ISM UNIVERSITY OF MANAGEMENT AND ECONOMICS
MASTER OF SCIENCE IN FINANCIAL ECONOMICS PROGRAMME

Darius Jurgilas

MASTER'S THESIS

MEASURING AND FORECASTING VALUE AT RISK FOR NORDIC AND BALTIC STOCK INDEXES

Supervisor
Dr. Dalia Kaupelytė

VILNIUS, 2012

Abstract

Jurgilas, D. Measuring and forecasting Value at Risk for Nordic and Baltic stock indexes.[Manuscript]: Master Thesis: economics. Vilnius, ISM University of Management and Economics, 2012.

The aim of the thesis is to evaluate performance of 17 Value at Risk (VaR) models in forecasting one-day VaR of four Nordic and three Baltic stock indexes and to choose best models for both low volatility period before financial crisis and high volatility period during and after financial crisis. Five unconditional volatility VaR models and VaR models with three volatility specifications – EWMA, GARCH, APARCH – with three return distribution (normal, Student's t and skewed Student's t) assumptions as well as extreme value theory VaR models are estimated at 1 percent and 5 percent level of significance and compared. Data consists of daily closing prices of four Nordic (OMX Stockholm, OMX Copenhagen, OMX Helsinki and Oslo Stock Exchange All Share) and three Baltic (OMX Vilnius, OMX Riga and OMX Tallinn) stock indexes from December 31, 1999 to December 30, 2011. Forecasts of all models are backtested and performance is compared using violation ratios, Kupiec unconditional coverage test and Christoffersen conditional coverage test. Initial statistical and graphical analysis of return series shows that all index return distributions have fat tails and are skewed. Another finding is that volatility models with Student's t distribution and skewed Student's t distribution better fit index return data than with normal return distribution. Backtesting indicate that many conditional volatility VaR models with non-normal return distribution produce good return forecasts during low volatility period of 2004-2007 for all stock indexes examined. However, only APARCH(1,1) VaR model with Student's t and skewed Student's t distribution assumption as well as EVT VaR models with fat-tailed GARCH filter are able to produce accurate VaR forecasts for Baltic stock markets during high volatility period of year 2008-2011. Nordic stock markets can be predicted using more VaR models, however, the same models as for Baltic stock markets, show best forecasting performance. Another finding is that all VaR models perform better at higher level of confidence and more VaR models accurately predict Nordic stock market returns than Baltic stock market returns. Unconditional VaR models failed most of the time due to volatility clustering present in historical stock index returns. Implication of the thesis is that VaR forecasting models are applicable to Baltic and Nordic stock markets and that risk managers should enhance their quantitative modeling capabilities to measure risk by using more advanced models which account for stylized facts of financial returns. Further research could examine VaR forecasting for longer period than one-day and apply VaR models and backtesting for other asset classes and their portfolios in financial markets of Nordic and Baltic countries.

Key words: Value at Risk, GARCH, APARCH, Extreme Value Theory, Nordic and Baltic stock indexes.

Santrauka

Jurgilas, D. Šiaurės Europos ir Baltijos šalių akcijų indeksų rizikos vertės matavimas ir prognozavimas [Rankraštis]: magistro baigiamasis darbas: ekonomika. Vilnius, ISM Vadybos ir ekonomikos universitetas, 2012.

Darbo tikslas yra įvertinti 17 rizikos vertės (angl.Value at Risk - VaR) modelių galimybes prognozuoti vienos dienos rizikos vertę keturiems Šiaurės Europos šalių ir trims Baltijos šalių akcijų indeksams ir surasti geriausius rizikos vertės modelius, pateikiančius tiksliausias prognozes ramiu prieškriziniu laikotarpiu ir didelių svyravimų (angl. volatility) finansų rinkose laikotarpiu finansų krizės metu ir po jos. Darbe modeliuojami ir įvertinami penki nesąlyginiai (angl. unconditional), t.y. neatsižvelgiantys į rinkų svyravimų dydį, rizikos vertės modeliai, taip pat rizikos vertės modeliai, kurie modeliuoja rinkų svyravimą naudodami tris ekonometrinius modelius EWMA, GARCH, APARCH su trimis grąžų pasiskirstymo prielaidomis – normaliuoju, Stjudento ir asimetrišku Stjudento skirstiniu. Ekstremaliųjų grąžų finansų rinkose modeliai taip pat naudojami apskaičiuoti rizikos vertę. Visi modeliai apskaičiuoti taikant vieno procento ir penkių procentų reikšmingumo lygmenį (vienpusiu 95 ir 99 procentų pasikliauties intervalu). Modeliavimui naudoti duomenys yra keturių Šiaurės Europos (OMX Stokholmas, OMX Kopenhaga, OMX Helsinkis ir Oslo) bei trijų Baltijos (OMX Vilnius, OMX Ryga ir OMX Talinas) akcijų indeksų vertės nuo 1999.12.31 iki 2011.12.30. Modeliai testuojami ir jų prognozavimo galimybės lyginamos remiantis pažeidimų santykiniais rodikliais (angl. violation ratios) bei statistiniais Kupiec ir Christoffersen tikėtinumo testų rezultatais. Pirminė statistinė ir grafinė indeksų grąžų laiko eilutės analizė parodė, kad visų analizuotų akcijų indeksų grąžų skirstiniai turi sunkiąsias uodegas ir yra asimetriški. Modelių prognozavimo testavimas parodė, kad dauguma rizikos vertės modelių, įvertinančių rinkų svyravimų dydį ir įvertindami prielaidą, kad indeksų grąžos pasiskirsčiusios ne pagal normalųjį skirstinį, sugebėjo neblogai prognozuoti indeksų grąžas ramiuoju 2004-2007 metų periodu visiems analizuotiems akcijų indeksams. Tačiau tik APARCH(1,1) rizikos vertės modelis su Stjudento ir asimetriško Stjudento skirstinio prielaida bei ekstremalių grąžų rizikos vertės modeliai su GARCH grąžų nepastovumo (angl. volatility) filtru sugebėjo pateikti priimtinas Baltijos šalių akcijų indeksų grąžų vienos dienos prognozes neramiu 2008-2011 metų laikotarpiu. Šiaurės Europos akcijų rinkoms prognozuoti finansų krizės laikotarpiu tiko daugiau rizikos vertės modelių, tačiau geriausias prognozes pateikė tie patys modeliai kaip ir Baltijos akcijų indeksams. Kita darbo išvada yra ta, kad modeliai pateikė geresnes prognozes esant aukštesniam (1% vietoj 5%) reikšmingumo lygmeniui ir kad Baltijos šalių akcijų rinkos yra sunkiau prognozuojamos nei Šiaurės Europos. Nesąlyginiai rizikos vertės modeliai pateikė prastus rezultatus, nes neatsižvelgė į grąžų kintamumo grupavimąsi (angl. volatility clustering). Darbas parodo, jog įmanoma taikyti sudėtingus rizikos vertės prognozavimo modelius Šiaurės Europos ir Baltijos akcijų indeksams ir kad rizikų valdytojai turėtų sustiprinti savo kiekybinio modeliavimo sugebėjimus matuoti rizikas panaudojant pažangius rizikos vertės modelius. Tolimesni tyrimai galėtų prognozuoti rizikos vertę ilgesniam nei vienos dienos periodui bei taikyti rizikos vertės modelius kitoms turto klasėms bei jų portfeliams Šiaurės Europos ir Baltijos šalių finansų rinkose.

Reikšminiai žodžiai: Rizikos vertė (VaR), GARCH, APARCH, Ekstremaliųjų reikšmių teorija, Šiaurės Europos ir Baltijos šalių akcijų indeksai

Table of Contents

List of Figures

List of Tables

Introduction

Value at Risk (VaR) is one of the most important tools used in risk management in the last two decades. It became even more important during recent years because trading activity in financial markets has grown tremendously and many well-known financial institutions faced big trading losses during recent financial crisis. Therefore financial regulators and supervisory committees favored using quantitative techniques such as VaR to evaluate possible losses that these institutions can incur. VaR provides answer to a question what is possible financial loss over given time horizon with given level of confidence? To put it in other words, usually Value at Risk of investment portfolio is a maximum amount possible to loose during short period of time such as one day with given level of confidence.

VaR modeling methodologies using GARCH volatility models (Bollerslev, 1986) and Extreme Value Theory (Embrechts, Kluppelberg, & Mikosch, 1997) were created to account for non-normality and skewness of return distribution as well as for stylized facts of financial returns such as volatility clustering and leverage effect. Many researchers compared forecasting performance of many of these models for developed and emerging stock markets around the world. However, Nordic and Baltic region had little attention. Furthermore, recent financial crisis have caused many models to fail. All these reasons and disagreement about which VaR model best accommodates market risk supports the relevance of further research into VaR. Therefore various VaR measurement techniques are evaluated particularly for Nordic and Baltic stock market indexes in order to find best models which could be used by financial institutions in those stock markets to manage market risk and to solve the following problem.

The **problem** to solve in this research is to find which VaR models provide best forecasts for particular stock markets during different time periods and financial conditions.

The **aim** of the thesis is to evaluate performance of 17 Value at Risk (VaR) models in forecasting one-day VaR of four Nordic and three Baltic stock indexes and to choose best models for both low volatility period before financial crisis and high volatility period during and after financial crisis. To reach the aim of the thesis, the following objectives are raised:

1. To analyze and compare various empirical researches on measuring and forecasting VaR for both developed and emerging stock markets and to find possible gaps which could be resolved in this thesis.

2. To measure properties of financial returns of Nordic and Baltic stock indexes such as normality and skewness and to check whether stylized facts such as volatility clustering are present in the data.
3. To present and estimate various volatility and distribution models. To find best fitting models which accurately describe financial return properties of Nordic and Baltic stock indexes.
4. To use many VaR forecasting models and create forecasts of possible losses with given levels of confidence and evaluate forecasting performance. To find best VaR models for Nordic and Baltic stock indexes before and after financial crisis.

The main aspects of **methodological approach** is that unconditional VaR models such as Historical simulation, normal VaR, Student's t VaR and conditional VaR models, where volatility is forecasted using Exponentially Weighted Moving Average (EWMA), Generalized Autoregressive Conditional Heteroskedastic (GARCH) and other models, are evaluated using daily return data of seven stock indexes in Nordic and Baltic stock exchanges for the period from year 2000 to the end of 2011. One-day forecasts are calculated assuming that stock index historical return data is sufficient information for model evaluation and forecasting. Models are estimated with 95% and 99% confidence level and assuming that Nordic stock markets are developed (less volatile), while in Baltic countries stock markets are small and still considered as developing (more volatile). In order to evaluate forecasting performance, several backtesting methods are performed including Kupiec test and Christoffersen test. Graphical analysis of forecasted returns compared with actual returns is presented to analyze forecasting behavior of the models.

The whole quantitative data analysis is performed using programming language and software environment for statistical computing called R. In order to estimate advanced distribution and volatility models, special packages are used: fGarch (Wuertz, Chalabi, & others, 2009) and fExtremes (Wuertz & others, 2009). Many good examples of applications of these packages can be found in Danielsson (2011).

The **implication** of the thesis is that it expands empirical research on VaR by measuring it for Nordic and Baltic stocks indexes and evaluates many simple and advanced VaR models for time periods before and after financial crisis in 2008, giving valuable insight about VaR model forecasting performance during calm and volatile periods. Findings of the thesis could also be useful for financial practitioners, particularly risk managers in Nordic and Baltic stock markets, who are responsible for market risk management.

Thesis consists of seven parts. First part introduces concept of Value at Risk and overviews main parts of the thesis. Second and third chapters summarize and critically review empirical research on modeling of Value at Risk and defines research problem. In Methodology part theory of VaR models are presented as well as methods for both VaR estimation and comparison are explained. Fifth section presents empirical results of measuring and forecasting VaR for Nordic and Baltic stock indexes. These empirical results are discussed in sixth section of the thesis, where implications and limitations are also presented. Finally, conclusions are stated in the last part of master thesis.

1. Review of empirical research

This section begins briefly introducing Value at Risk concept and presents review of literature on Value at Risk measurement and applications in financial markets. Firstly, overview of the literature which describe and implement more advanced VaR techniques with various volatility models are presented. After that, empirical research on Extreme Value Theory in VaR measurement is reviewed and critically evaluated. Finally, empirical research performed in various financial markets worldwide is shortly overviewed and summary of many analyzed empirical studies are provided.

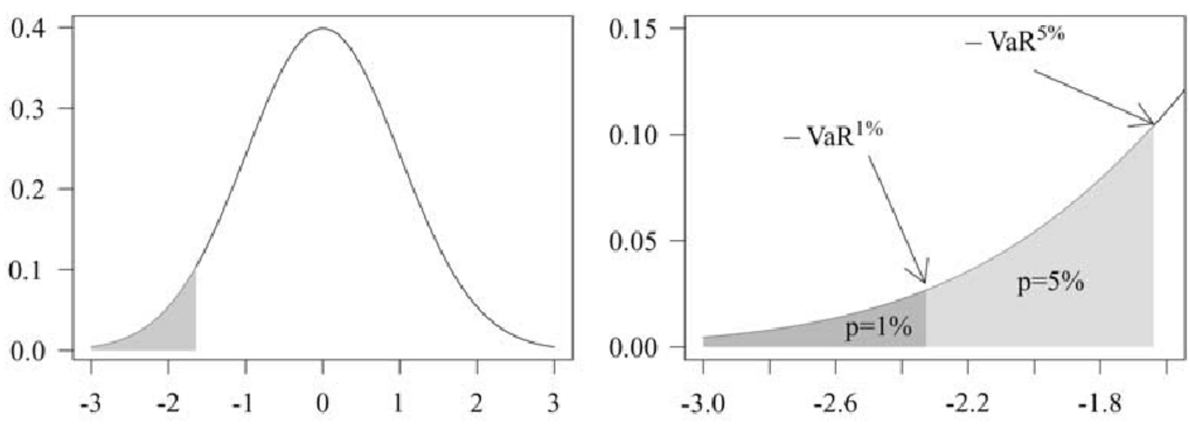
1.1. Definition of Value at Risk

Value at Risk is defined as “the worst loss over a target horizon that will not be exceeded with a given level of confidence” (Jorion, 2007, p. 17). In other words, it is some percentile of the distribution generated from the expected gains and losses of portfolio. If we denote α as significance level (for example 5% or 1%), V_t and V_{t-1} as portfolio values at time t and $t-1$ respectively, ΔV - as change in portfolio value, we can define VaR in statistical manner accordingly:

$$\begin{aligned} &\Delta V = V_t - V_{t-1} \\ &\Pr(\Delta V < -VaR) = \alpha \end{aligned} \qquad (1)$$

In Figure 1 definition of VaR is shown graphically, left panel shows the entire density of standardized return distribution assumed to be standard normal distribution. Right panel zooms in on the left tail, where the shaded areas identify the 1% and 5% probabilities (the area under the curve from negative infinity to negative VaR equals 0.01 and 0.05, respectively) (Danielsson, 2011, p. 77).

Figure 1: Return density function of st. normal distribution and VaR (left panel) and 1% VaR and 5% VaR of left tail (right panel).



Source: Danielsson (2011).

VaR depends on three parameters: confidence (or significance) level, time horizon and portfolio value. Time variable, usually measured in trading days, tells for what period is the prediction of VaR and measurement of potential losses is performed. The longer the time period is the higher the possible losses are. Horizon is chosen depending on the financial entity's activity type – traders are interested in one-day VaR, while for pension and investment fund managers it is more informative when measured for longer periods such as one year. According to Alexander (2008b), "often the significance level is set by external body, such as banking regulator. Under Basel II Accord, banks using internal VaR models to assess their market risk capital requirement should measure VaR at the 1% significance level, i.e. the 99% confidence level." However, for short data samples, confidence level can be lower, usually 95%, while sometimes, when very conservative risk management approach is applied, confidence level can be as high as 99.9% or even 99.99%.

Research on VaR can be dated back to the portfolio optimization theory by Markowitz (1952). However, it did not become a widely accepted measure until the end of 1980s. Stock crash in 1987 called Black Monday triggered financial world. Many academically-trained quantitative analysts started working on new risk management framework. VaR became popular measure of risk in 1994, when famous RiskMetrics were published by J.P. Morgan investment bank (Longerstaey & Zangari, 1996). Another big stimulus for research on VaR was created in 1999 when Basel II Accord (BIS, 2006) stated that preferred measure for market risk is VaR. Since then, research on VaR exploded and is still ongoing process of improving risk management.

1.2. Review of empirical research on modeling and measuring VaR

VaR models have been researched extensively for more than twenty years. Many different methodologies were created with their advantages and disadvantages. In order to improve predicting power and accuracy of different VaR models, many researchers worked on advanced approaches. Many advanced techniques such as Extreme Value Theory (EVT) or class of Autoregressive Conditional Heteroskedastic (ARCH) models were applied to various asset classes and different regions in the world; however, different results from these studies show that there is no best model for everything, so it is a useful contribution to test various methods in particular region and on specific asset class using most recent financial data from the market including recent financial crisis.

One of the first studies regarding VaR accuracy and usefulness for different assets was performed by Beder (1995). Author tested two classical VaR models – Historical simulation (HS) and

Monte Carlo simulation (MCS) – using two data sets with 100 and 250 trading days for HS model and two different correlation estimates for MCS model and applied it for two holding periods – one day and two weeks. Three portfolios consisting of bonds, stock index and options, as well as the mix of these assets, were used to calculate VaR. Author found that the results of VaR calculation was different up to 14 times for the same portfolio! Her study showed that VaR measures are very dependent on different parameters, data used for calculation, assumptions and methodology. Although her study used only two models, it revealed many VaR shortcomings.

One of the disadvantages of basic Historical simulation method is that all returns are equally weighted. It becomes especially important when choosing estimation window size. If long period of data is taken, very old returns have big impact on VaR estimation. Another thing is that in longer period probability of high negative returns is higher, so VaR can be overestimated. Furthermore, moving window eventually drops off old high fluctuations, which causes VaR estimate to jump around. However, the main advantage of HS is that it does not have parameters, which can be wrongly estimated, as sometimes occur in advanced parametric methods. The issue of equal weights is addressed in Boudoukh, Richardson and Whitelaw (1998), where they suggest to weight returns according to the time from their realization to VaR estimation, i.e., according to the "age" of the return. Another suggestion is to weight returns by volatility in order to account for latest volatility changes (Hull & White, 1998).

In one of the first works on the distribution of financial returns Mandelbrot (1963) showed that financial returns are not normally distributed and distribution usually have so called "fat tails", that is, distribution is leptokurtic. Therefore Bormetti *et al.* (2007) aimed to "present a non-Gaussian approach to market risk management and to describe its potentials, as well as limitations, in comparison with standard procedures used in financial analysis." Standard methods, according to authors, are three main approaches of VaR well known in the literature and used in practice: parametric approach – normal VaR, Historical simulation and Monte Carlo simulation. To capture the excess kurtosis of empirical data, they used historical returns and fitted Student's *t* distribution, which is known to be better for modeling fat tails in financial returns. It is also reminded in the paper, that normal distribution is a special case of Student's t distribution, when tail parameter, called degree of freedom, goes to infinity. Thus it allows finding at which degree of freedom best model is produced. Scholars used 1000 daily returns of two stocks and two indexes from Italian stock market. Authors found that "tail parameters fall in the range (2.9, 3.5), thus confirming the strong leptokurtic nature of the return distributions, both for single assets and market indexes" (Bormetti, Cisana, Montagna, & Nicrosini, 2007). Overall, the study performed by Bormetti et al. (2007) found that "at a 95 percent confidence level, the Student's *t* and normal distributions are approximately equivalent, while at a higher level the Student's *t* model outperforms the normal model" and state that "the approach may have helpful implications for practical applications in the field of financial risk management" (Bormetti, Cisana, Montagna, & Nicrosini, 2007).

The main limitation of normal and Student's t VaR measures is that they assume that volatility of financial returns is constant. However, another stylized fact states, as mentioned in Brooks (2008, p. 380), that financial returns follow volatility clustering. This means that "large returns are expected to follow large returns, and small returns to follow small returns" (Brooks, 2008), and it does not depend on the sign of the return. Bollerslev (1986) worked on estimation of generalized autoregressive conditional heteroskedastic (GARCH) process, which could capture fat tails and volatility clustering in the return distribution. Distribution of financial returns are changing over time, and this effect can be captured by GARCH process, which incorporates conditional volatility, that is, VaR estimates can be made conditional. Special case of this method, which uses exponentially-weighted variance of returns, is employed in RiskMetrics, created by corporation J.P. Morgan in 1993 (Longerstaey & Zangari, 1996). However, Bollerslev (1987) showed that the distribution of returns with conditional volatility is still not normal. What is more, GARCH does not account for third stylized fact of financial returns – leverage effects, or asymmetric effect of volatility, which means that negative returns tend to increase future volatility more than positive returns. To account for this effect group of researchers specified asymmetric power ARCH (APARCH) model, which could capture leverage effect (Ding, Granger, & Engle, 1993). There are many studies on VaR which employed GARCH model; however, only a few were found which used more advanced APARCH model for volatility.

One of such researches was performed for Taiwan Stock Index futures (Huang & Lin, 2004). In order to model two stylized facts of fat tails and volatility clustering, authors enhanced normal and Student's t VaR by modeling volatility with EWMA and APARCH models. Accuracy of models was backtested using four measures: binary and quadratic loss functions and two likelihood ratio tests of unconditional coverage and conditional coverage. Finally, model efficiency was measured using mean relative scaled bias (MRSB) calculation. Smallest relative bias pointed to the most efficient model. Data sample consisted of daily returns of two Taiwan stock index futures - SGX-DT and TAIFEX – for the period of May 7, 1998 – January 31, 2002 and July 21, 1998 – January 31, 2002, respectively.

EWMA model did not require estimation, because its decay factor was set in advance, according to RiskMetrics methodology, to 0.94. Model parameters where estimated for APARCH models with normal and Student's t distribution errors. Parameter estimates proved that empirical data exhibit volatility clustering and fat tails. Furthermore, power term, ranging between 1.6008 and 1.7148, show that ordinary GARCH model is not accurate, because it has power term set in advance in the model as 2. Models are estimated for confidence interval of 95 percent, 99 percent and 99.9 percent. Overall, Huang and Lin (2004) conclude that

> *for asset returns which exhibit fatter tails and volatility clustering, like the TAIFEX and SGX-DT futures, the VaR values produced by the Normal APARCH model are preferred at lower confidence levels. However, at high confidence levels, the VaR forecasts obtained by the Student APARCH model are more accurate than those generated using either the RiskMetrics or Normal APARCH models.*

One of the issues which raises question about the reliability of the study is the sample size. Especially, when complex model such as APARCH is measured, which requires a lot of data, and very high confidence intervals such as 99.9 percent are used. When models are backtested, in order to expect, for example, 5 violations at 99.9 percent confidence level, 5000 observations are needed for that.

In 2008, Angelidis and Degiannakis, employed similar approach to two stock indexes in the Athens Stock Exchange market. Instead of EWMA, they compared GARCH models with APARCH models. The main improvement compared with study performed by Huang and Lin (2004), was that they made assumption of distribution of returns to be not only as normal or Student's t, but also as skewed Student's t, which accounts for asymmetry of return distribution. In order to check how this approach works in local market, authors took daily returns of General and Bank index from Athens Stock Exchange for a period from April 25, 1994 to December 19, 2003, which gave 2412 observations. However, authors used 2000 observations for estimation of the models and only 412 were left for forecasting. Although the confidence level used in estimation was 95 percent and 99 percent, which is appropriate for the size of in-sample window, out-of-sample window is very small for ordinary back testing models such as Christoffersen test, raising a question of accuracy of the study results. By and large, Angelidis and Degiannakis (2008) found that GARCH-normal model is appropriate for Bank index, while APARCH-Skewed Student's t model is more appropriate for General index, suggesting that "the volatility of the Greek stock market, although it is predictable, there is not an explicit model, which is the most accurate for all the forecasting purposes."

Smith (2008) performed analysis regarding structural breaks in GARCH models. Author applied LM and CUSUM structural break tests for various financial time series starting in 1990 and found that many GARCH models pass standard specification tests, but are not able to pass structural break tests. However, in spite of these results, Degiannakis, Floros and Livada (2012) employed the EWMA of RiskMetrics, GARCH-normal and APARCH-Student's t models to estimate VaR forecasting performance before and after financial crisis in 2008. In their paper they used data from both developed (USA, United Kingdom, Germany) and emerging (Greece and Turkey) markets for the period October 3, 1988 – December 31, 2008. It gave from 4985 data points of Greek stock index GRAGENL to 5113 daily returns of UK stock index FTSE100. Researchers used 4000 returns estimation window to forecast returns for the rest of the total sample. Furthermore, authors state that "in order to investigate whether the models have the ability to forecast the next trading-day VaR even in periods as the year 2008, (…) we evaluate the models separately for year 2008." (Degiannakis, Floros, & Livada, 2012). They compared model forecasting performance for both periods using classical Kupiec and Christoffersen tests and found that for almost three-year period before the end of 2007, GARCH(1,1) VaR provides most satisfactory forecasting performance for four indexes out of five used. However, for the year 2008, only APARCH(1,1) with Student's t return distribution provided satisfactory forecasts. Authors of the paper discuss that APARCH model is the best for the crisis period, because volatility is high and return distribution is skewed and exhibit fat tails. Degiannakis, Floros and Livada (2012) conclude that their paper

provides evidence that the classic risk measurement technique of VaR estimation works satisfactorily even in periods such as the year 2008 with extreme highly volatility and strong down turn tendency of the markets.

1.3. Review of empirical research on extreme value theory VaR

Another strand of literature focusing on VaR modeling and estimation is about Extreme Value Theory (EVT), which is used to model risk of extremely rare events. EVT focuses on the distribution of extreme returns ("fat tails") instead of the whole distribution. As mentioned before, in most financial time series distribution of returns is fat-tailed and asymmetric. Therefore it is a good reason to compare this model against parametric volatility models. In fact, EVT is considered as a good alternative for GARCH models. Although EVT was known for several decades in other physical sciences such as engineering and hydrology, it became popular in finance in the end of 1980s, when unprecedented extreme event occurred – Black Monday. Hereafter several studies which evaluate VaR models using EVT and compare the results with other VaR models are reviewed.

Gencay and Selcuk (2004) compared relative performance of VaR model under Extreme Value Theory with variance-covariance method and Historical simulation in ten emerging markets – Argentina, Brazil, Hong Kong, Indonesia, Korea, Mexico, Philippines, Singapore, Taiwan and Turkey. Authors explain that emerging markets experience larger volatility and larger negative returns than developed economies. Furthermore, these financial markets have huge impact on economies of developed countries, because significant portion of investments from developed countries goes to mutual and hedge funds investing in those emerging markets. Therefore investors from developed countries can benefit from additional knowledge about the risks and reward dynamics in the emerging markets. High average daily returns in those countries are explained by high inflation and a risk premium for increased probability of devaluation of the currency. What is more, return distribution in those markets is highly fat-tailed and skewed. Researchers focused on those particular countries, because most of them suffered a lot during Asian financial crisis in 1997-1998. During similar period countries in South America devaluated their currencies and had huge impact on stock markets as well. Therefore, authors of the paper model these extreme events using EVT. They employ different time series lengths ending on December 29, 2000: from 1076 daily returns in Philippines stock market to 7305 in Hong Kong and Taiwan. Then Generalized Pareto distribution is fitted to empirical data and tail distributions are estimated. Using tail distribution parameters authors calculated that at 99.9% quantile 10% negative daily returns on average are possible in those stock markets, i.e., this means that one day in approximately 4 years can show -10% return. Gencay and Selcuk (2004) used 500, 1000 and 1500 days moving windows to estimate various parametric and non-parametric VaR models at several significance levels ranging from 5% to 0.1%. Comparison of violation ratios revealed that EVT VaR produces least underestimation and overestimation of the risk in tails, especially when significance level was high from 1% to 0.1%. Thus, Gencay and Selcuk (2004) conclude that "the GPD and the extreme value theory is an indispensable part of risk management, especially in the VaR calculations and stress testing."

There are several researches where EVT is compared to parametric VaR models based on GARCH volatility modeling. For example, 30 years of daily return data (for period February 8, 1971 – June 22, 2001) of NASDAQ Composite index were used to compare out-of-sample performance of parametric VaR models, EVT and quantile regression models (Kuester, Mittik, & Paolella, 2005).

Summary statistics of index return data show that there is considerable asymmetry and leptokurtosis in the distribution of returns suggesting strong violation of normality. Authors use 1000 returns moving window (approximately four years of data) to update model parameters for each window for every one day forecast. Thus they get 6681 one day VaR forecasts to study the predictive performance of the models. The improvement of their study on VaR measurement using EVT is that, unlike in other researches ((Diebold, Schuermann, & Stroughair, 1998) and (McNeil & Frey, 2000)), they use "skewed t instead of the normal distribution in order to better account for conditional asymmetry and heavy–tailedness" (Kuester, Mittik, & Paolella, 2005). First of their findings is that unconditional VaR models such as naïve Historical simulation fails the independence test. Furthermore, accounting for asymmetry using skewed Student's t distribution provides superior performance relative to normal return distribution assumption. Secondly, they found that volatility modeling significantly improves forecasting performance, and again, skewed Student's t distribution is best in all GARCH models relative to symmetric Student's t and normal distribution assumption. In contrast to some previous researches, which found that normal GARCH is performing well at 5% significance level, Kuester, Mittik and Paolella (2005) found that "5% quantile is still not large enough for the normal assumption to be adequate" in GARCH model. Finally they show that dynamic risk measurement using EVT with normal and skewed Student's t return distributions and filtering returns with GARCH volatility model provides best results at various confidence levels. To sum up, although they found that VaR estimates often underestimate the real negative returns in the market, the best method according to them is hybrid method, which uses EVT approach combined with fat-tailed GARCH volatility modeling.

More recent study performed by Ozun, Cifter, and Yilmazer (2010), employed eight different Extreme Value Theory models, which were enhanced by filtering returns, and applied it only to Istanbul Stock Exchange. Performance was compared to GARCH models with normal, Student's t and skewed Student's t distribution assumptions, as well as FIGARCH, using various backtesting methods such as classical Kupiec and Christoffersen, Lopez test, as well as Diebold and Mariano test, root mean squared error (RMSE) estimates. More than five years of data (for period January 2, 2002 – April 18, 2007) of Istanbul Stock index 100 is used as data for empirical research on emerging market. The main contribution of this study is that authors use not only VaR, but also Expected shortfall as coherent risk measure to predict portfolio returns. What is more, eight rolling quantiles of two, three, four, five, ten, 15, 20 and 50 days are used for estimation. The maximum quantile is chosen as 40 days, because "conditional EVT approximate to unconditional EVT more than 40 days rolling." (Ozun, Cifter, &

Yilmazer, 2010). The main findings of their results are that filtered expected shortfall models forecasting performance is better than that of GARCH models, however, it is difficult to choose best model in the group. Therefore they calculated number of violations for up to 70 days forecast and found that filtered expected shortfall model with 40 days rolling quantile is best performer. Thus they concluded that this model is most appropriate for forecasting returns for more than one month in the future. Overall, results of study performed by Ozun, Cifter and Yilmazer (2010) showed that filtered EVT models are able to capture risk in fat tails and perform better than any GARCH VaR model.

