if $\xi > 0$ then F belongs to MDA of the Fréchet distribution and $G_{\xi,\beta}$ is the Pareto distribution; and finally in the limiting case where $\xi = 0$ F belongs to MDA of the Gumbel distribution and $G_{\xi,\beta}$ is the exponential distribution.

Regarding the variation of ξ and β_u with respect to u, it has been already mentioned in the preceding subsection that the shape parameter ξ is intrinsic to the underlying process and therefore independent of u. In case of β_u, consider a limiting GPD with scale parameter β_{u_0} for an excess distribution F_{u_0} with threshold u_0. For an arbitrary threshold $u_1 > u_0$, the excess distribution F_{u_1} has a limiting GPD with scale parameter $\beta_{u_1} = \beta_{u_0} + \xi(u_1 - u_0)$ (Zivot and Wang 2006).

The Pickands, Balkema and de Haan Theorem can be simplified into a statement that for a large class of underlying distributions, a function β_u can be found such that

$$\lim_{u \to \bar{x}} \sup_{0 \le y < \bar{x} - u} \left| F_u(y) - G_{\xi,\beta_u}(y) \right| = 0. \tag{2.18}$$

In words, it says that for a certain threshold u, the excess distribution may be thought of as to be exactly GPD with certain values of ξ and β. The only problem is to find the proper threshold. In principle, there has to be a compromise in choosing a threshold which will be, on one hand, sufficiently high so that the asymptotic theorem can be considered essentially valid and, on the other hand, sufficiently low so that there is enough data left for the estimation.

One possible way of finding optimal threshold is to exploit the mean excess function of the GPD which has a very convenient form of a straight line with positive slope

$$e(u) = \frac{\sigma + \xi u}{1 - \xi} \quad , \tag{2.19}$$

where $\sigma + \xi u > 0$ (Schirmacher et al. 2005). Hence, when plotted as a function of threshold u, the empirical mean excesses should follow an upward sloping linear trend down to a point where the GPD ceases to be a suitable model.

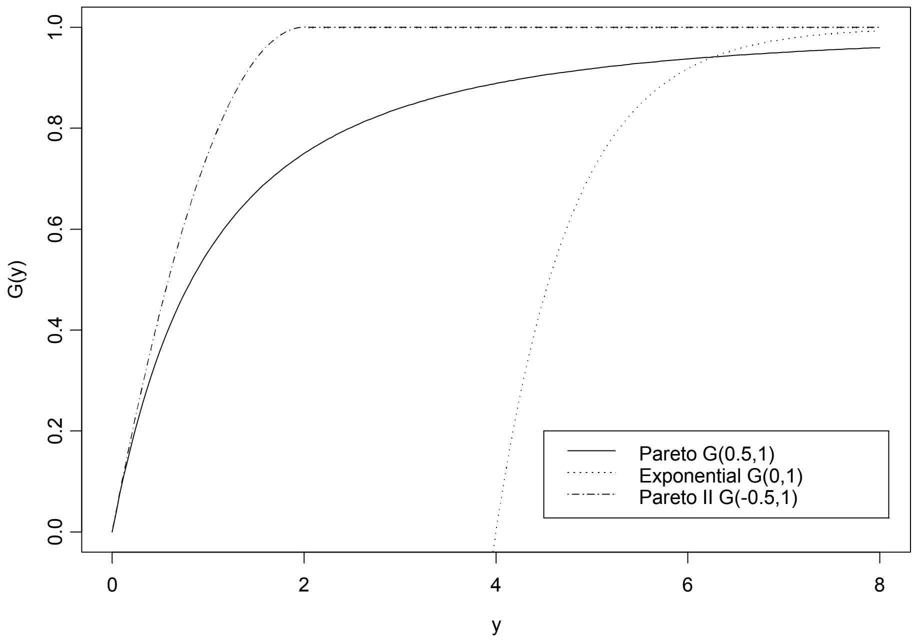

Figure 17 CDFs G(ξ, β) of the three types of the GPD. Source: Zivot and Wang (2006, p. 161).

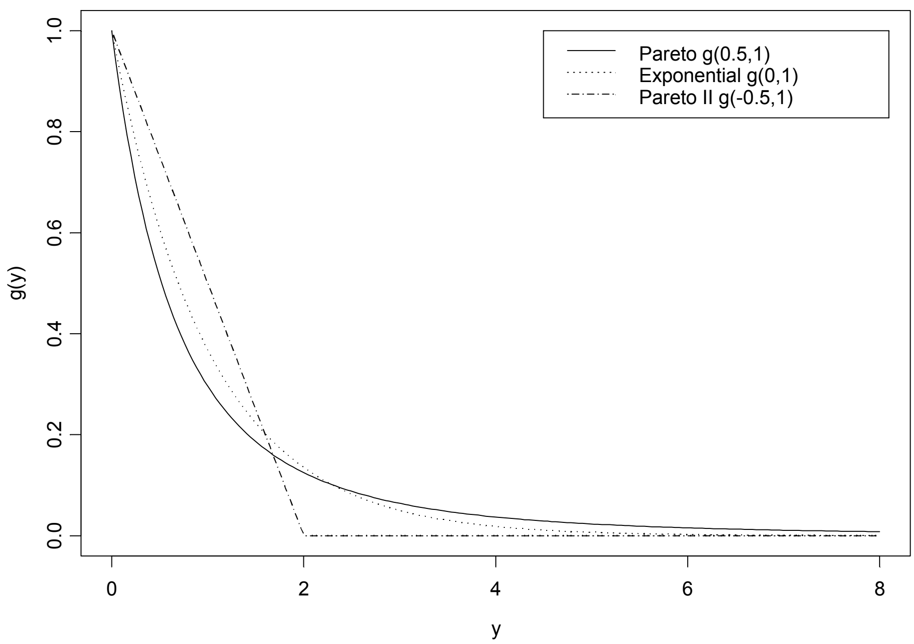

Figure 18 PDFs g(ξ, β) of the three types of the GPD. Source: Zivot and Wang (2006, p. 161).

Figure 19 shows the enlarged areas of interest from the mean excess plots shown in Figure 4 and Figure 5. In case of the left tail, there are several relatively mild kinks but in general the pattern is linear and any threshold between 0.0175 and 0.03 seems to be reasonable. The right tail proves more challenging as the upward trend is broken down at approximately 0.05 and the pattern seems to be slightly concave. Nevertheless, a region of similar threshold values to that of the left tail can be considered plausible but we cannot expect the fit of GPD to be perfect.

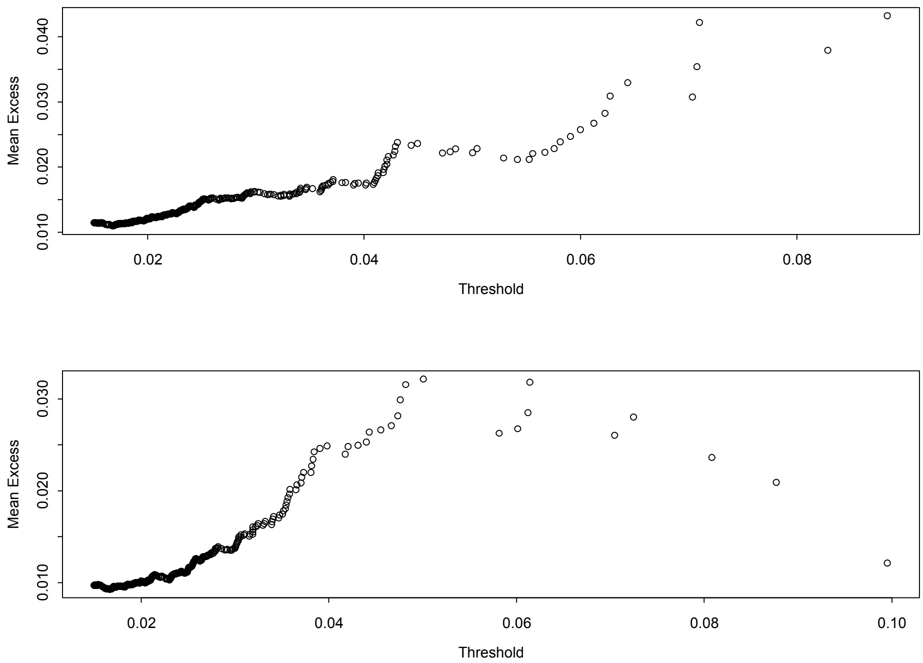

Figure 19 The mean excess plot of the left (upper plot) and the right tail (lower plot) of PX returns. Source: author's calculations.

An alternative way of finding the ideal number of order statistics is to plot the estimates of the shape parameter ξ as a function of number of observations exceeding u and then look for a region where these estimates are stable. Figure 20 and Figure 21 indicate the estimates of ξ are particularly stable in the region between 0.017 and 0.02 for either tail. Let us arbitrarily choose 0.019, corresponding to 274 and 265 observations from the left and the right tail, respectively.

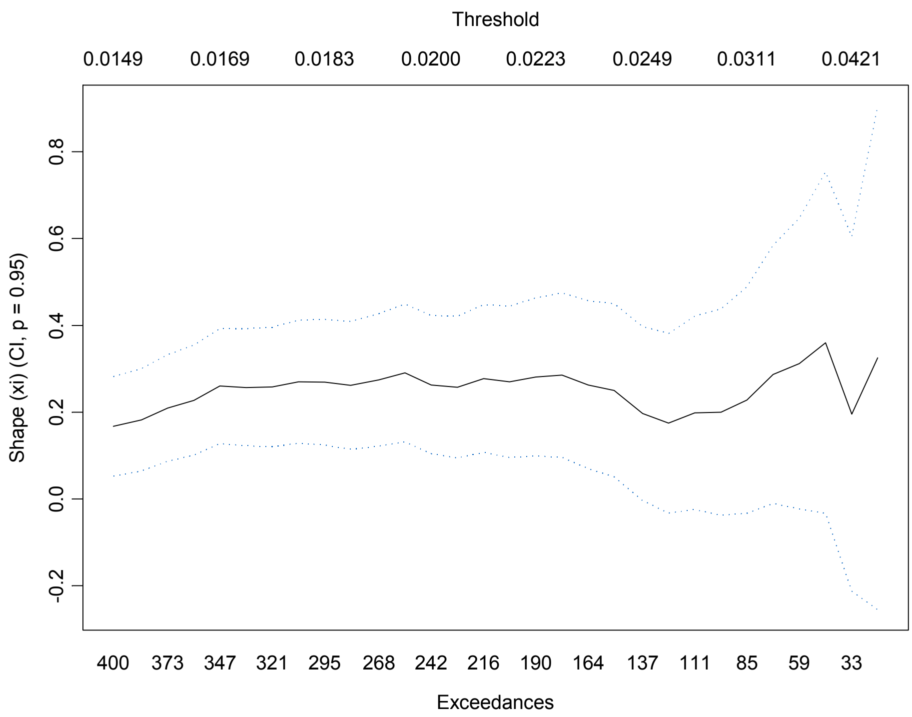

Figure 20 The ML estimates of shape parameter ξ of the left tail of PX returns based on fitting the GPD as a function of a number of exceedances. Source: author's calculations.

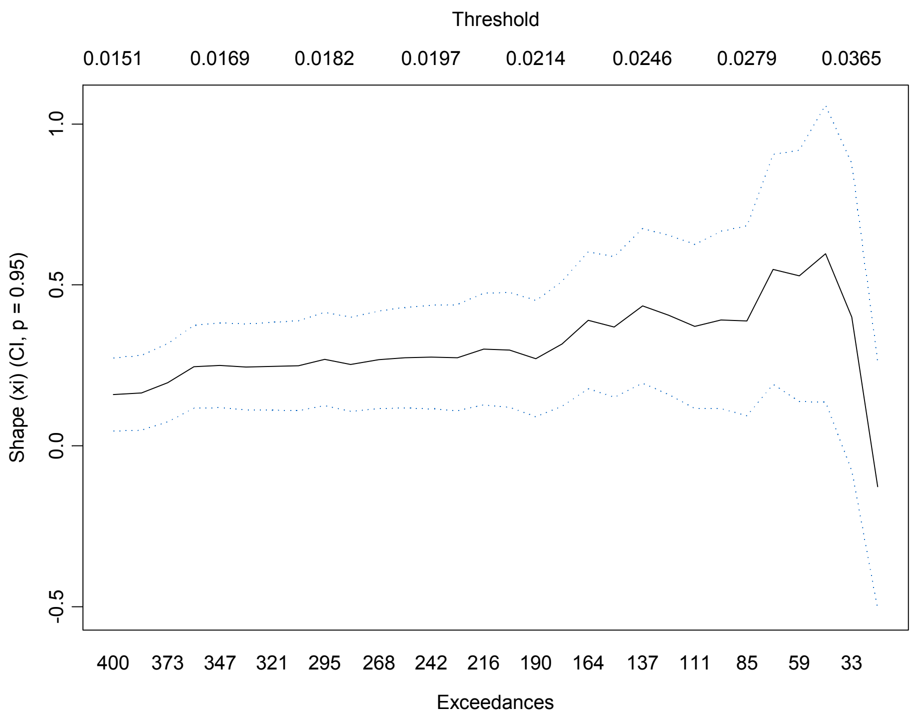

Figure 21 The ML estimates of shape parameter ξ of the right tail of PX returns based on fitting the GPD as a function of a number of exceedances. Source: author's calculations.

Once the optimal threshold u is determined, the GPD parameters can be estimated. Similarly to the Block Maxima method, the Maximum Likelihood estimation method can be applied. Let $x_1,...,x_n$ be an IID sample of returns with unknown CDF F and let $x_{(1)},...,x_{(n)}$ be the re-ordered sample in a way that $x_{(1)} \geq x_{(2)} \geq ... \geq x_{(n)}$. Let $k < n$ be a number of observations exceeding threshold u, i.e., $x_{(k)} \geq u > x_{(k+1)}$, and let $y_i = x_{(i)} - u$ be the excesses over the threshold for $i = 1,...,k$. The results of the Pickands, Balkema and de Haan Theorem imply that if u is large enough, then $\{y_1,...,y_k\}$ may be thought of as a random sample drawn from GPD with unknown parameters ξ and β_u. For $\xi \neq 0$, the log-likelihood function based on Equation (2.17) is

$$l(\xi, \beta_u) = -k \ln \beta_u - \left(1 + \frac{1}{\xi}\right) \sum_{i=1}^{k} \ln\left(1 + \frac{\xi y_i}{\beta_u}\right), \tag{2.20}$$

provided $y_i \geq 0$ when $\xi > 0$, and $0 \leq y_i \leq -\frac{\beta_u}{\xi}$ when $\xi < 0$. For $\xi = 0$, the log-likelihood function is

$$l(\beta_u) = -k \ln(\beta_u) - \frac{1}{\beta_u} \sum_{i=1}^{k} y_i \ . \tag{2.21}$$

For details, see Zivot and Wang (2006) and references therein.

