TURUN KAUPPAKORKEAKOULU
Turku School of Economics

PREDICTIVE PERFORMANCE OF VALUE AT RISK MODELS

Evaluation before, during and after market crisis

Master´s Thesis
in Accounting and Finance

Author:
Rami Katajisto 9642

Supervisors:
Prof. Hannu Schadéwitz
Senior Lecturer Tuula-Marja Vilja

23.4.2008
Turku

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF EQUATIONS

LIST OF ACRONYMS

BM	Block maxima
CDF	Cumulative probability density function
CRD	Capital Requirements Directive
EMH	Efficient market hypothesis
EVT	Extreme value theory
EQMA	Equally weighted moving average
EWMA	Exponentially weighted moving average
GARCH	General autoregressive conditional heteroskedasticity
GBM	Geometric Brownian motion
GEV	Generalized extreme value distribution
GPD	Generalized Pareto distribution
HS	Historical simulation
i.i.d.	Independent and identically distributed
LR_{cc}	Likelihood ratio test statistic for conditional coverage
LR_{ind}	Likelihood ratio test statistic for independence
LR_{uc}	Likelihood ratio test statistic for unconditional coverage
MA	Moving average
MC	Monte Carlo simulation
MEF	Mean excess function
PDF	Probability density function
POT	Peaks over threshold
RM	RiskMetrics
VaR	Value at risk

1 INTRODUCTION

1.1 Financial risks in the 21st century

The recent growth of the risk-management industry can be traced directly to the increased volatility of the financial markets since the early 1970s. This increased volatility first became apparent in the currency markets after the collapse of the Bretton Woods Agreement, when the fixed exchange rate system broke down in 1971, leading to flexible and volatile exchange rates. This development was soon followed by interest rates and commodity prices (Simons 1996; Jorion 2007, 4–5). Figures 1 and 2 below depict some examples of these movements in exchange rates and interest rates during the past 30 years.

Figure 1 plots the development of USD exchange rate to GBP (Great Britain Pound) and to JPY (Japanese Yen) starting from September 1978 to October 2007. The exchange rates are rebased to 100 at the starting date to fit the both time series in the same chart. Increased volatility in the late 1970s and throughout the 1980s can clearly be observed in Figure 1. Further examination of Figure 1 also reveals that volatility seems to dampen since the mid 1990s, especially in the USD to GBP time series. In between this time, the dollar has reached dizzying heights, such as in 1984 against GBP, just to fall to unpredicted lows, as is the case for the moment against euro (not presented in figure 1), in the process creating wild swings in the competitive advantage of nations – and nightmares for unhedged firms.

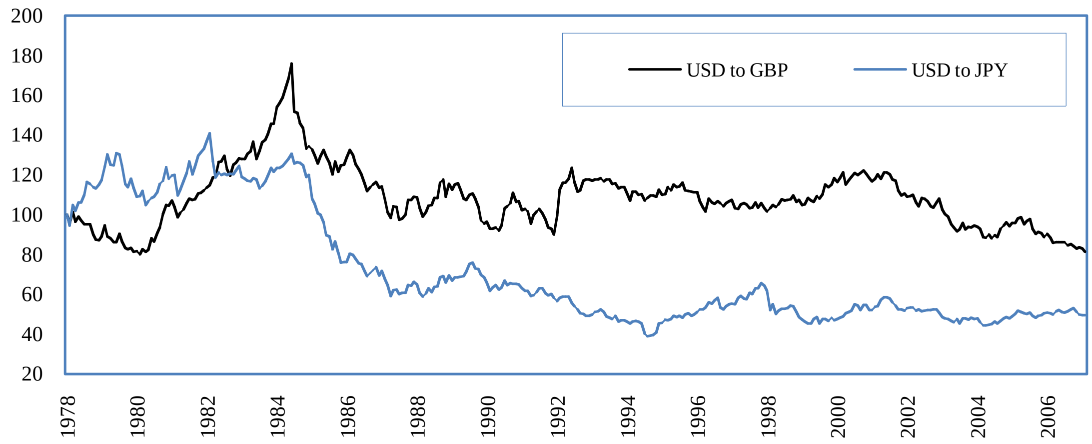


Figure 1 Movements in exchange rates (1978-2007)