1.4. Review of empirical research by financial market location

Researches can also be compared by the geographical distribution. Many studies were performed using data from developed countries, however, more and more researchers are working on data from emerging markets, because these markets are more volatile and extreme negative returns are usually much higher than in developed markets. Previously mentioned VaR models were tested in different emerging markets. One of the most comprehensive studies was performed by Huang and Tseng (2009). They compared the performance and reliability of five major VaR methodologies, using more than 26 years of return data on 37 equity indices from developed and emerging markets. The main originality of their study was that they used recently developed nonparametric kernel estimator (KE), which produced good VaR estimates and outperformed other methods. Other researchers worked on particular regions, for example, focused on South and Eastern European markets (Kalyvas, Steftos, Sriopoulos, & Georgopoulos, 2007). In order to assess market risk, they used two VaR methodologies – Historical simulation and EVT with both conditional and unconditional return distributions. They found that Hungary exhibits higher risk under extreme conditions than all other markets under study. Assaf (2009) used EVT to estimate losses predicted by VaR in Middle East and North Africa region (Egypt, Jordan, Morocco and Turkey). Turkey exhibited highest market risk among other countries in the study. Greek stock market also received attention from local researchers: Diamandis, Kouretas and Zarangas (2006) and Angelidis and Degiannakis (2008) applied GARCH and APARCH models for different stock indexes in Athens and found "there are adequate methods in predicting market risk, but it does not seem to be a specific model that is the most accurate for all the forecasting tasks" (Angelidis & Degiannakis, 2008).

Unfortunately, there is almost no research on VaR in Nordic and Baltic stock exchanges. One of the researches used VaR estimate for one of the performance measures used in that study, however, only one VaR methodology was used – Historical simulation (Mačiulis, Lazauskaitė, & Bengtsson, 2007). Since authors did not “seek to forecast future performances, but merely compare past performances of Baltic and Nordic exchanges” (Mačiulis, Lazauskaitė, & Bengtsson, 2007), they used only Historical simulation method for VaR, assuming that “historical market data is best estimator for future changes”. However, usually this assumption is thought as a shortcoming of VaR measurement. What is more, there is a gap in the literature regarding performance forecasting of Nordic and Baltic stock exchanges, which is attempted to be fulfilled in this thesis.

Table 1 presents the summary of analyzed empirical research, using methodology most related to the one used in thesis.

To summarize results of literature analysis we can conclude that there is no perfect VaR model for any dataset from any geographical region. Recent research shows that simple VaR models are not able to produce reliable VaR forecasts. However, many advanced models produces different results for different datasets from different stock markets. Finally, although there were many studies on VaR forecasting ability in Southern and Central Europe, research on Northern and Eastern Europe, particularly Nordic and Baltic stock markets are almost non-existing.

Table 1: Summary of empirical research on modeling and measuring various VaR models

Paper	Purpose	Methodology	Data	Results
Bormetti et al. (2007) "A non-Gaussian approach to risk measures"	Modeling fat tails of returns distribution with Student's t distribution and comparison with other VaR models	Student's t VaR, normal VaR, Historical simulation and RiskMetrics (EWMA) (as well as expected shortfall (ES) for all these methodologies)	1000 daily returns from the Italian stock market - Autostrade SpA and Telecom Italia (May 15th 2001 - May 5th 2005), financial indexes Mib30 and Mibtel (March 27th 2002 - March 13th 2006)	Student's t distribution is able to better capture fat tails or return distribution and VaR model has good practical implications compared with widely used normal VaR and RiskMetrics methodologies.
Huang and Lin (2004) "Value-at-Risk Analysis for Taiwan Stock Index Futures: Fat Tails and Conditional Asymmetries in Return Innovations"	Examine the forecasting performance of three value-at-risk (VaR) models	RiskMetrics, Normal APARCH and Student APARCH. Models evaluated using binary and quadratic loss functions, Kupiec and Christoffersen LR tests, MRSB measure.	Taiwan stock index futures SGX-DT (May 7, 1998–January 31, 2002) and TAIFEX (July 21, 1998–January 31, 2002) daily prices.	Normal APARCH VaR model is preferred at lower confidence levels. At high confidence levels, Student APARCH model is more accurate than RiskMetrics or Normal APARCH models.
Angelidis and Degiannakis (2008), "Forecasting one-day-ahead VaR and intra-day realized volatility in the Athens Stock Exchange Market"	Evaluate forecasting performance of symmetric and asymmetric ARCH VaR models.	GARCH and APARCH VaR model with normal, Student's t and skewed Student's t return distribution. Models evaluated using Christoffersen LR test and logarithmic error function.	Daily closing prices of the Athens Stock Exchange General and the Bank indices from 25 April 1994 to 19 December 2003.	The most appropriate method for the Bank index is normal GARCH VaR model, while skewed Student's t APARCH VaR model applies for the General index.
Degiannakis, Floros and Livada (2012) "Evaluating Value-at-Risk Models before and after the Financial Crisis of 2008: International Evidence"	Focus on the performance of three alternative Value-at-Risk (VaR) models before and after the financial crisis.	Riskmetrics (EWMA), normal GARCH(1,1) VaR, skewed Student's t APARCH(1,1) VaR. Models evaluated using Kupiec and Christoffersen LR tests.	Mature (US, UK, Germany) and emerging (Greece and Turkey) stock market indexes daily returns. Period: 1988.10.03-2008.12.31	Normal GARCH(1,1) VaR model has satisfactory performance for the period before the crisis of 2008, but not after. That is available by the skewed Student's t APARCH(1,1) VaR model.
Gencay and Selcuk (2004) "Extreme value theory and Value-at-Risk: Relative performance in emerging markets"	Investigate the relative performance of Value-at-Risk (VaR) models in 10 different emerging markets	Normal VaR (variance-covariance VaR), Historical simulation and Extreme value theory (EVT) VaR at 5%, 2.5%, 1%, 0.5% and 0.1% level of significance.	10 emerging markets in South America, Asia and Europe. From 1076 to 7305 daily returns for the period ending on December 29, 2000.	Results indicate that EVT-based VaR estimates are more accurate at higher quantiles. In addition, risk and reward are not equally likely in these economies.
Kuester, Mittik and Paolella (2005) "Value–at–Risk Prediction: A Comparison of Alternative Strategies"	Compare the out-of-sample performance of existing and some new models for predicting Value-at-Risk.	Historical simulation, unconditional, GARCH(1,1) and EVT VaR models with normal, Student's t and skewed Student's t distributions as well as 5 CAViaR models.	NASDAQ Composite Index daily returns for period February 8, 1971 - June 22, 2001.	Most VaR approaches underestimate risk. A hybrid method, combining a heavy-tailed GARCH filter with an extreme value theory-based approach performs best overall.
Ozun, Cifter, and Yilmazer (2010) "Filtered extreme-value theory for value-at-risk estimation: evidence from Turkey"	Compare the predictive performance of filtered extreme-value theory model with other conditional volatility models.	Eight filtered EVT models, GARCH(1,1) with normal, Student's t and skewed Student's t distribution, FIGARCH. Models compared using Kupiec test, Christoffersen test, Lopez test, Diebold and Mariano test, RMSE measures.	Istanbul Stock Exchange index daily returns for period January 2, 2002 - April 18, 2007.	The results indicate that filtered EVT performs better in terms of capturing fat-tails in stock returns than parametric VaR models.

2. Research problem definition

Empirical studies presented in review of empirical research part and summarized in Table 1 examined many different techniques to calculate VaR. Although simple parametric VaR models assume that asset returns follow normal (Gaussian) distribution, the problem is that many empirical researches show that financial returns are non-normal and asymmetrically distributed (Bormetti, Cisana, Montagna, & Nicrosini, 2007). Therefore empirical researches presented in literature review tried to evaluate new advanced VaR models, which account for these properties of financial returns as well as volatility clustering and other stylized facts of financial returns. Although many studies have been conducted, there are still different results from different stock markets as presented in Table 1. One of the "gaps" is that most of the studies presented in review of empirical research focused only on a few VaR models and estimated their forecasting performance only in one or several stock markets. Another "gap" in the literature is that financial markets in certain geographical regions are covered when testing advanced VaR approaches leaving many financial markets behind. Furthermore, to the thesis author's knowledge, many banks in Nordic and Baltic region, which use VaR methodology, rely on simple and basic VaR models, which have many shortcomings, especially in the volatile markets; while the scientific literature on more advanced VaR models for stock exchanges in these markets are almost non-existent. The problem to solve in this research is to find which VaR models provide best forecasts for particular stock markets during different time periods and financial conditions. The main motive of this thesis is to perform a study on Value at Risk forecasting performance in Nordic and Baltic stock exchanges by comparing many different, simple and advanced, VaR models combined from various researches presented in review of empirical research and choosing the best ones for each stock market, which could be used by financial institutions in these countries to manage financial risk. Finally, this comprehensive study tries to fill in the gaps in the literature as presented above. Thesis aims to answer these research questions:

1) Which VaR models – simple or advanced ones - have best one-day forecasting performance for each Nordic and Baltic stock index?

2) How does the various VaR models' forecasting performance change with respect to developed Nordic and emerging Baltic stock markets, calm and volatile time periods before and after financial crisis and different confidence levels?

In order to answer research questions, following hypotheses of a study are formulated:

H1. Index returns of emerging Baltic stock markets are non-normally and asymmetrically distributed, while index returns of developed Nordic stock markets are normally and symmetrically distributed.

First hypothesis comes from the assumption that Baltic stock markets are still emerging, therefore, as presented in Gencay and Selcuk (2004), such stock markets are usually non-normally and assymetrically distributed. Ragarding developed stock markets, empirical evidence is mixed, but the assumption is that Nordic stock markets are normally and symmetrically distributed.

H2. Changing assumption of stock index return distribution from normal to Student's t and skewed Student's t significantly improves volatility models' fit to Nordic and Baltic stock index return data.

Several empirical researches, for example, performed by Kuester, Mittik and Paolella (2005) and Huang and Lin (2004) measured volatility models using several return distribution assumptions such as Student‘s t and found out that the ones, which account for leptokurtosis and skewness, enhances volatility model‘s fit to the return data. Therefore this hypothesis is tested for return distributions of Nordic and Baltic stock indexes.

H3. Addition of power term and leverage effect parameters to volatility model significantly improves model fit to the whole return data sample.

Huang and Lin (2004) analyzed APARCH model applicability in modeling VaR and found out that ability to freely measure power term and accounting for leverage effect significantly improved volatility model fit to the sample of index futures returns. Therefore, third hypothesis in this thesis checks whether two additional parameters added to GARCH volatility model improves it statistically significantly when modeling volatility of Nordic and Baltic stock indexes.

H4. Statistically significantly more VaR models pass Christoffersen test when level of confidence is increased from 95% to 99%.

As presented in review of empirical research part, Gencay and Selcuk (2004) as well as Kuester, Mittik and Paolella (2005) found that increased level of confidence provides better VaR forecasting results. Thereofore, hypothesis that more VaR models provide accurate index return forecasts at higher level of confidence is tested for Nordic and Baltic stock markets as well.

H5. Statistically significantly more VaR models provide accurate forecasts in terms of Christoffersen test results in developed Nordic stock markets than in small emerging Baltic stock markets.

Fifth hypothesis involve assumption that more advanced VaR models are required to model stock market fluctuations in emerging and small stock markets which sometimes have unprecedented structural shocks not reflected in other developed stock markets and basic VaR models are not able to provide accurate VaR forecasts.

H6. It is possible to find one VaR model which has best forecasting performance for Nordic and Baltic stock markets during any financial market conditions.

Finally, many researchers tried to find one best VaR model for all stock markets and time periods, but usually it did not exist. But maybe there is only one VaR model which fits all seven Nordic and Baltic stock markets during all time periods and financial conditions? Last hypothesis, as popular among VaR modeling researchers, is based on this question.

Thesis will contribute to our knowledge by expanding empirical research on modeling and measuring value at risk for Nordic and Baltic stock markets, which had relatively low attention from researchers, and give evidence for financial practitioners how they could improve their risk management tools and provide better risk management within their financial institutions.

3. Methodology

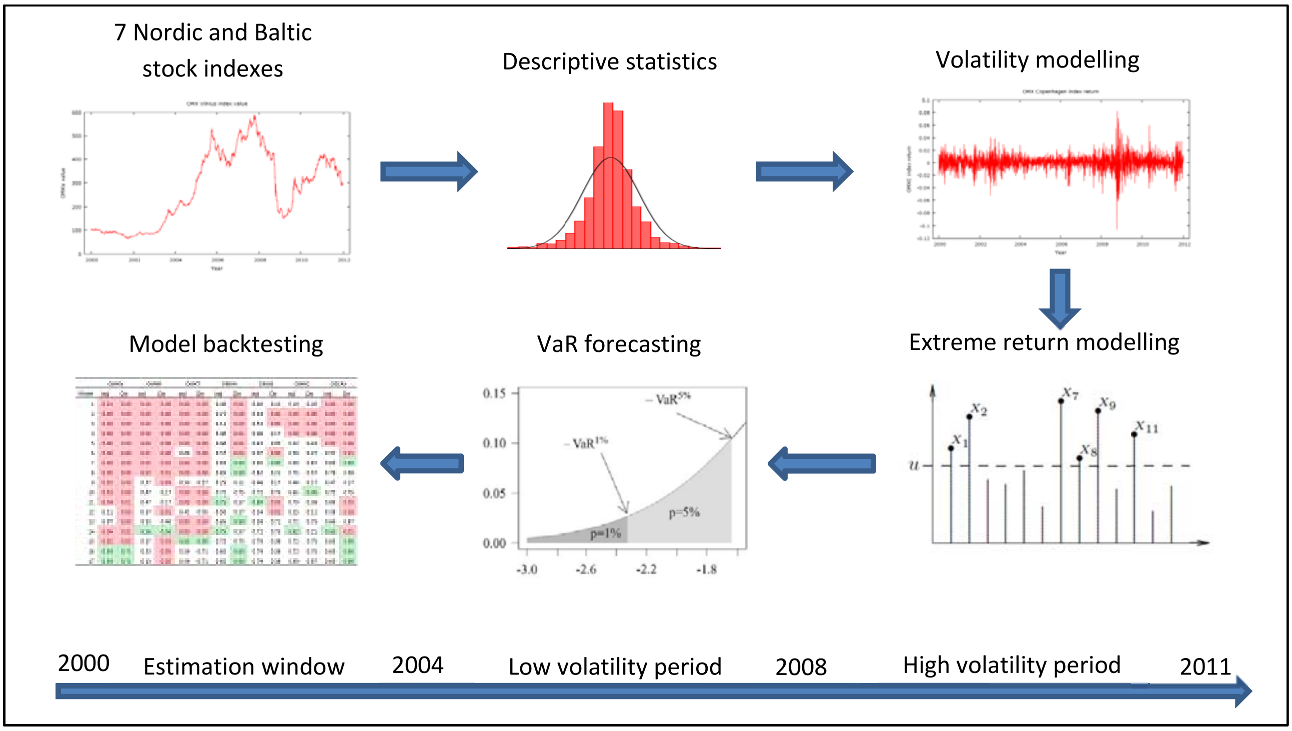
This section of the thesis presents research design, the description of methodological approach used in the research and clear explanation and justification of the selected models and methods.

3.1. Empirical research aim and research design

Empirical research is designed to solve research problem, answer research questions and verify formulated hypotheses. As already mentioned in previous parts of the thesis, problem of the thesis is to find which VaR models provide best forecasts for particular stock markets during different time periods and financial conditions. Thesis aims to answer research questions such as whether simple or advanced VaR models have best one-day forecasting performance for each Nordic and Baltic stock index and how does the various VaR models' forecasting performance change with respect to developed Nordic and emerging Baltic stock markets, calm and volatile time periods before and after financial crisis and different confidence levels? Thesis hypotheses, as formulated in research problem definition part, are verified by methods and procedures presented in this methodology part.

Figure 2 presents methodological approach and sequence of empirical research.

Figure 2: Sequence of empirical research



In order to measure and evaluate forecasting performance of various VaR models for four Nordic and three Baltic stock indexes and verify thesis hypotheses, following approach is implemented:

a) Log differences of daily values of Nordic (Helsinki, Stockholm, Oslo, Copenhagen) and Baltic (Vilnius, Riga, Tallinn) stock exchange indexes are taken for period 2000-2011.
b) Descriptive statistical measures are calculated and normality tests such as Doornik-Hansen test and Jarque-Bera test are performed. Descriptive graphs are depicted for graphical analysis.
c) For conditional VaR models conditional volatility is estimated using EWMA, GARCH(1,1) and APARCH(1,1) models, assuming that returns follow normal, Student's t and skewed Student's t distributions. Fitted models for the whole sample are compared using likelihood ratio tests.
d) Seventeen different VaR models, presented in Table 2 are estimated.
e) All models are tested with 1% and 5% significance level.
f) In order to assess model performance, 1000 daily returns as a moving window are used to estimate one-day-ahead forecasts for two periods – low volatility period of 2004-2007 and high volatility period of 2008-2011 – as well as their combination.
g) Likelihood ratio tests of unconditional coverage (Kupiec's test) and conditional coverage (Christoffersen's test) are estimated to evaluate and compare forecasting performance. Last three hypotheses are verified using results of model backtesting.
h) VaR forecasts and actual returns are plotted for the whole testing window in order to analyze models' forecasting behavior.

Table 2: Model number and model name

Model	Model name	Model	Model name	Model	Model name
1	Historical simulation	7	Student's t EWMA VaR	13	Student's t APARCH(1,1) VaR
2	Normal VaR	8	Skewed Student's t EWMA VaR	14	Skewed Student's t APARCH(1,1) VaR
3	Student's t VaR	9	Normal GARCH(1,1) VaR	15	EVT Normal GARCH(1,1) VaR
4	Skewed Student's t VaR	10	Student's t GARCH(1,1) VaR	16	EVT Student's t GARCH(1,1) VaR
5	EVT VaR	11	Skewed Student's t GARCH(1,1) VaR	17	EVT Skewed Student's t GARCH(1,1) VaR
6	Normal EWMA VaR	12	Normal APARCH(1,1) VaR		

In this thesis we focus on the assumption that financial returns are distributed normally or according to Student's t and skewed Student's t distribution. Unconditional VaR models do not account for volatility changes and clustering, therefore three volatility models – EWMA, GARCH and APARCH – are estimated in this thesis in order to compare its performance against simple unconditional VaR models. There are more volatility models created, however we use the ones which are most popular among VaR researchers. Finally, stock index return data includes recent financial crisis, which had extreme negative returns. Therefore VaR models based on Extreme Value Theory are also estimated and compared with all other models employed in the thesis. In Table 3 all VaR models used in this thesis are grouped and compared by presenting their advantages and disadvantages:

Table 3: VaR model advantages and disadvantages

VaR model (group)	Advantages	Disadvantages
Historical simulation	No distribution assumption, non-parametric, simple	Assume that historical returns are a good proxy for future returns
Unconditional VaR models	Simple to measure	Distribution assumption, do not account for volatility clustering
EWMA VaR models	Simple model accounting for volatility clustering	Does not account for mean-reverting of time series
GARCH VAR models	Mean-reverting to conditional volatility	Non-linear, restricted power term
APARCH VaR models	Non-restricted power term and account for leverage effect	Very complex, unable to account for structural breaks
EVT VaR models	Focus in the tail of distribution, capture extreme returns	Requires a lot of data

Note: Each of these models except Historical simulation and unconditional EVT are estimated using normal, Student's t and skewed Student's t distribution assumption.

One of the thesis limitations is that it is delimited by performing analysis using only stock indexes, which represent portfolios of only one asset class. Other asset classes were not considered in this thesis, because detailed portfolios of financial assets in financial institutions of Nordic and Baltic countries are not publicly available. Furthermore, derivatives markets are non-existent in stock exchanges of Baltic countries, so it is not possible to compare with Nordic derivatives markets. The assumption is that modeling and analysis of stock market index returns are good enough to provide empirical findings which could be applicable in other asset classes.

Although VaR is not a coherent risk measure and Expected Shortfall (ES) is a coherent risk measure (Dowd, 2005, pp. 32-37), ES did not receive the same attention as VaR due to other limitations. What is more, VaR is exclusively used for calculation of capital requirements of banks' portfolios, as emphasized in Basel II Accord (BIS, 2006). Therefore, this thesis focuses only on VaR forecasting performance.

In the rest of the section, detailed description of methodological approach is presented by giving detailed explanation of each model used in the thesis, providing description of backtesting methodology and hypothesis testing procedure.

3.2. Unconditional VaR models

In this section four unconditional VaR models are presented – non-parametric Historical simulation and three parametric VaR models with normal, Student's t and skewed Student's t return distribution assumption.

Historical simulation

One of the most well-known non-parametric VaR calculation methods is called historical simulation (HS). Historical simulation uses empirical distribution of data to calculate VaR and produce return forecasts. This means that it is assumed that historical returns are a good proxy for future returns. It is non-parametric method, because there is no need for statistical models, which require parameter estimates. If we have t historical observations, VaR forecast for $t+1$ is equal to α - quantile Q_α of distribution of these returns, where α is significance level (or $1-\alpha$ level of confidence):

$$VaR_{t+1} = Q_\alpha(r_t, r_{t-1}, ..., r_1) \qquad (2)$$

For example, for 100 historical daily returns, 5% VaR forecast for one day will be equal to 5th lowest return from a sorted sample.

Normal VaR

Parametric models are based on the assumption that returns follow some probability distribution function:

$$r_t = \mu_t + \varepsilon_t = \mu_t + \sigma_t z_t \qquad (3)$$

Where r_t is return, μ_t is mean, σ_t is standard deviation, i.e. both measurable parameters, and z_t is independently and identically distributed and $z_t \sim f_Z$, where f_Z is standardized probability density function.

In this section focus is on unconditional VaR models, thus $\mu_t \equiv \mu$ and $\sigma_t \equiv \sigma$. Normal VaR approach assumes that returns follow normal distribution. Returns are normally distributed if has the density function:

$$\phi(r;\mu,\sigma) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{1}{2}\left(\frac{r-\mu}{\sigma}\right)^2} \quad (4)$$

Returns with density function ϕ are denoted by:

$$r \sim \mathcal{N}(\mu, \sigma^2) \qquad (5)$$

Standardized returns follow standard normal distribution:

$$z = \frac{r-\mu}{\sigma} \sim \mathcal{N}(0,1) \qquad (6)$$

Here we derive VaR for simple returns. Simple return can be denoted as:

$$r_t = \frac{V_t - V_{t-1}}{V_{t-1}} \qquad (7)$$

From equation (1) we have

$$\Delta V = V_t - V_{t-1}$$

$$Pr(\Delta V < -VaR) = \alpha$$

Making several operations we obtain VaR:

$$\begin{aligned} \alpha &= Pr(\Delta V < -VaR) \\ &= \Pr(V_t - V_{t-1} < -VaR) \\ &= Pr(r_t V_{t-1} < -VaR) \\ &= Pr\left(\frac{r_t - \mu}{\sigma} < \frac{\frac{-VaR}{V_{t-1}} - \mu}{\sigma}\right) \\ &= Pr\left(z < \frac{\frac{-VaR}{V_{t-1}} - \mu}{\sigma}\right) \qquad (8) \end{aligned}$$

We denote the inverse distribution of standardized returns z as quantile function $\mathrm{Q}_z(\alpha)$, which can be seen as significance level.

$$Q_z(\alpha) = \frac{\frac{-VaR}{V_{t-1}} - \mu}{\sigma} \qquad (9)$$

$$VaR_t = -\big(\mu + \sigma Q_z(\alpha)\big)V_{t-1} \qquad (10)$$

For example, if returns are normally distributed, significance level α is 1% or 0.01, initial portfolio value V_{t-1} equal to 1, mean return is equal to 0 and standard deviation of returns is equal to 1%, we can obtain from statistical z tables that $Q_z(0.01) = -\Phi^{-1}(0.01) = -2.33$, and we get $\mathrm{VaR}(1\%) = -\big(0 + 1\% \times (-2.33)\big) \times 1 = 2.33\%$.

Student's t VaR

However, due to earlier mentioned stylized facts, financial returns have fat-tailed distribution. As suggested by Bollerslev (1987), fat-tailed returns can be better modeled with Student's t distribution denoted as $t(\nu)$ with ν degrees of freedom and with probability density function:

$$f(x;\nu) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{(\nu-2)\pi}\,\Gamma\left(\frac{\nu}{2}\right)}\left(1 + \frac{x^2}{\nu-2}\right)^{-\frac{\nu+1}{2}}, \nu > 2 \quad (11)$$

Here Γ is Gamma function, denoted as:

$$\Gamma(x) = \int_0^\infty e^{-t}\, t^{x-1}\, dt \quad (12)$$

Student's t distribution is fitted to the data using maximum likelihood estimation and parameter estimates for mean and volatility are used in the quantile function entering VaR equation:

$$VaR_t = -\big(\mu + \sigma Q_{student's\, t}(\alpha)\big)V_{t-1} \qquad (13)$$

Skewed Student's t VaR

In order to model the leptokurtosis as well as the asymmetry of the conditional distribution of innovations, the skewed student t density function can be used (Lambert & Laurent, 2000):

$$f(z_t;\nu,g) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{(\nu-2)\pi}\,\Gamma\left(\frac{\nu}{2}\right)}\left(\frac{2s}{g+g^{-1}}\right)\left(1 + \frac{sz_t + m}{\nu-2} g^{-d_t}\right)^{-\frac{\nu+1}{2}}, \nu > 2 \quad (14)$$

Where g is asymmetry parameter, Γ is Gamma function, mean m and standard deviation s of the non-standardized skewed Student's t distribution are denoted as:

$$m = \frac{\Gamma\left(\frac{\nu-1}{2}\right)\sqrt{(\nu-2)}}{\sqrt{\pi}\Gamma\left(\frac{\nu}{2}\right)}(g + g^{-1}) \qquad (15)$$

$$s = \sqrt{(g^2 + g^{-2} - m^2 - 1)} \qquad (16)$$

$$d_t = \begin{cases} 1, if\ z_t \geq -\frac{m}{s} \\ -1, if\ z_t < -\frac{m}{s} \end{cases} \qquad (17)$$

Skewed Student's t distribution is fitted to the data using maximum likelihood estimation and parameter estimates for mean and volatility are used in the quantile function entering VaR equation:

$$VaR_t = -\left(\mu + \sigma \mathrm{Q}_{skewed\ student's\ t}(\alpha)\right)V_{t-1} \qquad (18)$$

3.3. Volatility modeling using EWMA, GARCH and APARCH models

In previous section we assumed that volatility is not changing over time, i.e., it is constant. However, looking at real market data we observe that volatility is always changing and some patterns can be seen in the data. One of the well-researched properties is so called volatility clustering. This means that volatility tends to change its trend slowly, following previous observations quite persistently; statistically it means that there is high autocorrelation of absolute and squared returns (Kuester, Mittik, & Paolella, 2005). Another stylized fact is earlier mentioned leverage effect. To put simply, this means that volatility increases more when there are high negative returns than there are high positive returns. In order to account for these effects in VaR models, volatility should be modeled as a time-varying variable. In the next few sections several volatility models are presented which were created in order to account for these features of volatility and are used in this thesis.

EWMA model

One of the simplest linear models for conditional volatility σ_t is moving average (MA). Moving average is calculated by taking some number of historical variances (squared standard deviations) and calculating their average. Every other average discards last variance out of calculation and takes in to account the next one (first-in-first-out method). Because it is an arithmetic average, each historical observation has equal weight. This is a shortcoming of this model, because it is assumed that old observations have much lower effect on the current volatility than the most recent ones. This model was improved by weighing each observation with different weight, which is increasing exponentially as the "age" of observations is decreasing, and was called exponentially weighted moving average (EWMA). According to this model, forecast of variance $\sigma^2{}_{t+1}$ at period $t+1$ is equal

$$\sigma^2{}_{t+1} = \lambda\sigma^2{}_t + (1-\lambda)r_t^2 \quad (19)$$

Parameter λ is called decay factor and should be $0 < \lambda < 1$. Recursively calculating previous variances we can get such formula:

$$\sigma^2{}_t = \lambda\sigma^2{}_{t-1} + (1-\lambda)r_{t-1}^2$$
$$\sigma^2{}_{t+1} = \lambda(\lambda\sigma^2{}_{t-1} + (1-\lambda)r_{t-1}^2) + (1-\lambda)r_t^2 = (1-\lambda)(r_t^2 + \lambda r_{t-1}^2) +$$
$$+\lambda^2\sigma^2{}_{t-1} = \cdots = (1-\lambda)(r_t^2 + \lambda r_{t-1}^2 + \lambda^2 r_{t-2}^2 + \cdots)$$
$$\sigma^2{}_{t+1} = (1-\lambda)\sum_{i=0}^{\infty}\lambda^i r_{t-i}^2 \quad (20)$$

Decay factor shows how fast weight decreases for old squared returns. For example, in famous RiskMetrics (Longerstaey & Zangari, 1996) decay factor is set to 0.94 for daily returns. Therefore, in this thesis the same decay factor is used for daily returns volatility modeling. The main disadvantage of the EWMA model is that λ is constant and the same is applied for all asset classes and different time periods. It is not realistic to assume that λ is the same for all assets.

From the formula we see that number of observations goes to infinity. In practice, this number is truncated at some reasonable lag. There comes second limitation of this model. When the number of observations in the infinite sum is truncated to some finite number, sum of weights is not equal to 1 anymore. This could make a big difference, especially for small samples, so corrections have to be done in this model. The adjusted formula for n observations (Danielsson, 2011, pp. 33-34):

$$\sigma^2{}_{t+1} = \frac{(1-\lambda)}{(1-\lambda^n)}\sum_{i=0}^{n}\lambda^i r_{t-i}^2 \quad (21)$$

Another feature which is not accounted by EWMA model is mean-reverting of volatility time series. This means that volatility is either at high level or low level compared to historical average and it tend to move to unconditional volatility, i.e. that historical average. This property is accounted in GARCH volatility forecasting model presented in the next section.

Finally, in order to calculate EWMA VaR, we take square root of forecasted variance using EWMA and insert it into VaR equation:

$$VaR_t = -\left(\mu + \sigma_t \mathrm{Q}_{z_t}(\alpha)\right) V_{t-1} \qquad (22)$$

In this thesis three EWMA models are estimated. In formula (22) quantile function of normal distribution is also replaced to Student's t and to skewed Student's t distribution to obtain Student's t EWMA VaR estimate and skewed Student's t EWMA VaR estimate.