Table 5 and Table 6 the ML estimates of the GPD parameters for the left and the right tail, respectively. Fixed fractions of data as well as the exceedances over the optimal threshold found earlier were used. Again, the estimates of ξ based on the lowest numbers of observations are not statistically significant. As the amount of data available increases, estimates tend to be more reliable but their values decline as the observations in the extreme tail regions lose their sway in favour of moderate observations from more central regions of the distribution.

For the optimal threshold, the POT method yields realistic estimates of the tail index of 3.59 for the left and 3.69 for the right tail. The significance of these estimates is highlighted by the fact that their t-statistics are the highest from all reported subsamples.

Number of observations	10%	5%	2.5%	1%	**6.85%**
	399	199	99	39	**274**
xi	0.1657**	0.2764**	0.2042	0.3604	**0.2784****
SE	0.0528	0.0942	0.1224	0.2548	**0.0792**
alpha	6.0350	3.6179	4.8972	2.7747	**3.5920**
beta	0.0095**	0.0093**	0.0124**	0.0124**	**0.0084****
SE	0.0007	0.0011	0.0019	0.0037	**0.0008**

Table 5 ML estimates of GPD parameters of the left tail of PX returns. Source: author's calculations. (* and ** denote estimates significant at 5% and 1% level, respectively)

Number of observations	10%	5%	2.5%	1%	**6.63%**
	399	199	99	39	**265**
xi	0.1595**	0.2882**	0.3817**	0.6059	**0.2709****
SE	0.0489	0.0842	0.1357	0.3259	**0.0726**
alpha	6.2696	3.4698	2.6199	1.6504	**3.6914**
beta	0.0081**	0.0074**	0.0081**	0.0091**	**0.0071****
SE	0.0006	0.0008	0.0013	0.0032	**0.0007**

Table 6 ML estimates of GPD parameters of the right tail of PX returns. Source: author's calculations. (* and ** denote estimates significant at 5% and 1% level, respectively)

In a fashion similar to the Block Maxima method, the fit of the GPD to the underlying data can be assessed with a scatter plot and a QQ-plot of residuals. Figure 22 contains the respective plots for both tails. Residuals in both scatter plots are distributed relatively evenly. The QQ-plots suggest that the fit of the left tail is markedly better than the fit of the right one; this comes as no surprise given the shape of the mean excess function of the right tail.

2.3 SEMI-PARAMETRIC ESTIMATORS

Compared to the parametric methods of tail estimation, the family of semi-parametric estimators represents more modern developments in the field of statistics of extremes. However, the semi-parametric estimators are also grounded in the theoretical background of the Extreme Value Theory, since they are built on the elementary assumption that $F \in D_M\left(H_\xi\right)$, i.e., that the underlying distribution function F belongs to the Maximum Domain of Attraction of an Extreme Value Distribution H_ξ.

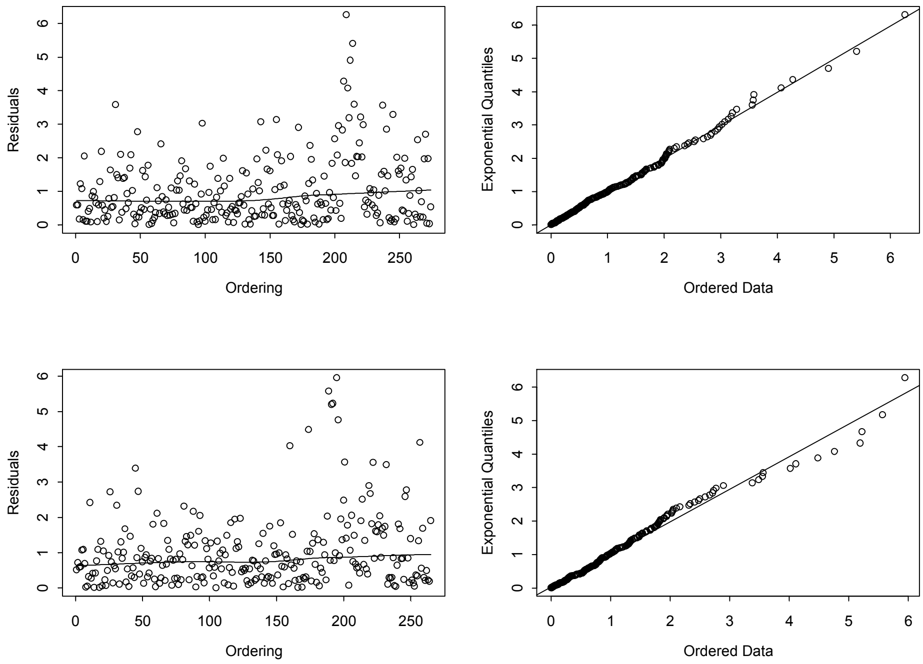

Figure 22 The scatter plot and the QQ-plot of residuals of GPD fitted to the left tail (upper plots) and the right tail (lower plots) of PX returns above the optimal threshold of 0.019. Source: author's calculations.

Naturally, estimates of shape parameter ξ, now called the Extreme Value Index (EVI),[6] are the most important but semi-parametric estimators of location and scale parameters were developed as well. The combined results may be used then to estimate complex distributional characteristics such as extreme quantiles, return periods etc. (see Gomes et al. 2008).

The intuition behind semi-parametric estimators of ξ, as presented by Kearns and Pagan (1997), is very simple. Let $x_1,...,x_n$ be the realizations of an underlying random variable X with Paretian tail behaviour. For $\alpha, c > 0$, the tail can be approximated by

$$\Pr\left(X > x\right) \approx cx^{-\alpha} \;\; as \;\; x \to \infty. \tag{2.22}$$

Taking logs yields

$$\log \Pr\left(X > x\right) = \log c - \alpha \log x \tag{2.23}$$

$$\log x = C - \xi \log \Pr\left(X > x\right), \tag{2.24}$$

[6] The Extreme Value Index is usually denoted by the Greek letter γ, but in order to keep the notation as simple as possible and consistent with the preceding sections we stick to the letter ξ.

where $C = \xi \log c$ and $\xi = \alpha^{-1}$ is the EVI. Now, let $x_{(1)}, ..., x_{(n)}$ be the sample re-ordered in a way that $x_{(1)} \geq x_{(2)} \geq ... \geq x_{(n)}$. Selecting a certain threshold u, $\Pr\left(X > x_{(j)}\right)$ in Equation (2.23) can be replaced by the empirical survivor function

$$\Pr\left(X > x_{(j)}\right) \approx \frac{j}{n}, \tag{2.25}$$

where j denotes the number of $x_{(i)}$'s such that $x_{(i)} > x_{(j)}$. Substituting into Equation (2.24) yields linear relationship in $\left[\log j; \log x_{(j)}\right]$ space with slope coefficient ξ

$$\log x_{(j)} = \left(C + \log n\right) - \xi \log j. \tag{2.26}$$

In general, two points are necessary and sufficient to pin the slope of a line down. A natural candidate is $x_{(1)}$, the largest tail observation. But how far into the centre of the distribution one should go to select the other point is unclear. Venturing too far will cause the linear approximation to lose it accuracy as the chosen observation will not come from the tail region. But reluctance to go far enough can distort the slope by resultant shortage of observations available for estimation. This problem will be thoroughly discussed a little later; for now, let us denote this point $x_{(k)}$. An obvious estimator of the slope is then

$$\hat{\xi}_{deHaan-Resnick} = \frac{\log x_{(1)} - \log x_{(k)}}{\log k - \log 1} = \frac{\log x_{(1)} - \log x_{(k)}}{\log k}. \tag{2.27}$$

This indeed is an estimator derived by de Haan and Resnick (1980) for IID random variables with distribution function F in the MDA of an extreme value distribution with tail index $0 < \alpha = \xi^{-1} < 2$. Unfortunately, this restriction on values of the tail index α makes this estimator inadequate for financial time series which typically exhibit tails thinner than that.

Pickands (1975) proposed a very general estimator

$$\hat{\xi}_{Pickands} = \frac{1}{\log 2}\left[\log\left(x_{(k/4)} - x_{(k/2)}\right) - \log\left(x_{(k/2)} - x_{(k)}\right)\right] \tag{2.28}$$

which is weakly consistent under the first order condition in Equation (2.6) for any value $\xi \in \mathbb{R}$ and for k intermediate, i.e., $k/n \to 0$ as $n \to \infty$ (Gomes et al. 2008). The universality of this estimator is, to a great extent, outweighed by its rather large asymptotic variance (Beirlant et al. 2004). Dekkers and de Haan (1989) show that

$$\sqrt{k}\left(\hat{\xi}_{Pickands}-\xi\right)\xrightarrow{d}N(0,\sigma_P^2), \tag{2.29}$$

where $\sigma_P^2=\xi^2\left(2^{2\xi+1}+1\right)\left[2\log 2\left(2^{\xi}-1\right)\right]^{-2}$. The distance between the respective order statistics $x_{(k/4)}$ and $x_{(k)}$ is troublesome as well. Unless a vast number of observations is available, k has to be fairly large to allow for reliable estimates but in that case $x_{(k)}$ possesses information about the distribution well away from the tails.

Apart from their individual downsides, the de Haan-Resnick and the Pickands estimators share one more weakness in that they gather information about the tails from only two and four order statistics, respectively. Statistical intuition suggests that an estimator based on more information, i.e., more order statistics, will be more reliable. Furthermore, as our results in the preceding subsections and literature survey indicate, we can restrict our attention to distributions in the MDA of the Fréchet Extreme Value Distribution, i.e., to distributions with $\xi>0$.

An estimator exploiting both these aspects was developed, as a conditional maximum likelihood estimator, by Hill (1975)

$$\hat{\xi}_{Hill}=\frac{1}{k-1}\sum_{i=1}^{k-1}\left(\log x_{(i)}-\log x_{(k)}\right). \tag{2.30}$$

The Hill estimator is well-established in financial applications, not least because for distributions with $\xi>0$ it is more efficient than the Pickands estimator (Longin 1996).

Under the same conditions as the Pickands estimator, the Hill estimator is weakly consistent and asymptotically normal

$$\sqrt{k}\left(\hat{\xi}_{Hill}-\xi\right)\xrightarrow{d}N\left(0,\xi^2\right) \tag{2.31}$$

(see de Haan and Resnick (1998) and references therein).

A little more intricate semi-parametric estimator was developed by Dekkers et al. (1989). It is called the Moment estimator and has the functional form

$$\hat{\xi}_{Moment}=M^{(1)}+1-\frac{1}{2}\left\{1-\frac{\left(M^{(1)}\right)^2}{M^{(2)}}\right\}^{-1}, \tag{2.32}$$

where

$$M^{(j)}=\frac{1}{k-1}\sum_{i=1}^{k-1}\left(\log x_{(i)}-\log x_{(k)}\right)^j,\ j\geq 1. \tag{2.33}$$

For $j=1$, Equation (2.33) is equivalent to the Hill estimator. Like the Pickands estimator, the Moment estimator is weakly consistent and asymptotically normal for all $\xi \in \mathbb{R}$

$$\sqrt{k}\left(\hat{\xi}_{Moment}-\xi\right)\xrightarrow{d}N\left(0,\sigma_M^2\right), \tag{2.34}$$

where

$$\sigma_M^2=\begin{cases}1+\xi^2 & \text{for } \xi \geq 0, \\ (1-\xi)^2(1-2\xi)\left\{4-8\dfrac{1-2\xi}{1-3\xi}+\dfrac{(5-11\xi)(1-2\xi)}{(1-3\xi)(1-4\xi)}\right\} & \text{for } \xi < 0.\end{cases} \tag{2.35}$$

For formal completeness, in the semi-parametric framework, apart from the first order condition in Equation (2.6), we have to include a second order condition, specifying the rate of convergence in Equation (2.6). It is common to assume the existence of a function A^*, not changing in sign and approaching zero as $t\to\infty$, such that

$$\lim_{t\to\infty}\frac{\dfrac{U(tx)-U(t)}{a(t)}-\dfrac{x^{\xi}-1}{\xi}}{A^*(t)}=H_{\xi,\rho}(x)=\frac{1}{\rho^*}\left(\frac{x^{\xi+\rho^*}-1}{\xi+\rho^*}-\frac{x^{\xi}-1}{\xi}\right) \tag{2.36}$$

for all $x>0$ where $\rho^*\leq 0$ is a second order parameter controlling the speed of convergence of maximum values, linearly normalized, towards the limit law in Fisher-Tippett Theorem. Then

$$\lim_{t\to\infty}\frac{A^*(tx)}{A^*(t)}=x^{\rho^*} \tag{2.37}$$

for every $x>0$.

For heavy tails, the convention is to assume that the rate of convergence towards zero of $\ln U(tx)-\ln U(t)-\xi\ln x$ as $t\to\infty$ is known. The second order condition then simplifies to

$$\lim_{t\to\infty}\frac{\ln U(tx)-\ln U(t)-\xi\ln x}{A(t)}=\frac{x^{\rho}-1}{\rho}\ , \tag{2.38}$$

where $\rho\leq 0$ and $A(t)\to 0$ as $t\to\infty$. For details see Gomes et al. (2008).

In Table 7 and Table 8 we report the estimates of extreme value index ξ (and tail index α) using the four presented estimators. To calculate the estimates, constant fractions of tail observations were used. It has to be noted that the ambition behind these tables is to illustrate the general patterns of behaviour of the individual estimators, rather than to provide accurate estimates of the tail index.

The performance of the individual estimators is fairly distinctive. The de Haan-Resnick estimator gives a consistent set of estimates for both tails. There are no wild swings in the values but there is nothing to be said about their reliability as the asymptotic properties are unknown. The tail index estimates between 2.51 and 3.01 for both tails seem to be plausible.

The Pickands estimator fails completely. Apart from reaching negative values in several cases, the estimates show wild swings in values across subsamples. Reasonable estimates are achieved only for 5% subsamples in both tails but standard errors are relatively large.

The Hill estimator yields fairly consistent and statistically significant tail index estimates. Across all values of k, the estimates range from 2.07 to 3.11 for the left tail and from 2.37 to 3.11 for the right one. What catches the eye is their counter-intuitive tendency to increase with decreasing number of underlying observations, suggesting relatively less probability mass in the tails in more extreme regions.