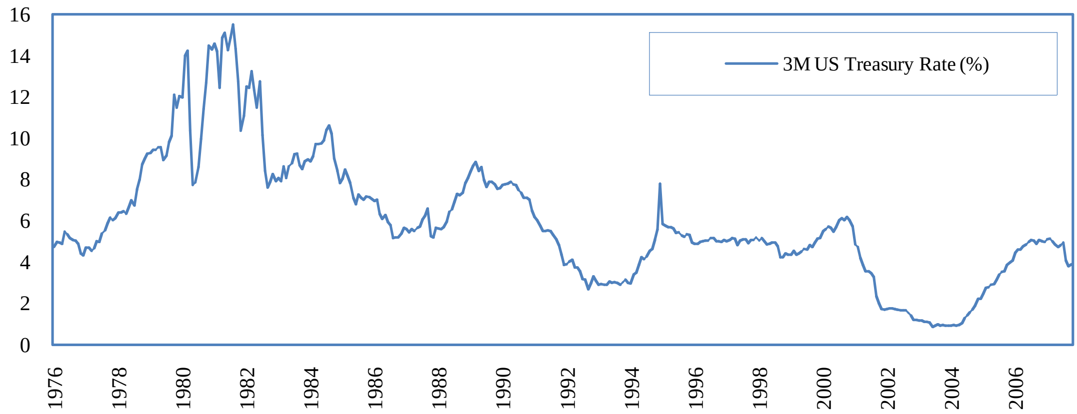


Figure 2 Movements in 3 month U.S. Treasury rate (1976-2007)

Figure 2 illustrates the development of 3 month U.S. Treasury rate from January 1976 to October 2007. An extremely volatile era can also be observed in this time series of bond prices starting from the late 1970s and lasting until the early 1980s, reflecting the inflationary pressures spreading throughout the U.S. economy. In October 1979, the Federal Reserve Bank forcefully attempted to squash inflation. Interest rates immediately shot up, became more volatile, and led to a sustained appreciation of the dollar (Jorion 2007, 6). Reflecting the impacts of this interest rate volatility, the largest bank failure in the U.S., Continental Illinois, was witnessed at this era. Poor interest rate risk management attributed to Continental eventually failing in 1984 and it was the subject of an expensive government bailout (Hull 2007, 22). In figure 2 it can also be seen how much interest rates can fluctuate over time. In this monthly data for about 30 years the 3 month U.S. Treasury rate ranges from a minimum of 0,86% to as high as 15,52%. These seasons of high volatility in the market variables since the 1970s spurred new emphasis on financial risk management.

In addition to this unleashed volatility, firms generally have become more sensitive to movements in financial variables. Prior to the 1970s, banks were heavily regulated or comfortably cartelized in most industrial countries. Industrial corporations were mainly selling in domestic markets, and therefore not too concerned about exchange rate movements. This all changed with deregulation and globalization. The 1970s witnessed a worldwide movement towards market-oriented policies and deregulation of financial markets. As a result of deregulation, financial institutions were forced to be more competitive and became more vulnerable to financial risks. As globalization became a popular buzzword, barriers to international trade and investments were lowered. As a result, firms had to recognize the truly global nature of competition and in the process they became more exposed to a greater variety of financial risks (Jorion 2007, 7).

This above described development created the need for new financial instruments and tools for risk management. Technological development, in both physical equipment and financial theory, made it possible for derivates to answer this call. Derivatives markets provide an essential mechanism to exchange financial risks, since they allow risks to be transferred to those best able to bear them. As a result the 1980s witnessed a surge in the size of the derivatives markets. However, derivatives are not a God-given solution to all our problems concerning the management of financial risks. The development of progressively complex derivatives instruments has made it more difficult to understand these instruments and if the risks embedded in them are not controlled and understood properly, derivatives possess the potential to increase risks exponentially – rather than to reduce them. This can be witnessed in the highly publicized derivatives disasters in the 1990s and 2000s. Companies, such as Orange County, Metallgesellschaft and Bearings Bank, just to name a few, suffered great losses in derivatives trading. Many presentations on these disasters have been written (for a comprehensive description see e.g. Partnoy 1999) and as a conclusion it can be said that these big losses were not simply due to derivatives, but due to criminal acts, lack of supervision and poor risk management. It is clear that derivatives played an important part in all this, but it cannot be said that they were in the leading role (Jorion 2007, 10, 15, xxii).