GARCH model

First conditional volatility forecasting model was autoregressive conditional heteroskedasticity (ARCH) model proposed by Engle (1982). This model states that conditional variance of tomorrow's return is equal to constant ω plus several previous squared returns (denoted by p lags).

$$\sigma^2{}_{t+1} = \omega + \sum_{i=0}^{p-1} \alpha_i \, r_{t-i}^2 \quad (23)$$

If we set number of lagged returns to 1, we have ARCH(1) model:

$$\sigma^2{}_{t+1} = \omega + \alpha r_t^2 \quad (24)$$

However, ARCH model is not well suited to capture volatility, because several hundred lags of squared returns, which are statistically significant, can be required to obtain several hundred parameters in order to capture entire volatility structure. Bollerslev proposed the solution to this problem (Bollerslev, 1986) by constructing generalized ARCH model called GARCH model. He included lagged volatilities (denoted by q lags), which allowed incorporating all historical returns in the ARCH model. This improvement resulted in GARCH(p,q) model for conditional volatility $\sigma^2{}_t$ at time t:

$$\mathrm{r}_t = \sigma_t z_t$$

$$z_t \sim \mathcal{N}(0,1)$$

$$\sigma^2{}_t = \omega + \sum_{i=1}^{p} \alpha_i \, r_{t-i}^2 + \sum_{j=1}^{q} \beta_j \, \sigma_{t-j}^2 \quad (25)$$

Here we set that z_t is an independently and identically distributed (i.i.d.) normal variable with zero mean and unit variance. However, standardized returns can be assumed to follow any other distribution, such as Student's *t* or skewed Student's *t*.

In practice, most popular version of this formula is taking one lag for both terms, which result in GARCH(1,1) model for $t + 1$:

$$\sigma^2{}_{t+1} = \omega + \alpha r_t^2 + \beta \sigma_t^2 \quad (26)$$

Expectation of squared returns is equal to unconditional volatility (follows from statistical properties of the moments):

$$\mathrm{E}(r^2) = \sigma^2 = \mathrm{E}(r_t^2) = \mathrm{E}(\sigma_t^2 z_t^2) = \mathrm{E}(\sigma_t^2) \quad (27)$$

Therefore we can obtain unconditional volatility from GARCH(1,1) model:

$$\sigma^2 = E(\omega + \alpha r_t^2 + \beta \sigma_t^2) = \omega + \alpha E(r_t^2) + \beta E(\sigma_t^2) = \omega + \alpha \sigma^2 + \beta \sigma^2$$

$$\sigma^2 = \frac{\omega}{1-\alpha-\beta} \quad (28)$$

To ensure positive volatility forecast, parameters must be restricted. First restriction is that all parameters must be positive: $\omega, \alpha, \beta > 0$. Another restriction comes from unconditional volatility formula: in order to ensure covariance stationarity, sum of parameters α and β must be lower than one:

$$\alpha + \beta < 1$$

However, this restriction is needed only when forecast of unconditional volatility is measured, for prediction of conditional volatility we should not restrict ourselves by this limitation.

The main drawback in the estimation of GARCH model parameters is that model is nonlinear. Parameters should be estimated using maximization of the log likelihood function. To do that numerical optimization methods are used. Under assumption of normal distribution of returns, GARCH likelihood function is (Alexander, 2008a, pp. 137-138):

$$\ln \mathcal{L}(\omega, \alpha, \beta) = -\frac{1}{2}\sum_{t=1}^{T}\left(\ln(2\pi) + \ln(\sigma_t^2) + \frac{\varepsilon_t^2}{\sigma_t^2}\right) \quad (29)$$

A certain minimum amount of data is necessary for the likelihood function to be well defined. Often several years of daily data are necessary to ensure proper convergence of the model. Therefore, in this thesis we use all available data for Baltic stock indexes and the same amount for Nordic indexes, i.e. 12 years of daily returns. For forecasting purposes, 1000 daily returns window is used to estimate models, which is equal to approximately 4 years of data.

After estimating parameters of GARCH(1,1) model, we calculate variance forecast and insert it into VaR formula (10). Again, three GARCH(1,1) VaR models are estimated, changing assumption of return distribution from normal to Student's t and skewed Student's t.

Jorion (2007) noted that EWMA is a special case of GARCH(1,1) model, were parameter ω is equal to 0 and sum of parameters α and β is equal to 1. Therefore EWMA model has persistence equal to 1 and can also be called integrated GARCH (IGARCH) model. Both models are used in this thesis to forecast one-day returns, therefore forecasting results of both models are expected to be quite similar if persistence of GARCH(1,1) model is also near 1.

APARCH model

Many different extensions to common GARCH(1,1) model have been created, but had little improvement compared to initial model. However, an extension to capture asymmetry and leverage effects in return distribution made model attractive for practical use, because it gave more flexibility in volatility calculation. In traditional GARCH model, returns and volatility are squared, most likely due to assumption of normality of the return distribution as stated in Huang and Lin (2004, p. 80). However, if we think that data is distributed non-normally, higher moments such as skewness and kurtosis can explain data better. In this case GARCH model is limited and model with different power terms could be more appropriate. Therefore, a new model was created to account for asymmetry and leverage effects, called asymmetric power GARCH (APARCH) model (Ding, Granger, & Engle, 1993). In this model, additional term for power is estimated, but not set by the researcher as in the GARCH model. Additional parameter $\delta > 0$ allows for infinite number of transformations. In the original article on the new model (Ding, Granger, & Engle, 1993), authors applied their model to US stock market data and found that model provides a good fit of data and optimal power term is 1.43.

Normal APARCH(p,q) model can be expressed:

$$\mathrm{r_t} = \sigma_t z_t$$

$$z_t \sim \mathcal{N}(0,1)$$

$$\sigma^{\delta}{}_{t} = \omega + \sum_{i=1}^{p} \alpha_i \left(|r_{t-i}| - \gamma_i r_{t-i}\right)^{\delta} + \sum_{j=1}^{q} \beta_j \, \sigma_{t-j}^{\delta} \qquad (30)$$

Parameters $\omega, \alpha_i(i = 1..p), \beta_j(j = 1..q), \delta$ and $\gamma_i(i = 1..p)$ have to be estimated. Parameter δ is a power term and parameters γ_i $(-1 < \gamma_i < 1)$ capture leverage effects. A positive (negative) value of γ_i means that past negative (positive) shocks have a deeper impact on current conditional volatility than past positive (negative) shocks (Pagan & Schwert, 1990).

Most common is APARCH(1,1) model:

$$\sigma^{\delta}{}_{t+1} = \omega + \alpha(|r_t| - \gamma r_t)^{\delta} + \beta \sigma_t^{\delta} \qquad (31)$$

Student's t APARCH(1,1) VaR model and skewed Student's t APARCH(1,1) VaR model is also estimated assuming that returns follow Student's t and skewed Student's t distribution, respectively. When $\gamma = 0$ and $\delta = 2$, we get one of several APARCH(1,1) model special cases, GARCH(1,1) model.

Parameters of APARCH model are estimated using maximum likelihood approach and conditional volatility forecast is entered into VaR formula to obtain APARCH VaR estimate.

APARCH model is very complex and has many parameters, which makes it difficult to estimate. Sometimes estimation fails due to short data sample or if data sample has structural breaks. Since data sample used in this thesis includes volatility clusters during financial crisis, this may complicate estimation of APARCH model.

3.4. Extreme Value Theory (EVT) application in VaR modeling

In previous sections we focused mainly on events which occur 2-3 days per year (at 1% level of significance). However, some events occur only once in several years, such as Black Monday in 1987, when Dow Jones Industrial Average index fell 22.61% in one day, or extreme losses during collapse of Lehman Brothers in 2008. In these cases, more advanced techniques such as Extreme Value Theory (EVT) are needed to evaluate risks.

Many VaR models are based on modeling the entire distribution of financial returns. Very often, Normal or Student's t distribution is good enough to approximate data for common events, when normal conditions are present in the market. However, these distributions produce inaccurate estimation of the distribution of the tails.

In contrast, Extreme Value Theory focuses on analyzing the tail areas of distributions, where uncommon events are present. Most of the EVT methods fully exploit extreme observations in the tail distribution. Detailed presentation of EVT and its application follows in next few sections.

Extreme Risk Modeling and Measuring

Extreme Value Theory is applied in all areas of financial risk management, whether it is a market, credit, operational or insurance risk. "One of the greatest challenges to the risk manager is to implement risk management models which allow for rare but damaging events, and permit the measurement of their consequences." (McNeil A. J., 1999). In this thesis we focus on market risk and particularly at calculation of VaR using EVT.

Financial returns follow some kind of probability distribution and extreme events lie in the tail of that distribution. In order to create EVT model for risk, we need to select particular probability distribution. Parameter estimates are obtained by fitting distribution to empirical data. EVT model tries to provide best estimate of the tail region of probability distribution.

In EVT, VaR is a risk measure, which attempts to describe the tail of the loss distribution. When EVT was developed, researchers focused on the upper tail of the distribution, so in order to analyze losses in the stock market, negative returns are multiplied by -1 to convert lower tail to upper tail.

In principle, there are two broad kind of extreme value models. One of the first models developed for extreme value measurement is *block maxima* model. This model is used when large samples of identically and independently distributed losses are available and many large observations (extremes) are present in those samples.

A more recent and modern group of models are the *peaks-over-threshold* (POT) models (McNeil, Frey, & Embrechts, 2005). These models focus on all losses which exceed some high threshold. “POT models are generally considered as most useful for practical applications, due to their more efficient use of return data on extreme values, which are often very limited” (McNeil, Frey, & Embrechts, 2005).

Peaks-over-threshold (POT) models

Peaks-over-threshold (POT) method uses all data that is considered extreme in the sense that it exceeds some designated high threshold.

Group of POT models can be further distinguished in two styles of analysis:

a) Semi-parametric models (e.g., Hill estimator),
b) Fully parametric models (e.g., generalized Pareto distribution or GPD).

In the thesis GPD distribution is used to estimate EVT VaR model, because, in order to measure extreme risk, rather simple parametric formula is used and estimates are obtained using maximum likelihood method.

We denote $\mathrm{X}_1, \mathrm{X}_2, \dots, \mathrm{X}_\mathrm{n}$ as identically distributed random variables with unknown underlying distribution function $\mathrm{F(x)} = \mathrm{Pr}(\mathrm{X_i} \leq \mathrm{x})$ (McNeil A. J., 1999). Here assumption that random variables are independently distributed is dropped, because it is clear that financial returns are dependent.

Generalized Pareto distribution (GPD) is a two parameter distribution with distribution function (McNeil, Frey, & Embrechts, 2005):

$$G_{\xi,\beta}(x) = \begin{cases} 1 - \left(1 + \frac{\xi x}{\beta}\right)^{\frac{-1}{\xi}}, \xi \neq 0 \\ 1 - e^{-\frac{x}{\beta}}, \qquad \xi = 0 \end{cases}, \quad (32)$$

where $\beta > 0$ and $\mathrm{x} \geq 0$ when $\xi \geq 0$ and $0 \leq \mathrm{x} \leq -\frac{\beta}{\xi}$ when $\xi < 0$. The parameters ξ and β are referred as *shape* and *scale* parameters, respectively. Three extreme value distributions are generalized under $G_{\xi,\beta}$ distribution, however, each of them have a different name according to the value ξ (McNeil A. J., 1999):

a) Re-parameterized version of the ordinary Pareto distribution, when $\xi > 0$, (famous in insurance mathematics for modeling large losses)
b) the exponential distribution, when $\xi = 0$,
c) Pareto type II distribution, when $\xi < 0$.

Differently parameterized Pareto distribution is most suitable to manage financial risks, because GPD is fat-tailed when $\xi > 0$. What is more, some moments for GPD distribution do not exist, in contrast to Gaussian distribution. When $\xi > 0$, GPD is fat-tailed and

$$E(X^k) = \infty \; for \; k \geq \frac{1}{\xi} \quad (33)$$

For example, when $\xi = \frac{1}{4}$, the GPD is a distribution with infinite fourth moment.

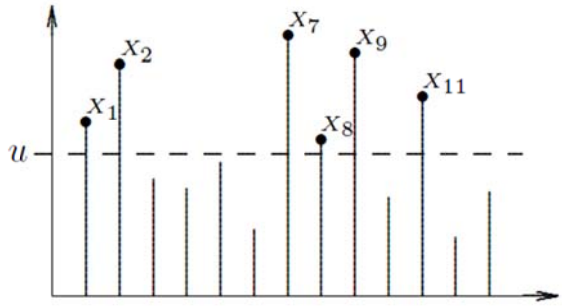
If X is a random variable with distribution function F, the excess distribution over the threshold u has a distribution function (McNeil A. J., 1999):

$$F_u(x) = Pr(X - u \leq x | X > u) = \frac{F(x+u) - F(u)}{1 - F(u)}, \quad (34)$$

for $0 \leq x \leq x_F - u$, where $x_F \leq \infty$ is the right endpoint of F. The excess distribution represents probability that a loss exceeds the threshold u by at most an amount x, focusing on the positive part of $X - u$ (McNeil, Frey, & Embrechts, 2005).

For example, in Figure 3 POT approach is shown, where focus is on returns exceeding a given high threshold. Observations X_1, X_2, X_7, X_8, X_9 and X_{11} all exceed threshold u and are considered as extreme events (Gilli & Kellezi, 2006).

Figure 3: Excesses over a threshold u.



Note: Observations X_1, X_2, X_7, X_8, X_9 and X_{11} all exceed threshold u and are considered as extreme returns.

Source: Gilli and Kellezi (2006).

According to the second theorem in Extreme Value Theory ((Pickands, 1975) and (Balkema & de Haan, 1974)), F_u is well approximated by GPD:

$$F_u(x) \rightarrow G_{\xi,\beta}(x), as\ u \rightarrow \infty \quad (35)$$

Therefore model for a risk X_i having distribution F assumes that, for a certain large u, the excess distribution above this threshold may be exactly generalized Pareto distribution (McNeil, Frey, & Embrechts, 2005):

$$F_u(x) = G_{\xi,\beta}(x) \quad (36)$$

Excess losses are modeled by fitting GPD and obtaining parameters using maximum log-likelihood method. Excess losses are obtained by counting losses above threshold u as N_u and subtracting threshold from each exceeding loss.

Threshold u should be chosen "sufficiently high, so that the asymptotic theorem can be considered essentially exact" (McNeil A. J., 1999), but at the same time, sufficiently low, so that sufficient amount of data for estimation of parameters is available. In this thesis threshold u is chosen as 95% percentile, thus 5% of lowest returns in the estimation window are defined as extreme returns.

When we have GPD model for excess losses, we can estimate the tail of the underlying loss distribution F and finally obtain VaR as EVT measure of tail risk. Formula for tail probabilities is

$$\hat{F}(x) = \frac{N_u}{n}\left(1 + \hat{\xi}\frac{x-u}{\hat{\beta}}\right)^{-\frac{1}{\hat{\xi}}} \quad (37)$$

where $\frac{N_u}{n}$ is empirical estimator of $F(u)$ and $\hat{\xi}$ and $\hat{\beta}$ are parameters from GPD model. Finally, inverting this formula, we obtain VaR – high quantile for the underlying distribution. For $\alpha \geq F(u)$:

$$VaR_\alpha = q_\alpha(F) = u + \frac{\hat{\beta}}{\hat{\xi}}\left(\left(\frac{n(1-\alpha)}{N_u}\right)^{-\hat{\xi}} - 1\right) \quad (38)$$

Conditional EVT VaR model

VaR obtained using methodology in previous chapter is unconditional. This means that it does not account for changing volatility and its clustering. In dynamic risk management, volatility is updated every time volatility model is reestimated. The estimation of conditional EVT VaR models in this thesis is performed using this approach (McNeil A. J., 1999):

1) One-day volatility forecast is created by using GARCH(1,1) model. All three distribution assumptions – normal, Student's t and skewed Student's t – are made, thus creating three different conditional EVT VaR models.
2) Standardized residuals are extracted from GARCH(1,1) model and are regarded as realizations of the unobserved, independent noise variables z_t ($t \in (1 \dots n)$).
3) EVT model is applied to these residuals. At 95% percentile level of threshold, generalized Pareto distribution (GPD) is fitted and $VaR_\alpha(Z)$ is obtained according to formula (40).

4) Finally, $VaR_{\alpha}(Z)$ is multiplied by volatility forecast performed by GARCH(1,1) model in step 1 to obtain one-day conditional EVT VaR forecast.

All volatility models are reestimated every 25 observations. This limitation is taken because it takes too long to fit non-normal distributions and to estimate six volatility models for seven stock indexes using 1000 returns and repeat it for 2000 times (size of testing window), because more than 100000 model estimations are needed. The assumption is that model parameters do not change very much when a few last observations are omitted and a few new are included. Analysis of results of backtesting one model for one index shows that there is very little difference of results obtained from models estimated for every forecast and models are reestimated only every 25th forecast.

3.5. Model backtesting and hypothesis testing

In order to choose best model for forecasting risk we choose backtesting methodology. Backtesting is a procedure, which allows comparing various VaR models. In this procedure, VaR forecasts are compared with realized historical returns. VaR violation occurs when forecasted returns, or losses, are lower (better) than the real ones. Backtesting procedure allows for identifying limitations of the VaR models and may help to improve the model. Backtesting is a practical application used in financial institutions, because it is a powerful tool to check model validity and prevent underestimation of VaR, i.e. that financial institution has sufficient capital set aside for unanticipated losses. On the other hand, VaR model backtesting helps to ensure that model does not overestimate the VaR and excessive capital is not set aside, which could be used for more profitable activities. After models are backtested, the statistical significance of the VaR violations is tested using unconditional coverage and conditional coverage tests. Finally, hypothesis testing is presented in the end of this section.

Procedure of backtesting

The best way to validate model, is to create a forecast of future performance using currently available data and to wait for realization of the actual result. However, violations in the model usually occur rarely, so many future realizations will be needed in order to test the significance of the violations that occurred. To avoid waiting for future performance, it is much better and easier to backtest a model and to evaluate VaR forecasting performance employing only historical realizations of returns.

In-sample or estimation window is a number of historical observations used to forecast risk. Out-of-sample or testing window is data sample, for which we create forecast and later compare forecasts with actual returns in the sample. The sum of these windows is equal to the size of the whole data sample. We denote in-sample window as W_E, out-of-sample window as W_T and total sample size is N, such that:

$$N = W_E + W_T \quad (39)$$

In this thesis, we have 12 years of data with 250 trading days in each year on average: in total 3000 observations. We set $W_E = 1000$. In order to estimate forecasts to the rest of the sample we set $W_T = 2000$. This means that we use 1000 observations, approximately four years of data, to estimate model and create first forecast for day 1001. In this thesis rolling windows are used. This means that second forecast is performed using 1000 observations, but first observation is dropped and the latest one is included. To illustrate, in order to create forecast for observation on day 1002, observations from day 2 to day 1001 are used. In this way, rolling over the rest of the data, we create 2000 one-day forecasts, which are used in the backtesting to evaluate performance of forecasting.

VaR violation occurs when actual return exceeds VaR forecast. One of the simplest tests is to calculate number of violations and take a ratio of it with an out-of-sample window size W_T. This violation ratio expressed in percent should not differ significantly from level of significance of VaR model, which is backtested. Violations can be expressed by a “hit” function, or indication function, as presented in Campbell (2005):

$$I_t(\alpha) = \begin{cases} 1, if\ r_t \leq VaR_t(\alpha) \\ 0, if\ r_t > VaR_t(\alpha) \end{cases} \quad (40)$$

We denote violation ratio as:

$$VR = \frac{\sum_{t=1}^{W_T} I_t(\alpha)}{\alpha \times W_T} \quad (41)$$

VR should not differ much from 1. If it is in the range $[0.8, 1.2]$, it can be considered as a good forecast (Danielsson, 2011, p. 147). However, this is simple violation ratio analysis. In order to precisely evaluate significance of violation ratio, formal tests of significance are employed. These tests are presented in the next two sections.

Christoffersen, one of the most famous backtesting researchers, stated that “problem of determining the accuracy of a VaR model can be reduced to the problem of determining whether the sequence of VaR violation satisfies two properties” (Christoffersen, 1998):

a) Unconditional Coverage Property: probability, that realized loss is lower than VaR forecast, must be equal to level of significance of VaR model, i.e., α, and can be denoted as $\Pr(I_t(\alpha) = 1) = \alpha$. If this probability is higher than α, this suggests that the reported VaR measure systemically understates portfolio's actual level of risk. In contract, if too few VaR violations occur, model overstates risk level; meaning that model is too conservative and require too much capital as a buffer.

b) Independence Property: "this property places a restriction on how often VaR violations may occur" (Campbell, 2005). This means that any two elements of sequence of VaR violations have to be independent of each other. Otherwise, reported VaR measure is inadequate. If at least two VaR violations are in a row in the sequence of violations, this represents clustering of violations and informs that VaR model lacks a response to changing market risks (Campbell, 2005).

"It is important to recognize that the unconditional coverage and independence properties of the hit sequence are separate and distinct and must both be satisfied by an accurate VaR model" (Campbell, 2005). These two properties of indicator function $I_t(\alpha)$ are expressed as:

$$I_t(\alpha) \overset{i.i.d}{\sim} B(\alpha) \qquad (42)$$

which reads that the indicator function, $I_t(\alpha)$, is identically and independently distributed as a Bernoulli random variable with probability α. These two properties of indicator function are enough to completely characterize accurate VaR model with chosen significance level α.

Kupiec's unconditional coverage test

One of the first and most popular tests for VaR backtesting is Kupiec's unconditional coverage test. VaR model is accurate if VaR estimates are not violated more or less than $\alpha \times 100$ times. If the observed number of violations $\hat{\alpha} \times 100$ significantly differs from the expected number $\alpha \times 100$, VaR model is inaccurate and produces poor results.

Kupiec test is a log-likelihood ratio test, which has asymptotic distribution of $\chi^2(1)$ (chi-square with 1 degree of freedom) (Kupiec, 1995). The same indicator function is used to count violations:

$$I_t(\alpha) = \begin{cases} 1, if\ r_t \leq VaR_t(\alpha) \\ 0, if\ r_t > VaR_t(\alpha) \end{cases} \qquad (43)$$

Null hypothesis for VaR violations are:

$$H_0 : I_t(\alpha) \sim B(\alpha) \qquad (44)$$

Here $B(\alpha)$ stands for Bernoulli distribution. The Bernoulli density function is denoted as:

$$(1-\alpha)^{1-I_t}(\alpha)^{I_t}, \ \ I_t = 0{,}1. \qquad (45)$$

Violation probability $\hat{\alpha}$ is estimated by:

$$\hat{\alpha} = \frac{\sum_{t=1}^{W_T} I_t(\alpha)}{W_T} \qquad (46)$$

In order to simplify formulas, we denote number of violations as $v_1 = \sum_{t=1}^{W_T} I_t(\alpha)$. Thus number of non-violated estimations is testing window size less number of violations, denoted as $v_0 = W_T - v_1$. Under H_0, $\alpha = \hat{\alpha}$, the restricted likelihood function is (Christoffersen, 1998):

$$\mathcal{L}_R(\alpha) = \prod_{t=1}^{W_T} (1-\alpha)^{1-I_t} (\alpha)^{I_t} = (1-\alpha)^{v_0} (\alpha)^{v_1} \qquad (47)$$

The unrestricted likelihood function uses estimated probability $\hat{\alpha}$:

$$\mathcal{L}_U(\hat{\alpha}) = \prod_{t=1}^{W_T} (1-\hat{\alpha})^{1-I_t} (\hat{\alpha})^{I_t} = (1-\hat{\alpha})^{v_0} (\hat{\alpha})^{v_1} \qquad (48)$$

Finally, we use log-likelihood ratio test to check, if $\mathcal{L}_R(\alpha) = \mathcal{L}_U(\hat{\alpha})$, which means, to check, if $\alpha = \hat{\alpha}$:

$$LR_{UC} = 2\big(ln\mathcal{L}_R(\alpha) - ln\mathcal{L}_U(\hat{\alpha})\big) = 2ln\left(\frac{(1-\hat{\alpha})^{v_0}(\hat{\alpha})^{v_1}}{(1-\alpha)^{v_0}(\alpha)^{v_1}}\right) \sim \chi^2(1) \qquad (49)$$

For example, choosing $\alpha = 1\%$, from chi-square statistical table we get that null hypothesis is rejected if $LR_{UC} > 6.63$.

Kupiec and other unconditional coverage tests have two shortcomings. “The first shortcoming is that these tests are known to have difficulty detecting VaR measures that systematically under report risk.” (Campbell, 2005). Under current regulatory framework (BIS, 2006), sample size is equal to one year, which is very small sample, therefore these tests exhibit low power.

Kupiec test checks unconditional coverage of VaR estimates. However, conditional coverage, or clustering of violations, is not taken into account. As reported by Campbell (2005), it might happen, that VaR model successfully passes Kupiec test, but suffers from dependent VaR estimates, which fall into clusters, while both conditions must be satisfied for model to be accurate. Therefore, conditional coverage test developed by Christoffersen should be performed, which is presented in the next section (Christoffersen, 1998).

Christoffersen’s conditional coverage test

In order to perform VaR model accuracy test for second coverage property, Christoffersen developed conditional coverage, or independence, test (Christoffersen, 1998). Christoffersen denoted sequence of violations $I_t(\alpha)$ as binary first-order Markov chain with transition probability matrix:

$$\Pi_1 = \begin{bmatrix} \pi_{00} & \pi_{01} \\ \pi_{10} & \pi_{11} \end{bmatrix} = \begin{bmatrix} 1-\pi_{01} & \pi_{01} \\ 1-\pi_{11} & \pi_{11} \end{bmatrix}, \qquad (50)$$

Where $\pi_{ij} = Pr(I_t = j|I_{t-1} = i)$, i,j = 0,1. For example, π_{01} means that on day $t-1$ there was no violation, but on day t violation occurred. Therefore restricted likelihood function is:

$$\mathcal{L}_R(\Pi_1) = (1-\pi_{01})^{v_{01}}(\pi_{01})^{v_{01}}(1-\pi_{11})^{v_{10}}(\pi_{11})^{v_{11}}, \qquad (51)$$

where $v_{ij}, i, j = 0,1$, is a number of observations with value i followed by j. So the previously denoted total number of violations is $v_1 = v_{01} + v_{11}$. The number of observations, which not exceeded the VaR forecast, is $v_0 = v_{00} + v_{10}$. Maximum likelihood method is used to solve $\mathcal{L}_R(\Pi_1)$ function and to obtain parameters:

$$\widehat{\Pi_1} = \begin{pmatrix} \frac{v_{00}}{v_{00}+v_{01}} & \frac{v_{01}}{v_{00}+v_{01}} \\ \frac{v_{10}}{v_{10}+v_{11}} & \frac{v_{11}}{v_{10}+v_{11}} \end{pmatrix} \qquad (52)$$

Now we can estimate Markov chain model on the sequence of violations obtained from VaR model. Under the null hypothesis that there is no clustering of violations, the probability of seeing violation tomorrow does not depend on today seeing a violation, therefore, $\pi_{01} = \pi_{11} = \pi_2$ and transition probability matrix is:

$$\Pi_2 = \begin{bmatrix} 1-\pi_2 & \pi_2 \\ 1-\pi_2 & \pi_2 \end{bmatrix}, \qquad (53)$$

The unrestricted likelihood function under null hypothesis is:

$$\mathcal{L}_U(\Pi_2) = (1-\pi_2)^{(v_{00}+v_{10})}(\pi_2)^{(v_{01}+v_{11})} \qquad (54)$$

And maximum likelihood estimate is $\widehat{\Pi_2} = \widehat{\pi_2} = \frac{v_{01}+v_{11}}{v_{00}+v_{10}+v_{01}+v_{11}} = \frac{v_1}{v_0+v_1} = \frac{v_1}{N}$

Finally, likelihood ratio test is performed using both likelihood functions:

$$LR_{ind} = 2\left(ln\mathcal{L}_R(\Pi_1) - ln\mathcal{L}_U(\Pi_2)\right) \sim \chi^2(1) \qquad (55)$$

However, this is only one part of conditional coverage test, the test for independences. Full conditional coverage test contains evaluation of both properties for accurate VaR model. Therefore likelihood ratio test for conditional coverage is a sum of Kupiec's unconditional coverage test and test for independences and this joint test is asymptotically $\chi^2(2)$ distributed:

$$LR_{CC} = LR_{UC} + LR_{ind} \sim \chi^2(2) \qquad (56)$$

Although it may seem, that this test should be preferred to tests which only test one property, but usually it is not the case, because this test has less power to reject a VaR model which only satisfies one of the two properties. For example, if VaR model passes unconditional coverage test by Kupiec, but violates independence property, test of independences is more powerful than joint test to reject such model (Danielsson, 2011).

Hypothesis testing procedure

Hypothesis testing is based on thesis methodology is explained in previous sections. First hypothesis is tested by measuring properties of financial returns of Nordic and Baltic stock indexes. Statistical measures of return distribution such as skewness and kurtosis and results of normality tests provide information about normality and skewness of return distribution and allow verifying first hypothesis. Second and third hypotheses are verified by measuring GARCH(1,1) and APARCH(1,1) volatility models with normal, Student's t and skewed Student's t return distribution assumption for the whole data sample of each Nordic and Baltic stock index and comparing models using likelihood ratio tests. Restricted and unrestricted volatility models are compared by imposing restrictions are presented in hypotheses. For example, GARCH(1,1) model with normal distribution is restricted and Student's t GARCH(1,1) model is unrestricted, therefore there is one restriction and if likelihood ratio is statistically significant at 5 percent, hypothesis is not rejected. Fourth and fifth hypotheses are verified using one-sided *t*-test. Percentage of VaR models which pass Christoffersen test at 5% level of significance are compared for both groups of VaR models measured at 95^{th} and 99^{th} quantile. In case of fourth hypothesis, *t*-test measures whether average percentage of VaR models which pass Christoffersen test and are measured at 99^{th} quantile is statistically significantly greater than average percentage of VaR models measured at 95^{th} quantile. If it is statistically significant at 5%, fourth hypothesis is not rejected. Similarly, percentage of VaR models which pass Christoffersen test at 5% level of significance for Nordic stock indexes are compared with percentage of VaR models which pass Christoffersen test at 5% level for Baltic stock indexes. If *t*-test shows that average percentage of VaR models pass Christoffersen test for Nordic stock indexes is statistically significantly higher than average percentage of VaR models for Baltic stock indexes at both levels of confidence of VaR models, fifth hypothesis is not rejected. Finally, last hypothesis is not rejected if only one model for each Nordic and Baltic stock index has highest p-value of Christoffersen test during both low and high volatility time periods.

4. Empirical Results

Empirical research part is constructed in following manner. In first section, index return data and its statistical properties are described and preliminary data analysis is performed. Then, GARCH and APARCH volatility models with three distribution assumptions are fitted to the whole return sample of each stock index and parameters are reported. Volatility models are compared using likelihood ratio tests. Generalized Pareto distribution from extreme value theory is fitted to returns exceeding 95 percentile threshold and estimated parameters are presented. In third section, forecasts of 17 VaR models are estimated for seven stock indexes in Nordic and Baltic stock exchanges during low volatility (year 2004-2007) and high volatility (year 2008-2011) periods at 95 percent and 99 percent level of confidence. Backtesting procedure is performed to find VaR models with best forecasting performance during both periods with both levels of significance. Finally, VaR forecasts are graphically compared with actual returns for selected stock markets to analyze model forecasting behavior. In all parts, hypotheses are verified.