The performance of the Moment estimator is patchy. The 10% tail index estimates for both tails are unreasonably high while those based on smaller subsamples struggle to be significant. In contrast with the Hill estimates, the tail index estimates tend to decrease with decreasing size of subsample. They reach values between 3.35 and 5.12 for the left tail and 2.70 and 5.87 for the right one. For the left tail, the estimates are higher on average than in case of the Hill and the de Haan-Resnick estimators, but standard errors are relatively high.

All in all, the Hill estimator seems to be best suited for the purposes of inference on tails of financial time-series. The de Haan-Resnick estimator is ruled out by the absence of asymptotic properties, the Pickands estimator is wildly inconsistent and the Moment estimator is held back by high standard errors. For a more thorough comparison of the Hill, the Pickands and the Moment estimators see de Haan and Peng (1998), who explicitly define conditions which favour the particular estimators.

Number of tail observations		10% 399	5% 199	2.5% 99	1% 39
de Haan and Resnick	xi	0.3984	0.3784	0.3745	0.3727
	SE	N/A	N/A	N/A	N/A
	alpha	2.5103	2.6429	2.6703	2.6829
Pickands	xi	0.0345	0.5033**	-0.0975	-0.6171*
	SE	0.0906	0.1381	0.1793	0.2786
	alpha	28.9656	1.9870	-10.2569	-1.6206
Hill	xi	0.4820**	0.3830**	0.3681**	0.3220**
	SE	0.0241	0.0272	0.0370	0.0516
	alpha	2.0746	2.6108	2.7164	3.1053
Moment	xi	0.1955**	0.2983**	0.2277*	0.2977
	SE	0.0510	0.0740	0.1031	0.1671
	alpha	5.1155	3.3519	4.3918	3.3591

Table 7 The extreme value index estimates (with asymptotic std. errors) of the left tail of PX returns based on respective fractions of tail observations. Source: author's calculations.
(* and ** denote estimates significant at 5% and 1% level, respectively)

Number of tail observations		10% 399	5% 199	2.5% 99	1% 39
de Haan and Resnick	xi	0.3505	0.3336	0.3327	0.3408
	SE	N/A	N/A	N/A	N/A
	alpha	2.8529	2.9980	3.0057	2.9339
Pickands	xi	-0.0865	0.1885	0.0352	0.9454**
	SE	0.0894	0.1310	0.1819	0.3422
	alpha	-11.5613	5.3046	28.3984	1.0578
Hill	xi	0.4216**	0.3367**	0.3216**	0.3473**
	SE	0.0211	0.0239	0.0323	0.0556
	alpha	2.3716	2.9700	3.1095	2.8795
Moment	xi	0.1704**	0.3020**	0.3608**	0.3697*
	SE	0.0508	0.0741	0.1068	0.1707
	alpha	5.8669	3.3116	2.7714	2.7049

Table 8 The extreme value index estimates (with asymptotic std. errors) of the right tail of PX returns based on respective fractions of tail observations. Source: author's calculations.
(* and ** denote estimates significant at 5% and 1% level, respectively)

Actually, the Pickands, the Hill and the Moment estimators belong to a broader family of classical semi-parametric estimators which share a set of common properties. Apart from being weakly consistent and asymptotically normal under the same condition, the empirical patterns of all three estimators exhibit two trends:

- They have high variance for high thresholds $x_{(k)}$, i.e., for small values of k;
- They have high bias for low thresholds $x_{(k)}$, i.e., for large values of k.

Following Gomes et al. (2008), two other common features, which relate to the problem of selection of optimal value of k, can be added:

- There is only a narrow region of stability, when the estimates are plotted as a function of k, making the adaptive choice of the threshold on the basis of a sample path stability criterion difficult;
- The mean-squared error function is of a very peaked nature, making the choice of k_0, where MSE attains its minimum, also difficult.

As a possible remedy to the latter two difficulties with finding the optimal k, the modified Hill estimator was proposed by Huisman et al. (1997). Luckily, it eventually solves the first two drawbacks as well.

2.3.1 The Modified Hill Estimator

The modified Hill estimator is based on the conventional Hill estimator which was shown to be consistent and asymptotically normal under certain conditions. The consistency property holds not only for IID samples but for a wide class of dependent stochastic processes as well (see Hsing (1991) and Resnick and Starica (1998)). The problems come with smaller samples which are drawn from not exactly Paretian distribution (i.e., the underlying CDF F does not belong to the MDA of the Fréchet distribution) where the estimator suffers from severe bias and tends to overestimate the value of ξ.

Dacorogna et al. (1995) show that for the class of distributions satisfying Equation (1.3) and for a given k the asymptotic expected value of the Hill estimator is approximated by

$$E\left(\xi(k)\right) \approx \frac{1}{\alpha} - \frac{b\beta}{\alpha(\alpha+\beta)} a^{-\frac{\beta}{\alpha}} \left(\frac{k}{n}\right)^{\frac{\beta}{\alpha}} \tag{2.39}$$

and the asymptotic variance by

$$\operatorname{var}\left(\xi\left(k\right)\right) \approx \frac{1}{k\alpha^2} . \tag{2.40}$$

These approximations facilitate good understanding of the principal problem of semi-parametric estimators, namely determining the optimal number of the order statistics k. As k increases, the variance of the estimator decreases and it becomes more efficient but at the same time the bias of the estimator worsens. Similar trade-offs related to goodness of fit of the underlying theoretical model were observed also in case of the Block Maxima and the Peaks over Threshold parametric methods.

Obviously, k should rise with the sample size n. Unfortunately, there is no simple rule or reliable diagnostic procedure which would hint at the optimal k. Hall (1982) provides inconclusive formulas showing the rate at which k should theoretically change with n. DuMouchel (1983) specifically suggests using data outside the 10th and 90th percentile to let the tails speak for themselves. Loretan and Phillips (1994) state as well that in principle k should not exceed 10% of the underlying sample and add that it is advisable to estimate ξ for a variety of trial values of k. However, they fail to clarify how to eventually detect the right value.

Figure 23 and Figure 24 show the estimates of ξ for the left and the right tail, respectively, of PX returns as a function of number of order statistics for $k \leq 0.1n$. Clearly, the Hill estimates decrease with the decreasing number of order statistics used in the estimation but at the same time the 95% confidence interval bands widen. Although the trend is broken for small values of k (particularly in case of the right tail where it can be attributed to its extraordinary structure), the confidence interval in this region is already so wide that it does not represent serious counter-factual evidence.

Huisman et al. (1997) further expose this problem through a simulation experiment. They report average Hill estimates as a function of k from 1,500 simulated samples with $n = 250$ drawn from Student-t distributions with 1 to 5 degrees of freedom. The conclusions are straightforward: for small values of k the average Hill estimate $\hat{\xi}\left(k\right)$ is almost equal to the true value but its variance is sky-high. The variance declines for intermediate values of k, but the bias increases. As k approaches n [7] the bias explodes and, by implication, so does variance. Two conclusions are drawn from this exercise:

[7] They use absolute values of observations to better utilize the limited information in such a small sample. The symmetry of the Student-t distribution allows this.

firstly, high values of k are necessarily sub-optimal, and secondly, there is a bias for any k exceeding zero.

In fact, literature offers a plethora of possible solutions to the problem of finding an optimal value of k for small samples. The most trivial one is to increase the number of observations. High-frequency data are often used for inference on behaviour of tails but in some areas they are not available or do not suit the purpose of inference.

Then there is a couple of more sophisticated methods. Jansen and de Vries (1991) used a Monte Carlo simulation, where n innovations are drawn from a known distribution (with a known tail index) and the value of k is then determined by minimizing MSE of the Hill estimates. The results, however, depend on the underlying unknown distribution and have to be interpreted carefully. Another, and rather complex, method was offered by Beirlant et al. (1996). It includes a weighted least squares algorithm to obtain the slope at the upper right tail of a Pareto quantile plot. But as their optimal k exceeds zero, some bias is still present, although their procedure is heavily weighted towards minimizing the bias. Moreover, the small-sample properties of this estimator remain unknown.

A completely different approach was proposed by Huisman et al. (1997) who observed that for values of k lower than some threshold κ, the Hill estimates of ξ seem to increase virtually linearly in k, implying that the exponent β / α in Equation (2.39) is very close to one. Figure 25 and Figure 26 confirm the validity of this observation for extended regions of both tails of PX returns.

After replacing the bias term with $f(k)$, Equation (2.39) becomes

$$\xi(k) = \beta_0 + f(k) + \varepsilon(k) \tag{2.41}$$

and using the linear function approximation[8] for $k \leq \kappa$ yields

$$\xi(k) = \beta_0 + \beta_1 k + \varepsilon(k)\,. \tag{2.42}$$

Having estimates of $\xi(k)$ for $k = 1,...,\kappa$ at disposal, one can estimate Equation (2.42) for $k \to 0$ in order to get an optimal unbiased estimate of ξ in the form of intercept β_0. The trick is that instead of searching for optimal k, this approach exploits the information embodied in a whole range of Hill estimates and smartly avoids the bias-variance trade-off.

[8] According to Huisman et al. (1997), results of non-linear regression models proved to be worse than those obtained by the linear specification.

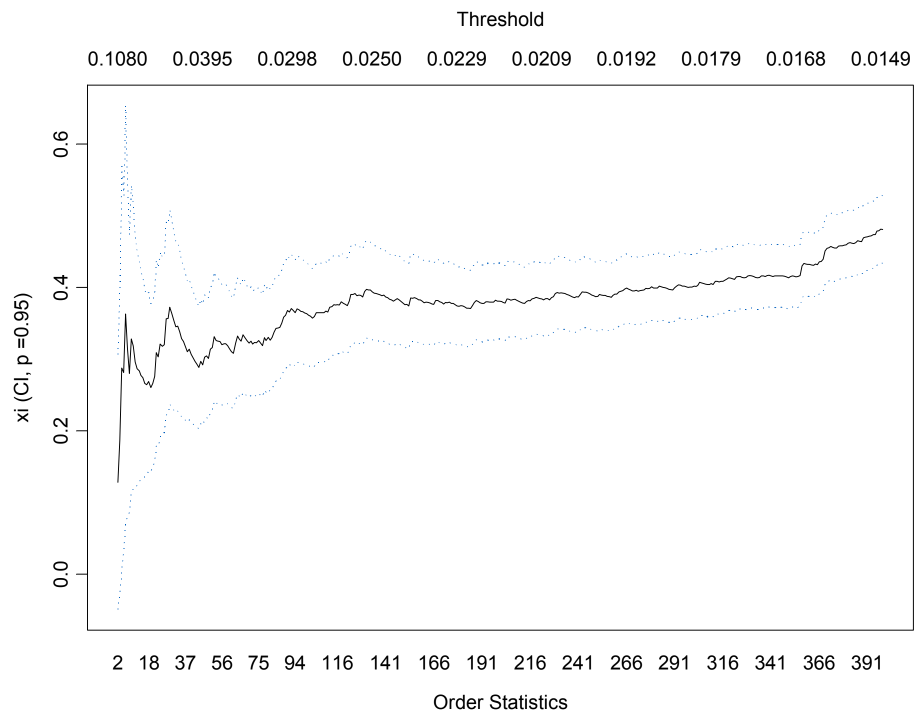

Figure 23 The Hill estimates of ξ of the left tail of PX returns with 95% confidence interval as a function of k. Source: author's calculations.

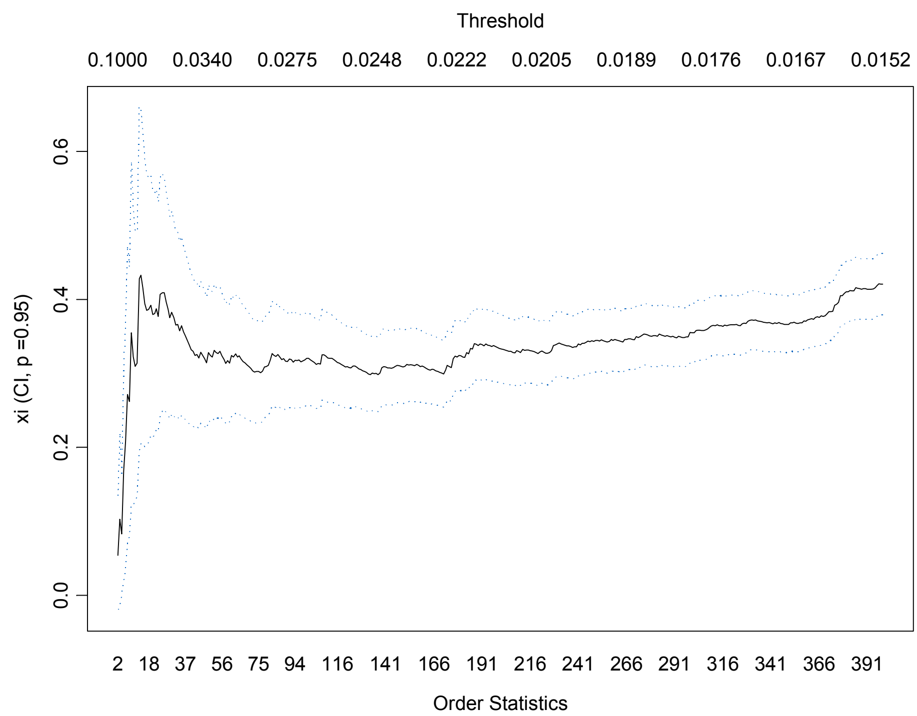

Figure 24 The Hill estimates of ξ of the right tail of PX returns with 95% confidence interval as a function of k. Source: author's calculations.

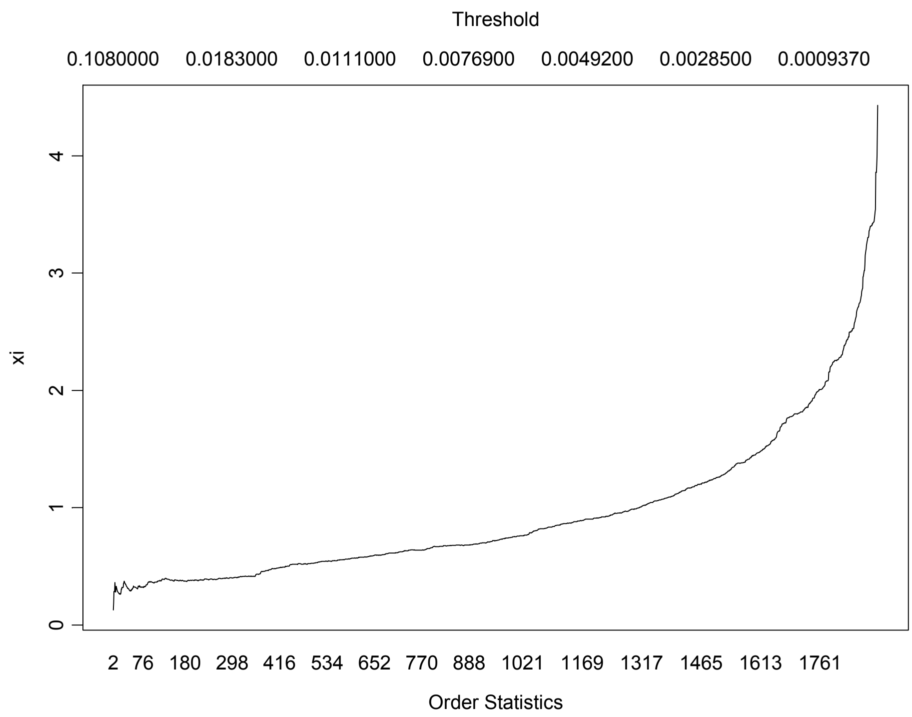

Figure 25 The Hill estimates of ξ of the left tail of PX returns as a function of k. Source: author's calculations.