Spurred into action, financial institutions and regulators focused on finding new reliable risk measurement techniques to address these problems. This is when they turned to value at risk (VaR), an easy-to-understand method for quantifying market risk.

1.2 Brief introduction to value at risk and rise of the method

What is this value at risk then that has caused so much commotion among regulators, practitioners and scholars? Basically the VaR concept is concerned with market risk, but it is also applicable to other types of financial risk as well, namely credit risk and operational risk (see e.g. Basel Committee 2005). Tsay (2005, 287–288) conveniently defines VaR as a "*single estimate of the amount by which an institution's position in a risk category could decline due to general market movements during a given holding period*". At the heart of producing this kind of estimate is the probability distribution function (PDF) of the position's return. There are various approaches to producing and estimating this distribution. We can, for example, use the empirical distribution itself or try to fit some kind of parametric distribution, such as the normal distribution or the Student-t distribution, to the data and estimate its parameters. Yet another alternative is to concentrate directly on the tails of the distribution and to model the tail behaviour of the distribution parametrically, rather than modelling the whole distribution. Hence, there are numerous approaches to calculating VaR. The most popular of these amongst practitioners have been historical simulation (HS) (nonparametric), variance-covariance

(VC) (parametric), and Monte-Carlo simulation (MC) (parametric) approaches. Only recently the extreme value theory (EVT) approach for calculating VaR (parametric or semiparametric) has gained increased interest in academic fronts. Practitioners, however, have shown only limited interest in building models based on this particular approach (see e.g. Rogachev 2007). An approach is referred to as parametric if it involves estimating the parameters of a certain probability distribution.

To better illustrate the concepts embodied in this brief definition of VaR by Tsay presented above, Hull (2007, 196) is quoted: *"When using the value-at-risk measure we are interested in making a statement of the following form: We are X percent certain that we will not lose more than V dollars in the next N days."* In this case the variable *V* is the VaR of the portfolio and it is determined by a single point of the estimated distribution. The other two parameters in the quotation above are *X*, which denotes the confidence level (i.e. the coverage probability) and *N*, which denotes the time horizon (i.e. the holding period). In other words, VaR is the loss corresponding to the (100–*X*)th percentile of the distribution of the change in the value of the portfolio over the next *N* days. For example, if the coverage probability (*X*%) is specified to be 95%, then the VaR is the fifth percentile (or 0,05 quantile) of the estimated distribution for the portfolio return *N* days hence. In this case with a continuous distribution, there is exactly a 5% probability that the actual observed return will fall below the fifth percentile (Sullivan, Stoumbos & Brooks 2007, 324). In this thesis the mnemonic α is used to describe the probability of VaR being exceeded, i.e. $\alpha = 1 - X$. Figure 3 below illustrates these parameters involved in determining the VaR for a portfolio. In this case the return distribution of the portfolio is assumed to be normal, hence the Gaussian shape of the distribution. *X* once again denotes the coverage probability. One thing clearly pointed out by figure 3 is that VaR only measures possible losses, not gains. Hence, when a long (short) position is assumed we concentrate on the left (right) tail of the distribution.