4.1. Descriptive statistics

The data used in this thesis consist of daily stock index values from Nordic (OMX Helsinki in Finland, OMX Stockholm in Sweden, OMX Copenhagen in Denmark and OSEAX in Oslo stock exchange in Norway) and Baltic (OMX Vilnius in Lithuania, OMX Riga in Latvia and OMX Tallinn in Estonia) stock exchanges. The largest possible period of data is taken for all stock indexes. Since OMX Vilnius and OMX Riga stock indexes were introduced on December 31st, 1999, the whole data sample consists for period 1999-12-31 – 2011-12-30. Due to different holiday systems in all analyzed countries, size of data sample for each country is slightly different, ranging from 2988 daily returns in Vilnius Stock Exchange to 3030 in Tallinn Stock Exchange, as presented in Table 4. All indexes are market capitalization indexes and include all shares traded on respective stock exchanges and represent market portfolios of these countries. Data is obtained from Bloomberg. For all non-stationary price time series p_t, daily returns r_t are defined as log differences of daily prices: $r_t = ln(p_t) - ln(p_{t-1})$, in order to obtain stationary index return time series.

Table 4 summarizes the basic statistical characteristics of the return series. The highest averages of daily returns are in Estonia (0.0476%), Latvia (0.0436%) and Lithuania (0.0366%). In contrast, average daily return for stock index in Finland is negative (-0.0332%), meaning that index value was much lower at the end of sample than in the beginning of the sample. The most volatile markets for the period were in Finland (2.047%), Latvia (1.597%) and Sweden (1.551%). Overall, daily return averages of all indexes are so small compared with daily volatility, that daily return mean is effectively, without loss of generality, assumed to be zero in this Thesis and risk measures are constructed accordingly.

Sample kurtosis estimates show that daily returns are not normally distributed. Normal distribution has kurtosis 3, or excess kurtosis equal to zero. Nordic countries have excess kurtosis from 2.88 to 5.5, while Baltic countries have excess kurtosis from 9.44 to 16.22. Although, all countries have non-normally distributed returns, Baltic countries have more leptokurtic return distributions and "fatter tails" than Nordic countries. Especially high excess kurtosis is in Lithuania, indicating highly peaked return distribution and very fat tails.

According to sample skewness, daily returns have almost symmetric distribution only in Sweden. Tallinn stock index have a positive skewness, while the rest exhibit negative skewness, which indicate that asymmetric fat tail of return distribution extends more towards negative returns than positive values, again indicating non-normality. Therefore, first hypothesis is partially rejected, because these statistical measures show that Nordic countries, although assumed to be developed, have stock markets, which index returns are non-normally and asymmetrically distributed. Hypothesis that Baltic stock indexes are non-normally and asymmetrically distributed is not rejected.

Index returns of Baltic stock exchanges have daily first order autocorrelation from 5.31 percent to 15.16 percent, while index returns of Nordic stock exchanges have daily first order autocorrelation from 0.3 percent to 6.2 percent. First order autocorrelation of squared returns for all indexes ranges from 13.43 percent to 47.43 percent. Squared returns are a good proxy for volatility. Therefore, high autocorrelation of squared returns for all indexes provides very strong evidence of the predictability of volatility and volatility clustering. Ljung-Box test of daily return autocorrelation with 20 lags is statistically significant at 1 percent level of significance for all indexes except OSEAX index, which is significant at 5 percent level of significance. Finally, the same test for squared returns shows that all indexes are statistically significant at 1 percent level of significance and this also confirms strong autocorrelation and predictability of volatility. Likewise, Jarque-Bera and Doornik-Hansen test results in the bottom of Table 4 indicate that returns of all stock indexes do not have normal distributions.

Descriptive charts (stock index value, one-day returns, their density and QQ (quantile-quantile) plot against Gaussian distribution) for each stock index are given in Figure 11 to Figure 17 in appendix. Volatility clustering is immediately apparent from the charts of daily returns. Although main clustering occurred in 2008-2009 for each index, pattern of clustering for the whole period is quite different in every stock exchange. The return density graphs and the QQ-plot against the normal distribution show that return distributions for all indexes exhibit fat tails.

Table 4: Summary statistics for Nordic and Baltic stock index returns for period of year 2000-2011.

	OMXV	OMXR	OMXT	OMXH	OMXS	OMXC	OSEAX
No of daily returns	2988	3010	3030	3015	3016	3007	3013
Mean (%)	0.036631	0.04357	0.047614	-0.033218	-0.00205	0.013937	0.029672
Median (%)	0.052703	0.029609	0.065121	0.054739	0.062949	0.052583	0.12425
Minimum (%)	-11.938	-14.705	-7.0459	-17.425	-8.0689	-10.583	-9.7089
Maximum (%)	11.001	10.18	12.094	14.563	8.6289	8.2013	9.1881
Standard deviation (%)	1.1871	1.597	1.2551	2.0474	1.5507	1.2094	1.5434
C.V.	32.407	36.653	26.36	61.635	754.65	86.775	52.014
Skewness	-0.47341	-0.58519	0.37683	-0.2133	-0.003724	-0.35049	-0.61357
Ex. Kurtosis	16.218	13.419	9.4425	5.2658	2.8797	5.5017	5.3665
ACF (1 lag) of returns (%)	14.36	5.31	15.16	0.30	0.90	6.20	0.50
ACF (1 lag) of squared returns (%)	37.86	47.43	15.11	13.43	17.04	23.19	27.10
Ljung–Box 20 lags (p-value, %)	0.00	0.00	0.00	0.98	0.00	0.03	1.63
Ljung–Box squared returns 20 lags (p-value, %)	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Jarque-Bera test	32858.5*	22754.4*	11328.3*	3506.33*	1042.09*	3853.93*	3804.54*
Doornik-Hansen test	5065.18*	3870.51*	2680.08*	1280.56*	552.937*	1279.62*	1006.66*

Note: * Significant at 1% level.

4.2. GARCH, APARCH and EVT model estimation

In order to perform VaR analysis, the first step is to estimate parameters of volatility models: GARCH(1,1) and APARCH(1,1). EWMA model does not require any estimation for the conditional volatility specification, because it is equivalent to IGARCH(1,1) model with predefined parameters ($\omega = 0, \alpha = 0.06, \beta = 0.94$) (Jorion, 2007).

Table 5 and Table 6 present the results for maximum likelihood estimation of GARCH(1,1) and APARCH(1,1) models with Normal, Student's t and skewed Student's t return distributions on seven Nordic and Baltic stock index return data. Table 5 shows that under normal return distribution assumption, the autoregressive effect in the volatility specification is strong for those models which have β around 0.90-0.95. Therefore models on Helsinki and Stockholm stock indexes suggest strong memory effects. Autoregressive effect for Vilnius stock index is weakest compared with other models, but still statistically significant at 1% significance level. For the GARCH(1,1) models with Student's t and skewed Student's t distribution assumption, parameter β estimates are lower for most of the indexes, but they are still very significant. The estimated degrees of freedom parameters ν, which show

Table 5: Parameter estimates of the GARCH(1,1) model for Nordic and Baltic stock index returns

Parameter:	OMXV	OMXR	OMXT	OMXH	OMXS	OMXC	OSEAX
	Normal distribution						
ω	0.0000096*	0.0000062*	0.0000013*	0.0000008*	0.0000020*	0.0000027*	0.0000050*
	(0.0000021)	(0.0000011)	(0.0000003)	(0.0000003)	(0.0000005)	(0.0000006)	(0.0000011)
α	0.1298976*	0.1441769*	0.1409514*	0.0492883*	0.088954*	0.1061331*	0.1187764*
	(0.0185331)	(0.0174751)	(0.0133201)	(0.0056151)	(0.0105177)	(0.0123752)	(0.0131080)
β	0.8014517*	0.832231*	0.8704028*	0.9499815*	0.9044451*	0.8743004*	0.8575055*
	(0.0309328)	(0.0186654)	(0.0101184)	(0.0052244)	(0.0107216)	(0.0146911)	(0.0150225)
LLH	-9430.47	-8976.27	-9415.20	-8060.64	-8756.03	-9491.08	-8846.33
	Student's t distribution						
ω	0.0000143*	0.0000067*	0.0000013**	0.0000009**	0.0000013*	0.0000025*	0.0000045*
	(0.0000026)	(0.0000018)	(0.0000005)	(0.0000004)	(0.0000005)	(0.0000007)	(0.0000011)
α	0.3125728*	0.2033738*	0.1879142*	0.054206*	0.0805323*	0.1098617*	0.111103*
	(0.0463418)	(0.0375895)	(0.0272677)	(0.0076525)	(0.0109899)	(0.0146258)	(0.0144156)
β	0.6182793*	0.7996223*	0.8439306*	0.944826*	0.916508*	0.8736064*	0.8699995*
	(0.0453460)	(0.0327887)	(0.0179657)	(0.0071679)	(0.0109133)	(0.0164991)	(0.0161693)
ν	3.98696*	3.656209*	3.940997*	6.916458*	10.000000*	10.000000*	10.000000*
	(0.3027936)	(0.2698196)	(0.3034089)	(0.7715570)	(1.5642040)	(1.6186150)	(1.3831160)
LLH	-9791.15	-9204.48	-9657.68	-8148.92	-8781.44	-9517.11	-8854.91
	Skewed Student's t distribution						
ω	0.0000156*	0.0000067*	0.0000013**	0.0000009**	0.0000013*	0.0000025*	0.0000047*
	(0.0000029)	(0.0000018)	(0.0000005)	(0.0000004)	(0.0000005)	(0.0000007)	(0.0000011)
α	0.3346866*	0.2054071*	0.1926987*	0.0536774*	0.0790159*	0.1055994*	0.1050128*
	(0.0498621)	(0.0374433)	(0.0283762)	(0.0075425)	(0.0106637)	(0.0142854)	(0.0136078)
β	0.5995265*	0.7966791*	0.8424632*	0.9452121*	0.9185011*	0.8783623*	0.8759345*
	(0.0467804)	(0.0328430)	(0.0182674)	(0.0070891)	(0.0105615)	(0.0161604)	(0.0154256)
Skew (g)	0.9298621*	1.022021*	0.9519507*	0.9272443*	0.895657*	0.9142917*	0.8223197*
	(0.0200621)	(0.0206608)	(0.0192837)	(0.0230477)	(0.0228121)	(0.0229556)	(0.0221201)
ν	3.900258*	3.697988*	3.864305*	6.95901*	10.000000*	10.000000*	10.000000*
	(0.2964330)	(0.2773171)	(0.2977031)	(0.7842807)	(1.5515910)	(1.6268650)	(1.4310000)
LLH	-9796.81	-9205.07	-9660.59	-8153.53	-8790.96	-9523.51	-8882.65

Note: Parameters for both GARCH(1,1) and APARCH(1,1) models estimated using the entire dataset and assuming three different distributions for the standardized residuals. Standard errors are presented in parentheses. *** - 10% significance level, ** - 5% significance level, * - 1% significance level. LLH – Log likelihood estimate.

the shape of the Student's t distribution, are from 3.66 to 3.99 for GARCH(1,1) models on Baltic stock indexes and from 6.92 to 10 for Nordic stock indexes, suggesting that return distribution have fat tails. Furthermore, skewness parameter g in many of the skewed Student's t GARCH(1,1) models is lower than one and tells that returns distributions are negatively skewed.

Table 6 has two additional parameters compared to Table 5. Parameter δ is a power term and parameter γ measures for leverage effects in APARCH(1,1) model. For all stock indexes, parameter γ

Table 6: Parameter estimates of the APARCH(1,1) model for Nordic and Baltic stock index returns

Parameter:	OMXV	OMXR	OMXT	OMXH	OMXS	OMXC	OSEAX
				Normal distribution			
ω	0.0004971*	0.0000162*	0.0002936*	0.0002767*	0.0001544*	0.0001812*	0.0001006*
	(0.0001467)	(0.0000029)	(0.0000616)	(0.0000544)	(0.0000258)	(0.0000299)	(0.0000146)
α	0.1354369*	0.150965*	0.1323739*	0.0521239*	0.0644528*	0.0879038*	0.1014603*
	(0.0207723)	(0.0181260)	(0.0113256)	(0.0052272)	(0.0107300)	(0.0105752)	(0.0125766)
β	0.8334958*	0.8361393*	0.8865088*	0.9533765*	0.9271551*	0.8998334*	0.8716271*
	(0.0351833)	(0.0173828)	(0.0094099)	(0.0044868)	(0.0082895)	(0.0116927)	(0.0137624)
γ	0.1612439*	0.0544921***	0.0500761	0.6759732*	0.9219406*	0.5483461*	0.5161255*
	(0.0483078)	(0.0326633)	(0.0335949)	(0.0955565)	(0.1443326)	(0.0861271)	(0.0807870)
δ	1.097148*	1.787502*	0.9351369*	0.8306564*	1.116259*	1.127871*	1.382512*
	(0.1625543)	(0.1975944)	(0.1202137)	(0.1429197)	(0.1400371)	(0.1685291)	(0.2069016)
LLH	-9440.34	-8973.42	4206.72	3986.73	-8521.47	-9315.82	-8844.92
				Student's t distribution			
ω	0.000206*	0.0000705*	0.0002654*	0.0003246*	0.0001294*	0.0001876*	0.0001248*
	(0.0000381)	(0.0000180)	(0.0000843)	(0.0000742)	(0.0000253)	(0.0000353)	(0.0000216)
α	0.3010511*	0.2062891*	0.1624235*	0.0593151*	0.0618438*	0.0959006*	0.1009755*
	(0.0404101)	(0.0314490)	(0.0198694)	(0.0068587)	(0.0073168)	(0.0128942)	(0.0138328)
β	0.6501865*	0.8176698*	0.8711491*	0.947392*	0.9310791*	0.8936568*	0.87768*
	(0.0472031)	(0.0284460)	(0.0153639)	(0.0059632)	(0.0074363)	(0.0141639)	(0.0152082)
γ	0.1184564*	0.1073159**	0.0849066**	0.6656353*	1.0000000*	0.5414026*	0.5224194*
	(0.0430958)	(0.0476506)	(0.0447076)	(0.1023759)	(0.0326925)	(0.0960129)	(0.0933175)
δ	1.419038*	1.463731*	0.9298443*	0.8067619*	1.129028*	1.127075*	1.326462*
	(0.2190472)	(0.2146634)	(0.1406413)	(0.1528611)	(0.1426493)	(0.1882001)	(0.2266318)
ν	4.065528*	3.632837*	4.00831*	7.708*	10.000000*	10.000000*	10.000000*
	(0.3134043)	(0.2665128)	(0.3104730)	(0.9128563)	(1.3596430)	(1.4515970)	(1.2681820)
LLH	-9797.41	-9205.22	-9663.76	-8099.82	-8796.98	-9518.52	-8869.63
				Skewed Student's t distribution			
ω	0.0001874*	0.0000666*	0.0002691*	0.0003276*	0.0001301*	0.0001964*	0.0000991*
	(0.0000346)	(0.0000168)	(0.0000854)	(0.0000737)	(0.0000253)	(0.0000372)	(0.0000171)
α	0.3162918*	0.2080701*	0.1640035*	0.0586136*	0.061557*	0.0929667*	0.0960267*
	(0.0431780)	(0.0314446)	(0.0202239)	(0.0067238)	(0.0072423)	(0.0126286)	(0.0134837)
β	0.6354247*	0.8150314*	0.8707758*	0.947919*	0.9316907*	0.8970455*	0.8807939*
	(0.0487933)	(0.0286495)	(0.0154293)	(0.0058594)	(0.0073458)	(0.0138705)	(0.0149917)
γ	0.0916579**	0.1095147**	0.0617984	0.6686179*	1.0000000*	0.5373109*	0.4836866*
	(0.0433001)	(0.0470741)	(0.0457296)	(0.1022519)	(0.0307284)	(0.0971771)	(0.0949450)
δ	1.453155*	1.476299*	0.9280954*	0.8044165*	1.12625*	1.112592*	1.382881*
	(0.2236819)	(0.2152935)	(0.1398219)	(0.1524513)	(0.1401759)	(0.1871556)	(0.2370525)
Skew (g)	0.9393227*	1.023226*	0.9547141*	0.9222414*	0.8901335*	0.9262014*	0.8362157*
	(0.0208978)	(0.0207899)	(0.0201856)	(0.0234574)	(0.0234289)	(0.0238330)	(0.0227453)
ν	3.965635*	3.679127*	3.92503*	7.817513*	10.000000*	10.000000*	10.000000*
	(0.3057253)	(0.2749248)	(0.3044327)	(0.9460631)	(1.3545590)	(1.4729850)	(1.3251360)
LLH	-9801.22	-9206.06	-9666.42	-8101.60	-8806.17	-9521.57	-8894.47

indicate leverage effect for negative returns in the conditional variance model APARCH(1,1). For most indexes this estimate is statistically significant at 1% significance level, however, for OMX Riga stock index this term is significant at 5% level with Student's t and skewed Student's t distributions and at 10% with normal distribution assumption.

For OMX Tallinn stock index γ is significant at 10% only for Student's t distribution, for other two distributions, this term is insignificant suggesting that there was no leverage effect in that stock market during the period 2000-2011. In contrast, strongest leverage effect was present in the OMX Stockholm stock market. The power terms δ, estimated for the APARCH(1,1) models, were fitted for all stock index returns and are from 0.80 to 1.79 and are all statistically significant at 1% significance level. The results suggest the optimal power term is lower than two, which would seem to support the use of a APARCH(1,1) model rather than GARCH(1,1) model, because it allows the power term to be estimated (GARCH(1,1) model has fixed power term $\delta = 2$) (Huang & Lin, 2004).

In order to compare models, likelihood ratio tests were performed. It allows testing significance of multiple parameters in those models. Table 7 presents results of likelihood ratio tests for the model estimates in Table 5 and Table 6. Second hypothesis H2 is not rejected, because skewed Student's t GARCH(1,1) model is better than ordinary GARCH(1,1) model for all stock indexes. This suggests that skewed Student's t distribution significantly improves model, assuming it instead of normal distribution. Additionally, even symmetric Student's t distribution is better in APARCH(1,1) model than the normal distribution. However, third hypothesis H3 is partially rejected, because normal APARCH(1,1) is better than normal GARCH(1,1) only for OMX Vilnius stock index return modeling. For all other stock markets, normal GARCH(1,1) is better. However, when skewed Student's t distribution is assumed in APARCH(1,1) model and compared with skewed Student's t GARCH(1,1) model, it becomes better for much more stock indexes. It is not significantly better for OMX Riga stock index modeling, but for other markets is significantly better at 1% significance level. Finally, accounting for return asymmetry in Student's t APARCH(1,1) model significantly improves it for OMX Vilnius, OMX Stockholm, OMX Copenhagen and Oslo Stock exchange All Share indexes at 1% significance level, but for OMX Helsinki and OMX Tallinn is significantly better at 5% significance level, while for OMX Riga stock index this gives no improvement.

Additionally, VaR is estimated using Extreme Value Theory (EVT). In this thesis generalized Pareto distribution is used to model extreme values. This distribution belongs to Peaks-over-Threshold (POT) modeling described in section 4.3.3. Firstly we choose a threshold which is equal to 95th percentile of inverse return distribution. Return distribution are inversed, because traditionally EVT focuses on upper tail and we want to measure VaR which focuses on negative returns. Table 8 presents results of initial data preparation and model estimation. For my stock return data, 5% lowest returns make 149-151 return samples below threshold. Figure 18(a) to Figure 24(a) show threshold for each stock index.

Table 7: Likelihood ratio tests for the results in Table 5 and Table 6

Index	Unrestricted model:	Restricted model:	LR statistic:	Restrictions	P-value:
OMXV	sstd garch(1, 1)	norm garch(1, 1)	732.69	2	0.0000
	norm aparch(1, 1)	norm garch(1, 1)	19.75	2	0.0000
	std aparch(1, 1)	norm aparch(1, 1)	714.14	1	0.0000
	sstd aparch(1, 1)	sstd garch(1, 1)	8.82	2	0.0061
	sstd aparch(1, 1)	std aparch(1, 1)	7.62	1	0.0032
OMXR	sstd garch(1, 1)	norm garch(1, 1)	457.60	2	0.0000
	norm aparch(1, 1)	norm garch(1, 1)	-5.69	2	0.0000
	std aparch(1, 1)	norm aparch(1, 1)	463.60	1	0.0000
	sstd aparch(1, 1)	sstd garch(1, 1)	1.99	2	0.1844
	sstd aparch(1, 1)	std aparch(1, 1)	1.69	1	0.1313
OMXT	sstd garch(1, 1)	norm garch(1, 1)	490.78	2	0.0000
	norm aparch(1, 1)	norm garch(1, 1)	-27243.83	2	0.0000
	std aparch(1, 1)	norm aparch(1, 1)	27740.95	1	0.0000
	sstd aparch(1, 1)	sstd garch(1, 1)	11.66	2	0.0015
	sstd aparch(1, 1)	std aparch(1, 1)	5.33	1	0.0120
OMXH	sstd garch(1, 1)	norm garch(1, 1)	185.79	2	0.0000
	norm aparch(1, 1)	norm garch(1, 1)	-24094.74	2	0.0000
	std aparch(1, 1)	norm aparch(1, 1)	24173.11	1	0.0000
	sstd aparch(1, 1)	sstd garch(1, 1)	-103.87	2	0.0000
	sstd aparch(1, 1)	std aparch(1, 1)	3.56	1	0.0357
OMXS	sstd garch(1, 1)	norm garch(1, 1)	69.86	2	0.0000
	norm aparch(1, 1)	norm garch(1, 1)	-469.11	2	0.0000
	std aparch(1, 1)	norm aparch(1, 1)	551.02	1	0.0000
	sstd aparch(1, 1)	sstd garch(1, 1)	30.42	2	0.0000
	sstd aparch(1, 1)	std aparch(1, 1)	18.37	1	0.0000
OMXC	sstd garch(1, 1)	norm garch(1, 1)	64.85	2	0.0000
	norm aparch(1, 1)	norm garch(1, 1)	-350.52	2	0.0000
	std aparch(1, 1)	norm aparch(1, 1)	405.39	1	0.0000
	sstd aparch(1, 1)	sstd garch(1, 1)	-3.88	2	0.0000
	sstd aparch(1, 1)	std aparch(1, 1)	6.10	1	0.0077
OSEAX	sstd garch(1, 1)	norm garch(1, 1)	72.66	2	0.0000
	norm aparch(1, 1)	norm garch(1, 1)	-2.81	2	0.0000
	std aparch(1, 1)	norm aparch(1, 1)	49.41	1	0.0000
	sstd aparch(1, 1)	sstd garch(1, 1)	23.64	2	0.0000
	sstd aparch(1, 1)	std aparch(1, 1)	49.69	1	0.0000

Note: norm stands for normal distribution, std – Student's t and sstd – skewed Student's t distribution.

These charts are inverted, therefore positive returns and threshold in the chart are actually negative returns and negative threshold. Exact threshold return is shown in Table 8. Figure 18(b) to Figure 24(b) show only peaks over threshold. This excess distribution is used to fit generalized Pareto distribution and parameter estimates are obtained, which are presented in Table 8. In the Figure 18(c) to Figure 24(c) we clearly see that model fit excess losses very well. When value of ξ is positive, it means that excess return distribution have fat tails. From Table 8 we observe that OMX Riga and OMX Vilnius has most heavy-tailed excess distributions, while OMX Tallinn stock index return excess distribution follows exponential distribution, because ξ is almost zero. Finally, in order to calculate EVT VaR, tails of distributions should be estimated. These tails are presented in Figure 18(d) to Figure 24(d). This graph is also inverted and on x axis negative returns are presented as positive. Y axis in this chart indicates tail probabilities. For example, top left corner of tail graph in Figure 18(d) shows that threshold return of -0.016247 corresponds to tail probability of 0.05. Points in the graph represent 149 extreme losses and "the curve going through these points allows extrapolating into the area where the data become a sparse and unreliable guide to their unknown parent distribution" (McNeil A. J., 1999).

Table 8: 95th percentile threshold exceedances, return over threshold and generalized Pareto distribution (GPD) estimates

95th quantile threshold	OMXV	OMXR	OMXT	OMXH	OMXS	OMXC	OSEAX
Threshold exceedances	149	150	151	150	150	150	150
Threshold return	-0.016247	-0.021468	-0.018209	-0.032956	-0.025762	-0.019315	-0.024851
GPD estimate xi (ξ)	0.260699	0.380114	-0.001247	0.093269	-0.073641	0.059992	0.196075
	(0.105358)	(0.113688)	(0.096667)	(0.081154)	(0.076844)	(0.076470)	(0.116739)
GPD estimate beta (β)	0.009315	0.010601	0.011321	0.014753	0.011581	0.009517	0.011444
	(0.001192)	(0.001416)	(0.001395)	(0.001670)	(0.001263)	(0.001029)	(0.001587)

4.3. Analysis of models' forecasting performance

This section is split into three parts. First part presents model forecasting performance for the period 2004-2007 approximately. 1000 returns of each index (starting using data from period 2000-2003 and moving by one day to the end of 2007) are used to make a forecast for another 988-1030 data points depending on the index. Violation ratios, Kupiec's, independence and Christoffersen's tests are used to evaluate performance of model forecasting for low volatility period before financial crisis in 2008. Second part presents next four years (around year 2008-2011, 1000 data points), which exhibit high volatility cluster and extreme returns are observed at that time. Final part uses forecasts from both periods together to backtest models estimated in both low and high volatility periods. In each part models are evaluated at 1 percent and 5 percent significance level and best model is chosen for each index, each period and at each significance level using results of backtesting procedure. Although the performance is different across VaR models using different assumptions of distributions and levels of confidence, some clear patterns are observed.

4.3.1. Low volatility period before financial crisis

First of all we present Table 9, which shows violation ratios and Kupiec test p-values for the forecasting period of year 2004-2007 at 5 percent significance level. Violation ratio shows whether VaR forecasts exceeded actual returns more times than expected according to significance level. For example, if model forecasted 65 VaR values, which exceeded actual losses for 1000 returns forecasting window, while at 5 percent the expected number of violations was 50, we get that violation ratio is 1.3 and risk is underestimated by the model. Kupiec test examines "whether the observed percentage of violations is statistically equal to expected percentage of violations" (Kupiec, 1995). If violation ratio is close to 1, it means that number of violations is very close to the expected number under significance level of the model.

Table 9: Violation ratios and Kupiec test p-values for period 2004-2007 at 5% significance level

	OMXV		OMXR		OMXT		OMXH		OMXS		OMXC		OSEAX	
Model	VR	Kupiec	VR	Kupiec	VR	Kupiec	VR	Kupiec	VR	Kupiec	VR	Kupiec	VR	Kupiec
1	1.21	0.13	0.93	0.61	0.89	0.42	0.59	0.00	0.77	0.08	0.87	0.35	1.24	0.09
2	0.93	0.62	0.53	0.00	0.54	0.00	0.51	0.00	0.77	0.08	0.79	0.12	1.13	0.37
3	1.48	0.00	0.73	0.04	1.55	0.00	0.63	0.00	0.87	0.32	0.87	0.35	1.20	0.15
4	1.26	0.08	0.77	0.08	1.38	0.01	0.59	0.00	0.77	0.08	0.83	0.21	1.07	0.63
5	1.23	0.10	0.95	0.72	0.89	0.42	0.59	0.00	0.77	0.08	0.85	0.28	1.26	0.06
6	0.73	0.04	0.73	0.04	0.82	0.16	1.06	0.64	1.14	0.31	1.07	0.60	1.07	0.63
7	1.03	0.82	4.61	0.00	1.05	0.72	1.38	0.01	1.44	0.00	1.27	0.06	1.26	0.06
8	1.05	0.71	4.65	0.00	1.18	0.19	1.32	0.03	1.34	0.02	1.17	0.22	1.05	0.74
9	0.89	0.42	0.57	0.00	0.58	0.00	0.99	0.91	1.08	0.55	0.95	0.73	1.01	0.96
10	1.03	0.82	4.36	0.00	0.85	0.27	1.36	0.01	1.38	0.01	1.13	0.35	1.16	0.24
11	1.01	0.93	4.42	0.00	0.85	0.27	1.20	0.15	1.08	0.55	0.99	0.96	0.97	0.81
12	0.87	0.34	0.53	0.00	0.60	0.00	0.99	0.91	1.10	0.46	0.95	0.73	1.16	0.24
13	1.03	0.82	4.34	0.00	0.83	0.21	1.32	0.03	1.32	0.03	1.25	0.08	1.28	0.05
14	1.01	0.93	4.38	0.00	0.82	0.16	1.22	0.12	1.20	0.15	0.99	0.96	0.99	0.93
15	1.15	0.28	1.09	0.52	0.99	0.94	1.10	0.46	1.10	0.46	1.01	0.93	1.01	0.96
16	1.11	0.42	1.11	0.43	1.09	0.53	1.18	0.19	1.14	0.31	0.97	0.84	0.99	0.93
17	1.11	0.42	1.09	0.52	1.09	0.53	1.20	0.15	1.14	0.31	1.01	0.93	0.99	0.93

Note: VR denotes violation ratio, Kupiec – Kupiec test p-values. In Table 9 to Table 20 red cells indicate models which fail likelihood ratio test at 5% significance level. Green cells indicate best model for each index, i.e. it is highest value in each column for likelihood ratio test p-value.

Table 10: Independence test and Christoffersen test p-values for period 2004-2007 at 5% significance level

	OMXV		OMXR		OMXT		OMXH		OMXS		OMXC		OSEAX	
Model	ind	Chr	ind	Chr	ind	Chr	ind	Chr	ind	Chr	ind	Chr	ind	Chr
1	0.00	0.00	0.00	0.00	0.00	0.00	0.29	0.00	0.07	0.04	0.00	0.00	0.01	0.01
2	0.00	0.00	0.00	0.00	0.00	0.00	0.17	0.00	0.07	0.04	0.00	0.00	0.01	0.03
3	0.00	0.00	0.00	0.00	0.00	0.00	0.37	0.01	0.01	0.02	0.00	0.00	0.01	0.01
4	0.00	0.00	0.00	0.00	0.00	0.00	0.29	0.00	0.07	0.04	0.00	0.00	0.02	0.07
5	0.00	0.00	0.00	0.00	0.00	0.00	0.29	0.00	0.07	0.04	0.00	0.00	0.01	0.00
6	0.00	0.00	0.00	0.00	0.01	0.01	0.08	0.20	0.15	0.22	0.00	0.00	0.23	0.43
7	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.15	0.06
8	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.04	0.01	0.00	0.00	0.20	0.42
9	0.00	0.01	0.01	0.00	0.06	0.00	0.00	0.01	0.10	0.21	0.03	0.09	0.15	0.36
10	0.00	0.02	0.00	0.00	0.01	0.02	0.05	0.01	0.15	0.01	0.00	0.00	0.18	0.20
11	0.00	0.01	0.00	0.00	0.00	0.00	0.03	0.04	0.10	0.21	0.01	0.04	0.11	0.28
12	0.01	0.02	0.00	0.00	0.00	0.00	0.01	0.04	0.29	0.43	0.28	0.52	0.40	0.35
13	0.00	0.02	0.00	0.00	0.00	0.00	0.01	0.00	0.22	0.04	0.05	0.03	0.67	0.13
14	0.00	0.01	0.00	0.00	0.00	0.00	0.01	0.01	0.23	0.18	0.35	0.64	0.13	0.32
15	0.00	0.00	0.26	0.43	0.00	0.01	0.01	0.03	0.29	0.43	0.00	0.01	0.15	0.36
16	0.00	0.01	0.04	0.08	0.00	0.00	0.03	0.04	0.02	0.04	0.01	0.03	0.13	0.32
17	0.00	0.01	0.10	0.21	0.00	0.00	0.03	0.04	0.02	0.04	0.00	0.01	0.13	0.32

Note: ind denotes independence test p-values, Chr – Christoffersen test p-values.