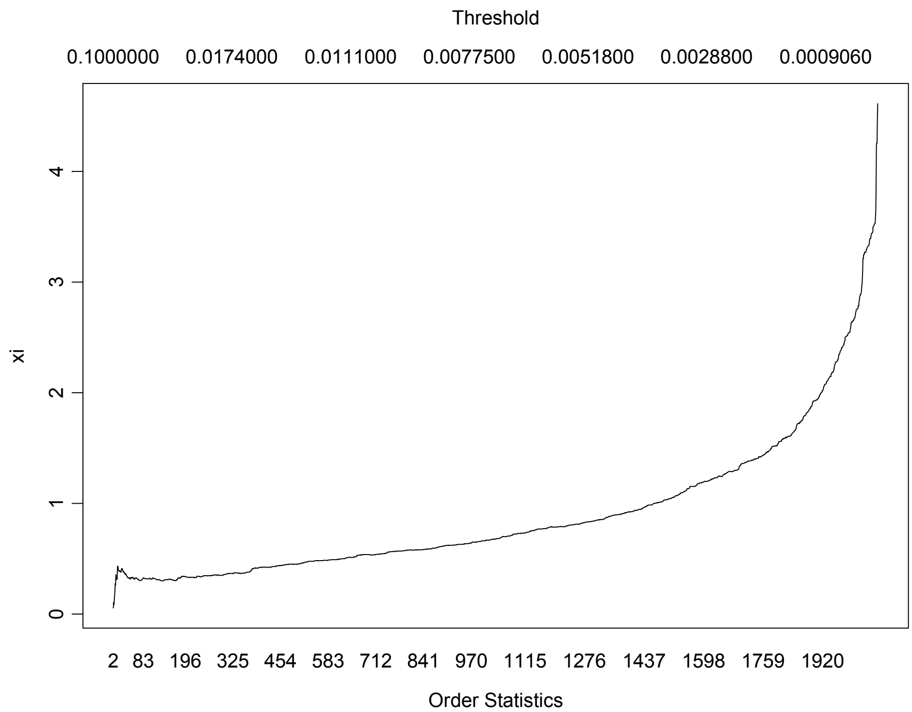

Figure 26 The Hill estimates of ξ of the right tail of PX returns as a function of k. Source: author's calculations.

While the way in which this method is constructed ensures solid robustness with regard to the value of κ, the trouble arises at the other end of the tail. As can be seen in Figure 23 and Figure 24, the behaviour of the Hill estimator for very small values of k tends to be rather erratic, leading to the problem of setting reasonable lower threshold value of k. Theoretically, the Hill estimate can be computed already for $k = 2$, but based on Figure 23 and Figure 24 using 1% and 10% of the data as the lower and the upper threshold, respectively, looks sensible.

However, there is one technical caveat – Equation (2.42) cannot be estimated by the Ordinary Least Squares (OLS) method. Firstly, the variance of the Hill estimates is not constant over different values of k, as Equation (2.40) shows, and therefore the assumption of homoscedasticity is breached. This can be corrected for by using Weighted Lest Squares (WLS) method instead. And secondly, as the estimates of $\xi(k)$ are cumulatively computed from the same underlying observations, they are likely to be auto-correlated. As a consequence, standard error computations for both OLS and WLS methods are not suitable. This problem can be possibly overcome by using re-sampling methods as jack-knife or bootstrap.

In Table 9 the modified Hill estimates for both tails are reported. We used the OLS as well as the WLS regression methods but the results do not differ very much. Regarding the standard errors, both re-sampling methods gave very similar results. The main conclusion is that in contrast with classical semi-parametric estimators, the modified Hill estimator provides possibly unbiased EVI estimates with much reduced uncertainty.

	Method	xi	Bootstrap SE	Jackknife SE	alpha
Left Tail	OLS	0.3170	0.0025	0.0025	3.1546
	WLS	0.3183	0.0025	0.0024	3.1417
Right Tail	OLS	0.2864	0.0022	0.0022	3.4916
	WLS	0.2785	0.0022	0.0023	3.5907

Table 9 The modified Hill estimates based on OLS and WLS method for the left and the right tail of PX returns. Source: author's calculations.

2.4 LITERATURE SURVEY

To put our results into a broader perspective, let us review some of the empirical results provided in literature. We restrict our attention to works related to stock market tail behaviour.

In their study of covariance stationarity of heavy-tailed time series, Loretan and Phillips (1994) estimated the tail index of two series of US stock market returns. They report the conventional Hill estimates for several values of $k < 0.1n$. For monthly returns on a broad index of US stocks from January 1834 to December 1987, the estimates range from 2.95 to 3.55 for the left tail and from 2.46 to 2.95 for the right tail. For daily returns on the S&P 500 stock market index from July 1962 to December 1987, the estimates range from 3.44 to 3.80 for the left tail and 3.08 to 3.86 for the right tail.

An index of the most traded stocks on the New York Stock Exchange covering period 1885-1990 was investigated by Longin (1996). The workhorse in this case is the parametric Block Maxima method with block sizes ranging from one month to two years. The tail index estimates range from 2.268 to 3.509 for the left tail and from 2.786 to 3.610 for the right one. As a point of reference, the Pickands and Hill semi-parametric estimates based on specific values of k found by simulations for each estimator separately are reported. The Pickands estimator gives the tail index estimates of 2.410 for the left and 2.632 for the right tail and the corresponding Hill estimates are 2.770 and 3.030, respectively.

Huisman et al. (1997) report the conventional and the modified Hill estimates of monthly returns on the S&P Composite Price Index, covering period from January 1965 to December 1987. The optimal number of tail observations included for the conventional Hill estimator is determined by a Monte Carlo simulation. For the left tail, the conventional Hill estimator yields 2.994 and the modified Hill estimator 3.509 (OLS) and 4.219 (WLS). For the right tail, the respective estimates are 2.519 by the conventional estimator and 3.745 (OLS) and 4.000 (WLS) by the modified estimator.

Gilli and Këllezi (2006) use the two parametric methods to estimate the tail shape parameters of daily returns on stock indices of six major markets: European (ES50), British (FTSE100), Chinese (HS), Japanese (Nikkei), Swiss (SMI) and American (S&P500). Despite having the same end date in August 2004, the length of period covered differs for each of the series with the longest one, S&P500 (starting in January 1960), containing almost three times as many observations as the shortest one, SMI (starting in July 1988).

The Block Maxima estimates are based on yearly maxima and the only results worth reproducing here, due to lack of date in the other cases, are those of S&P500 and Nikkei: the former one yielding tail index estimates of 1.887 for the left and 10.000 for the right tail and the latter one of 3.984 and 10.417, respectively.

With regard to Peaks over Threshold method, with the exception of ES50, the estimates of the tail index range from 2.577 to 5.525 for the left tail and from 5.405 to 10.753 for the right one. ES50 estimates exhibit strange behaviour (in particular the left tail estimate of 22.222 is far-fetched). The optimal thresholds were determined by the mean excess plot.

Jondeau and Rockinger (2003) used a Datastream database of daily stock market returns for 20 countries to estimate parameters of GEVD. In general, their estimates of the shape parameter range from 2.874 to 7.576 for the left tail and from 2.421 to 5.495 for the right tail. Moreover, they divide their sample to four geographical areas: for Eastern Europe, the range of estimates is from 2.421 to 4.785 for the left tail and from 2.874 to 5.181 for the right tail; for mature markets, the estimates range from 3.425 to 5.495 and from 3.030 to 7.576, respectively; for Asia, they range from 2.457 to 5.000 and from 3.401 to 7.299, respectively; and finally, for Latin America, they range from 3.571 to 4.444 and from 4.167 to 5.587.

LeBaron and Samanta (2004) set out to find answers to questions such as whether market booms are more or less likely than market crashes and whether crashes in emerging countries are more frequent than in developed countries. Their dataset comprises of standardized daily returns on the Morgan Stanley Capitalization International (MSCI) country price indices of 8 emerging countries from Asia and South America and 12 developed countries and covers period from January 1, 1993 to May 1, 2003.

Although they report Pickands, conventional Hill and Dekkers, de Haan point estimates as well, the modified Hill estimator is central to their analysis. In case of emerging countries, the modified Hill estimator, based on 10% of order statistics and using WLS, yields tail index estimates ranging from 2.81 to 7.80 for the left and from 2.44 to 7.91 for the right tail. In case of the developed countries, the estimates range from 2.84 to 6.39 for the left tail and from 3.82 to 6.34 for the right one.

Regarding the Pickands, conventional Hill and Dekkers, de Haan estimates, they report results for a set of three values of k . The only point worth taking from them is that the Pickands estimates exhibit similarly volatile patterns as our results reported in Table 7 and Table 8.

And finally, Horák and Šmíd (2009) used daily, weekly and monthly returns of 22 stocks listed on three Central European stock exchanges (Czech, Polish and Hungarian) and three major western stock markets (NASDAQ, LSE, XETRA) for their estimations by the conventional Hill and the modified Hill estimators for 3-10% of the observations from the left, the right and both tails. The results for daily returns on the eight Czech shares are of special interest to us: for the left tail, the conventional Hill estimates range from 2.81 to 4.68 and the modified Hill estimates from 1.95 to 5.46; for the right tail, the estimates range from 2.77 to 4.79 and from 3.00 to 5.67, respectively; and for both tails, the estimates range from 2.91 to 4.60 and from 3.08 to 5.44, respectively.

It seems that our estimates are in line with the literature in all important aspects: while the magnitude of tail index estimates is hardly extraordinary for an emerging market, they are on average lower than those of mature markets. At the same time, left tail estimates are higher than the right ones.

2.5 SUMMARY

To conclude this section, we picked one parametric and one semi-parametric estimator to represent the respective approaches and reproduced their results in Table 10. The choice between the two parametric methods was straightforward as the performance of the Peaks over Threshold method was vastly superior to the data-intensive Block Maxima method. Concretely, estimates based on data exceeding the optimal threshold found through the mean excess plot are reported.

The family of semi-parametric estimators is represented by the modified Hill estimator which is not strictly semi-parametric in its very nature but is very closely related and gives the most reliable estimates. To be precise, the estimates by WLS method with bootstrap standard errors are reported.

		Peaks over Threshold (u=0.019)	Modified Hill Estimator (WLS)
Left Tail	xi	0.2784	0.3183
	SE	0.0792	0.0025
	alpha	3.5920	3.1417
Right Tail	xi	0.2709	0.2785
	SE	0.0726	0.0022
	alpha	3.6914	3.5907

Table 10 Comparison of the Peaks over Threshold and the Modified Hill estimates. Source: author's calculations.

There are not substantial numerical differences in the tail index estimates. Both methods give results in the interval from 3 to 4, in line with the conclusion from the

Mandelbrot's method of sample moments. Comparison of standard errors favours the Modified Hill estimator but it is misleading given the way the estimator is constructed. Nonetheless, even when compared to the classical Hill estimates, the POT estimates tend to have higher standard errors.

What is definitely worth mentioning is the fact that both sets of estimates suggest that the left tail is heavier than the right one. To further elaborate on this point, a simple test of the hypothesis that the distribution of PX returns is symmetric will be performed. Following Horák and Šmíd (2009), let us define the test statistic

$$\overline{\xi} = \hat{\xi}_L - \hat{\xi}_R , \tag{2.43}$$

where $\hat{\xi}_L$ and $\hat{\xi}_R$ are the conventional Hill estimates of the left and the right tail of the distribution. The asymptotic normality of the Hill estimator implies that the statistic should be approximately normally distributed with zero mean and variance $s_L^2 + s_R^2$, where s_L and s_R stand for asymptotic standard deviations of the respective Hill estimates. Notice that the bias of the Hill estimates is of little importance here as it simply cancels out. Equivalently, we can write

$$\overline{\overline{\xi}} = \frac{\hat{\xi}_L - \hat{\xi}_R}{\sqrt{s_L^2 + s_R^2}} \to N(0,1), \tag{2.44}$$

i.e., the standardized test statistic should follow the standard normal distribution.

Formally, we test the null hypothesis

H_0: the distribution is symmetric

against the alternative

H_1: the left tail is heavier than the right one

by checking for significantly positive $\overline{\xi}$.

In Table 11 we report the values of the standardized test statistic $\overline{\overline{\xi}}$ for the Hill estimates based on different numbers of tail observations. Having in mind that for the null hypothesis to be rejected on the 5% significance level the test statistic would have to surpass 1.96, we can conclude that there is no statistically significant evidence that the distribution of PX returns is asymmetric.

Number of observations	10%	5%	2.5%	1%
	399	199	99	39
Test statistic	1.8833	1.2816	0.9474	-0.3330

Table 11 Values of the standardized test statistic for Hill estimates based on different numbers of tail observations. Source: author's calculations.

The final point to be made is the observation that the results imply that one can comfortably rule out the possibility of either tail index of PX returns being lower than two. Three conclusions follow from this statement. Firstly, the tail index is finite and therefore the tails are not Gaussian. Secondly, the variance of the PX returns is finite and well defined. And thirdly, it allows us to make a judgement on the nature of the Czech stock market behaviour: it supports neither the discontinuous stable Paretian hypothesis nor the continuous Gaussian hypothesis.

These two extremes of market behaviour were discussed by Fama (1963). In a Gaussian market, a large price change is likely to be a result of a consequence of very small price changes and the path of an asset price is continuous. By contrast, in a stable Paretian market a large price change over a long interval is likely to be a result of one or a very few huge price changes that occurred during narrow subintervals and the price path exhibits discontinuities. As the Gaussian market hypothesis relies on the normal distribution, it requires the tail index of returns to be infinite. The stable Paretian hypothesis is based on the stable Paretian distribution which is characterised by tail index values in the region between 0 and 2.