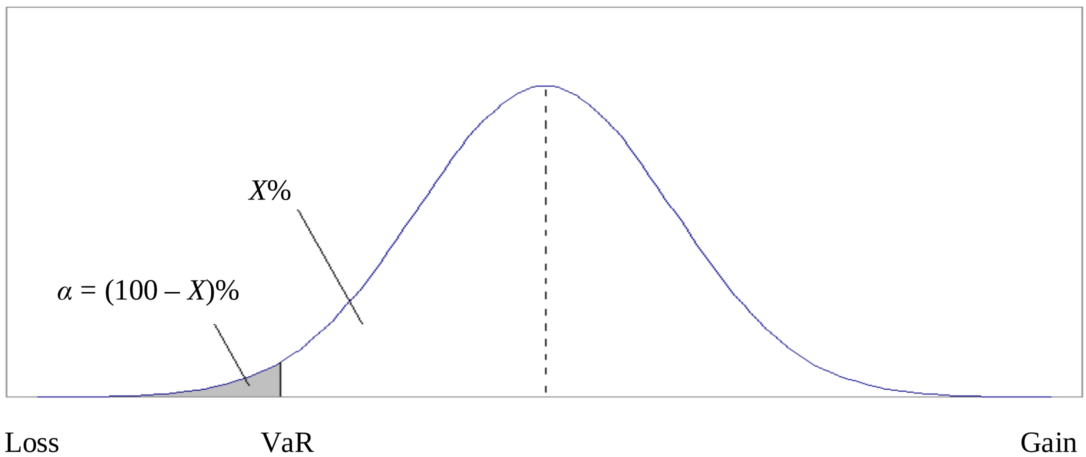


Figure 3 Simple illustration of value at risk with normal distribution

A practical example of these parameters involved in VaR calculation is given by the Amendment of 1996 by the Basel Committee (1996b), which dictates VaR of banks' trading books to be calculated using a 10-day holding period (N=10) and 99% confidence level (X=99). This means that it focuses on the possible loss over a 10-day period that is expected to be exceeded only 1% of the time under normal market conditions. The capital that Basel Committee requires banks to hold is this VaR measure times k (with an adjustment for what are termed as specific risks). This multiplicative factor k ranges from 3 to 4, depending on the performance of the bank's model in so called backtests. Furthermore, the Basel Committee does not specify a particular VaR model that should be used in market risk calculations. Hence, banks have the liberty of choosing a VaR model that best suits their needs.

Value at risk first received wide representation in July 1993 in *the Group of Thirty* (G-30) *report* (1993). The Group of Thirty is a consultative group of top bankers, financiers and academics from leading industrial nations. This landmark report recommends guidelines for managing derivatives and in particular it advices to value positions using market prices and to asses financial risks with VaR. Other financial and regulatory communities in addition have recognized the potential in VaR, and set the scene for its dispersal. For instance, the Bank for International Settlements *Fisher report* (1994) urged financial intermediaries to disclose measures of value at risk publicly. The Derivatives Policy Group, affiliated with six large U.S. securities firms, has also advocated the use of VaR models in their *Framework for Voluntary Oversight* (1995) as an important way to measure market risk. The introduction of the *RiskMetrics* database compiled by J.P. Morgan in 1994 is perhaps the most notable private sector initiative to pave the way for the popularization of VaR. Essentially, RiskMetrics represents an elaborate variance-covariance matrix of risk and correlation measures that evolve through time. To produce their own VaR measures, users need computer software to integrate the RiskMetrics database with their own asset positions.

However, perhaps the most significant breakthrough for VaR in the banking sector occurred when it was proposed by the Basel Committee on Banking Supervision (1995) to be used as a method to measure banks' market risks when determining their regulatory minimum capital requirements. This Basel Committee's amendment saw daylight in 1996 and it was the first time banks were allowed to use their own risk measurement models to determine their capital charge. This decision stemmed from the recognition that many banks had developed sophisticated in-house risk management systems based on VaR that were in many cases far more complex than regulators could ever dictate.

Furthermore, the European Union (2006) has passed legislation, the Capital Requirements Directive (CRD), comprising Directive 2006/48/EC and Directive 2006/49/EC, that will require the regulatory capital for securities firms to be calculated in a similar manner that for banks. Another initiative by the European Union, Solvency II, is likely to lead to the capital for insurance companies in Europe being calculated in a broadly similar way to that for banks. VaR is therefore becoming, and with hastening speed, the industry standard for measuring risk in the financial sector.

1.3 Previous research on value at risk methods

Since the first appearance of value at risk its performance has been under serious examination in a myriad of academic studies. The performance of VaR methodology per se has not been under scrutiny as the emphasis of these studies has been on the relative performance of various VaR approaches (VC, HS, MC etc.). Generally these studies differ in terms of tested VaR models, time span and composition of data, and implemented comparison methods. Next, a general characterisation of these studies and their results is given to offer perspectives on possible areas that still may need further studying in this field. This presentation is not by any means exhaustive due to the vast number of papers covering this topic.