Table 10 shows the results of independence and Christoffersen tests at 5 percent significance level. From Table 9 it is clearly seen that many models are performing well when forecasting returns of almost all stock indexes (except OMX Riga and OMX Helsinki) in low volatility period according to Kupiec test results. However, Table 10 shows that most of the models fail independence and Christoffersen tests for Baltic stock indexes.

OMX Vilnius stock index is best forecasted using GARCH(1,1) and APARCH(1,1) models under skewed Student's t distribution assumption. Only unconditional Student's t VaR and normal EWMA VaR fails Kupiec test. However, only skewed Student's t GARCH(1,1) model passes independence and Christoffersen tests, but independence and conditional coverage is not rejected only at 1% level of significance. Quite different situation is for OMX Riga stock index, where best performer is unconditional EVT VaR model according to Kupiec test results in Table 9. Other extreme value theory VaR models also perform quite well. This might be explained that during the examined period as seen in Figure 12(b) there were several extreme negative returns. This huge shock was caused by the privatization of 2 percent of Latvijas Gaze shares in the auction in OMX Riga stock exchange on

Table 11: Violation ratios and Kupiec test p-values for period 2004-2007 at 1% significance level

	OMXV		OMXR		OMXT		OMXH		OMXS		OMXC		OSEAX	
Model	VR	Kupiec	VR	Kupiec	VR	Kupiec	VR	Kupiec	VR	Kupiec	VR	Kupiec	VR	Kupiec
1	1.42	0.22	1.09	0.78	1.17	0.60	0.39	0.03	0.98	0.96	0.79	0.50	1.38	0.25
2	2.02	0.00	1.19	0.56	1.46	0.17	1.28	0.39	2.26	0.00	2.38	0.00	2.27	0.00
3	1.52	0.13	0.50	0.07	3.59	0.00	0.49	0.07	1.38	0.25	1.49	0.15	1.68	0.05
4	1.42	0.22	1.09	0.78	3.40	0.00	0.30	0.01	1.18	0.57	0.79	0.50	1.38	0.25
5	1.62	0.07	1.09	0.78	1.26	0.42	0.49	0.07	1.18	0.57	1.09	0.77	1.28	0.39
6	2.02	0.00	1.58	0.09	1.65	0.06	1.77	0.03	2.66	0.00	2.48	0.00	1.97	0.01
7	1.62	0.07	17.43	0.00	1.07	0.83	1.18	0.57	1.97	0.01	2.18	0.00	1.38	0.25
8	1.62	0.07	17.92	0.00	1.36	0.27	0.99	0.96	1.67	0.05	1.59	0.08	0.99	0.97
9	1.82	0.02	0.79	0.49	1.17	0.60	1.97	0.01	2.46	0.00	2.09	0.00	1.78	0.03
10	1.21	0.51	16.34	0.00	0.78	0.45	0.99	0.96	1.77	0.03	1.19	0.55	1.09	0.79
11	1.21	0.51	16.73	0.00	0.78	0.45	0.89	0.71	1.38	0.25	0.79	0.50	0.79	0.49
12	1.72	0.04	0.59	0.16	1.46	0.17	1.67	0.05	2.36	0.00	1.79	0.02	1.78	0.03
13	1.21	0.51	16.24	0.00	0.97	0.92	1.28	0.39	1.57	0.09	1.19	0.55	1.38	0.25
14	1.11	0.73	16.73	0.00	0.97	0.92	0.99	0.96	1.18	0.57	1.09	0.77	0.99	0.97
15	1.72	0.04	0.89	0.72	1.07	0.83	0.99	0.96	1.67	0.05	1.09	0.77	1.18	0.57
16	1.82	0.02	0.69	0.30	1.17	0.60	0.99	0.96	1.67	0.05	1.09	0.77	1.09	0.79
17	1.62	0.07	0.79	0.49	1.17	0.60	0.99	0.96	1.67	0.05	1.09	0.77	1.09	0.79

Note: VR denotes violation ratio, Kupiec – Kupiec test p-values

July 19, 2001, when shares was sold at three times higher price than in the market (Baltic Stock Exchanges, 2001). In four-week period index value increased by more than 100 percent. After that, it dropped by half during the next three weeks reaching levels before this volatility shock. Therefore many conditional volatility models failed to produce reasonable forecasts due to this structural break. As seen in Table 10, only extreme value models with GARCH volatility filters passed independence test as well as Christoffersen test. EVT VaR model with normal return distribution assumption was best. Although many models forecasting OMX Tallinn stock index returns performed well according to Kupiec test results, best performance was shown by EVT VaR model with normal GARCH filter, as presented in the Table 9. However, most models faced rejection of independence, which means that violations are clustered and models are inappropriate. Only GARCH(1,1) model under normal distribution passed independence test as well as Kupiec test, suggesting this model as the best at 5% significance level. For OMX Helsinki, none of the unconditional VaR models passed Kupiec test. This can be explained looking at Figure 14(b), which shows that models were estimated under the volatile period of year 2000-2004 for that index, therefore unconditional models overestimated risk and number

Table 12: Independence test and Christoffersen test p-values for period 2004-2007 at 1% significance level

	OMXV		OMXR		OMXT		OMXH		OMXS		OMXC		OSEAX	
Model	ind	Chr	ind	Chr	ind	Chr	ind	Chr	ind	Chr	ind	Chr	ind	Chr
1	0.53	0.38	0.00	0.02	0.01	0.02	0.86	0.09	0.66	0.90	0.72	0.74	0.18	0.21
2	0.06	0.00	0.01	0.02	0.00	0.00	0.15	0.25	0.54	0.00	0.02	0.00	0.11	0.00
3	0.22	0.15	0.82	0.20	0.00	0.00	0.82	0.19	0.18	0.21	0.22	0.16	0.29	0.08
4	0.53	0.38	0.00	0.02	0.00	0.00	0.89	0.03	0.59	0.74	0.72	0.74	0.18	0.21
5	0.47	0.15	0.00	0.02	0.00	0.00	0.82	0.19	0.59	0.74	0.62	0.85	0.15	0.25
6	0.06	0.00	0.02	0.02	0.28	0.09	0.42	0.06	0.20	0.00	0.15	0.00	0.41	0.02
7	0.26	0.10	0.00	0.00	0.10	0.26	0.59	0.74	0.37	0.02	0.09	0.00	0.53	0.42
8	0.26	0.10	0.00	0.00	0.18	0.22	0.66	0.90	0.45	0.11	0.25	0.12	0.66	0.90
9	0.41	0.05	0.72	0.74	0.13	0.27	0.41	0.02	0.64	0.00	0.07	0.00	0.32	0.05
10	0.59	0.70	0.00	0.00	0.72	0.71	0.66	0.90	0.42	0.06	0.59	0.73	0.62	0.85
11	0.59	0.70	0.00	0.00	0.72	0.71	0.69	0.86	0.53	0.43	0.72	0.74	0.72	0.74
12	0.44	0.09	0.79	0.36	0.21	0.18	0.29	0.08	0.28	0.00	0.04	0.01	0.32	0.05
13	0.59	0.70	0.00	0.00	0.08	0.22	0.56	0.58	0.47	0.18	0.59	0.73	0.18	0.21
14	0.62	0.83	0.00	0.00	0.08	0.22	0.66	0.90	0.59	0.74	0.62	0.85	0.66	0.90
15	0.44	0.09	0.75	0.56	0.10	0.26	0.66	0.90	0.45	0.11	0.11	0.26	0.59	0.73
16	0.33	0.04	0.75	0.56	0.59	0.76	0.66	0.90	0.45	0.11	0.11	0.26	0.62	0.85
17	0.26	0.10	0.75	0.56	0.59	0.76	0.66	0.90	0.45	0.11	0.11	0.26	0.62	0.85

Note: ind denotes independence test p-values, Chr – Christoffersen test p-values

of violations was much lower than expected under 5% level of significance. Surprisingly, most of these models passed independence test according to Table 9. Nevertheless, only GARCH and APARCH model under normal distribution assumption performed best for Finland's all share stock index according to Kupiec test, but all of them failed independence test at 5 percent level of significance of independence test. Combining Kupiec and independent test results, Christoffersen test results in Table 10 show that best model was normal EWMA VaR at 5 percent level of significance for VaR models. OMX Stockholm stock index returns were quite well forecasted by many models in first period, however, only GARCH conditional volatility models showed best performance under normal and skewed Student's t distribution according to Kupiec test results in Table 9. However, independence and Christoffersen test p-values suggest that normal APARCH(1,1) and normal EVT GARCH(1,1) VaR models performed best. All VaR forecasting models passed Kupiec test for OMX Copenhagen stock index. Best performing models were the same as for OMX Vilnius index: GARCH(1,1) and APARCH(1,1) under skewed Student's t distribution. However, only normal APARCH(1,1) and skewed Student's t APARCH(1,1) VaR model did not reject both independence and conditional coverage null hypothesis, thus suggesting them as best models for OMXC at 5 percent level of significance. This is not surprising, because these models account not only for fat tails of return distribution, but also volatility clustering and asymmetry in return distribution. Finally, Oslo Stock

Table 13: Violation ratios and Kupiec test p-values for period 2008-2011 at 5% significance level

	OMXV		OMXR		OMXT		OMXH		OMXS		OMXC		OSEAX	
Model	VR	Kupiec	VR	Kupiec	VR	Kupiec	VR	Kupiec	VR	Kupiec	VR	Kupiec	VR	Kupiec
1	1.18	0.20	1.58	0.00	1.68	0.00	1.74	0.00	1.42	0.00	1.56	0.00	1.36	0.01
2	1.12	0.39	1.60	0.00	1.46	0.00	1.72	0.00	1.52	0.00	1.62	0.00	1.34	0.02
3	1.52	0.00	1.88	0.00	2.82	0.00	1.96	0.00	1.68	0.00	1.88	0.00	1.54	0.00
4	1.40	0.01	1.92	0.00	2.64	0.00	1.86	0.00	1.58	0.00	1.72	0.00	1.32	0.03
5	1.24	0.09	1.62	0.00	1.74	0.00	1.74	0.00	1.42	0.00	1.56	0.00	1.36	0.01
6	0.96	0.77	1.00	1.00	1.10	0.47	1.34	0.02	1.26	0.07	0.98	0.88	1.44	0.00
7	2.02	0.00	2.12	0.00	2.80	0.00	1.76	0.00	1.90	0.00	1.92	0.00	1.94	0.00
8	1.94	0.00	2.18	0.00	3.00	0.00	1.66	0.00	1.74	0.00	1.66	0.00	1.72	0.00
9	1.20	0.16	1.18	0.20	0.96	0.77	1.38	0.01	1.24	0.09	0.96	0.77	1.46	0.00
10	2.30	0.00	2.32	0.00	2.56	0.00	1.80	0.00	1.80	0.00	1.76	0.00	1.84	0.00
11	2.12	0.00	2.16	0.00	2.66	0.00	1.64	0.00	1.66	0.00	1.56	0.00	1.64	0.00
12	1.24	0.09	1.16	0.26	1.00	1.00	1.40	0.01	1.42	0.00	1.18	0.20	1.42	0.00
13	2.30	0.00	2.42	0.00	2.64	0.00	1.84	0.00	1.98	0.00	1.90	0.00	1.98	0.00
14	2.10	0.00	2.34	0.00	2.74	0.00	1.66	0.00	1.72	0.00	1.68	0.00	1.58	0.00
15	1.22	0.12	1.24	0.09	1.06	0.67	1.30	0.04	1.10	0.47	0.92	0.56	1.32	0.03
16	1.16	0.26	1.26	0.07	1.10	0.47	1.28	0.05	1.04	0.77	0.96	0.77	1.32	0.03
17	1.16	0.26	1.28	0.05	1.10	0.47	1.28	0.05	1.06	0.67	0.98	0.88	1.32	0.03

Note: VR denotes violation ratio, Kupiec – Kupiec test p-values

Exchange All share index OSEAX returns were best forecasted by VaR models, which filtered returns with normal GARCH(1,1) VaR model and using EVT approach under normal distribution assumption. Not surprisingly, independence was rejected for all unconditional VaR models. According to Christoffersen test results, all conditional VaR and EVT VaR models produced reasonable forecasts.

All previous models predicted return at 5 percent significance level; however, Bank of International Settlements in its Basel II Accord (BIS, 2006) suggested that banks and other financial institutions should use 99 percent confidence level, or 1 percent level of significance. Thus we also compare model forecasting performance before financial crisis under 1 percent level of significance. It is immediately seen in Table 11 that results under higher level of confidence are different for all stock indexes. 1 percent level of significance means that around 11 violations are expected for a backtesting window. This is 5 times less expected violations than at 5 percent level of significance, suggesting that independence of returns should be much lower. Actually, this is the main reason, why many models fail independence test at 5 percent significance level as presented in Table 10. Much more violations are allowed and probability of violation clustering is much higher under higher level of significance. In general, models which have a smaller probability of Type I error are more preferred, because number of expected violations is lower.

Table 14: Independence test and Christoffersen test p-values for period 2008-2011 at 5% significance level

	OMXV		OMXR		OMXT		OMXH		OMXS		OMXC		OSEAX	
Model	ind	Chr	ind	Chr	ind	Chr	ind	Chr	ind	Chr	ind	Chr	ind	Chr
1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
3	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
4	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
5	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.00	0.00	0.00	0.00	0.00
6	0.00	0.00	0.00	0.01	0.00	0.00	0.24	0.03	0.02	0.01	0.11	0.28	0.21	0.01
7	0.00	0.00	0.01	0.00	0.00	0.00	0.39	0.00	0.98	0.00	0.01	0.00	0.03	0.00
8	0.00	0.00	0.01	0.00	0.00	0.00	0.40	0.00	0.56	0.00	0.00	0.00	0.08	0.00
9	0.00	0.00	0.01	0.01	0.01	0.03	0.30	0.02	0.52	0.20	0.27	0.52	0.52	0.01
10	0.00	0.00	0.29	0.00	0.00	0.00	0.73	0.00	0.71	0.00	0.11	0.00	0.36	0.00
11	0.00	0.00	0.29	0.00	0.00	0.00	0.91	0.00	0.94	0.00	0.04	0.00	0.45	0.00
12	0.01	0.01	0.16	0.20	0.00	0.01	0.65	0.02	0.63	0.01	0.80	0.43	0.60	0.01
13	0.02	0.00	0.69	0.00	0.00	0.00	0.84	0.00	0.53	0.00	0.70	0.00	0.77	0.00
14	0.05	0.00	0.69	0.00	0.00	0.00	0.97	0.00	0.90	0.00	0.68	0.00	0.30	0.00
15	0.00	0.00	0.00	0.00	0.00	0.01	0.18	0.05	0.55	0.65	0.52	0.69	0.85	0.08
16	0.06	0.09	0.02	0.01	0.00	0.01	0.16	0.06	0.41	0.69	0.27	0.52	0.85	0.08
17	0.06	0.09	0.02	0.01	0.00	0.01	0.16	0.06	0.46	0.69	0.11	0.28	0.85	0.08

Note: ind denotes independence test p-values, Chr – Christoffersen test p-values

Table 12 show that all models for OMX Vilnius stock index underestimated risk, because there were more violations than expected (VR > 1). One of the most advanced models - skewed Student's t APARCH(1,1) - managed to keep the title of best model at higher level of confidence. Other GARCH and APARCH models accounting for fat tails also performed reasonably well. All models did not reject independence hypothesis and Christoffersen tests ranked best models the same models as Kupiec test. Three unconditional VaR models were best performers for OMX Riga stock index according to Kupiec test. However, most of them failed violation independence test. Due to structural break, even at 1 percent level of significance, many conditional volatility models failed joint Christoffersen test, only EVT VaR model, accounting for these extreme returns managed to produce good forecasts. Most of the models for OMX Tallinn passed Kupiec test, but best performer was APARCH(1,1) model under Student's t and skewed Student's t distribution. However, independence and Christoffersen test results point to four other models. Ordinary GARCH(1,1) and EVT VaR models under Student's t and skewed Student's t distributions produced best forecasts for Estonia's stock index. OMX Helsinki stock index was quite different compared with others because was modeled during very volatile period. Therefore under high level of confidence of 99 percent, many advanced models passed Kupiec test. Six models slightly overestimated risk producing 0.99 p-values for Kupiec test of unconditional coverage. The

Table 15: Violation ratios and Kupiec test p-values for period 2008-2011 at 1% significance level

	OMXV		OMXR		OMXT		OMXH		OMXS		OMXC		OSEAX	
Model	VR	Kupiec	VR	Kupiec	VR	Kupiec	VR	Kupiec	VR	Kupiec	VR	Kupiec	VR	Kupiec
1	2.10	0.00	1.90	0.01	1.90	0.01	2.10	0.00	1.70	0.04	1.40	0.23	1.80	0.02
2	3.20	0.00	3.20	0.00	3.90	0.00	3.90	0.00	3.70	0.00	4.10	0.00	3.40	0.00
3	2.70	0.00	2.20	0.00	6.20	0.00	2.30	0.00	2.30	0.00	2.30	0.00	2.60	0.00
4	2.40	0.00	2.50	0.00	5.10	0.00	2.10	0.00	1.60	0.08	1.80	0.02	2.10	0.00
5	2.00	0.01	2.10	0.00	2.00	0.01	2.10	0.00	1.80	0.02	1.50	0.14	1.90	0.01
6	2.20	0.00	2.00	0.01	2.00	0.01	1.90	0.01	2.10	0.00	1.50	0.14	2.00	0.01
7	1.90	0.01	1.60	0.08	7.10	0.00	1.00	1.00	1.00	1.00	0.90	0.75	1.00	1.00
8	1.60	0.08	1.70	0.04	6.70	0.00	1.00	1.00	0.50	0.08	0.70	0.31	0.60	0.17
9	2.50	0.00	1.90	0.01	1.50	0.14	1.60	0.08	1.60	0.08	1.60	0.08	1.60	0.08
10	2.60	0.00	1.60	0.08	6.50	0.00	0.80	0.51	0.80	0.51	1.00	1.00	0.80	0.51
11	1.80	0.02	1.60	0.08	6.30	0.00	0.70	0.31	0.30	0.01	0.60	0.17	0.30	0.01
12	2.30	0.00	2.00	0.01	1.80	0.02	1.50	0.14	2.10	0.00	1.60	0.08	1.90	0.01
13	2.10	0.00	1.40	0.23	6.50	0.00	1.00	1.00	1.20	0.54	0.80	0.51	0.90	0.75
14	1.80	0.02	1.30	0.36	6.20	0.00	0.70	0.31	0.80	0.51	0.50	0.08	0.20	0.00
15	1.50	0.14	1.90	0.01	1.00	1.00	0.80	0.51	0.60	0.17	0.80	0.51	1.00	1.00
16	1.20	0.54	1.80	0.02	1.20	0.54	1.00	1.00	0.60	0.17	0.80	0.51	1.00	1.00
17	1.20	0.54	1.80	0.02	1.20	0.54	1.00	1.00	0.60	0.17	0.90	0.75	1.00	1.00

Note: VR denotes violation ratio, Kupiec – Kupiec test p-values

same models, according to Table 12, passed independence and Christoffersen tests. OMX Stockholm was quite hardly predictable at 1 percent level of significance and only historical returns contained enough information for future forecasting. In addition, this model was best according to Table 12, but also Student's t VaR as well as most advanced APARCH(1,1) VaR models was only slightly worse than Historical simulation model. Only extreme value models and most advanced APARCH model managed to perform best forecasts for OMX Copenhagen stock index. This is in line with extreme value theory, which states that under higher levels of confidence EVT models are performing better. However, extreme returns were clustered and EVT models show worse results for independence tests than GARCH(1,1) and APARCH(1,1) model under Student's t and skewed Student's t distribution. These models were also chosen as best by Christoffersen test results. OSEAX index returns where best predicted by EWMA and APARCH volatility models incorporated in VaR under skewed Student's t distribution according to results in Table 11 and Table 12. EVT VaR models also showed good results. From Table 11 it is clearly seen that EWMA, GARCH and APARCH (model 6, 9 and 12) under normal return distribution failed for almost all stock indexes. This proves that normality assumption of returns are highly discouraged because these stock indexes are fat tailed.

Table 16: Independence test and Christoffersen test p-values for period 2008-2011 at 1% significance level

	OMXV		OMXR		OMXT		OMXH		OMXS		OMXC		OSEAX	
Model	ind	Chr	ind	Chr	ind	Chr	ind	Chr	ind	Chr	ind	Chr	ind	Chr
1	0.01	0.00	0.00	0.00	0.00	0.00	0.46	0.01	0.46	0.10	0.19	0.20	0.00	0.00
2	0.00	0.00	0.00	0.00	0.00	0.00	0.27	0.00	0.19	0.00	0.00	0.00	0.00	0.00
3	0.00	0.00	0.00	0.00	0.00	0.00	0.11	0.00	0.53	0.00	0.00	0.00	0.00	0.00
4	0.00	0.00	0.00	0.00	0.00	0.00	0.46	0.01	0.48	0.17	0.00	0.00	0.00	0.00
5	0.00	0.00	0.00	0.00	0.00	0.00	0.46	0.01	0.43	0.05	0.22	0.16	0.00	0.00
6	0.00	0.00	0.01	0.00	0.06	0.00	0.37	0.03	0.07	0.00	0.50	0.27	0.37	0.01
7	0.00	0.00	0.02	0.02	0.00	0.00	0.65	0.90	0.65	0.90	0.69	0.87	0.65	0.90
8	0.00	0.00	0.03	0.01	0.00	0.00	0.65	0.90	0.82	0.21	0.75	0.57	0.79	0.38
9	0.02	0.00	0.37	0.03	0.50	0.27	0.25	0.11	0.48	0.17	0.48	0.17	0.47	0.17
10	0.03	0.00	0.47	0.17	0.00	0.00	0.72	0.75	0.72	0.75	0.65	0.90	0.72	0.75
11	0.04	0.01	0.47	0.17	0.00	0.00	0.75	0.57	0.89	0.03	0.79	0.38	0.89	0.03
12	0.11	0.00	0.37	0.01	0.42	0.05	0.50	0.27	0.34	0.01	0.25	0.11	0.39	0.03
13	0.07	0.00	0.53	0.40	0.00	0.00	0.65	0.90	0.59	0.71	0.72	0.75	0.69	0.87
14	0.04	0.01	0.56	0.56	0.00	0.00	0.75	0.57	0.72	0.75	0.82	0.21	0.93	0.01
15	0.02	0.02	0.37	0.03	0.65	0.90	0.72	0.75	0.79	0.38	0.72	0.75	0.65	0.90
16	0.59	0.71	0.33	0.05	0.59	0.71	0.65	0.90	0.79	0.38	0.72	0.75	0.65	0.90
17	0.59	0.71	0.33	0.05	0.59	0.71	0.65	0.90	0.79	0.38	0.69	0.87	0.65	0.90

Note: ind denotes independence test p-values, Chr – Christoffersen test p-values

4.3.2. High volatility period during and after financial crisis

In this part every model is estimated using 1000 days moving window and 1000 forecasts are obtained. For all stock indexes years 2008 and 2009 are most volatile, however, volatility in 2010 and 2011 is also higher compared to calm growth period before crisis. Initially, first estimations and forecasts of the volatile period are estimated using data from relatively less volatile first period. Therefore, forecasts are expected to underestimate real risk in the market. However, model performance starts fluctuating when high volatility enters estimation window. Overall backtesting results for the period of years 2008-2011 at 5 percent level of significance are presented in Table 13 and Table 14.

More models fails Kupiec test for second period compared with results from first period in Table 9. Conditional volatility EWMA, GARCH and APARCH models with normal return distribution as well as conditional EVT models shows best results according to Table 13. It is not surprising that almost all unconditional volatility models fail all three back tests trying to forecast during the crisis period, because they do not account for very high fluctuations is return volatility. Although there is different best model for each stock index, Table 14 shows that extreme value models are best for five

Table 17: Violation ratios and Kupiec test p-values for period 2004-2011 at 5% significance level

	OMXV		OMXR		OMXT		OMXH		OMXS		OMXC		OSEAX	
Model	VR	Kupiec	VR	Kupiec	VR	Kupiec	VR	Kupiec	VR	Kupiec	VR	Kupiec	VR	Kupiec
1	1.20	0.05	1.25	0.01	1.28	0.01	1.16	0.10	1.09	0.35	1.22	0.03	1.30	0.00
2	1.03	0.79	1.06	0.51	1.00	0.96	1.11	0.26	1.14	0.16	1.21	0.04	1.23	0.02
3	1.50	0.00	1.30	0.00	2.18	0.00	1.29	0.00	1.27	0.01	1.38	0.00	1.37	0.00
4	1.33	0.00	1.34	0.00	2.00	0.00	1.22	0.03	1.17	0.09	1.28	0.01	1.19	0.05
5	1.24	0.02	1.28	0.01	1.31	0.00	1.16	0.10	1.09	0.35	1.21	0.04	1.31	0.00
6	0.85	0.10	0.87	0.16	0.96	0.64	1.20	0.04	1.20	0.05	1.03	0.79	1.25	0.01
7	1.53	0.00	3.37	0.00	1.91	0.00	1.57	0.00	1.67	0.00	1.59	0.00	1.60	0.00
8	1.50	0.00	3.42	0.00	2.08	0.00	1.49	0.00	1.54	0.00	1.42	0.00	1.38	0.00
9	1.05	0.64	0.88	0.19	0.77	0.01	1.18	0.07	1.16	0.11	0.96	0.65	1.23	0.02
10	1.67	0.00	3.34	0.00	1.69	0.00	1.58	0.00	1.59	0.00	1.44	0.00	1.50	0.00
11	1.57	0.00	3.29	0.00	1.74	0.00	1.42	0.00	1.37	0.00	1.28	0.01	1.30	0.00
12	1.06	0.57	0.85	0.10	0.80	0.03	1.19	0.06	1.26	0.01	1.07	0.50	1.29	0.00
13	1.67	0.00	3.38	0.00	1.72	0.00	1.58	0.00	1.65	0.00	1.57	0.00	1.63	0.00
14	1.56	0.00	3.36	0.00	1.76	0.00	1.44	0.00	1.46	0.00	1.34	0.00	1.28	0.01
15	1.19	0.06	1.16	0.10	1.02	0.80	1.20	0.04	1.10	0.30	0.97	0.73	1.16	0.10
16	1.14	0.17	1.18	0.07	1.09	0.34	1.23	0.02	1.09	0.35	0.97	0.73	1.15	0.12
17	1.14	0.17	1.18	0.07	1.09	0.34	1.24	0.02	1.10	0.30	1.00	0.97	1.15	0.12

Note: VR denotes violation ratio, Kupiec – Kupiec test p-values

indexes, while GARCH and APARCH VaR models are also providing good forecasts for Nordic countries' stock indexes. According to Christoffersen tests results in Table 14, almost none of the models are able to withstand violation ratio clustering at 5 percent significance level for Baltic countries' stock indexes. Table 15 and Table 16 show model forecasting performance for period 2008-2011 at reduced Type I error, i.e., increased level of confidence to 99 percent. Compared with Table 13 and Table 14, results are much better at higher level of confidence. Best models for five indexes (excluding only OMX Vilnius and OMX Riga) are able exactly to forecast returns and to provide number of violations as expected. Returns of OMX Tallinn stock index, OMX Helsinki stock index and OSEAX stock index is exactly forecasted by EVT VaR model with GARCH(1,1) filter according to Kupiec test results. EWMA, GARCH(1,1) and APARCH(1,1) VaR models with Student's t distribution are best forecasters for Nordic stock indexes. Due to reduced level of significance, probability of violation clustering is lowered and this significantly improves forecasting performance. The same models as pointed by Kupiec test results in Table 15, are chosen as best forecasting models by Christoffersen test in Table 16. Therefore, the answer to a research question, which ask about model forecasting performance changes when level of significance increases, is that model forecasting performance increases a lot at higher level of significance, because important property of independent violations is rejected less often.

Table 18: Independence test and Christoffersen test p-values for period 2004-2011 at 5% significance level

	OMXV		OMXR		OMXT		OMXH		OMXS		OMXC		OSEAX	
Model	ind	Chr	ind	Chr	ind	Chr	ind	Chr	ind	Chr	ind	Chr	ind	Chr
1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
3	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
4	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
5	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00
6	0.00	0.00	0.00	0.00	0.00	0.00	0.04	0.02	0.01	0.00	0.00	0.00	0.07	0.01
7	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.05	0.00	0.00	0.00	0.00	0.00
8	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.07	0.00	0.00	0.00	0.02	0.00
9	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.11	0.08	0.02	0.06	0.61	0.06
10	0.00	0.00	0.00	0.00	0.00	0.00	0.11	0.00	0.21	0.00	0.00	0.00	0.09	0.00
11	0.00	0.00	0.00	0.00	0.00	0.00	0.12	0.00	0.24	0.00	0.00	0.00	0.60	0.01
12	0.00	0.00	0.00	0.00	0.00	0.00	0.15	0.06	0.28	0.02	0.58	0.68	0.83	0.02
13	0.00	0.00	0.00	0.00	0.00	0.00	0.06	0.00	0.70	0.00	0.11	0.00	0.85	0.00
14	0.00	0.00	0.00	0.00	0.00	0.00	0.08	0.00	0.47	0.00	0.71	0.00	0.79	0.02
15	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.24	0.30	0.01	0.03	0.39	0.18
16	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.02	0.04	0.01	0.03	0.36	0.20
17	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.02	0.05	0.00	0.01	0.36	0.20

Note: ind denotes independence test p-values, Chr – Christoffersen test p-values

4.3.3. Both low and high volatility periods

In this part every model is estimated using 1000 days moving window and 1988-2030 forecasts are obtained. Therefore it is a combination of results presented in previous parts, in order to test how models perform using data from both relatively calm and volatile periods and how it affects overall forecasting performance. Higher forecasting window ensures that backtesting procedure is more accurate because more violations are expected. Table 17 presents results of model backtesting for the whole period of approximately year 2004-2011. This period includes high volatility cluster during the years 2008-2009. Firstly, extremely high negative and positive returns in year 2008-2009 are forecasted using rather calm period of years 2004-2007. Later, these extreme returns of year 2008-2009 fall into estimation window for forecasting of years 2009-2011 when volatility was much lower. These conditions puts extraordinary challenge for all our models tested. The results in Table 17 show the performance under these conditions. It is clearly seen that fewer models passed Kupiec test of unconditional coverage compared to the results in Table 9 and Table 12, where normal market conditions were separated from volatile markets. Accidentally normal VaR model provided best forecasts for all three Baltic stock index returns and OMX Helsinki. For almost all stock indexes extreme value theory models were able to capture extreme returns during financial meltdown. Overall,

Table 19: Violation ratios and Kupiec test p-values for period 2004-2011 at 1% significance level

	OMXV		OMXR		OMXT		OMXH		OMXS		OMXC		OSEAX	
Model	VR	Kupiec	VR	Kupiec	VR	Kupiec	VR	Kupiec	VR	Kupiec	VR	Kupiec	VR	Kupiec
1	1.76	0.00	1.49	0.04	1.53	0.03	1.24	0.30	1.34	0.15	1.10	0.67	1.59	0.01
2	2.62	0.00	2.19	0.00	2.66	0.00	2.58	0.00	2.98	0.00	3.24	0.00	2.83	0.00
3	2.11	0.00	1.34	0.14	4.88	0.00	1.39	0.10	1.84	0.00	1.89	0.00	2.14	0.00
4	1.91	0.00	1.79	0.00	4.24	0.00	1.19	0.40	1.39	0.10	1.30	0.20	1.74	0.00
5	1.81	0.00	1.59	0.01	1.63	0.01	1.29	0.21	1.49	0.04	1.30	0.20	1.59	0.01
6	2.11	0.00	1.79	0.00	1.82	0.00	1.84	0.00	2.38	0.00	1.99	0.00	1.99	0.00
7	1.76	0.00	9.55	0.00	4.04	0.00	1.09	0.68	1.49	0.04	1.54	0.02	1.19	0.40
8	1.61	0.01	9.85	0.00	3.99	0.00	0.99	0.97	1.09	0.68	1.15	0.52	0.79	0.34
9	2.16	0.00	1.34	0.14	1.33	0.15	1.79	0.00	2.03	0.00	1.84	0.00	1.69	0.00
10	1.91	0.00	9.00	0.00	3.60	0.00	0.89	0.62	1.29	0.21	1.10	0.67	0.94	0.80
11	1.51	0.03	9.20	0.00	3.50	0.00	0.79	0.34	0.84	0.47	0.70	0.15	0.55	0.03
12	2.01	0.00	1.29	0.21	1.63	0.01	1.59	0.01	2.23	0.00	1.69	0.00	1.84	0.00
13	1.66	0.01	8.86	0.00	3.69	0.00	1.14	0.53	1.39	0.10	1.00	0.99	1.14	0.53
14	1.46	0.05	9.05	0.00	3.55	0.00	0.84	0.47	0.99	0.97	0.80	0.34	0.60	0.05
15	1.61	0.01	1.39	0.09	1.03	0.88	0.89	0.62	1.14	0.53	0.95	0.81	1.09	0.68
16	1.51	0.03	1.24	0.29	1.18	0.42	0.99	0.97	1.14	0.53	0.95	0.81	1.04	0.85
17	1.41	0.08	1.29	0.21	1.18	0.42	0.99	0.97	1.14	0.53	1.00	0.99	1.04	0.85

Note: VR denotes violation ratio, Kupiec – Kupiec test p-values

forecasting results presented in Table 17 and Table 18 are quite poor for the whole period at 5 percent level of significance. However, extreme value theory tells that these models perform best when level of confidence is high. Therefore we turn to Table 19, which shows results of backtesting using violation ratios and Kupiec test are presented at 1 percent level of significance.