The Czech stock market seems to be an intermediate case, since it exhibits more extremes and thus more risk for investors than a Gaussian market but fewer extremes and thus less risk than a stable Paretian market. Under the estimated values of the tail index of returns, market price may or may not exhibit discontinuities, depending on the process governing returns (Longin 1996).

3. APPLICATIONS TO RISK MANAGEMENT

The previous section was devoted to modelling the behaviour of extreme financial returns. One term, which is commonly associated with extremes, is risk. Jorion (2001) defines risk as the volatility of unexpected outcomes. In an ideal world, financial risk could be completely diversified away or totally eliminated by a perfect hedge. In the real world, however, the best one can do is to strive to manage it. Risk management, therefore, seems to be a perfect area where to apply insights offered by the statistics of extremes.

Each and every financial transaction comes with various risks attached. There is the risk of unfavourable movements in the level or volatility of market prices (Market Risk); the risk that counterparty will be unwilling or unable to fulfil its contractual obligations (Credit Risk or Default Risk); the risk that intended transactions will not be able to be carried out at prevailing market prices due to the excessive size of the position (Liquidity Risk); the risk of human and technical failures like frauds and trading errors (Operational Risk); and finally the risk that it will not be possible to legally enforce the transaction (Legal Risk).

Given our underlying dataset of PX returns, the focus in the remainder of this section will be exclusively on the market risk. In order to manage any risk, one should be able to measure it. Following Dhaene et al. (2003), a one-sided risk measure can be defined as a measure of the distance between a risky situation and the corresponding risk-free situation when only unfavourable discrepancies contribute to the risk. The opposite is a two-sided risk measure which considers also favourable discrepancies. While two-sided risk measures are useful in pricing the risk, from risk management point of view, the downside risks are dominant and therefore the analysis will be centred on one-sided measures. However, as an investor may hold long as well as short position in an asset, neither tail will be omitted.

There have been various attempts at finding the appropriate measure of market risk. For example, Roy (1952) proposed to use the maximum probability of a return being at least as high as some predetermined minimum tolerable threshold to quantify the risk, while Markowitz (1952) assigned this role to the variance of returns. Yet, due to their nature, neither of these measures is actually capable of capturing the risk of occurrence of the most adverse events. For this purpose, more advanced measures were designed.

Value at Risk and Expected Shortfall may be considered the most prominent among the present day risk measures. In essence, they are based on modelling the underlying loss distribution over some predetermined time horizon. This approach is vindicated by several facts. Firstly, losses are central to risk management and thus it is natural to base the measurement of the implied risk on their statistical distribution. Secondly, the concept of a loss distribution makes sense on all levels of aggregation, from a single asset portfolio to overall position of a financial institution. Thirdly, loss distributions are capable of accounting for netting and diversification effects. And finally, loss distributions are comparable across portfolios, regardless of their composition, given the time horizon is identical.

However, apart from these unquestionable positives, loss distributions have two weaknesses. One problem is their reliance on the past data. It is not guaranteed that the laws governing financial markets will not suddenly change and if they do, the historical data are not suitable for estimating the future risk any more. The other difficulty pertains to the complexity of estimating the loss distribution in practice. Even in stationary environment, it is very challenging to estimate the loss distribution accurately, above all in the tail regions which possess crucial information for risk management. McNeil et al. (2005) point out that, as a consequence, many seemingly sophisticated risk management systems are based on relatively crude statistical models of the loss distribution, incorporating unrealistic assumptions such as normality.

One more issue is associated with risk forecasting in general. That is whether to use a distribution conditioned on current market conditions or an unconditional one. Both approaches have their pros and cons which mostly depend on the time horizon, size and composition of portfolio at hand. Put simply, the choice between the two approaches is up to the user and the purpose of her enquiry. This issue is discussed in more depth, for example, by Danielsson and de Vries (2000) and McNeil and Frey (2000). Since our inference does not pursue any specific goal, and also for the sake of simplicity, the unconditional approach will be used in the remainder of this section.

3.1 VALUE AT RISK

Consider a portfolio of assets and let F be the distribution function of its returns. The task is to find a statistic based on F which would characterize the risk of holding the portfolio over a time period t. The most straightforward candidate could be the maximum loss. However, it falls short of its purpose on at least two grounds. Firstly, it is common in finance to work with distributions with unbounded support, i.e., the maximum is infinity. And secondly, it ignores the valuable information about the behaviour of returns incorporated in F.

Taking into account these remarks and adjusting the previous reasoning accordingly, a much better candidate for a risk measure is the maximum loss which is not exceeded with a given probability. That exactly is the Value at Risk (VaR) metric, a dominant market risk measure of today. It made its way into the first pillar of Basel II accord as a tool for determining regulatory capital requirements.

Formally, for a given time horizon t and confidence level p, VaR is defined as the loss in market value that is exceeded with probability $1-p$ over the time horizon t. In probabilistic terms, VaR is simply the $(1-p)$-th quantile of the underlying distribution of returns.

Setting values of t and p is a matter of judgement and largely depends on the purpose of its use. Duffie and Pan (1997) report that the Derivatives Policy Group proposed two-week time horizon and 99% confidence level for over-the-counter derivatives broker-dealer reports to the Securities and Exchange Commission; the Bank for International Settlements set p to 99% and t to 10 days for purposes of measuring the adequacy of bank capital; JP Morgan discloses its daily VaR at 95% confidence level; and Bankers Trust prefers daily VaR at 99% confidence level. In general, the most widely used confidence intervals are 95% and 99% with time horizons between 1 and 10 trading days.

Despite its popularity, VaR suffers from relatively severe deficiencies for which it was frequently criticised. One stream of criticism was directed at the fact that VaR is not a coherent risk measure. The concept of coherent risk measures was introduced by Artzner et al. (1999) in the form of four properties required from a risk measure. Let $\rho(W)$ be a risk measure viewed as a function of the distribution of portfolio value W, then it should satisfy:

- Monotonicity: $W_1 \leq W_2 \Rightarrow \rho(W_1) \geq \rho(W_2)$, i.e., portfolio with lower returns in all possible states of world has to be riskier;
- Translation invariance: $\rho(W+k) = \rho(W) - k$, i.e., cash injection k into the portfolio reduces the risk accordingly;

- Homogeneity: $\rho(aW) = a\rho(W)$, i.e., scaling the portfolio up or down increases or decreases its riskiness in the same proportionately;
- Sub-additivity: $\rho(W_1 + W_2) \leq \rho(W_1) + \rho(W_2)$, i.e., merging two portfolios does not increase the overall riskiness.

It is the fourth point on which VaR fails and thus counters intuition that there should be a diversification benefit associated with merging portfolios. Practically, it also means that one cannot be sure that aggregating VaR values of different portfolios or business units will result in a bound for the overall risk of the enterprise.

However, this argument was weakened by Danielsson et al. (2011) who proved that VaR is sub-additive in the relevant tail region if asset returns are multivariate regularly varying, for both independent and cross-sectionally dependent returns, provided the mean is finite. Yet, this asymptotic result may not hold in practice due to the tail coarseness problem which arises when only a handful of observations are used in the empirical estimation and the location of a specific quantile is greatly uncertain.

The other criticised problem with VaR is that it reveals nothing about the severity of possible losses when the confidence level is breached, what is implied by its very construction. Its suitability for setting regulatory capital requirements is therefore questionable.

To estimate VaR two types of methods were developed: parametric and non-parametric. Non-parametric estimation draws on the historical distribution of returns. Empirical VaR is equivalent to the corresponding percentile of the underlying distribution. It has three obvious weaknesses. Firstly, to obtain accurate estimates a large sample of observations is required. Secondly, using empirical distribution does not allow for conditionality of the parameters over time. And thirdly, empirical estimates depend on the chose quantile and thus do not allow for conversion of parameters.

To overcome these flaws, a parametric approach is often preferred. Since the distribution is approximated by a parametric distribution, parameters can be allowed to change over time. Also conversion of VaR values with respect to its parameters is possible, enabling comparisons of VaR estimates across various institutions. Nonetheless, parametric conversion is meaningful only if the parametric approach accurately reflects the underlying distribution at all quantiles in the tail. The choice of parametric distribution is therefore crucial.

The simplest parametric approach assumes that the expected returns are normally distributed with the mean and variance derived from the past data. Suppose that the return distribution F is normal with mean μ and variance σ^2. Then for $p \in (0,1)$

$$\mathrm{VaR}_p = \mu + \sigma\Phi^{-1}(1-p), \tag{3.1}$$

where Φ denotes the standard normal CDF and $\Phi^{-1}(1-p)$ is the $(1-p)$-th quantile of Φ. The normal distribution is a popular choice for VaR calculations since it allows for easy conversions across confidence intervals and time horizons. For example, for conversion of 1-day $\mathrm{VaR}_{0.95}$ to 10-days $\mathrm{VaR}_{0.99}$, the following formula holds

$$\mathrm{VaR}_{0.99}^{10-days} = \mathrm{VaR}_{0.95}^{1-day} \frac{\Phi^{-1}(0.01)}{\Phi^{-1}(0.05)}\sqrt{10}. \tag{3.2}$$

Nevertheless, normal distribution does not describe the distribution of returns very well, especially in tail regions, and there might be relatively large discrepancies between the tails of the theoretical and actual distribution, leading to flawed and unreliable VaR estimates. Thus, a fat-tailed distribution ought to be a better alternative.

Huisman et al. (1998) proposed to use Student-t distribution and designated the improved approach as VaR-x. As Student-t comes from the same family of distributions as the normal distribution, the procedure of estimating VaR is very similar with the proviso that a standard Student-t distribution has variance defined as $\alpha/(\alpha-2)$ where $\alpha > 2$ denotes degrees of freedom. As noted in subsection 2.1, α is equal to the tail index. Suppose that returns are distributed such that standardized returns $\tilde{X} = (X-\mu)/\sigma$ follow standard Student-t distribution with α degrees of freedom, then

$$\mathrm{VaR}_p = \mu + \sigma\sqrt{\frac{\alpha-2}{\alpha}}t_v^{-1}(1-p), \tag{3.3}$$

where t_v denotes the CDF of standard Student-t.

Although Student-t distribution is likely to capture the thickness of tails better than normal distribution, more precise models are available. Tail index estimation in section 2.2.2 was more or less based on modelling the tail behaviour of returns by the specific distributional form of Generalized Pareto Distribution (GPD). Recall that the conditional distribution of observations exceeding a predetermined high threshold u is defined as

$$F_u(y) = P\{X - u \le y | X > u\} \tag{3.4}$$

for $0 \le y < \bar{x} - u$, where $\bar{x} \le +\infty$ is the right endpoint of F. In words, the excess distribution expresses the likelihood that when the threshold is exceeded, it is not exceeded by an amount larger than y. It can be re-written in terms of the underlying F as

$$F_u(y) = \frac{F(y+u) - F(u)}{1 - F(u)} . \tag{3.5}$$

Let $x = u + y$ and recall the result of the Pickands, Balkema and de Haan theorem that for certain u, the excess distribution above this threshold may be approximated by GPD for some ξ and β

$$F_u(y) = G_{\xi,\beta}(y) . \tag{3.6}$$

Combining Equation (3.5) and Equation (3.6) yields

$$F(x) = (1 - F(u)) G_{\xi,\beta}(x-u) + F(u) \tag{3.7}$$

for $x > u$. This formula provides an explicit distributional form of the tail of the underlying distribution $F(x)$ for $x > u$.

The only unknown quantity is $F(u)$. For this purpose, a simple empirical estimator $(n-k)/n$, where n denotes the size of the sample and k denotes the number of observations exceeding u, is suitable and adds an element of historical simulation to this concept. Replacing $F(u)$ with the empirical estimator and the $G_{\xi,\beta}$ with its functional form with ML estimates of the parameters results in

$$\hat{F}(x) = 1 - \frac{k}{n}\left(1 + \hat{\xi}\frac{x-u}{\hat{\beta}}\right)^{-1/\hat{\xi}} , \tag{3.8}$$

where $\hat{\xi}$ and $\hat{\beta}$ are ML estimates of the respective parameters and $x > u$. This estimator may be looked upon as a historical simulation estimator augmented by the Extreme Value Theory. McNeil (1999) notes that such estimator can be constructed whenever there is reason to believe that the data come from a common distribution but its statistical properties are best understood in situations where they can also be assumed to be independent or only weakly dependent.

Finally, the VaR estimate for a given probability $p > F(u)$ is obtained by inverting the tail estimation formula in Equation (3.8)

$$\widehat{\text{VaR}}_p = u + \frac{\hat{\beta}}{\hat{\xi}}\left(\left(\frac{n}{k}(1-p)\right)^{-\hat{\xi}} - 1\right). \quad (3.9)$$

In Table 12 we report 1-day VaR estimates on confidence levels 95% and 99% for both tails, based on non-parametric empirical method and parametric methods using normal, Student-t and GPD. In calculations of Student-t and GPD, the results from preceding sections were used. Concretely, the modified Hill estimates were used for determining the number of DF of Student-t distribution and GPD was fitted to observations exceeding threshold $u = 0.019$.

The first look at the results reveals striking similarity between empirical and GPD estimates, especially on 95% confidence level. The reasons are simple. Firstly, the underlying sample is quite large and therefore contains enough extreme observations for the empirical method to be fairly accurate. And secondly, the fit of GPD to the extreme observations is very good and therefore the empirical and theoretical distributions are close to each other.

The performance of the other two parametric estimates based on the normal and Student-t distributions reflects the fact that these two distributions provide only very crude approximations of the underlying distribution. The normal estimates on 95% confidence level are the largest, and thus overestimate risk, while on 99% confidence level they are the lowest, and thus underestimate risk.

The Student-t estimates differ for each tail. For the left tail, they tend to underestimate risk on both confidence levels. By contrast, for the right tail, they tend to overestimate risk on both confidence levels. Surely, Student-t estimates very much depend on accuracy of the tail index estimates.

	P	Empirical	Normal	Student-t	GPD
Left tail	95%	0.0218	0.0236	0.0200	0.0218
	99%	0.0413	0.0335	0.0380	0.0405
Right tail	95%	0.0211	0.0240	0.0214	0.0211
	99%	0.0354	0.0339	0.0387	0.0365

Table 12 1-day VaR estimates with confidence levels 95% and 99% for both tails of PX returns based on different methods. Source: author's calculations.