In the late 1990s and early 2000s, when VaR still was a relatively new concept, the vast majority of these examinations on VaR models' performance studied different variations of the variance-covariance approach, sometimes accompanied by a simple historical simulation method (see e.g. Hendricks 1996; Jackson & Perraudin 1998; Wong, Cheng & Wong 2003). Using different variations of the variance-covariance method basically means that different models – such as EWMA, ARCH or GARCH – are used in estimating the volatility parameter (standard deviation) for the return distribution, in attempt to compensate for the much too fast decay of the tails in the Gaussian distribution. Therefore, comparing different approaches of the variance-covariance method usually boils down to volatility estimation. As a general conclusion from these studies it can be said that the variance-covariance approaches that incorporate time varying volatility, especially ARCH and GARCH, consistently fail to meet the Basel backtesting criteria, and are outperformed by a simple historical simulation method. However, previous studies on volatility forecasting, such as Baillie and Bollerslev (1992), Brooks (1998) and Taylor (1999), have concluded that GARCH-based models tend to have better performance in forecasting volatility than other time series models. In light of this fact, the poor performance of GARCH-based variance-covariance VaR models might appear somewhat surprising. The reasoning behind this poor performance is that these VaR models lead to satisfactory quantile estimates once a disaster has already hit the system. This means that GARCH-type models are suitable in signalling the continuation of a high risk regime, but what we are actually after is a model that predicts the occurrence of extreme events.

Eventually, the simple HS method was also extended to incorporate various volatility models (see Hull & White 1998 and Boudoukh, Richardson & Whitelaw 1998) and the VC approach no longer was the only VaR approach with the ability to handle time-varying volatility. Empirical comparisons showed that the performance of this extension to historical simulation was superior to simple historical simulation (and thus also in many cases superior to VC as well). Therefore, for quite some time the available VaR methods for practitioners were VC (with different distributions and volatility models), HS (with or without volatility updating) and MC. The MC approach per se has not gained a lot of interest in academic studies as it produces similar results to the VC method when linear positions are used.

The popularity of the variance-covariance VaR approach among practitioners and researchers stems mainly from its intuitive appeal and simplicity. However, most conventional variance-covariance approach models and some nonparametric distributions, mainly the historically simulated distribution, failed in capturing some rare events that took place in emerging financial markets over the last decade. This inadequacy has led researchers to model directly the tail behaviour of the return distribution parametrically rather than the whole distribution. This extreme value theory approach to VaR has been under serious research recently and its performance against other VaR approaches has also been tested. For instance, Longin (2000) and Bekiros and Georgoutsos (2005) report results according to which at very high confidence levels the EVT-based methodology produces more accurate forecasts of extreme losses than VC and HS methods. Furthermore, the results indicate that the VC and HS methods under- and over-predict losses, respectively.

Recently, papers that rather conclusively compare the performance of various VaR approaches and extensions have emerged. Kalyvas and Sfetsos (2006), for example, compare the performance of VC, HS, HS with volatility updating, and EVT VaR approaches. Adopting the Basel Committee's backtesting scheme they conclude that the EVT model performs superiorly to the other models in this particular case as it produces the least amount of violations. Similarly Bao, Lee and Saltoglu (2006) comprehensively compare the performance of 16 different VaR models and conclude that some of the models based on extreme value theory perform well under turbulent periods, whereas the more conventional approaches (VC, HS and MC) are unable to adapt to this unusual market situation. However, they find that most VaR models generally appear to behave similarly during tranquil periods.

1.4 Objective of the thesis

As described above, during a short span of time numerous papers have studied various aspects of VaR methodology. The recent research in this field has progressed so rapidly that comparing the relative predictive performance of different VaR models has not yet been fully matched. This thesis contributes to this particular problem by comparing the predictive performance of various traditional and novel VaR models from different aspects, providing valuable information for risk practitoners and regulators.