Table 19 shows that much more models passed Kupiec test for Nordic stock indexes than for Baltic stock indexes. This can be explained by the characteristics of return distribution of these stock indexes. Baltic stock indexes have much higher excess kurtosis than Nordic countries, indicating fatter tails, or simply, much more highly negative returns.

For OMX Vilnius stock index, only two most advanced VaR models - EVT and APARCH(1,1) - with skewed Student's t distribution passes Kupiec test. However violation ratio is still quite high and shows large underestimation of real risk in the market. According to Table 20, only EVT VaR with skewed Student's t return distribution assumption is able to produce significant forecasts. Combining results from Table 12, where best models was from ARCH class, and Table 16, where best models where EVT VaR, it is seen that forecasting power of EVT VaR models is reduced than adding calm period to the volatile one. This can be explained that EVT models are not designed for calm periods,

Table 20: Independence test and Christoffersen test p-values for period 2004-2011 at 1% significance level

	OMXV		OMXR		OMXT		OMXH		OMXS		OMXC		OSEAX	
Model	ind	Chr	ind	Chr	ind	Chr	ind	Chr	ind	Chr	ind	Chr	ind	Chr
1	0.02	0.00	0.00	0.00	0.00	0.00	0.32	0.35	0.39	0.24	0.24	0.46	0.00	0.00
2	0.00	0.00	0.00	0.00	0.00	0.00	0.05	0.00	0.14	0.00	0.00	0.00	0.00	0.00
3	0.00	0.00	0.00	0.00	0.00	0.00	0.06	0.04	0.18	0.00	0.00	0.00	0.00	0.00
4	0.01	0.00	0.00	0.00	0.00	0.00	0.29	0.40	0.37	0.17	0.00	0.00	0.00	0.00
5	0.00	0.00	0.00	0.00	0.00	0.00	0.35	0.29	0.34	0.08	0.35	0.29	0.00	0.00
6	0.00	0.00	0.00	0.00	0.03	0.00	0.71	0.00	0.03	0.00	0.25	0.00	0.82	0.00
7	0.00	0.00	0.00	0.00	0.00	0.00	0.49	0.72	0.34	0.08	0.09	0.02	0.45	0.53
8	0.00	0.00	0.00	0.00	0.00	0.00	0.53	0.82	0.49	0.72	0.27	0.44	0.61	0.56
9	0.08	0.00	0.38	0.23	0.37	0.24	0.16	0.00	0.86	0.00	0.18	0.00	0.60	0.02
10	0.04	0.00	0.00	0.00	0.00	0.00	0.57	0.75	0.41	0.33	0.48	0.72	0.55	0.81
11	0.08	0.02	0.00	0.00	0.00	0.00	0.61	0.55	0.59	0.66	0.66	0.32	0.73	0.08
12	0.25	0.00	0.41	0.32	0.56	0.03	0.53	0.04	0.15	0.00	0.02	0.00	0.71	0.00
13	0.12	0.01	0.00	0.00	0.00	0.00	0.47	0.63	0.37	0.17	0.53	0.82	0.26	0.44
14	0.07	0.03	0.00	0.00	0.00	0.00	0.59	0.67	0.53	0.82	0.61	0.56	0.70	0.13
15	0.11	0.01	0.41	0.18	0.21	0.46	0.57	0.75	0.47	0.63	0.17	0.38	0.49	0.72
16	0.47	0.08	0.32	0.35	0.45	0.54	0.53	0.82	0.47	0.63	0.17	0.38	0.51	0.79
17	0.41	0.16	0.35	0.29	0.45	0.54	0.53	0.82	0.47	0.63	0.19	0.43	0.51	0.79

Note: ind denotes independence test p-values, Chr – Christoffersen test p-values

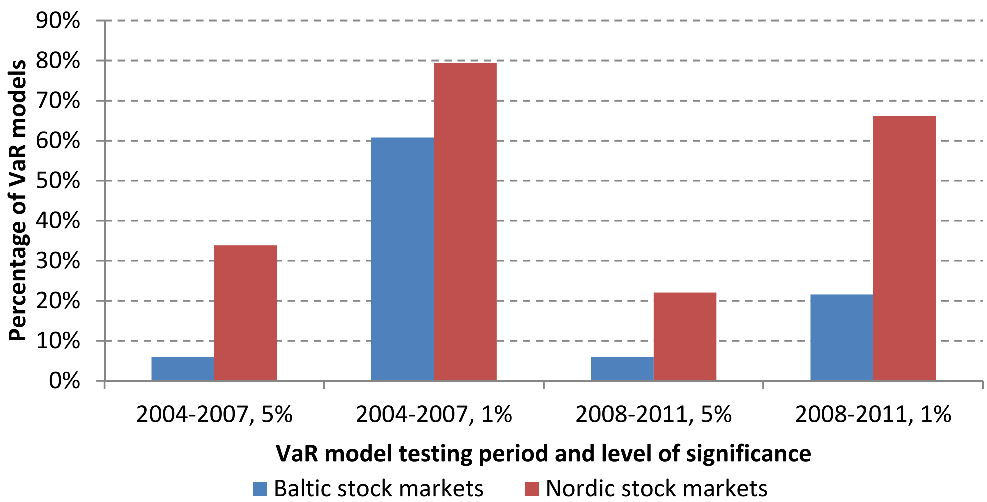
because number of extreme returns is very low. Similar results are obtained for OMX Riga and OMX Tallinn stock indexes.

Different situation is seen in Nordic stock markets, where financial crisis also affected stock market, but caused much lower volatility. Thus we see that more models produced reasonable forecasts for the whole period. In Table 19 we see that many models with conditional volatility and fat tailed distribution assumption even overestimated risk. Especially skewed Student's t GARCH(1,1) and APARCH(1,1) VaR models produced too high negative returns, causing violation ratio to be less than 1. Due to this reason, these models even failed Kupiec test, because violation ratio was too low, or risk was overestimated too much. In spite of a few exceptions, EVT models with skewed Student's t GARCH filter provided best forecasts, proving the ability of these models to capture extreme returns. Another remark about results in Table 19 is that unconditional, EWMA, GARCH(1,1) and APARCH(1,1) models under normal return distribution assumption did not pass Kupiec test for all Nordic stock indexes indicating wrong assumption of return distribution. Only EVT VaR model with normal GARCH(1,1) filter was able to produce reasonable forecasts, but still not the best one.

Again, forecasting power of EVT VaR models were reduced due to inclusion of calm period into estimation and forecasting windows. Thus advanced EWMA, GARCH and APARCH models performed better for most Nordic stock indexes.

Model overall forecasting performance results according to Christoffersen test are presented in Figure 4. Number of VaR models passing Christoffersen test at 5% level of significance is divided by total number of models (17). Models are grouped by three parameters – stock market, volatility period and level of significance. According to Figure 4, fourth hypothesis H4 is not rejected, because it is clear that average of first two columns is much lower than average of third and fourth columns. The same situation is with last four columns. This means that on average, more models pass Christoffersen test when VaR models are estimated at 1% level of significance (or 99th quantile) than at 5% level of significance (or 95th quantile) in spite of the time period. This is also proved statistically, because p-value of one-sided *t*-test that percentage of VaR models at 1 percent level of significance is higher than percentage of VaR models measured at 5 percent level of significance is 0.0008 and 0.0065 for periods 2004-2007 and 2008-2011 respectively. Fifth hypothesis H5 is also not rejected according to Figure 4, which shows that blue column for Baltic stock markets is always lower than red column on the right side of blue column. It is also statistically significant at 5 percent level according to one-sided t-test, because p-values for all four groups of VaR models grouped by volatility period and level of significance are lower than 0.05 (0.0004, 0.0352, 0.0469 and 0.0001 respectively).

Figure 4: Percentage of VaR models passing Christoffersen test at 5% level of significance during low and high volatility periods, for different significance levels of VaR models and both Nordic and Baltic stock markets



Overall, combining results from Table 12, Table 16 and Table 20, best VaR models measured at 99th percentile according to Christoffersen test are presented in Table 21. Although models differ across stock markets and volatility periods, conditional volatility VaR models with Student's t or skewed Student's t return distribution and conditional EVT models dominate in Table 21 and this means that these models provide best forecasts. Difference of significance (p-value) of back tests is rather small. Therefore, answer to first research question is obvious. Financial risk managers are advised to employ these models and compare them to get best picture of the market risks of their portfolios.

Table 21: Best models according to Christoffersen test at 1% level of significance

Index	Low volatility period (2004-2007)	High volatility period (2008-2011)	Both periods (2004-2011)
OMXV	Skewed Student's t APARCH(1,1) VaR, Student's t APARCH(1,1) VaR, Skewed Student's t GARCH(1,1) VaR	EVT Skewed Student's t GARCH(1,1) VaR, EVT Student's t GARCH(1,1) VaR	EVT Skewed Student's t GARCH(1,1) VaR, EVT Student's t GARCH(1,1) VaR
OMXR	Normal GARCH(1,1) VaR, EVT Skewed Student's t GARCH(1,1) VaR, EVT Student's t GARCH(1,1) VaR	Skewed Student's t APARCH(1,1) VaR, Student's t APARCH(1,1) VaR, Skewed Student's t GARCH(1,1) VaR	EVT Student's t GARCH(1,1) VaR, Normal APARCH(1,1) VaR, EVT Skewed Student's t GARCH(1,1) VaR
OMXT	EVT Skewed Student's t GARCH(1,1) VaR, EVT Student's t GARCH(1,1) VaR, Skewed Student's t GARCH(1,1) VaR	EVT Normal GARCH(1,1) VaR, EVT Skewed Student's t GARCH(1,1) VaR, EVT Student's t GARCH(1,1) VaR	EVT Skewed Student's t GARCH(1,1) VaR, EVT Student's t GARCH(1,1) VaR, EVT normal GARCH(1,1) VaR
OMXH	EVT Skewed Student's t GARCH(1,1) VaR, EVT Student's t GARCH(1,1) VaR, EVT normal GARCH(1,1) VaR, Skewed Student's t APARCH(1,1) VaR, Student's t GARCH(1,1) VaR, Skewed Student's t EWMA VaR	EVT Skewed Student's t GARCH(1,1) VaR, EVT Student's t GARCH(1,1) VaR, Student's t APARCH(1,1) VaR, Skewed Student's t EWMA VaR, Student's t EWMA VaR	EVT Skewed Student's t GARCH(1,1) VaR, EVT Student's t GARCH(1,1) VaR, Skewed Student's t EWMA VaR
OMXS	Historical simulation, Skewed Student's t APARCH(1,1) VaR, Skewed Student's t VaR	Student's t EWMA VaR, Skewed Student's t APARCH(1,1) VaR, Student's t GARCH(1,1) VaR	Skewed Student's t APARCH(1,1) VaR, Skewed Student's t EWMA VaR, Skewed Student's t GARCH(1,1) VaR
OMXC	Skewed Student's t APARCH(1,1) VaR, EVT VaR, Skewed Student's t GARCH(1,1) VaR	Student's t GARCH(1,1) VaR, EVT Skewed Student's t GARCH(1,1) VaR, EVT Student's t GARCH(1,1) VaR	Student's t APARCH(1,1) VaR, Student's t GARCH(1,1) VaR, Skewed Student's t APARCH(1,1) VaR
OSEAX	Skewed Student's t APARCH(1,1) VaR, Skewed Student's t EWMA VaR, EVT Skewed Student's t GARCH(1,1) VaR	EVT Skewed Student's t GARCH(1,1) VaR, EVT Student's t GARCH(1,1) VaR, EVT normal GARCH(1,1) VaR, Student's t EWMA VaR	Student's t GARCH(1,1) VaR, EVT Skewed Student's t GARCH(1,1) VaR, EVT Student's t GARCH(1,1) VaR

Last hypothesis H6 is rejected, because from Table 9 to Table 21, usually more than one VaR model has highest p-value according to Christoffersen test, and if there is one model with highest p-value, second best model is usually very close to the best one. Therefore, rejection of last hypothesis is consistent with many researches, which also were not able to find one best VaR model for all stock indexes, during low and high volatility periods and various levels of significance.

4.3.4. Graphical analysis of model forecasting behavior

Graphical analysis helps better understand empirical results presented in previous sections. Due to high number of VaR models tested, We present results by groups of models with similar characteristics. First we discuss the performance of the unconditional models, then the EWMA-, GARCH– and APARCH–based, and, finally, EVT models.

Figure 5 presents forecasting behavior of unconditional VaR models at 1 percent significance level for OMXV stock index. The graphs for all other indexes exhibit similar properties. It is clearly seen in the Figure 5 that all unconditional VaR models does not account for volatility clustering, therefore forecast adjustment to extreme negative returns is very slow and very persistent. When a period of low return volatility comes, models also very slowly decrease their forecasts to lower returns.

Therefore we make a conclusion, that all results of violation ratios and Kupiec test are pure coincidence for these models, because they are far from providing reasonable forecasts.

Completely different result is seen in Figure 6, which plots EWMA VaR forecasts of OMX Copenhagen index returns under three distribution assumptions. Student's t distribution account for fat

Figure 5: Unconditional VaR model forecasting performance of OMXV index returns at 1% significance level

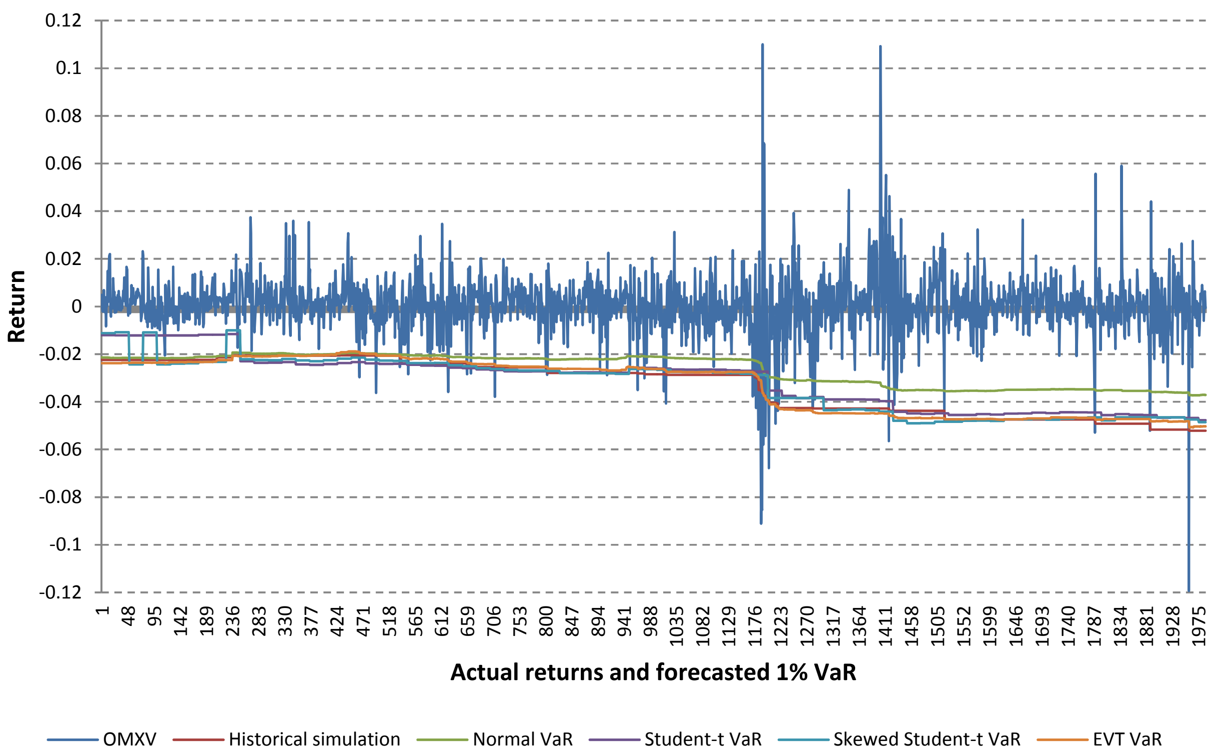
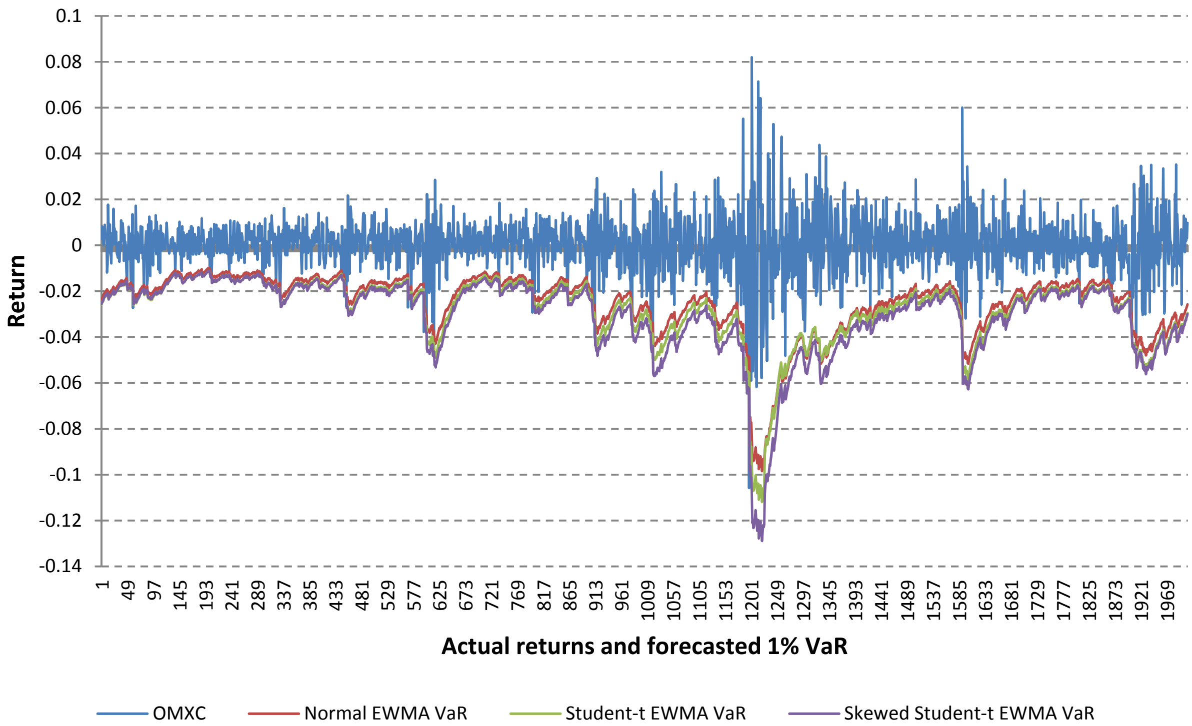


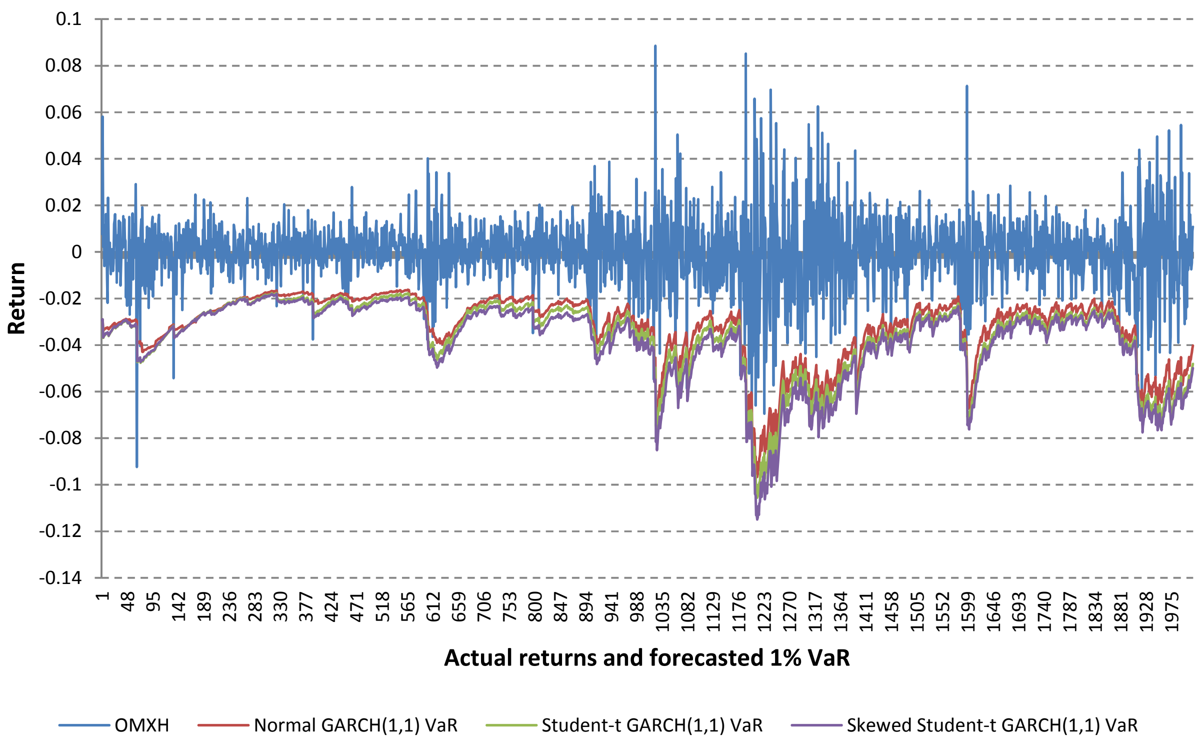
Figure 6: EWMA VaR model forecasting performance of OMXC index returns at 1% significance level



tails, therefore, when high negative return occurs, Student's t EWMA VaR model accounts for that and produces more negative forecast compared to normal EWMA VaR forecast. Even higher losses are forecasted by skewed Student's t EWMA VaR model, because this model not only accounts for fat tails, but also for asymmetric return distribution, which exhibit more negative returns. According to Table 19, only last EWMA VaR model passed Kupiec test, but still is not the best compared with other models. Although EWMA VaR models passed Kupiec test during the crisis period, as presented in Table 15, EWMA VaR models with Student's t and skewed Student's t distributions forecasted higher losses than actual negative returns, therefore, overestimating risk. The limitation of these models is that there is no need to estimate parameters of EWMA model, because they are set in advance.

Figure 7 presents GARCH VaR forecasting performance for OMX Helsinki stock index at 1 percent level of significance. Here the results are similar to the Figure 6, but the main difference is that model has less parameter restrictions than in EWMA model. Combining results from Table 19, we see that normal GARCH(1,1) VaR model produces smallest forecasted losses, which is not enough to account for actual market movement and therefore risk is underestimated. This model also did not pass Kupiec test. In contrast, GARCH(1,1) VaR models with Student's t and skewed Student's t distributions are too risky, because number of forecasted negative returns exceeding actual losses is

Figure 7: GARCH VaR model forecasting performance of OMXH index returns at 1% significance level



higher than expected. However, risk overestimation is in a reasonable limits and both models pass Kupiec test. Nevertheless, these models are still not best forecasters for OMX Helsinki stock index.

Figure 8 present forecasting results of most advanced model from class of autoregressive heteroskedastic models, because APARCH model is least restricted compared to GARCH and EWMA models. Again, we see that changing return distribution assumption from normal to Student's t and from Student's t to skewed Student's t, additional features of stock market data is accounted, which lead to forecasts of higher losses. Normal APARCH model is not accounting for fat tails, thus, according to Table 19, highly underestimates risk and produces too many forecasting violations and fails Kupiec test. Changing distribution assumption to Student's t, forecasting performance improves, but violation ratio is still high. Finally, skewed Student's t APARCH(1,1) model is best model for OMX Stockholm return forecasts, because only slightly overestimates risk.

Figure 9 exhibits forecasting results of special VaR model, which focuses not on the whole distribution, but on the tail of return distribution. Oslo stock exchange All share index returns are modeled with GARCH(1,1) volatility model with three distribution assumptions. Then residuals are extracted from the model, and tail of residual distribution is modeled with generalized Pareto distribution. This allows capturing extreme returns in the tails. Finally modeled loss is updated with

Figure 8: APARCH VaR model forecasting performance of OMXS index returns at 1% significance level

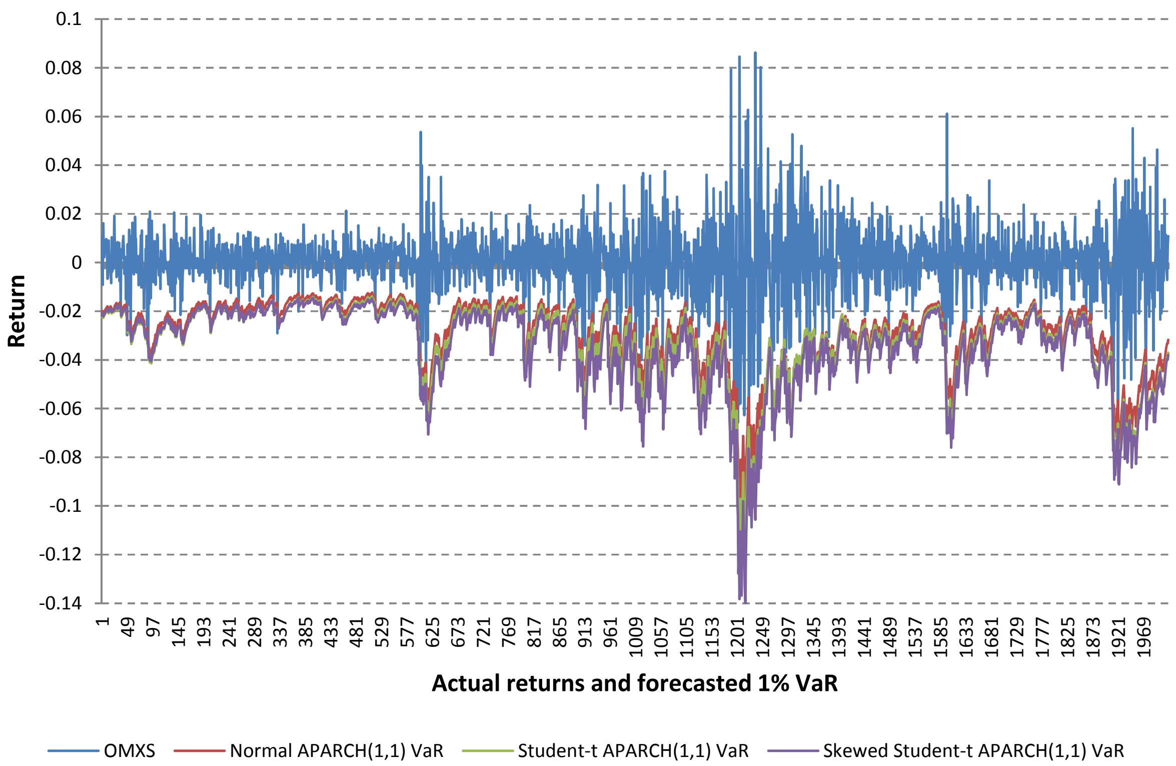
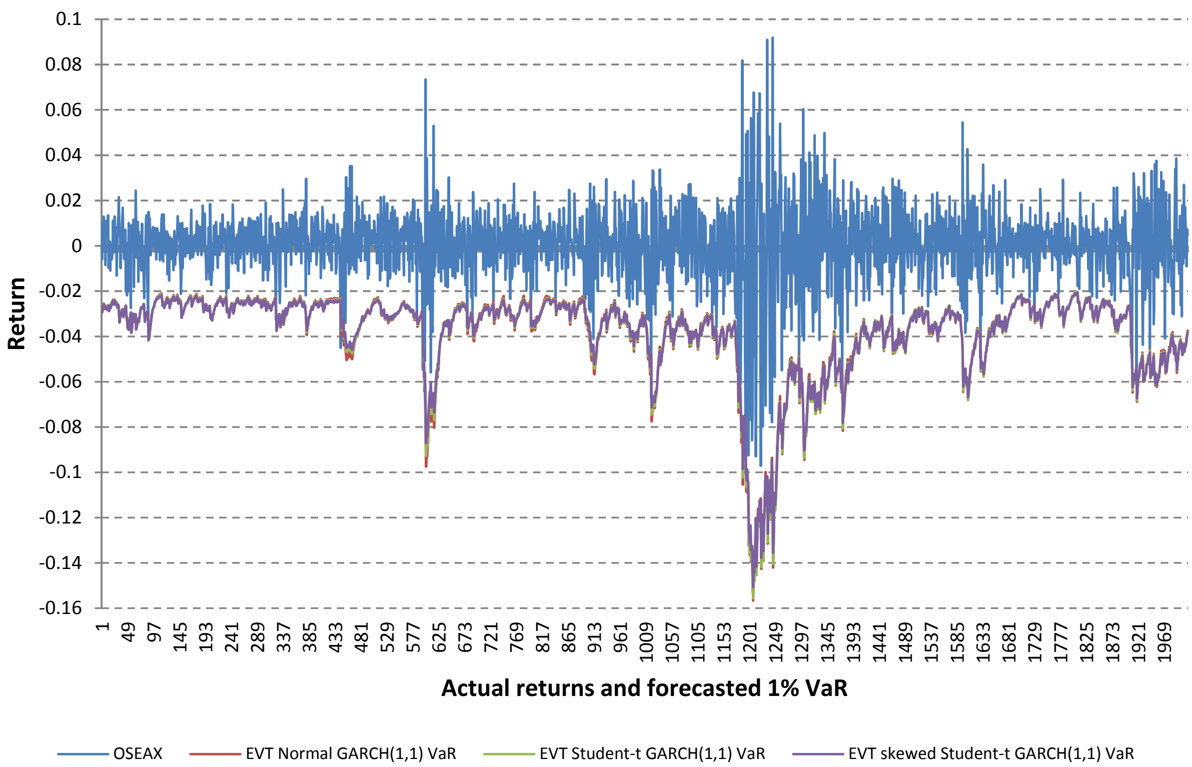


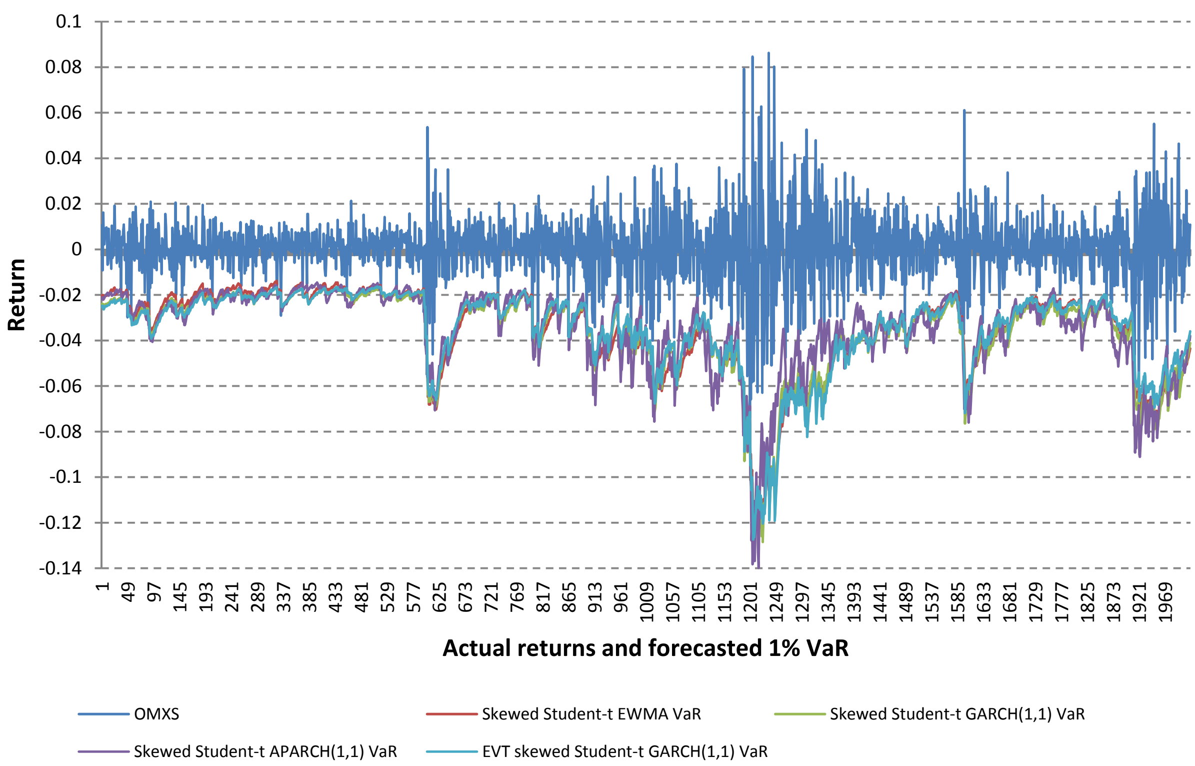
Figure 9: Conditional EVT VaR model forecasting performance of OSEAX index returns at 1% significance level



forecast of volatility from GARCH(1,1) model. Figure 9 shows that change of distribution does not affect very much the forecasts. However, looking at Table 19, we see that change from normal to Student's t distribution improves forecasts and provides best results among other models. Allowing for skewness in return distribution does not affect result.