To conclude, using simple distributional assumptions does not provide realistic estimates of VaR and allows for a considerably large scale of results. The GPD method is visibly superior to the other parametric methods and can provide guidance in case of smaller samples, where the empirical estimation is unreliable. Moreover, Danielsson et al. (2011) show that GPD-based estimator substantially reduces the likelihood of sub-additivity failures as it replaces coarse empirical estimates with smooth power law. However, if the goal is to face as low regulatory capital requirements as possible, approximation by the normal distribution on 99% confidence level seems to be a rational choice.

3.2 EXPECTED SHORTFALL

Expected Shortfall (ES) was developed as an improved alternative of VaR. Not only that it meets all four requirements of a coherent risk measure, it also possesses vital information about severity of losses when the confidence level is breached. It is defined as the expected value of loss, given that VaR value is exceeded, i.e.,

$$\mathrm{ES}_p = E\left[X \middle| X > \mathrm{VaR}_p \right]. \tag{3.10}$$

It shares with VaR all its advantages. It is universal as it can be applied to any financial instrument and to any underlying source of risk. It is complete as it produces a comprehensive assessment for portfolios exposed to various sources of risk. And finally, it is intuitive and straightforward as it answers a natural and legitimate question about the riskiness of the portfolio.

To estimate ES, methods similar to VaR can be used. The non-parametric empirical estimate is simply equivalent to the average value of returns larger than VaR. In case of parametric methods, it can be shown (see McNeil et al. 2005) that for normally distributed returns

$$ES_p = \pm\mu + \sigma \frac{\phi\left(\Phi^{-1}(1-p)\right)}{1-p}, \tag{3.11}$$

where ϕ is the density of the standard normal distribution. When the standardized returns follow Student-t distribution,

$$ES_p = \pm\mu + \sigma \sqrt{\frac{\alpha-2}{\alpha}} \frac{g_\alpha\left(t_\alpha^{-1}(1-p)\right)}{1-p} \left(\frac{\alpha + \left(t_\alpha^{-1}(1-p)\right)^2}{\alpha-1} \right), \tag{3.12}$$

where t_α denotes the distribution function and g_α the density of standard Student-t.

Derivation of GPD-based estimator starts from observation that ES is related to VaR by

$$\mathrm{ES}_p = \mathrm{VaR}_p + E\left[X - \mathrm{VaR}_p \middle| X > \mathrm{VaR}_p\right], \tag{3.13}$$

where the expectation term is the mean of the excess distribution $F_{\mathrm{VaR}_p}(y)$ over the threshold VaR_p. In subsection 2.2.2 variation of GPD parameters with regard to threshold was mentioned. Recall that shape parameter ξ is invariant and variation of scale parameter is governed by $\beta_{u_1} = \beta_{u_0} + \xi(u_1 - u_0)$ for $u_1 > u_0$. Thus, for $\mathrm{VaR}_p > u$ Equation (3.6) implies

$$F_{\mathrm{VaR}_p}(y) = G_{\xi, \beta_u + \xi(\mathrm{VaR}_p - u)}(y). \tag{3.14}$$

Having an explicit distributional description of observations exceeding VaR_p allows for investigation of losses beyond VaR_p. For $\xi < 1$ the mean of $G_{\xi,\beta}$ is given as

$$E[X] = \frac{\beta}{1-\xi}. \tag{3.15}$$

Substituting Equations (3.14) and (3.15) into Equation (3.13) yields

$$\mathrm{ES}_p = \mathrm{VaR}_p + \frac{\beta_u + \xi(\mathrm{VaR}_p - u)}{1-\xi}, \tag{3.16}$$

which can be re-arranged into

$$\frac{\mathrm{ES}_p}{\mathrm{VaR}_p} = \frac{1}{1-\xi} + \frac{\beta_u - \xi u}{(1-\xi)\mathrm{VaR}_p}, \tag{3.17}$$

which describes the relation between ES and VaR. Assuming underlying distribution with unbounded support, the weight of the second term on the right hand side is reduced as probability p approaches zero and VaR_p becomes large. The ratio is then largely determined by the factor $1/(1-\xi)$. Intuitively, for thicker tails with larger ξ the probability of an extreme event is higher and therefore the gap between VaR and ES is wider.

The corresponding ES estimator, for a given probability $p > F(u)$ has the form

$$\widehat{\mathrm{ES}}_p = \frac{\widehat{\mathrm{VaR}}_p}{1-\hat{\xi}} + \frac{\hat{\beta}_u - \hat{\xi} u}{1-\hat{\xi}}, \tag{3.18}$$

where $\hat{\xi}$ and $\hat{\beta}_u$ are ML estimates of GPD parameters for threshold u.

In Table 13 ES estimates based on different methods are reported. Naturally, these estimates are larger than those of VaR but conclusions are almost identical. The empirical and the GPD estimates are very close to each other although in case of the right tail, its poorer GPD fit manifests itself once more, causing minor risk underestimation. In case of the left tail, the difference is nearly indistinguishable.

Regarding the other two parametric approximations, the normal estimates underestimate risk on both confidence levels. There is again the dichotomy in the Student-t estimates when left tail estimates tend to underestimate risk while right tail estimates tend to overestimate it. On average, Student-t estimates are clearly much better than their normal counterparts.

	P	Empirical	Normal	Student-t	GPD
Left tail	95%	0.0345	0.0297	0.0324	0.0345
	99%	0.0604	0.0384	0.0576	0.0605
Right tail	95%	0.0316	0.0301	0.0331	0.0315
	99%	0.0534	0.0388	0.0556	0.0526

Table 13 1-day ES estimates with confidence levels 95% and 99% for both tails of PX returns based on different methods. Source: author's calculations.

Table 12 and Table 13 imply that for estimation of quantile-based market risk measures GPD-based approximation of tails is superior to approximations by simple distributions such as normal and Student-t. To have an idea about their accuracy, Figure 27 and Figure 28 show 99% confidence intervals for VaR and ES GPD-based estimates. The left vertical dotted line corresponds to VaR estimate, the right one to ES estimate. The horizontal dotted lines determine 95% and 99% confidence intervals at points where they intersect with the bended curves. The profile log-likelihood function was used to compute these intervals.

To facilitate understanding of these graphs, let us make an example. Suppose one is interested in checking, for instance, 95% confidence interval of the left tail VaR estimate. The initial step is to locate the intersection of the left vertical dotted line (representing VaR estimate equal to 0.0405) and the upper horizontal dotted line (representing 95% confidence intervals). Now, moving along the horizontal line to the left, the point where it is vertically crossed by the bended curve specifies the left bound (0.0368) of the confidence interval. Mirror-like, one can find the right bound (0.0456).

The confidence intervals of VaR are logically narrower than those of ES as they do not have to deal with the handful of most extreme observations which are spread over a very long interval. The very same fact also causes the asymmetry of ES confidence intervals (Zivot and Wang 2006).

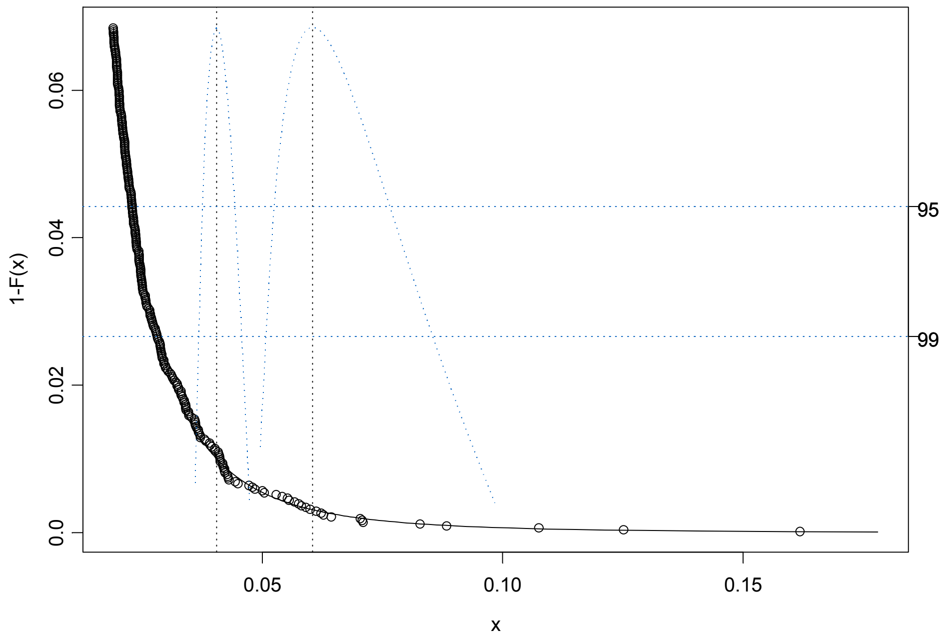

Figure 27 GPD fit to the left tail of PX returns with estimates and confidence intervals of 1-day VaR and ES on 99% confidence level. Source: author's calculations.

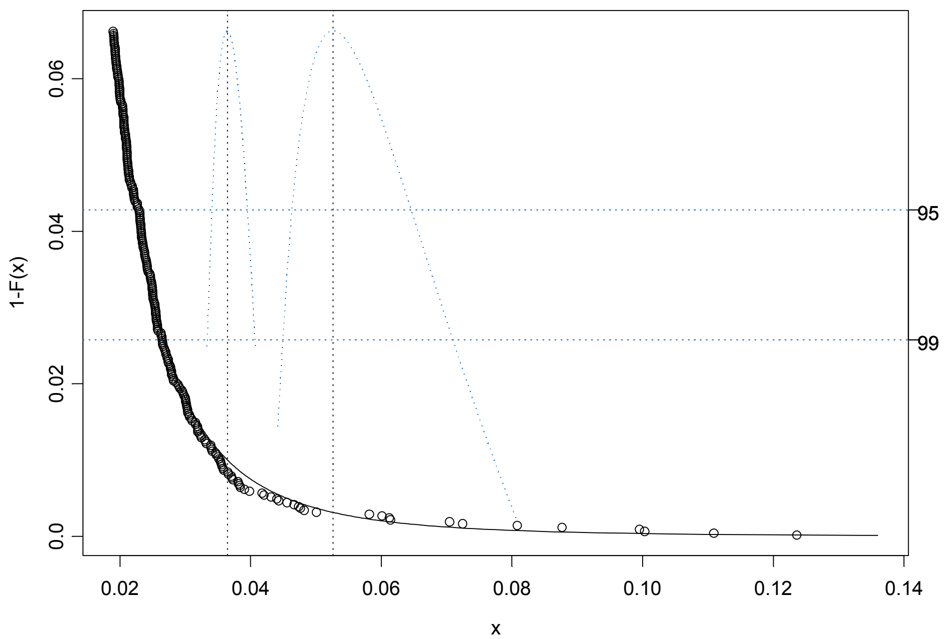

Figure 28 GPD fit to the right tail of PX returns with estimates and confidence intervals of 1-day VaR and ES on 99% confidence level. Source: author's calculations.

CONCLUSION

The main developments in the field of finance, such as the Modern Portfolio Theory of Markowitz (1952), the Capital Asset Pricing Model of Sharpe (1964) and Lintner (1965), or the option pricing formula of Black and Scholes (1973), rely on the hypothesis of normally distributed returns of financial assets. There is ample evidence in the literature that this assumption does not reflect reality on several grounds.

One of the departures from normality is the tail behaviour of returns which exhibits more extreme events than would be suggested by the Gaussian distribution. The phenomenon of heavy tails has a long history in finance, dating back to Mandelbrot's (1963) paper on the behaviour of cotton prices, but it took several decades, and at least as many market crashes, before it was adopted by and incorporated into mainstream thinking.

The first aim of this thesis was to infer on the tail behaviour of the Czech stock market index PX which, to our best knowledge, has not been tracked yet. An arsenal of methods was used to estimate the all-important measure of tail fatness, the tail index, of both tails of PX returns. This included parametric as well as semi-parametric techniques.

Parametric estimation methods rely on distributional assumptions about the behaviour of certain extreme events. In case of the Block Maxima method, it is the maximum observations from blocks of data which are assumed to follow the Generalized Extreme Value Distribution. Unfortunately, this method proved ill-suited for the estimation of our relatively short time series, as it demands very large blocks of data for the underlying theorem to hold.

The results of the second parametric method, the Peaks over Threshold method, were more satisfactory. This method is more efficient in subtracting information from extreme observations as it works with data exceeding some predetermined high threshold. The Generalized Pareto Distribution (GPD) is then fitted to the data. The caveat is to find the proper threshold. For optimal thresholds detected from mean excess plots tail index estimates of 3.5920 and 3.6914 were obtained for the left and the right tail, respectively.

The family of semi-parametric estimators was represented by classical estimators developed by Hill (1975), Pickands (1975), de Haan and Resnick (1980) and Dekkers et al. (1989). Similarly to the Peaks over Threshold model, these estimators are based on a subsample of tail observations whose size has to be determined. Unfortunately, in finite samples these estimators share a set of inconvenient properties which make this task if not impossible then certainly very difficult. To get round this problem, the modified Hill estimator due to Huisman et al. (1997) was embraced. It yielded tail index estimates of 3.1417 for the left and of 3.5907 for the right tail.

Brief literature survey put our tail index estimates into a broader perspective. They were found to be more or less in line with results obtained for an array of financial assets from all around the world. The Czech stock market behaves in a way one would expect – the tail index estimates are in general lower than in the developed markets and the left tail estimates tend to be lower than the right tail ones. However, after performing a simple test based on normality of conventional Hill estimator, we were not able to reject the hypothesis that the tails are symmetric and thus support the widely spread notion that the left tail of the distribution of returns should be heavier than the right one.

Risk management is considered to be the area of finance which should benefit the most from the insights on tail behaviour. Therefore, the second aim of the thesis was to illustrate how the presence of fat tails influences two common market risk measures, Value at Risk and Expected Shortfall.

Value at Risk (VaR) is a prominent quantile-based measure of market risk which made its way into the Basel II capital adequacy framework. In its essence, it is nothing more than a corresponding quantile of the loss distribution of returns on an asset. The normal distribution is often used as a theoretical proxy for the actual distribution of returns in computations of VaR.

To expose the implications of using the normal distribution, two alternative approaches were used. In the first, the normal distribution was replaced with a heavy-tailed Student-t distribution with a number of degrees equal to the estimated tail index of the empirical distribution. The other one was based on the approximation of the tails of the empirical distribution by GPD. The latter approach came as a clear winner from the comparison with estimates very close to their empirical counterparts, followed by the Student-t distribution. The normal distribution proved to be inadequate, in particular for higher confidence levels.