The variance-covariance and historical simulation approaches are currently the most widely adopted methods for calculating VaR (see e.g. Rogachev 2007). The popularity of the RiskMetrics approach of J.P. Morgan (1996), originally introduced in 1994, probably constitutes a great deal of the popularity of the variance-covariance approach as it falls under this particular category. The number of Monte Carlo adopters can be expected to rise as computers become more powerful and less expensive. None of the banks in Rogachev's (2007) survey reported using models based on the extreme value theory, which is probably due to the method's novelty and difficulties in harnessing it for practical use. Nevertheless, as the EVT approach has showed promising results in previous studies it is a natural object of further examination.

To offer a comprehensive picture of the predictive ability of various value at risk models this thesis studies several VaR approaches and their extensions. As the variance-covariance approach with normal distribution and EWMA volatility, namely the RiskMetrics 1994 approach (see J.P. Morgan 1996), is probably the most widely adopted VaR model it is chosen to serve as a benchmark for the other models. Additional VaR models, which are compared against this benchmark, fall under the categories of VC, HS, MC and EVT. Besides considering the standard forms of these approaches, extensions and variations to the standard methods, presented in previous literature, are considered. Essentially these variations entail filtering the return series in question in different ways and checking if this enhances the model's performance. A nearly endless number of different variations to these models could be achieved through modelling the conditional variance σ_t^2 with a variety of volatility models. In light of the fact that VaR essentially is a quantile of some distribution, various volatility models are not included in this thesis, and instead it focuses on the return distribution. González-Rivera, Lee and Mishra (2004) and Hansen and Lunde (2005) found that some simple volatility models often perform as well as more complex models which offers support to this choice. Furthermore, according to Alexander (2001, 64) the majority of practitioners do not use GARCH models at all, and if they do it is in a relatively basic form. Therefore, only a simple GARCH(1,1) model and EWMA for the RiskMetrics approach are considered for conditional variance σ_t^2 in this thesis. Table 1 below summarizes the approaches, denotations and distributions implemented in this thesis.

Table 1 Value at risk models and denotations used in this thesis

Approach	Denotation	Distribution used in this thesis
Variance-covariance	VC	Normal distribution
Historical simulation	HS	Historical distribution
Monte Carlo simulation	MC	Monte Carlo distribution
Extreme value theory (EVT)	GPD	Generalized Pareto distribution

Notes:
(1) This table defines the denotations and the distributions used in this thesis.
(2) In this thesis the implemented extreme value theory approach is based on the generalized Pareto distribution and, thus, this particular model is referred to as GPD. When referring more generally to the extreme value theory the denotation EVT is implemented (compare e.g. tables 1 and 4).
(3) These approaches are all calculated with a range of variations that are presented in latter parts of the thesis.

These approaches and their variations are presented in more detail in latter parts of the thesis. As can be seen in table 1 the generalized Pareto distribution is chosen to represent the extreme value theory approach in this thesis. Alternative EVT distributions (and approaches) also exist but the generalized Patreto distribution can be regarded as the most modern EVT approach and it is therefore applied. As mentioned above, extreme value theory approaches are scarcely adopted by financial institutions and therefore the comparison of this particular approach with the approaches currently adopted by financial institutions has an interesting feature: if the extreme value theory approach performs superiorly to the other approaches then practical adopters of VaR models should seriously consider updating the theoretical backgrounds of their current models (VC, HS and MC).

The ultimate objective of the thesis is to evaluate the predictive ability of these VaR approaches and their variations during an out-of-sample period in order to gain insights into which models are most appropriate for financial institutions to use as their in-house models. This evaluation of models is based on their accuracy, size of the estimates and bunching of exceptions. Accuracy is evaluated via empirical coverage probability, for which confidence limits are generated by the likelihood ratio test statistic for unconditional coverage (Kupiec 1995). Size of the VaR estimatess is evaluated by the predictive quantile loss function of Koenker and Basset (1978). Finally, bunching is evaluated by observing the quantile loss plots produced by this quantile loss function. Eventually, these tests will enable for ranking the models into preferential order. Preferred models can be found within all the models and within approaches, to see, for example, if variations within a certain approach are advantageous. The suitability of the different approaches to real-life surroundings is also given some consideration as introduction to each of these approaches is given.