It is also worth to compare different models under the same return distribution. We chose skewed Student's t, because it accounts for most market imperfections, which are usually incorporated in best VaR forecasting models. Figure 10 compares forecasting performance of EWMA, GARCH, APARCH and conditional EVT VaR models under skewed Student's t distribution for OMX Stockholm index returns. We see that there is no clear pattern of model forecasting behavior and in different time periods different models are producing highest loss forecasts. This is because every model has different parameter estimates due to different restrictions and this has strong impact on further model behavior. Therefore only backtesting results allow finding best model. Although these four models provides reasonable forecasts according to Kupiec test results in Table 19, best VaR model is skewed Student's t APARCH(1,1). EVT model is underestimating risk too much in this case.

Figure 10: EWMA, GARCH, APARCH and EVT VaR models under skewed Student's t distribution (OMXS index, 1% sign. level)



However, for all other Nordic indexes this model is the best forecaster. Due to violation clustering, as presented in Table 20, EVT VaR model losses forecasting power for more indexes. When relatively calm period is included the estimation and forecasting windows, this also causes a reduction in the significance of back tests for these models. This lead to the conclusion that most advanced models, which account for well-known market properties, should be used to forecast expected one day losses in the stock market. EVT models are better suited for extremely volatile markets conditions with structural breaks, while other conditional volatility models, such as GARCH and APARCH with fat tail distributions are good for stock market return prediction, including even still developing and small stock markets.

5. Discussion

This part of the thesis gives main findings of empirical research and compares it with existing researches overviewed in review of empirical research part. Furthermore, it presents implications and limitations of the study for current theory and practical applications. Finally, implications for further research and professional practice are discussed.

Descriptive statistics for return series of all stock indexes reveal that return distributions have fat tails and are skewed. High positive excess kurtosis and non-zero skewness as well as two normality tests show that data is not normally distributed. Returns series of stock indexes from Baltic States have much higher excess kurtosis than in Nordic stock exchanges indicating higher risk in those markets. Another measure giving the opportunity to model forecasts is high autocorrelation of squared returns, which is a good proxy for volatility, and allows predicting it. Ljung-Box test results support this finding. Graphical analysis of return plot and return distribution shows that volatility clustering is present in the return series. Quantile-quantile graph again proves that returns are non-normally distributed. First hypothesis is partially rejected, because all this information shows that Nordic stock index return distributions are also non-normally distributed and skewed, however, at the lower scale than Baltic stock markets.

Analysis of parameters of volatility models fitted to the whole data sample of each stock index confirms previous findings and gives additional information. Highly statistically significant parameter β estimates for both GARCH and APARCH models show that autoregressive affect is very strong in return series for all stock indexes. Degree of freedom estimates for Student's t distribution are from four to ten, indicating that return distribution is non-normal and much better fitted by Student's t distribution. Skewed Student's t distribution model has a skewness parameter and its estimate for return series of almost all stock indexes shows that distributions are negatively skewed, which means that high negative returns occur more often than high positive returns. Less restricted APARCH(1,1) volatility model is able to account for leverage effects and reveal that this effect is present in almost all return series and its parameter is highly statistically significant. Only returns of Tallinn stock index for the period 2000-2011 do not contain leverage effect according to APARCH(1,1) model parameter estimates. Another finding is that APARCH(1,1) model should be preferred against GARCH(1,1) model, because power term in APARCH(1,1) model is lower than two (power term is set to two in GARCH(1,1)) and this allows a better fit to the return data. Comparison of all these six volatility models using log likelihood ratio tests shows that changing return distribution assumption from normal to Student's t significantly improves model and its ability to fit the data. Going from Student's t distribution assumption to skewed Student's t distribution gives moderate improvement. Therefore, second hypothesis H2 is not rejected. However, third hypothesis H3 is partially rejected, because changing GARCH(1,1) model to APARCH(1,1) model using normal return distribution assumption gives negative improvement for several stock indexes. However, when this change is done using Student's t or skewed Student's t distribution assumption, APARCH(1,1) model is better fitting data compared to GARCH(1,1) model according to likelihood ratios and their statistical significance.

Extreme value theory models are also fitted to excess return data in order to model extreme returns in the left tail of return distribution of each stock index. Estimation results and graphical analysis show that these models are able to fit extreme returns; therefore it is worth to incorporate them in VaR model in order to account for extreme returns during financial crisis and other extreme stock market events.

Forecasting performance of 17 VaR models is evaluated for seven Nordic and Baltic stock indexes for low volatility and high volatility periods using 95% and 99% level of confidence. 1000 daily returns were used to estimate each model and one-day VaR forecasts were computed for the rest of return sample. Analysis of violation ratios and backtesting procedure using unconditional and conditional likelihood ratio tests allowed comparing and choosing best forecasting models. Answering to the research questions one of the main findings is that all models show better forecasting performance at 1 percent level of significance than at 5 percent level of significance. Therefore this verifies fourth hypothesis of the thesis. Results are presented in Figure 4. Another finding is that many VaR models, which forecast index returns during the low volatility period of year 2004-2007 pass Christoffersen test of conditional coverage at 1 percent level of significance. However, more models pass this test for Nordic stock indexes than for Baltic stock indexes giving first sign of risk differences in Nordic and Baltic stock markets. For example, many unconditional VaR models and simple conditional volatility VaR models are not able to pass Christoffersen test indicating model inadequacy in forecasting index returns in Baltic stock markets. Worse situation is when models try to forecast high volatility period of year 2008-2011, which include recent financial crisis. Only most advanced VaR models such as extreme value theory VaR models and APARCH(1,1) model with skewed Student's t distribution assumption are able to provide reasonable forecasts for Baltic stock markets. According to Figure 4, better situation is in Nordic stock markets, where more VaR models were able to provide statistically significant forecasts. However, best forecasts for many indexes were calculated by those EVT, GARCH and APARCH VaR models, which accounted for non-normal return distribution, i.e., assumed that returns are distributed according to Student's t or skewed Student's t distribution. The results are that even the most advanced models were struggling to predict index returns in Baltic stock markets suggest that risk is much higher in these markets compared to Nordic stock markets, because Baltic markets are small and probability of extreme negative returns is higher. This is in line with fifth thesis hypothesis, which state that more VaR models are able to produce accurate forecasts for Nordic stock indexes than for Baltic stock indexes. Although some unconditional VaR models were able to be best performers for some stock indexes during low volatility period according to Kupiec and Christoffersen test results, graphical analysis reveal that this is pure coincidence, because these models are not able to account for volatility clustering present in return series. The problem is that these models are not able to predict high returns when are estimated during low volatility period. In addition, these models are overestimating risk when extreme returns are incorporated into estimation.

Overall, summary of best models according to Christoffersen test at 1 percent level of significance presented in Table 21 shows that there is no particular model which is best for all stock indexes during all times and different levels of significance, which rejects last hypothesis of the thesis. Therefore, risk managers should estimate different models for their stock market and often check their performance through backtesting procedure, in order to use most accurate and most relevant models at the current stock market conditions. Nevertheless, the question is whether to use models which account for extreme returns and/or other stylized facts present in the data, not paying attention to simple models whose weaknesses were revealed in this thesis.

Most of the studies presented in the literature review compared only a few VaR models on one or two stock indexes (or futures), therefore none of them are completely comparable with the thesis. The closest paper with respect to number of models used and type of methodology was written by Kuester, Mittik and Paolella (2005). Although they used only one stock index, they evaluated the same five unconditional VaR models as in the thesis, as well as GARCH(1,1) model with three distribution assumptions and two EVT models. They found that most models underestimate risk, but skewed Student's t GARCH(1,1) VaR and both normal and Student's t EVT VaR models with GARCH(1,1) filtered returns are best performers for NASDAQ Composite Index. This is in line with the findings in this thesis, because these and similar models are among best models for seven Baltic and Nordic stock indexes. Although Huang and Lin (2004) compared only three conditional volatility VaR models for two stock index futures, they found that at higher level of significance, Student's t APARCH(1,1) VaR model outperforms APARCH(1,1) with normal return distribution assumption and normal EWMA (RiskMetrics) VaR models. The results obtained in the thesis coincide with their findings, because

APARCH(1,1) with Student's t distribution is among best models found in the thesis. What is more, even more advanced APARCH(1,1) model, with skewed Student's t return distribution assumption was evaluated in the thesis, and proved to be one of the best models. This result is consistent with the findings of research performed by Angelidis and Degiannakis (2008). They compared GARCH(1,1) and APARCH(1,1) with three distribution assumptions, the same as used in the thesis, and found out that skewed Student's t APARCH(1,1) VaR model is best for one index, while normal GARCH(1,1) VaR model is best for another index. Therefore they conclude that "as in the case of the VaR predictions, there is not a particular model that can be applied for both indices" (Angelidis & Degiannakis, 2008). The same result is obtained in the thesis, because several different advanced VaR models present equal or only slightly different forecasting performance, and rankings depend on the market situation (low or high fluctuations), data sample and other factors. Degiannakis, Floros and Livada (2012) compared three VaR models for five stock indexes for period before financial crisis and in year 2008. Approach is very similar and results are also similar. They got that at 5 percent level of significance for pre-crisis period, best model is normal EWMA VaR (RiskMetrics) for two indexes and normal GARCH(1,1) VaR model for the other three indexes. During the crisis period skewed Student's t APARCH(1,1) VaR model is best for four indexes out of five. Similarly, these three models are among best models in the thesis for different stock indexes for both low volatility (analogy to pre-crisis) and high volatility (crisis period is included) periods. Alhough it was not tested in their study, this thesis shows that better performance is shown when EWMA and GARCH VaR models are enhanced by changing return distribution assumption from normal to Student's t or skewed Student's t.

One of the limitations of the thesis already mentioned in the Methodological part is that non-normal distributions are fitted and volatility models are estimated every 25 observations. The assumption is that model parameters do not change much when window with 1000 data points is moved by one day dropping last observation and adding new one. However, difference in the results can be larger for models which have many parameters such as APARCH model. Another limitation of the thesis is rather small estimation sample for fitting extreme value theory models. Although 1000 data points are used for each index, number of extreme negative returns are much lower in this sample, therefore it can be difficult to fit GPD distribution to small sample of extreme returns. Finally, average return for the whole sample of each index was very close to zero, therefore it was assumed that mean return is zero and this allowed to simplify volatility models used for estimation.

Implications for further research are related to the limitations of the study. First of all, this study could be repeated after several years when more data is available for all stock markets analysed in this thesis. Another implication would be to perform multi-period forecasting and apply these VaR models not only to equity indexes but also other asset classes such as fixed income instruments or the ones offering non-linear payoffs, for example, options. Furthermore, various portfolios of these intruments could be created in order to apply multivariate volatility models such as orthogonal GARCH and calculate portfolio VaR.

One of the implications of the study for professional practice is that there is no particular model for each stock market during all time periods which is always the best because different models provide most accurate VaR forecasts, therefore, portfolio and risk managers should regularly perform backtesting of their models and choose the best ones and most suited for current financial market situation. Although simple VaR models are preferred by financial institutions, more advanced models should be considered for risk management, because simple models are not able to give accurate VaR forecasts, especially during financial market turmoil.

6. Conclusions

Thesis aimed to perform analysis on Value at Risk forecasting performance for Nordic and Baltic stock indexes in order to choose best VaR models during different time periods and at different levels of significance. As the analysis of relevant literature showed, the problem was that different VaR models show different results during calm and volatile time periods in the financial markets and results depend on properties of stock market. Therefore 17 VaR models were evaluated for their forecasting performance during low and high volatility periods using backtesting and Baltic stock market index data since inception in 2000 till the end of 2011 as well as Nordic stock index returns for the same period. Following objectives were raised in the thesis and such conclusions can be drawn from performed study:

1. Having analyzed and compared various theoretical and empirical researches on measuring and forecasting VaR for both developed and emerging stock markets, such gaps were found and resolved in this thesis:
   - Many researches focused on only one or a few stock indexes and compared only several VaR models. Furthermore, Nordic and Baltic region have faced little attention regarding VaR modeling compared with other financial markets worldwide. Therefore, this thesis contributed to theoretical and empirical research by comparing even 17 VaR models for Nordic and Baltic stock indexes as well as giving other useful insight regarding assumptions of return distributions and parameters of volatility models.
2. Properties of financial returns of Nordic and Baltic stock indexes such as normality and skewness were measured and to stylized facts such as volatility clustering were check in the data leading to such findings:
   - First hypothesis that returns of Baltic stock indexes are non-normally and asymmetrically distributed, while returns of Nordic stock indexes are normally and symmetrically distributed is partially rejected, because return distributions of Nordic stock indexes are also leptokurtic and asymmetric, but at lower scale than Baltic stock markets.
3. Various volatility and distribution models were presented and estimated to find best fitting models which accurately describe financial return properties of Nordic and Baltic stock indexes and such conclusions were made:

- Student's t and skewed Student's t distribution assumption is better than normal distribution because volatility model fits return data more accurately using first two distribution assumptions. Therefore, second hypothesis is not rejected.
- APARCH(1,1) model should be preferred instead of GARCH(1,1) model, because it has two additional parameters, which estimates power term and leverage effect, and it was found that APARCH(1,1) model has better fit to the index return data than GARCH(1,1) model when distribution of returns is assumed as Student's t and skewed Student' t. However, when APARCH(1,1) volatility model has normal return distribution assumption, GARCH(1,1) model has better fit and, therefore, third thesis hypothesis is only partially rejected.

4. Many VaR forecasting models were estimated, forecasts of possible losses with given levels of confidence were created for low and high volatility period and forecasting performance was evaluated to find best VaR models for Nordic and Baltic stock indexes before and after financial crisis. The empirical findings of backtesting procedure are as follows:
    - One-sided *t*-test, which compared percentage of VaR models which passed Christoffersen test when VaR models were estimated with 1% level of significance and with 5% level of significance, shows that statistically significantly more VaR models produce accurate VaR forecasts when models are measured at higher level of confidence. Therefore, fourth thesis hypothesis is not rejected.
    - The same statistical test compared model forecasting performance in Nordic and Baltic stock markets and found out that more VaR models produce accurate VaR forecasts for Nordic stock indexes than for Baltic stock indexes. This is due to the stock market properties: Baltic stock market is small and probability of extreme negative returns is higher than in more developed Nordic stock market. Therefore, fifth thesis hypothesis is also not rejected.
    - Finally, results of Christoffersen test show that usually more than one VaR model provides best VaR forecasts and that best models for the same index are different at different stock markets conditions and different levels of confidence. If there is only one VaR model, which has highest p-value in Christoffersen test, most of the time, second best model is very close to the best model's statistical significance. Therefore, last hypothesis of the thesis is strongly rejected and is in line with other empirical research, which also was not able to find one best VaR model for many stock markets during

different stock market conditions and with different levels of confidence. However, most advanced VaR models such as EVT VaR with GARCH(1,1) volatility filter as well as APARCH(1,1) VaR model with skewed Student's distribution assumption are among best models for many stock indexes as presented in Table 21.

Overall, thesis made a contribution to theoretical and empirical research by estimated many VaR models and found best ones for Nordic and Baltic stock indexes, as well as evaluated properties of return distributions of these stock markets and showed that more advanced volatility models better fit return data of stock indexes. Finally, thesis showed that more VaR models provide accurate VaR forecasts according to Christoffersen test at higher level of confidence and for Nordic stock markets than for Baltic stock markets.

The implication is that VaR forecasting models are applicable to Baltic and Nordic stock markets and that risk managers should enhance their quantitative modeling capabilities to measure risk by using more advanced VaR models which account for stylized facts of financial returns. Further research could examine multi-period VaR forecasting and apply VaR models and backtesting for other asset classes and their portfolios in financial markets of Nordic and Baltic countries.

Reference list

1. Alexander, C. (2008a). *Market Risk Analysis (Volume II): Practical Financial Econometrics.* John Wiley & Sons Ltd.

2. Alexander, C. (2008b). *Market Risk Analysis (Volume IV): Value-at-Risk Models.* John Wiley & Sons Ltd.

3. Angelidis, T., & Degiannakis, S. (2008). Forecasting One-day-ahead VaR and Intra-Day Realized Volatility in the Athens Stock Exchange Market. *Managerial Finance, 34*(7), 489-497.

4. Assaf, A. (2009). Extreme observations and risk assessment in the equity markets of MENA region: Tail measures and Value-at-Risk. *International Review of Financial Analysis, 18*, 109-116.

5. Balkema, A., & de Haan, L. (1974). Residual life time at great age. *Annals of Probability, 2*, 792–804.

6. Baltic Stock Exchanges. (2001, 07). *Baltic Monthly Statistical bulletins.* Retrieved March 24, 2012, from NASDAQ OMX Baltic Web site: http://www.nasdaqomxbaltic.com/bulletins/pdf/2001/2001ms07.pdf

7. Beder, T. S. (1995). VAR: Seductive but Dangerous. *Financial Analysts Journal, 51*(5), 12–24.

8. BIS. (2006). *Basel II: International Convergence of Capital Measurement and Capital Standards.* Bank for International Settlements, Basel Committee on Banking Supervision, Basel, Switzerland. Retrieved January 30, 2012, from http://www.bis.org/publ/bcbs128.htm

9. Bollerslev, T. (1986). Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics, 31*(3), 307–327.

10. Bollerslev, T. (1987). A Conditionally Heteroskedastic Time Series Model for Speculative Prices and Rates of Return. *Review of Economics and Statistics, 69*, 542–547.

11. Bormetti, G., Cisana, E., Montagna, G., & Nicrosini, O. (2007). A non-Gaussian approach to risk measures. *Physica A, 376*, 532-42.

12. Boudoukh, J., Richardson, M., & Whitelaw, R. (1998). The best of both worlds: a hybrid approach to calculating value at risk. *Risk, 11(May)*, 64–67.

13. Brooks, C. (2008). *Introductory Econometrics for Finance* (2nd ed.). Cambridge University Press.

14. Campbell, S. (2005). *A Review of Backtesting and Backtesting Procedures, Technical Report.* Federal Reserve staff working paper in the Finance and Economics Discussion Series.

15. Christoffersen, P. (1998). Evaluating Interval Forecasts. *International Economic Review, 39*, 841-862.

16. Christoffersen, P., & Jacobs, K. (2004). Which GARCH model for option valuation? *Management Science, 50*, 1204-1221.

17. Danielsson, J. (2011). *Financial Risk Forecasting.* John Wiley & Sons Ltd.

18. Degiannakis, S., Floros, C., & Livada, A. (2012). Evaluating Value-at-Risk Models before and after the Financial Crisis of 2008: International Evidence. *Managerial Finance, 38*(4).

19. Diamandis, P., Kouretas, G., & Zarangas, L. (2006). *Value-at-Risk for Long and Short Trading Positions: The Case of the Athens Stock Exchange., Working Paper.* University of Crete, Department of Economics.

20. Diebold, F. X., Schuermann, T., & Stroughair, J. D. (1998). *Pitfalls and Opportunities in the Use of Extreme Value Theory in Risk Management. Working paper.* The Wharton School, University of Pennsylvania.

21. Ding, Z. C., Granger, W. J., & Engle, R. F. (1993). A Long Memory Property of Stock Market Returns and a New Model. *Journal of Empirical Finance, 1*, 83–106.

22. Dowd, K. (2005). *Measuring Market Risk* (2nd ed.). John Wiley and Sons, Ltd.

23. Embrechts, P., Kluppelberg, C., & Mikosch, T. (1997). *Modelling extremal events for insurance and finance.* Berlin: Springer.

24. Engle, R. F. (1982). Autoregressive conditional heteroskedasticity with estimates of the variance. *Econometrica, 50*, 987–1007.

25. 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, 12*, 180–190.

26. Gencay, R., & Selcuk, F. (2004). Extreme value theory and Value-at-Risk: Relative performance in emerging markets. *International Journal of Forecasting, 20*, 287– 303.

27. Gilli, M., & Kellezi, E. (2006). An Application of Extreme Value Theory for Measuring Financial Risk. *Computational Economics, 27*(1), 1-23.

28. Gnedenko, B. (1943). Sur la distribution limite du terme maximum d'une serie aleatoire. *Ann. of Math., 44*, 423–453.

29. Huang, A.-H., & Tseng, T.-W. (2009). Forecast of value at risk for equity indices: an analysis from developed and emerging markets. *Journal of Risk Finance, 10*(4), 393-409.

30. Huang, Y. C., & Lin, B.-J. (2004). Value-at-Risk Analysis for Taiwan Stock Index Futures: Fat Tails and Conditional Asymmetries in Return Innovations. *Review of Quantitative Finance and Accounting, 22*(2), 79-95.

31. Hull, J., & White, A. (1998). Incorporating volatility updating into the historical simulation method for value-at-risk. *Journal of Risk, 1(Fall)*, 5–19.

32. Jorion, P. (2007). *Value at Risk: The New Benchmark for Managing Financial Risk.* (3rd ed.). New York: McGraw Hill.

33. Kalyvas, L., Steftos, A., Sriopoulos, C., & Georgopoulos, A. (2007). An investigation of riskiness in South and Eastern European markets. *International Journal of Financial Services Management, 2*(1/2), 13-21.

34. Kuester, K., Mittik, S., & Paolella, M. (2005). Value-at-risk prediction: a comparison of alternative strategies. *Journal of Financial Econometrics, 4*(1), 53-89.

35. Kupiec, P. (1995). Techniques for Verifying the Accuracy of Risk Management Models. *Journal of Derivatives, 3*, 73-84.

36. Lambert, P., & Laurent, S. (2000). *Modeling Skewness Dynamics in Series of Financial Data, Discussion Paper.* Louvain-la-Neuve: Institut de Statistique.

37. Longerstaey, J., & Zangari, P. (1996). *RiskMetrics™--Technical Document* (4th ed.). New York: Morgan Guaranty Trust Co.

38. Mačiulis, N., Lazauskaitė, V., & Bengtsson, E. (2007). Evaluating performance of Nordic and Baltic stock exchanges. *Baltic Journal of Management, 2*(2), 140–153.

39. Mandelbrot, B. (1963). The variation of certain speculative prices. *Journal of Business, 36*, 394–419.

40. Markowitz, H. (1952). Portfolio Selection. *The Journal of Finance, 7*(1), 77–91.

41. McNeil, A. J. (1999). *Extreme Value Theory for Risk Managers.* Zurich: Departement Mathematik, ETH Zentrum.

42. McNeil, A. J., & Frey, R. (2000). Estimation of Tail-related Risk Measures for Heteroscedastic Financial Time Series: An Extreme Value Approach. *Journal of Empirical Finance, 7(3/4)*, 271–300.

43. McNeil, A., Frey, R., & Embrechts, P. (2005). *Quantitative Risk Management: Concepts, Techniques and Tools.* Princeton, NJ.: Princeton University Press.

44. Ozun, A., Cifter, A., & Yilmazer, S. (2010). Filtered extreme-value theory for value-at-risk estimation: evidence from Turkey. *Journal of Risk Finance, 11*(2), 164-179.

45. Pagan, A., & Schwert, G. (1990). Alternative Models for Conditional Stock Volatility. *Journal of Econometrics, 45*, 267–290.