VaR has been frequently criticised on two grounds. That it says nothing about the severity of losses when it is exceeded and that it is not a coherent risk measure as defined by Artzner et al. (1999). To deal with these shortcomings, a quantile-based alternative, Expected Shortfall (ES), was developed. It is defined as the expected value of loss, given that VaR is exceeded.

Similar exercise as in case of VaR was performed in order to compare estimates based on the respective distributional assumptions. Unsurprisingly, ES estimates also confirmed the superiority of GPD in approximating the tail behaviour. The Student-t distribution too delivered reasonable estimates, especially compared to the normal distribution which is absolutely unsuitable for estimation of ES.

Although we made a strong case against the use of normal distribution in financial applications in this thesis, it is fair to acknowledge that it still has a role to play in finance. As Costa et al. (2005) show, for investors who do not resort to derivatives or particularly risky asset management strategies and hold mainly Treasury Bills and a few equities it remains a reliable and useful statistical tool for managing financials risks.

The bottom line is that one should be very cautious when she intends to use the normal distribution to approximate the behaviour of returns. It can make her life much easier but at the same time it can lead to a serious miscalculation of risks she is faced with. Recent financial crisis provides the best warning.

REFERENCES

Acerbi, C., Tasche, D. (2001): Expected Shortfall: A Natural Coherent Alternative to Value at Risk; Economic Notes by Banca Monte dei Paschi di Siena S.p.A., Vol. 31, No. 2-2002, pp. 379–388.

Artzner, P., Delbaen, F., Eber, J.-M., Heath, D. (1999): Coherent Measures of Risk; Mathematical Finance, Vol. 9, No. 3, pp. 203–228.

Bachelier, L. (1900): Théorie de la Spéculation; Annales Scientifiques de l'École Normale Supérieure, Vol. 3, No. 17, pp. 21-86. Translated in Cootner, P. H. (1964): The Random Character of Stock Market Prices; MIT Press, Cambridge, MA, USA.

Balkema, A. A., de Haan, L. (1974): Residual Life Time at Great Age; The Annals of Probability, Vol. 2, No. 5, pp. 792-804.

Barndorff-Nielsen, O. E. (1997): Normal Inverse Gaussian Distributions and Stochastic Volatility Modelling; Scandinavian Journal of Statistics, Vol. 24, No. 1, pp. 1-13.

Beirlant, J., Vynckier, P., Teugels, J. (1996): Tail Index Estimation, Pareto Quantile Plots, and Regression Diagnostics; Journal of the American Statistical Association, Vol. 91, pp. 1659-1667.

Beirlant, J., Goegebeur, Y., Segers, J., Teugels, J. (2004): Statistics of Extremes: Theory and Applications; John Wiley & Sons Ltd., Chichester, West Sussex, England.

Black, F., Scholes, M. (1973): The Pricing of Options and Corporate Liabilities; Journal of Political Economy, Vol. 81, No. 3, pp. 637-654.

Blattberg, R. C., Gonedes, N. J. (1974): A Comparison of the Stable and Student Distributions as Statistical Models for Stock Prices; Journal of Business, Vol. 47, No. 2, pp. 244-280.

Campbell, J. Y., Hentschel, L. (1992): No News is Good News: An Asymmetric Model of Changing Volatility in Stock Returns; Journal of Finance, Vol. 31, No. 3, pp. 281-318.

Cont, R. (2001): Empirical Properties of Asset Returns: Stylized Facts and Statistical Issues; Quantitative Finance Volume I, pp. 223-236.

Cont, R., Potters, M., Bouchaud, J.-P. (1997): Scaling in Stock Market Data: Stable Laws and Beyond; In: Dubrulle, B., Graner, F., Sornette, D. (eds, 1997): Scale Invariance and Beyond: Les Houches Workshop, March 10-14, 1997; Springer, Berlin, Germany.

Costa, M., Cavaliere, G., Iezzi, S. (2005): The Role of the Normal Distribution in Financial Markets; In: Vichi, M., Monari, P, Mignani, S., Montanavi, A. (2005): New Developments in Classification and Data Analysis, pp. 343-350.

Dacorogna, M. M., Müller, U. A., Pictet, O. V., de Vries, C. G. (1995): The Distribution of Extremal Foreign Exchange Rate Returns in Extremely Large Data Sets; Tinbergen Institute Discussion Paper No. 95-70.

Danielsson, J., de Vries, C. G. (1997): Beyond the Sample: Extreme Quantile and Probability Estimation; Tinbergen Institute Discussion Paper No. 98-016/2.

Danielsson, J., de Vries, C. G. (2000): Value at Risk and Extreme Returns; FMG-Discussion Paper No. 273, Financial Markets Group, London School of Economics.

Danielsson, J., de Vries, C. G., Jorgensen, B. N., Mandira, S. (2006): Comparing Downside Risk Measures for Heavy Tailed Distributions; Economic Letters, Vol. 92, No. 2, pp. 202-208.

Danielsson, J., de Vries, C., Jorgensen, B., Samorodnitsky, G., Sarma, M. (2011): Fat Tails, Value-at-Risk and Subadditivity; available at http://www.riskresearch.org/?papid=35&catid=0#paper_detail.

de Haan, L., Ferreira, A. (2006): Extreme Value Theory: An Introduction; Springer Science+Business Media, LLC, New York, NY, USA.

de Haan, L., Peng, L. (1998): Comparison of Tail Index Estimators; Statistica Neerlandica, Vol. 52, No. 1, pp. 60-70.

de Haan, L., Resnick, S. I. (1980): A Simple Asymptotic Estimate for the Index of Stable Distribution; Journal of the Royal Statistical Society, Series B, Vol. 42, No. 1, pp. 83-87.

de Haan, L., Resnick, S. I. (1998): On Asymptotic Normality of the Hill Estimator; Stochastic Models, Vol. 14, No. 4, pp. 849–866.

Dekkers, A. L. M., de Haan, L. (1989): On the Estimation of the Extreme-value Index and Large Quantile Estimation; The Annals of Statistics, Vol. 17, No. 4, pp. 1795-1832.

Dekkers, A. L. M., Einmahl, J. H. J., de Haan, L. (1989): A Moment Estimator for the Index of an Extreme-Value Distribution; The Annals of Statistics, Vol. 17, No. 4, pp. 1833-1855.

Dhaene, J., Goovaerts, M. J., Kaas, R. (2003): Economic Capital Allocation Derived from Risk Measures; North American Actuarial Journal, Vol. 7, No. 2, pp. 44-59.

Diebold, F. X., Schuermann, T., Stroughair, J. D. (1998): Pitfalls and Opportunities in the Use of Extreme Value Theory in Risk Management; In: Refenes, A.-P. N., Moody, J. D., Burgess, A. N. (eds.): Advances in Computational Finance; Kluwer Academic Publishers, Amsterdam, pp. 3-12.

Duffie, D., Pan, J. (1997): An Overview of Value at Risk; Journal of Derivatives, Vol. 4, No. 3, pp. 7-49.

DuMouchel, W. H. (1983): Estimating the Stable Index α in Order to Measure Tail Thickness: A Critique; The Annals of Statistics, Vol. 11, No. 4, 1019-1031.

Eberlein, E., Keller, U., Prause, K. (1998): New Insights into Smile, Mispricing and Value at Risk: The Hyperbolic Model; Journal of Business, Vol. 71, No. 3, pp. 371-405.

Embrechts, P., Klüppelberg, C., Mikosch, T. (1997): Modelling Extremal Events for Insurance and Finance; Springer-Verlag, Berlin.

Engle, R. F. (1982): Autoregressive Conditional Heteroskedastic Models with Estimates of the Variance of United Kingdom Inflation; Econometrica, No. 50, pp. 987-1007.

Fama, E. (1963): Mandelbrot and the Stable Paretian Hypothesis; Journal of Business, Vol. 36, No. 4, pp. 420-429.

Fisher, R., Tippett, L. (1928): Limiting Forms of the Frequency Distribution of the Largest or Smallest Member of a Sample; Proceedings of the Cambridge Philosophical Society, Vol. 24, No. 2, pp. 180-190.

Ghose, D., Koroner, K. F. (1995): The Relationship between GARCH and Symmetric Stable Processes: Finding the Source of Fat Tails in Financial Data; Journal of Empirical Finance, Vol. 2, No. 3, pp. 225-251.

Gilli, M., Këllezi, E. (2006): An Application of Extreme Value Theory for Measuring Financial Risk; Computational Economics, Vol. 27, No. 2-3, pp. 207-228.

Gnedenko, B.V. (1943): Sur La Distribution Limite Du Terme Maximum D'une Série Aléatoire; Annals of Mathematics, Vol. 44, No. 3, pp. 423-453.

Gomes, M. I., Canto e Castro, L., Fraga Alves, M. I., Pestana, D. (2008): Statistics of Extremes for IID Data and Breakthroughs in the Estimation of the Extreme Value Index: Laurens de Haan Leading Contributions; Extremes, Vol. 11, No. 1, pp. 3-34.

Gumbel, E. J. (1958): Statistics of Extremes; Columbia University Press, New York.

Hall, P. (1982): On Some Simple Estimates of an Exponent of Regular Variation; Journal of the Royal Statistic Society, Series B, Vol. 44, No. 1, pp. 37-42.

Hill, B. M. (1975): A Simple General Approach to Inference about the Tail of a Distribution; Annals of Statistics, Vol. 3, No. 5, pp. 1163-1174.

Horák, J., Šmíd, M. (2009): On Tails of Stock Returns: Estimation and Comparison between Stocks and Markets; available at SSRN: http://ssrn.com/abstract=1365229.

Hsing, T. (1991): On Tail Index Estimation Using Dependent Data; The Annals of Statistics, Vol. 19, No. 3, pp. 1547-1569.

Huisman, R., Koedijk, K., Kool, C., Palm, F. (1997): Fat Tails in Small Samples; available at SSRN: http://ssrn.com/abstract=51141.

Huisman, R., Koedijk, K. G., Pownall, A. J. (1998): VaR-x: Fat Tails in Financial Risk Management; Journal of Risk, Vol. 1, No. 1, pp. 47-61.

Huisman, R., Koedijk, K., Kool, C., Palm, F. (2002): The Tail Fatness of FX Returns Reconsidered; De Economist, No. 150, pp. 299-312.

Jansen, D. W., de Vries, C. G. (1991): On the Frequency of Large Stock Returns: Putting Booms and Busts into Perspective; The Review of Economics and Statistics, Vol. 73, No. 1, pp. 18-24.

Jondeau, E., Rockinger, M. (2003): Testing for Differences in the Tails of Stock-Market Returns; Journal of Empirical Finance, Vol. 10, No. 5, pp. 559-581.

Jorion, P. (2001): Value at Risk: the New Benchmark for Managing Financial Risk, 2nd Edition; McGraw-Hill, New York, USA.

Kearns, P., Pagan, A. (1997): Estimating the Density Tail Index for Financial Time Series; Review of Economics and Statistics, Vol. 79, No. 2, pp. 171-175.

Kon, S. (1984): Models of Stock Returns – A Comparison; Journal of Finance, Vol. 39, No. 1, pp. 147-165.

LeBaron, B., Samanta, R. (2004): Extreme Value Theory and Fat Tails in Equity Markets; available at SSRN: http://ssrn.com/abstract=873656.

Lintner, J. (1965): The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolio and Capital Budgets; Review of Economics and Statistics, Vol. 47, No. 1, pp. 13-37.

Longin, F. M. (1996): The Asymptotic Distribution of Extreme Stock Market Returns; Journal of Business, Vol. 69, No. 3, pp. 383-408.

Loretan, M., Phillips, P. C. B. (1994): Testing the Covariance Stationarity of Heavy-tailed Time Series; Journal of Empirical Finance, Vol. 1, pp. 211-248.

Mandelbrot, B. (1963): The Variation of Certain Speculative Prices; Journal of Business, Vol. 36, No. 4, pp. 394-419.

Mandelbrot, B., Hudson, R. L. (2008): The (Mis)Behaviour of Markets: A Fractal View of Risk, Ruin and Reward, Profile Books Ltd., London, UK.

Markowitz, H. (1952): Portfolio Selection; Journal of Finance, Vol. 7, No. 1, pp. 77-91.

McNeil, A. J. (1998): Calculating Quantile Risk Measures for Financial Return Series Using Extreme Value Theory; ETH Department Mathematik Report, available at http://e-collection.ethbib.ethz.ch/view/eth:25077.

McNeil, A. J. (1999): Extreme Value Theory for Risk Managers; In: Internal Modelling and CAD II: Qualifying and Quantifying Risk within a Financial Institution, RISKbooks, pp. 93-113.

McNeil, A. J., Frey, R. (2000): Estimation of Tail-related Risk Measures for Heteroskedastic Financial Time Series: An Extreme Value Approach; Journal of Empirical Finance, Vol. 7, No. 3-4, pp. 271-300.

McNeil, A. J., Frey, R., Embrechts, P. (2005): Quantitative Risk Management: Concepts, Techniques and Tools; Princeton University Press, Princeton, New Jersey, USA.

Mikosch, T., Starica, C. (2000): Limit Theory for the Sample Autocorrelations and Extremes of a GARCH(1,1) Process; The Annals of Statistics, Vol. 28, No. 5, pp. 1427-1451.

Pickands III, J. (1975): Statistical Inference Using Extreme Order Statistics; The Annals of Statistics, Vol. 3, No. 1, pp. 119-131.

Press, S. J. (1967): A Compounds Events Model for Security Prices; Journal of Business, Vol. 40, No. 3, pp. 317-335.

Resnick, S. (2007): Heavy-Tail Phenomena: Probabilistic and Statistical Modeling. Springer Science+Business Media, LLC, New York.

Resnick, S., Starica, C. (1998): Tail Index Estimation for Dependent Data; The Annals of Applied Probability, Vol. 8, No. 4, pp. 1156-1183.

Roy, A. D. (1952): Safety First and Holding of Assets; Econometrica, Vol. 20, No. 3, pp. 431-449.

Sahinler, S., Topus, D. (2007): Bootstrap and Jackknife Resampling Algorithms for Estimation of Regression Parameters; Journal of Applied Quantitative Methods, Vol. 2, No. 2, pp. 188-199.

Schirmacher, D., Schirmacher, E., Thandi, N. (2005): Stochastic Excess-of-Loss Pricing within a Financial Framework; In: Casualty Actuarial Society Forum, Spring 2005 Edition, pp. 297-352; available at http://www.casact.org/pubs/forum/05spforum/.

Sharpe, W. F. (1964): Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk; Journal of Finance, Vol. 19, No. 3, pp. 237-260.

Zivot, E., Wang J. (2006): Modeling Financial Time Series with S-PLUS®, Second Edition; Springer Science+Business Media, Inc., New York, USA.