1.5 Data

One of the objectives of this thesis is to compare the predictive ability of VaR models on before, during and after market crisis periods. This is the main driver for the choice of data. The impacts of market crises are most pronounced and simultaneous for stock indices, making them a natural source of data for this particular examination. The bursting of the so called technology bubble in 2000 is chosen to represent market crisis in this thesis and, therefore, the examined market indices are chosen in a way that they all were impacted by this crisis relatively simultaneously but with different intensity, allowing for the division of the time series into distinct periods reflecting prevailing market conditions. Following these criteria the different approaches and possible variations to VaR modelling are applied to the stock markets of four different economies. Namely, the used stock indices are S&P 500 Composite (United States), CDAX General 'Kurs' (Germany), HEX General (Finland), and India BSE National (India), all retrieved from the Thomson Financial database. The S&P 500 and CDAX General are chosen to represent *developed stock markets*, whereas the OMX Helsinki and BSE National represent *emerging markets*. This choice grants the opportunity of observing the performance of different VaR models in different market contexts. To be more precise, it is expected that especially the emerging market time series do not behave normally and exhibit some degree of skewness and kurtosis. That is, their probability distributions might not be symmetric and exhibit more extreme realisations than expected according to the normal distribution.

The data comprises daily closing prices of these indices starting from 1.1.1988. As mentioned the performance of the models is evaluated during three two-year out-of-sample periods: before the bursting of the so called technology bubble in 2000 (representing normal market conditions), during the market crisis period caused by this speculative bubble bursting (representing market crisis period), and after the markets had settled from this shock (representing normal market conditions). Thus, insights into the performance of the tested VaR models under different market conditions can be gained. Though two of these periods represent normal market conditions VaR models are expected to perform divergently during these particular periods. This is due to the fact that the predictions of VaR models in the after crisis period are to some extent affected by the crisis period, whereas the before crisis period was preceded by a less turbulent period, presumably leading to more conservative VaR predictions (see appendices 1 and 2 for the market indices' price and return series). The three evaluation periods are presented in table 2 below.

Table 2 Three evaluation periods

	Period 1 (before crisis)	Period 2 (during crisis)	Period 3 (after crisis)
Out-of-sample period	1.1.1997−31.12.1998	1.1.2000−31.12.2001	1.1.2003−31.12.2004
(sample lenght)	P ≈ 504	P ≈ 504	P ≈ 504
In-sample period	1.1.1988−31.12.1996	1.1.1988−31.12.1999	1.1.1988−31.12.2002
(sample length)	R ≈ 2268	R ≈ 3024	R ≈ 3780

There are marginal differences in the lenths of these out-of-sample and in-sample periods for each index due to e.g. national holidays and data availability. Hence, the numbers presented in table 2 are not exact. The primary reason for the data starting as early as in 1.1.1988 is due to the fact that the EVT approach requires a relatively long time series to produce satisfactory quantile estimates. The sample for each index consists of total T observations and it is split into an in-sample part of size R and an out-of-sample part (evaluation period) of size P so that $T = R + P$. For these samples a rolling window scheme is implemented. That is, the $(t - R)$th one-step-ahead VaR prediction is based on observations $t - R$ through $t - 1$, where $t = R + 1, \ldots, T$.

1.6 Methodology and methods

The established accounting research methodologies can be classified under *conceptual*, *nomothetical*, *decision-oriented* and *action-oriented approaches* (Kasanen, Lukka & Siitonen 1993). This thesis is empirical and descriptive by nature as conclusions are deducted from empirical results, thus implying that this thesis follows the nomothetical reseach approach.

The predictions of each VaR model under scrutiny are generated assuming a long position in each of the indices with two of the most common confidence levels implemented by practitioners in market risk VaR calculation, i.e. $X = 99\%$ and $X = 95\%$, which translates to $\alpha = 0{,}01$ and $\alpha = 0{,}05$, respectively. The predictive performance of different VaR models is compared against actual outcomes of the market variables and competing models' results. Testing how well the VaR model would have performed in the past is referred to as backtesting. The models in this thesis are backtested during three out-of-sample periods (i.e. before, during and after market crisis) using three different evaluation criteria: empirical coverage probability, predictive quantile loss and bunching of violations. The test statistic on the statistical significance of the empirical coverage probability is based on Kupiec (1995) and the predictive quantile loss function is based on Koenker and Basset (1978). Finally, bunching of violations is examined by observing the plots produced by the predictive quantile loss function separately for each VaR model and stock index.