46. Pickands, J. (1975). Statistical inference using extreme order statistics. *Annals of Statistics, 3*, 119–131.

47. Smith, D. (2008). Testing for structural breaks in GARCH models. *Applied Financial Economics, 18*(10), 845-862.

48. Wuertz, D., & others. (2009). fExtremes: Rmetrics - Extreme Financial Market Data. *R package version 2100.77.* Retrieved January 14, 2012, from http://CRAN.R-project.org/package=fExtremes

49. Wuertz, D., Chalabi, Y., & others. (2009). fGarch: Rmetrics - Autoregressive Conditional Heteroskedastic Modelling. *R package version 2110.80.* Retrieved January 14, 2012, from http://CRAN.R-project.org/package=fGarch

Appendix

Figure 11: OMX Vilnius stock index value (a), one-day returns (b), their density (c) and QQ-plot against Gaussian distribution (d)

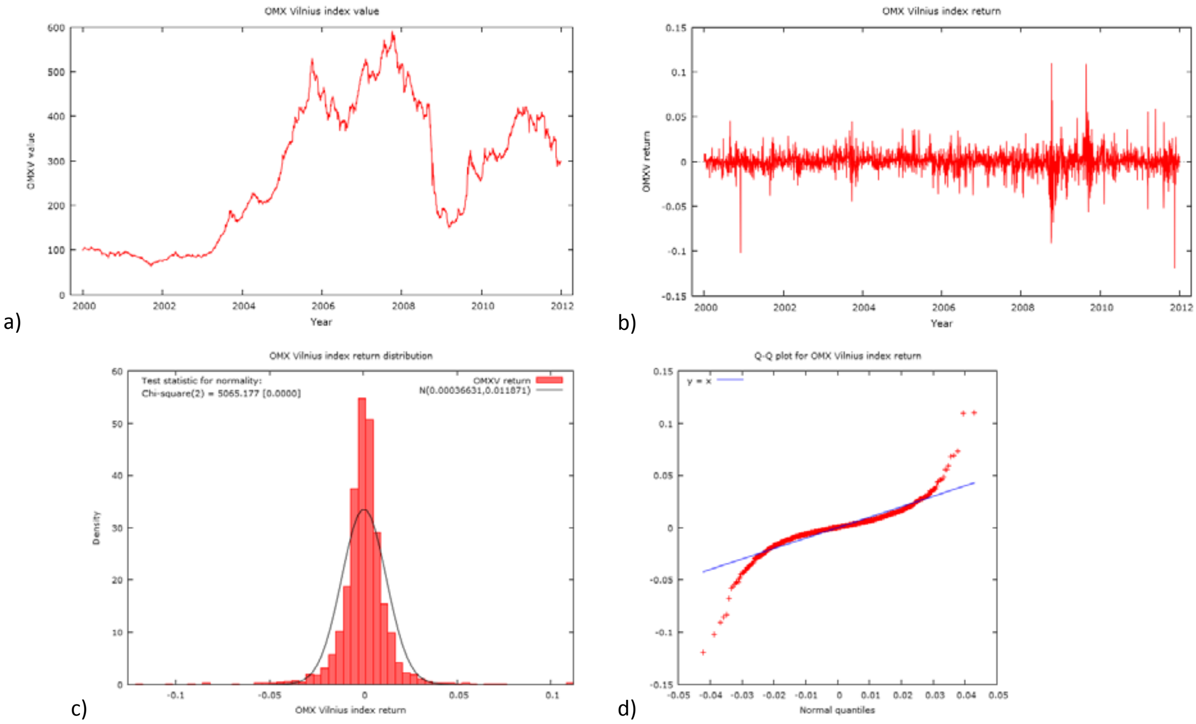


Figure 12: OMX Riga stock index value (a), one-day returns (b), their density (c) and QQ-plot against Gaussian distribution (d)

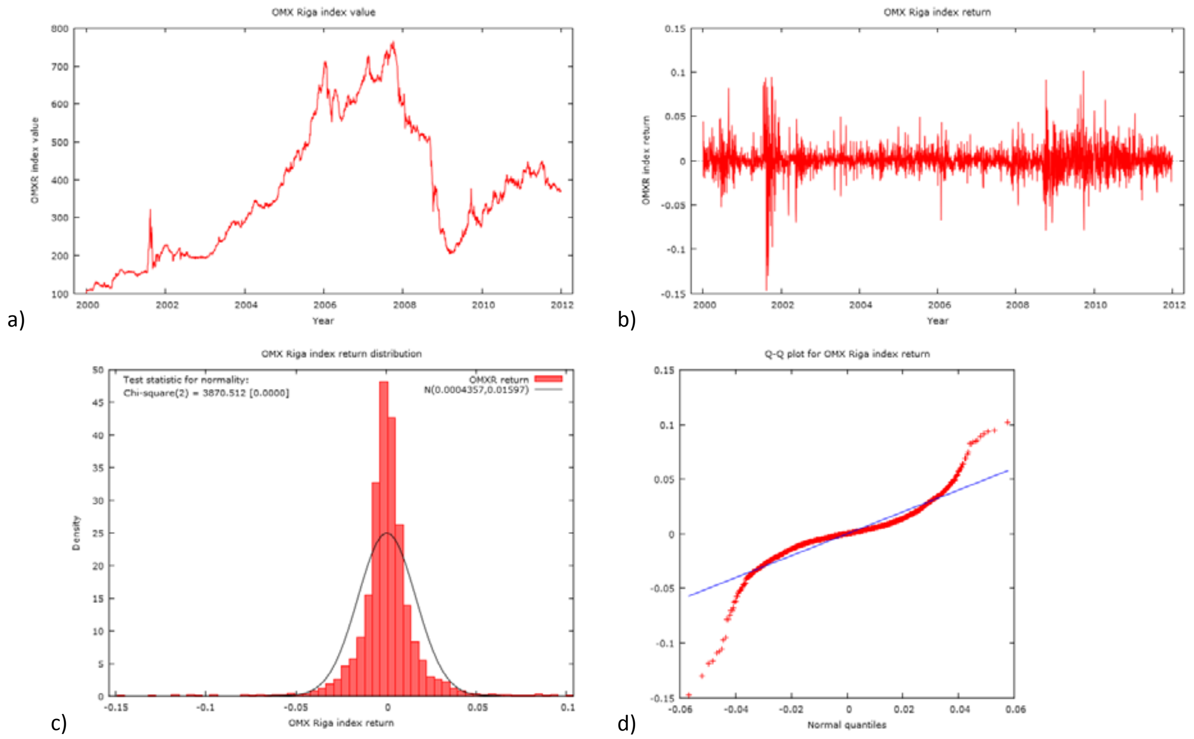


Figure 13: OMX Tallinn stock index value (a), one-day returns (b), their density (c) and QQ-plot against Gaussian distribution (d)

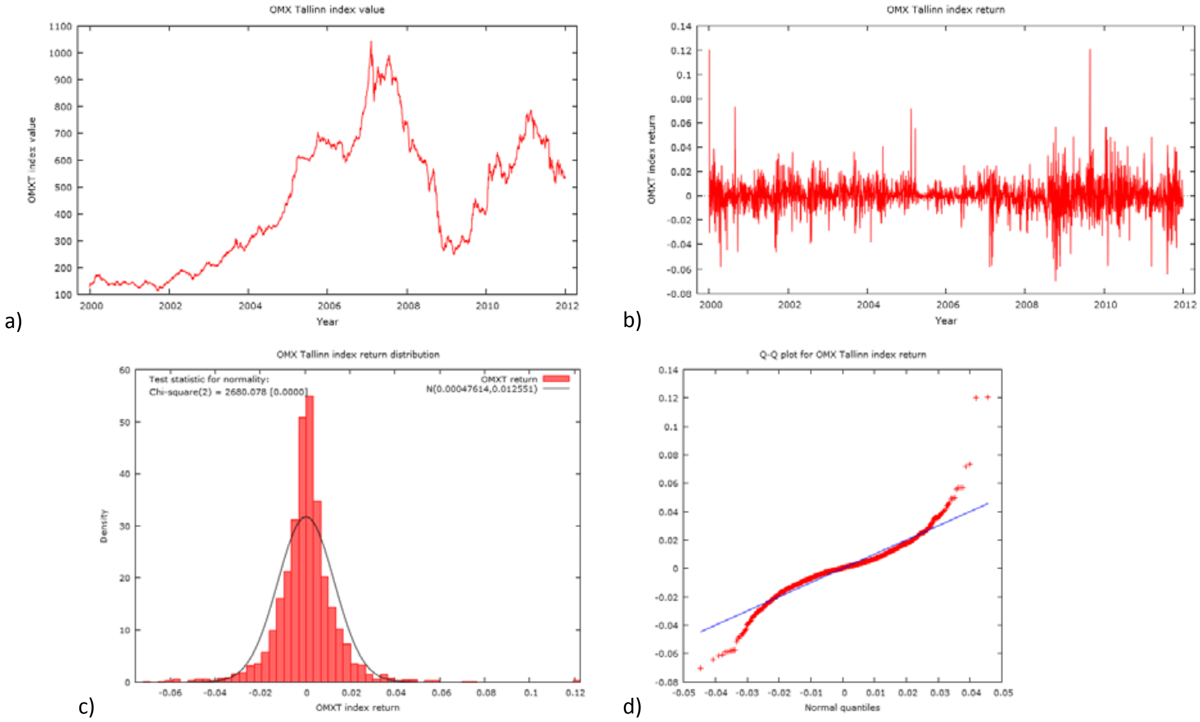


Figure 14: OMX Helsinki stock index value (a), one-day returns (b), their density (c) and QQ-plot against Gaussian distribution (d)

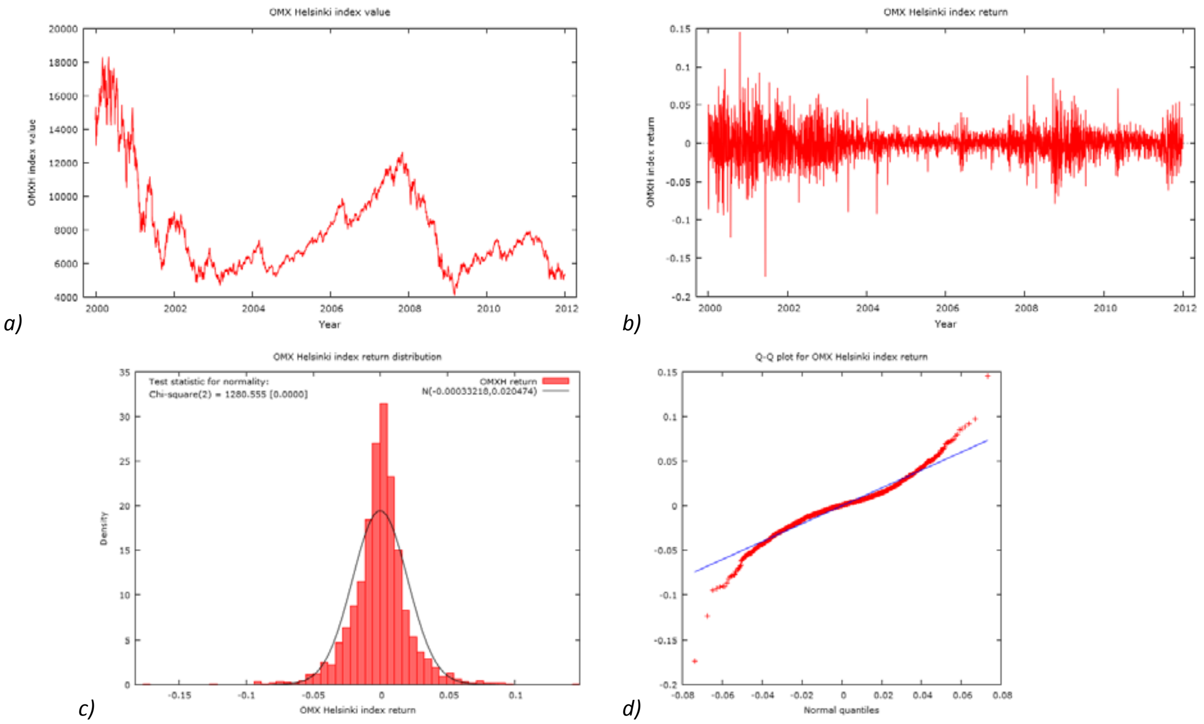


Figure 15: OMX Stockholm stock index value (a), one-day returns (b), their density (c) and QQ-plot against Gaussian distribution (d)

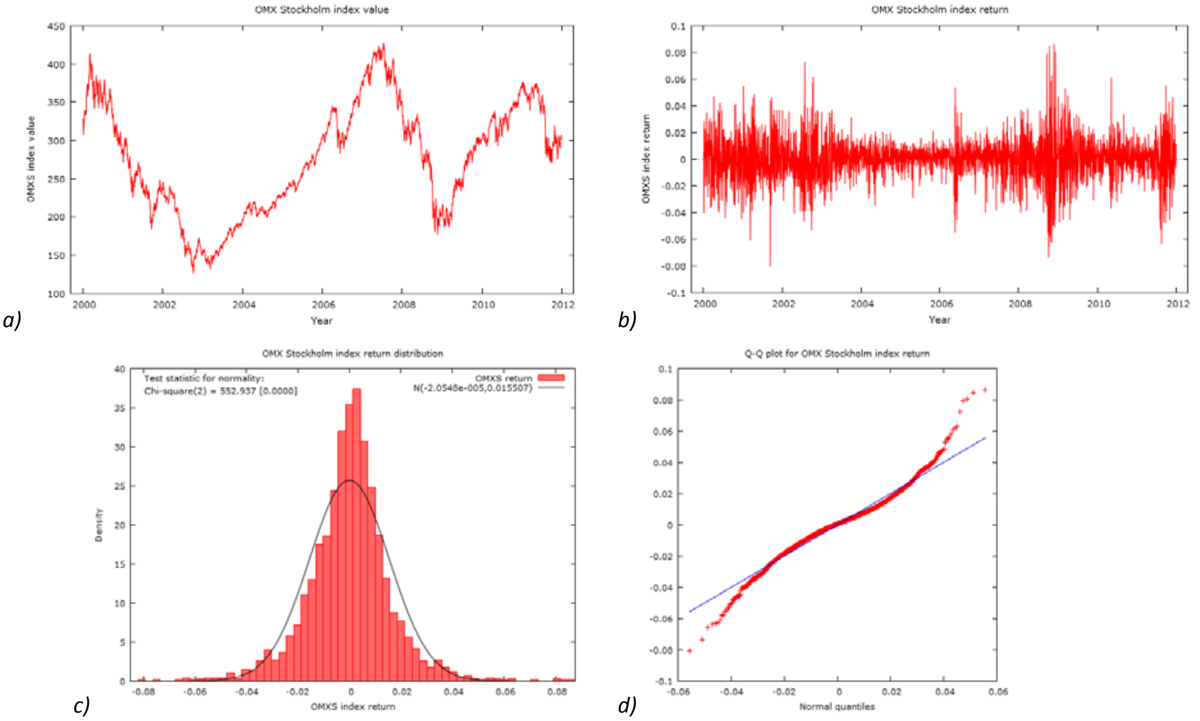


Figure 16: OMX Copenhagen stock index value (a), one-day returns (b), their density (c) and QQ-plot against Gaussian distribution (d)

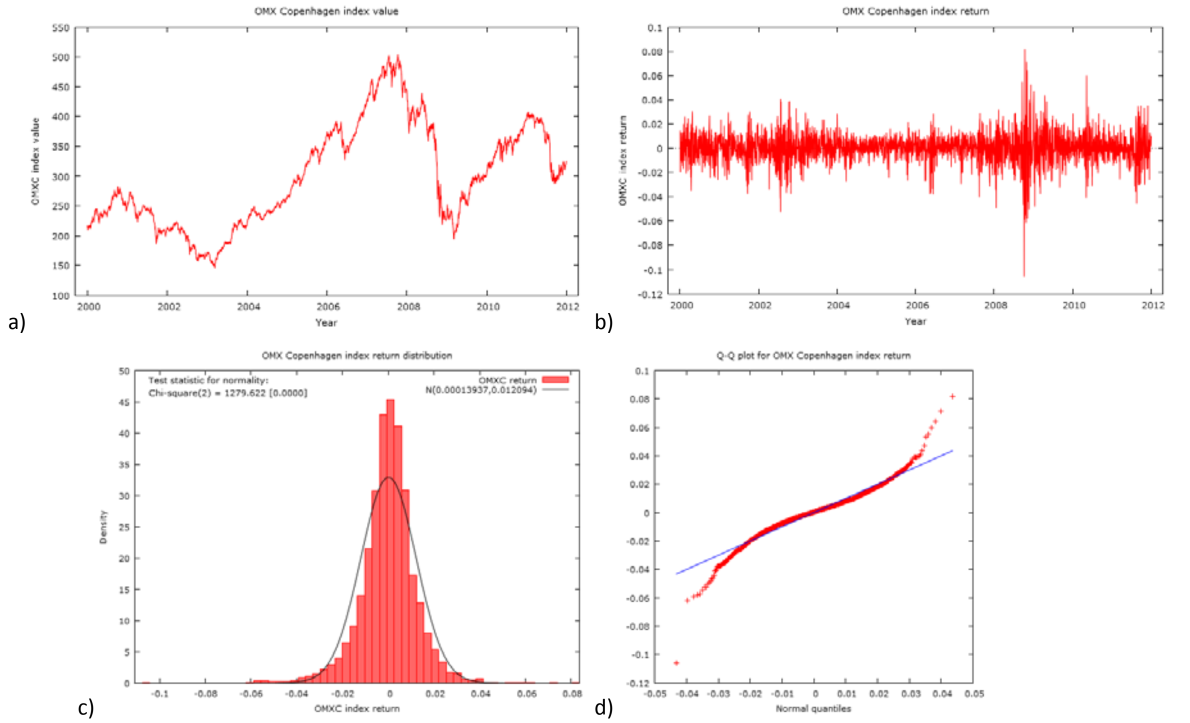


Figure 17: OSEAX Oslo stock index value (a), one-day returns (b), their density (c) and QQ-plot against Gaussian distribution (d)

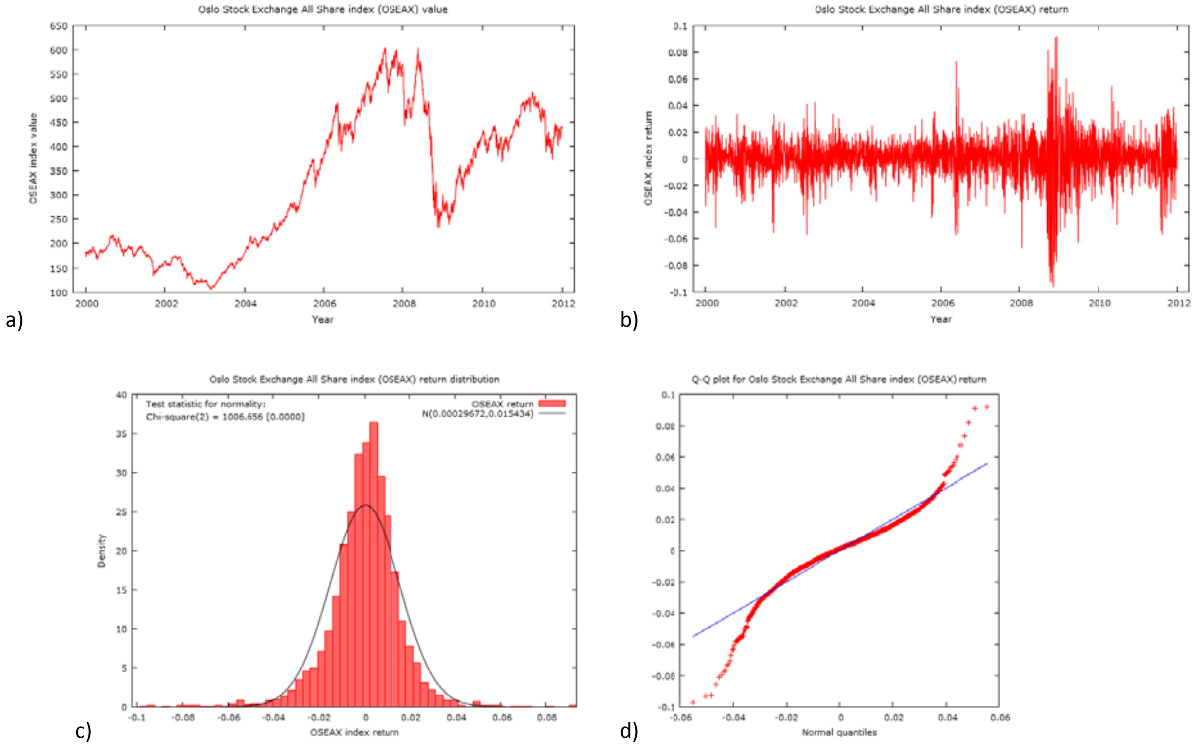


Figure 18: OMX Vilnius stock index returns beyond a threshold value, peaks over a threshold, excess distribution and its tail.

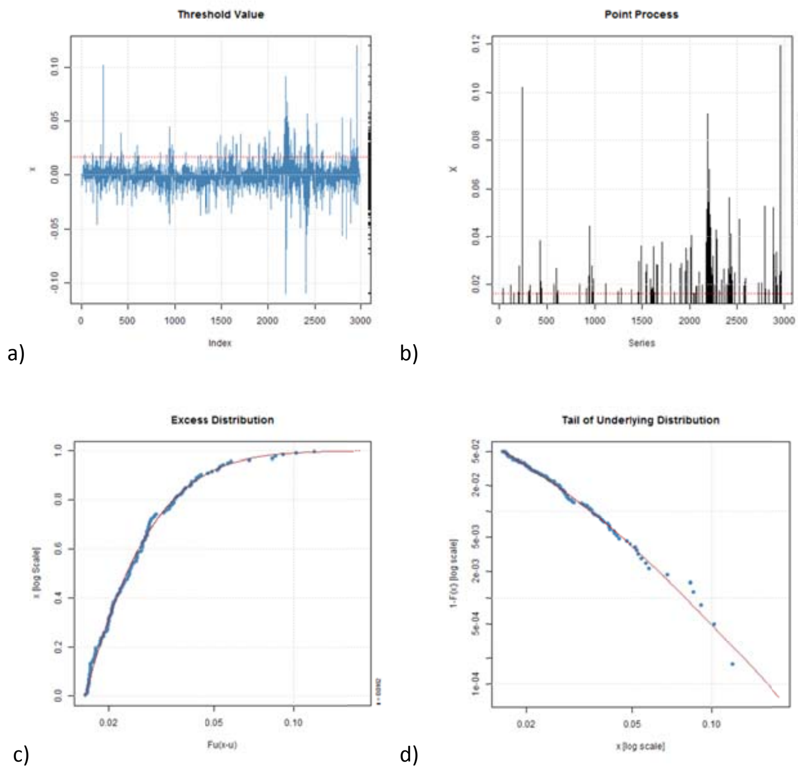


Figure 19: OMX Riga stock index returns beyond a threshold value, peaks over a threshold, excess distribution and its tail.

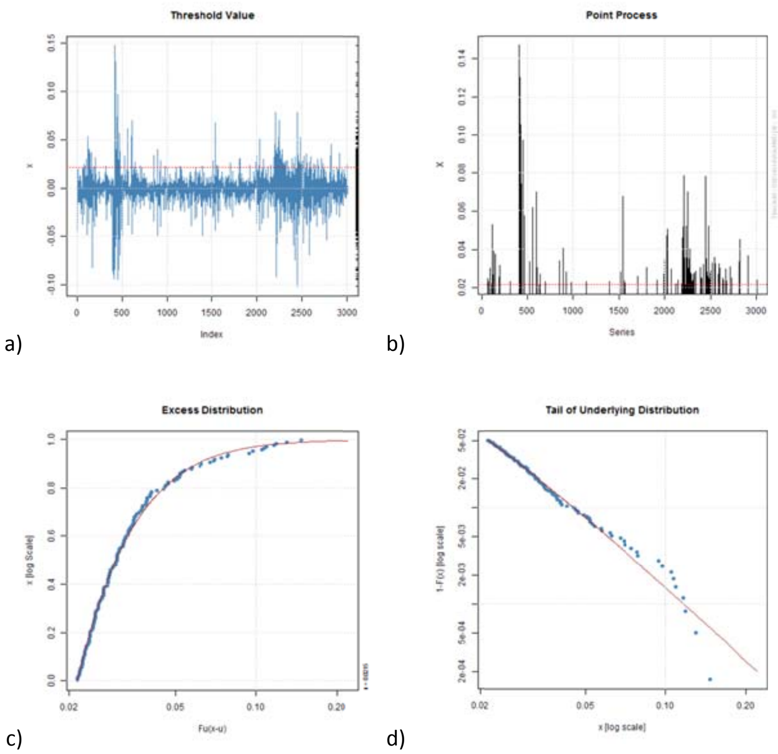


Figure 20: OMX Tallinn stock index returns beyond a threshold value, peaks over a threshold, excess distribution and its tail.

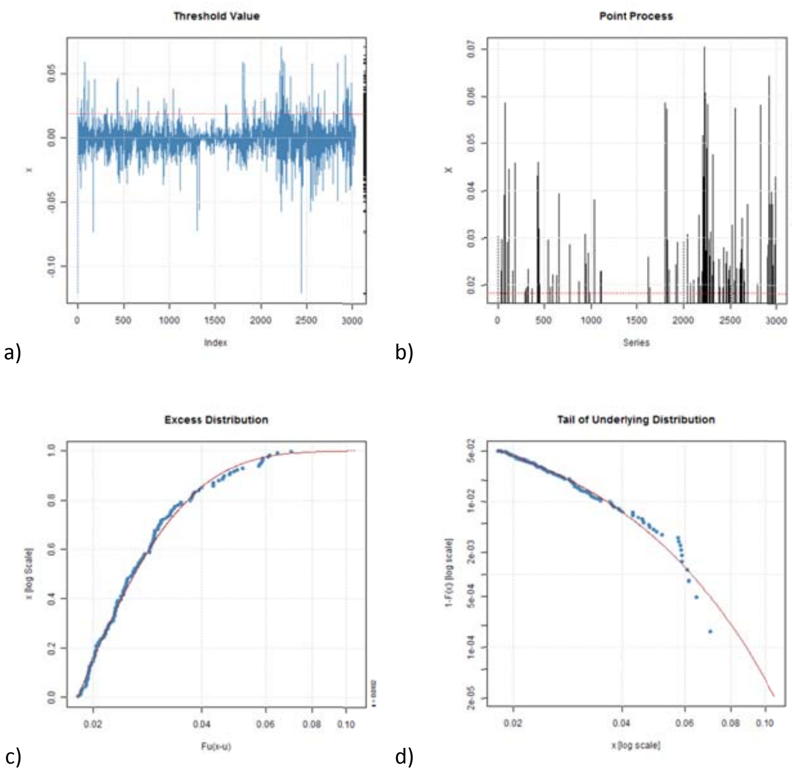


Figure 21: OMX Helsinki stock index returns beyond a threshold value, peaks over a threshold, excess distribution and its tail.

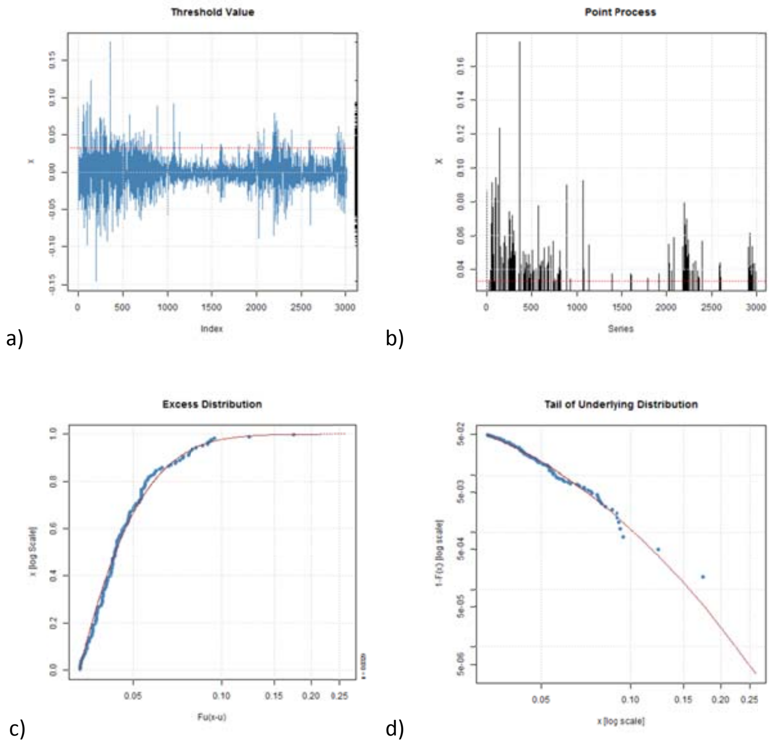


Figure 22: OMX Stockholm stock index returns beyond a threshold value, peaks over a threshold, excess distribution and its tail.

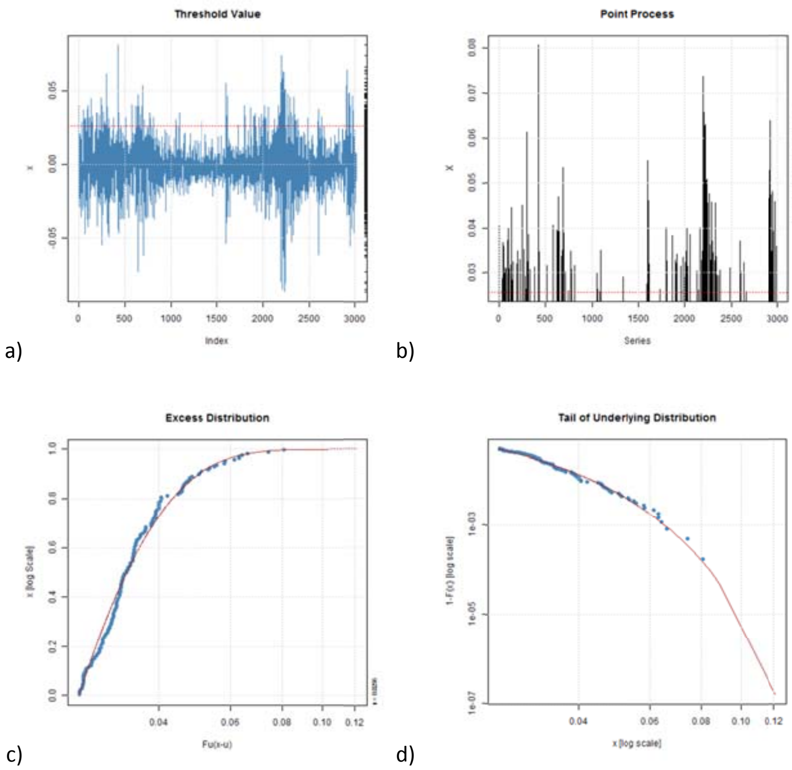
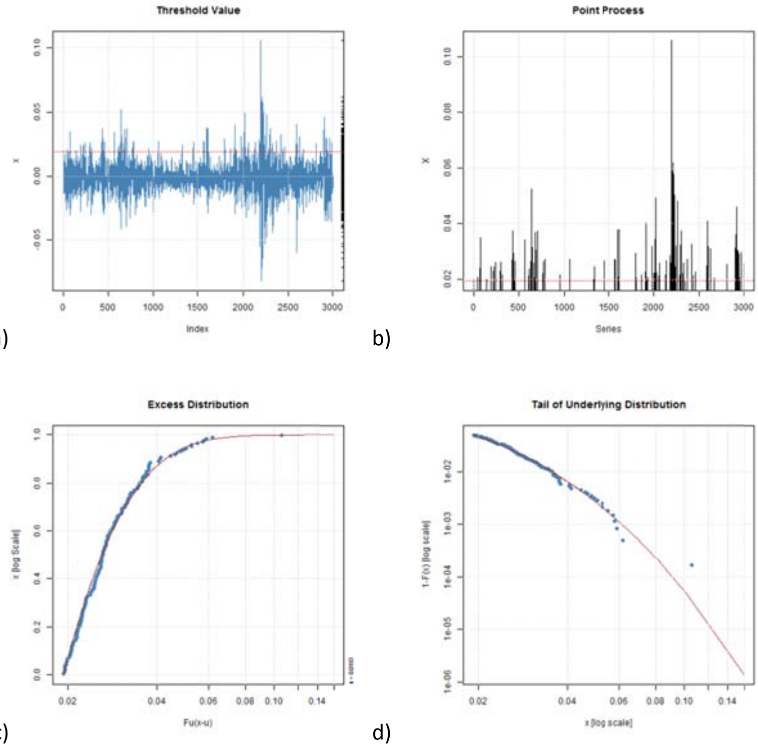


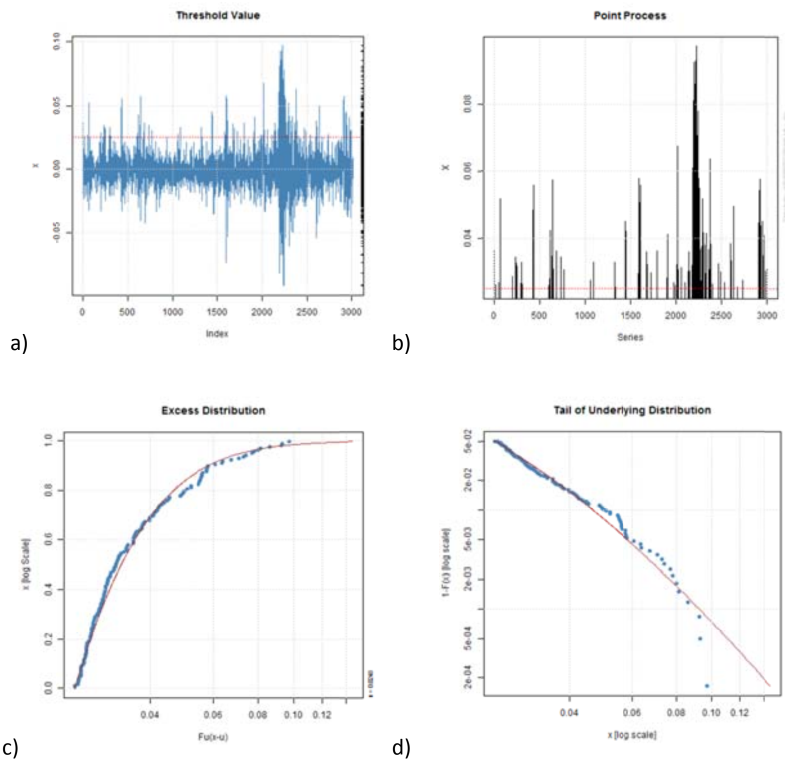
Figure 23: OMX Copenhagen stock index returns beyond a threshold value, peaks over a threshold, excess distribution and its tail.



a) b)

c) d)

Figure 24: OSEAX Oslo stock index returns beyond a threshold value, peaks over a threshold, excess distribution and its tail.



a) b)

c) d)