APPENDIX A

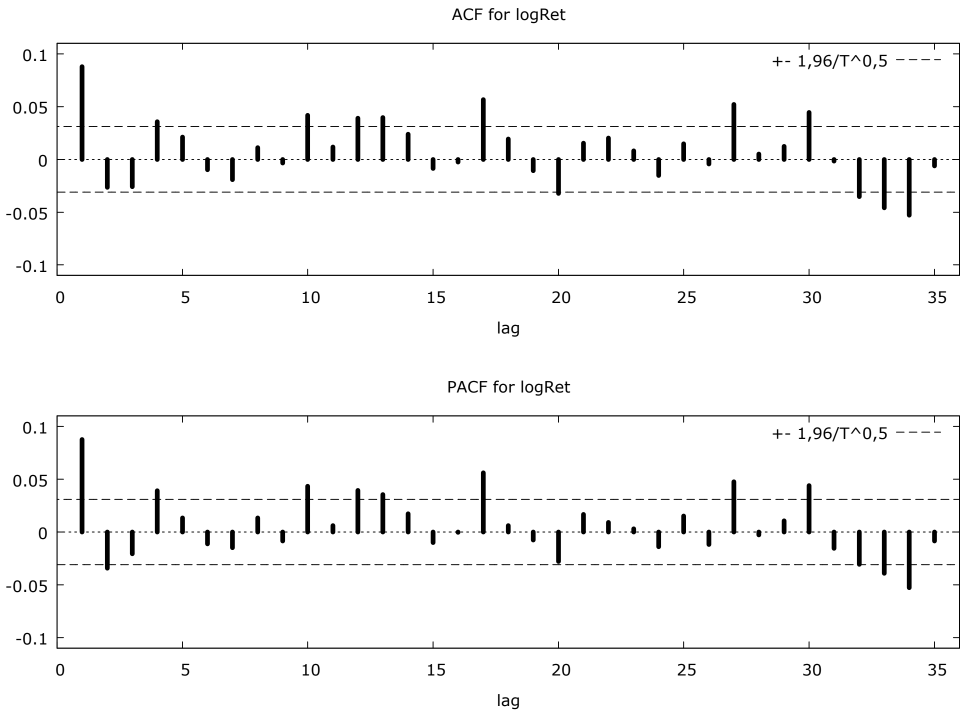

Figure 29 Correlogram of PX returns. Source: author's calculations.

LAG	ACF		PACF		Q-stat.	[p-value]
1	0,0878	***	0,0878	***	30,8143	[0,000]
2	-0,0264	*	-0,0344	**	33,6047	[0,000]
3	-0,0258		-0,0206		36,2664	[0,000]
4	0,0356	**	0,0392	**	41,3439	[0,000]
5	0,0212		0,0132		43,1411	[0,000]
6	-0,0097		-0,0114		43,5150	[0,000]
7	-0,0193		-0,0147		45,0008	[0,000]
8	0,0112		0,0133		45,5002	[0,000]
9	-0,0036		-0,0086		45,5508	[0,000]
10	0,0417	***	0,0435	***	52,5091	[0,000]
11	0,0117		0,0061		53,0628	[0,000]
12	0,0390	**	0,0395	**	59,1683	[0,000]
13	0,0397	**	0,0355	**	65,4914	[0,000]
14	0,0238		0,0172		67,7690	[0,000]
15	-0,0086		-0,0101		68,0684	[0,000]
16	-0,0025		-0,0003		68,0930	[0,000]
17	0,0565	***	0,0561	***	80,9331	[0,000]
18	0,0193		0,0061		82,4326	[0,000]
19	-0,0108		-0,0076		82,9057	[0,000]
20	-0,0322	**	-0,0279	*	87,0842	[0,000]
21	0,0153		0,0168		88,0203	[0,000]
22	0,0202		0,0090		89,6556	[0,000]
23	0,0081		0,0030		89,9187	[0,000]
24	-0,0151		-0,0141		90,8381	[0,000]
25	0,0149		0,0153		91,7294	[0,000]
26	-0,0043		-0,0119		91,8053	[0,000]
27	0,0519	***	0,0477	***	102,6699	[0,000]
28	0,0052		-0,0027		102,7788	[0,000]
29	0,0122		0,0107		103,3786	[0,000]
30	0,0443	***	0,0438	***	111,2993	[0,000]
31	-0,0015		-0,0155		111,3087	[0,000]
32	-0,0352	**	-0,0306	*	116,3126	[0,000]
33	-0,0460	***	-0,0390	**	124,8576	[0,000]
34	-0,0530	***	-0,0528	***	136,1898	[0,000]
35	-0,0061		-0,0084		136,3383	[0,000]

Table 14 Autocorrelation and partial autocorrelation function of PX returns. Source: author's calculations.

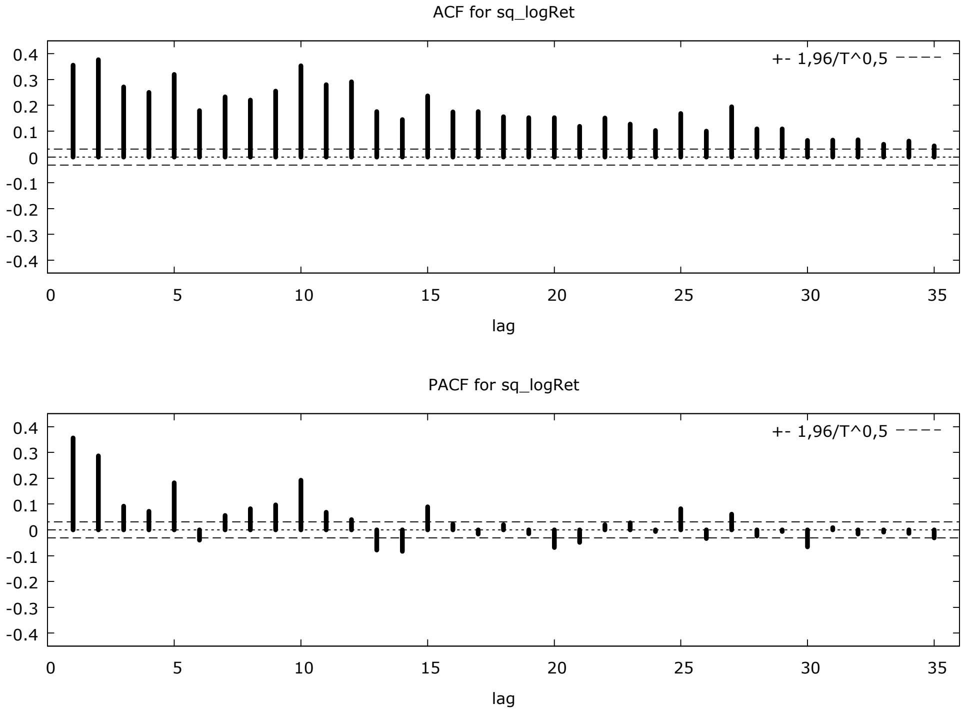

Figure 30 Correlogram of squared PX returns. Source: author's calculations.

LAG	ACF	Sig.	PACF	Sig.	Q-stat.	[p-value]
1	0,3559	***	0,3559	***	506,8874	[0,000]
2	0,3772	***	0,2869	***	1076,3237	[0,000]
3	0,2712	***	0,0915	***	1370,7134	[0,000]
4	0,2506	***	0,0715	***	1622,0700	[0,000]
5	0,3201	***	0,1828	***	2032,4272	[0,000]
6	0,1794	***	-0,0394	**	2161,3948	[0,000]
7	0,2328	***	0,0564	***	2378,5847	[0,000]
8	0,2201	***	0,0813	***	2572,7510	[0,000]
9	0,2556	***	0,0973	***	2834,5729	[0,000]
10	0,3527	***	0,1922	***	3333,4990	[0,000]
11	0,2795	***	0,0683	***	3646,9442	[0,000]
12	0,2907	***	0,0398	**	3986,0878	[0,000]
13	0,1756	***	-0,0782	***	4109,8059	[0,000]
14	0,1452	***	-0,0830	***	4194,4397	[0,000]
15	0,2362	***	0,0895	***	4418,4448	[0,000]
16	0,1739	***	0,0227		4539,9504	[0,000]
17	0,1760	***	-0,0165		4664,4419	[0,000]
18	0,1565	***	0,0195		4762,8716	[0,000]
19	0,1526	***	-0,0144		4856,4068	[0,000]
20	0,1521	***	-0,0685	***	4949,3588	[0,000]
21	0,1185	***	-0,0488	***	5005,8210	[0,000]
22	0,1512	***	0,0204		5097,7230	[0,000]
23	0,1270	***	0,0278	*	5162,6301	[0,000]
24	0,1022	***	-0,0058		5204,6768	[0,000]
25	0,1681	***	0,0817	***	5318,3806	[0,000]
26	0,1001	***	-0,0336	**	5358,7478	[0,000]
27	0,1944	***	0,0611	***	5510,9466	[0,000]
28	0,1090	***	-0,0231		5558,8075	[0,000]
29	0,1090	***	-0,0060		5606,6953	[0,000]
30	0,0637	***	-0,0652	***	5623,0328	[0,000]
31	0,0650	***	0,0091		5640,0734	[0,000]
32	0,0670	***	-0,0160		5658,1811	[0,000]
33	0,0496	***	-0,0086		5668,1148	[0,000]
34	0,0616	***	-0,0135		5683,4447	[0,000]
35	0,0427	***	-0,0307	*	5690,8098	[0,000]

Table 15 Autocorrelation and partial autocorrelation function of squared PX returns. Source: author's calculations.

APPENDIX B

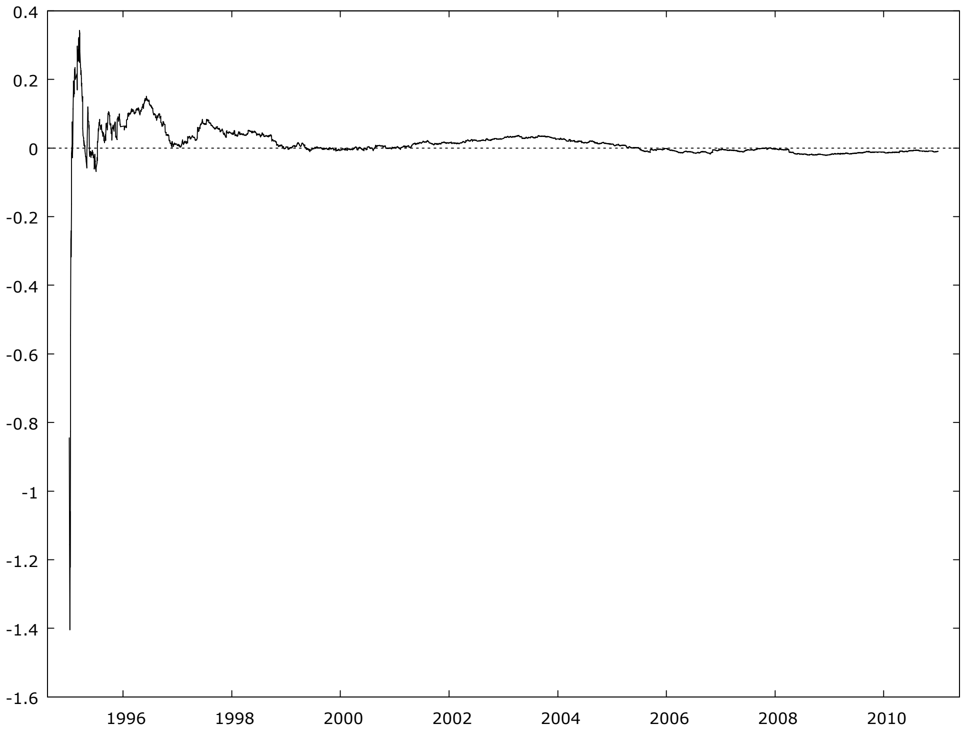

Figure 31 Cumulative mean of a random sample drawn from Student-t distribution with 3 degrees of freedom. Source: author's calculations.

Figure 32 Cumulative variance of a random sample drawn from Student-t distribution with 3 degrees of freedom. Source: author's calculations.

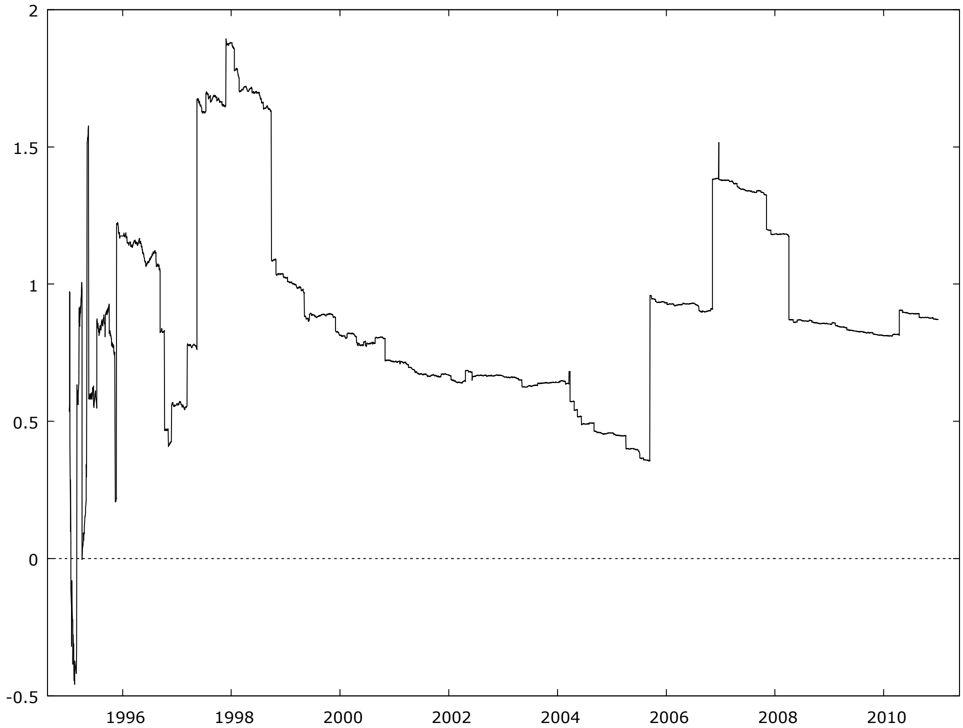

Figure 33 Cumulative skewness of a random sample drawn from Student-t distribution with 3 degrees of freedom. Source: author's calculations.

Figure 34 Cumulative kurtosis of a random sample drawn from Student-t distribution with 3 degrees of freedom. Source: author's calculations.