1.7 Structure of the thesis

The remainder of this thesis is structured as follows. Section 2 provides the reader with the appropriate tools to understand the methods and concepts used in latter parts of the thesis when actual VaR estimates are generated. Section 2 begins by contemplating the concept of risk and presents how risks are typically classified. After this, the statistical concepts included in measuring risk and time-varying volatility are presented. A discussion on the statistical behaviour of actual asset returns concludes section 2.

Section 3 opens with a more detailed description of the VaR measure, its parameters and additional features included in VaR modelling. After this, the most common VaR approaches (VC, HS, MC and EVT) and their advantages and disadvantages are introduced. A description of the possible extensions to the standard versions of these approaches is also given. The approaches and variations that are actually calculated in the empirical part of this thesis are discussed in more detail. Section 3 concludes with a comparison of different VaR approaches' assumptions and abilities. Section 4 introduces the comparison methods implemented in this thesis and section 5 continues with presentation of the data and research results. Conclusions are given at the end of section 5. Section 6 summarizes the thesis.

2 RISKS AND THE ART OF MEASUREMENT

2.1 Nature of risk

2.1.1 Concept of risk and financial risks

If one were to ask people on their concepts of risk, one presumably would receive as many varying answers as the number of people interviewed. The term risk is typically associated with the volatility of unexpected outcomes, and especially with the negative outcomes, that is losses. This one sided view of risk only as a possibility of losses is rather common among executives, as can be seen, for example, from the interviews carried out by Kasanen, Lundström, Puttonen and Veijola (1996, 55). However, in financial theory risk is not only understood as the possibility of losses, but as the possibility of gains as well. Therefore, the results of exposing to risk can be either negative or positive. The key point of risk is the present uncertainty about outcomes in the future. When talking about financial risks, this volatility of unexpected outcomes usually refers to the volatility of the value of assets or liabilities of interest (Jorion 2007, 3).

Truth of the matter is that when doing business there are risks everywhere: we cannot anticipate with certainty which turmoil affects our calculations, or what price we will get from our products, and so forth. Examples of factors that cause risk in everyday business are natural hazards, strikes, interest rate movements, competitors' investments and exchange rate movements (Kasanen et. al. 1996). These various types of risks that face firms can be broadly classified into *business* and *nonbusiness* risks. Business risks are those which the firm willingly assumes to create a competitive advantage and add value for shareholders. Exposure to business risk in different ways creates the core competency of all business activity. Other risks, over which a firm has no control, can be classified as nonbusiness risks. These include strategic risks which result from fundamental shifts in the economy or political environment. These nonbusiness risks are difficult to hedge, except by diversifying across business lines and countries (Jorion 2007, 3–4).

Finally, *financial risks* can be defined as those which relate to possible losses in financial markets. Financial risks are traditionally classified into four categories: (1) *market risk*, (2) *credit risk*, (3) *operational risk* and (4) *liquidity risk*. In some applications a fifth class of financial risks is also considered: (5) *model risk*. Market risk (or price risk) is the risk of a change in the market price. Credit risk originates from the fact that counterparties may be unable or unwilling to fulfil their contractual obligations.

A change in the credit rating of a bond issuer can be also viewed as credit risk. Operational risks are in general defined as risks arising from human and technical errors or accidents. These include frauds, inadequate controls, problems in back-office operations and so forth. Liquidity risk refers to not being able to sell or buy the necessary amount due to lack of buyers or sellers. Last, model risk arises when risks are measured and priced using a flawed mathematical model (Kahra 2007; Jorion 2007, 22–27; Knüpfer & Puttonen 2004, 183–185).

However, this is just a one way of classifying financial risks, and hence some of the categories presented here might be missing or some additional categories might be presented in other context. A comprehensive and detailed description of different types of financial risks is given in Basel Committee's (1994) *Risk Management Guidelines for Derivatives*. Table 3 below depicts the above presented categorization of financial risks. In table 3 market risks are further divided into (1.1) *interest rate risk*, (1.2) *exchange rate risk*, (1.3) *equity risk*, and (1.4) *commodity risk*. These market risks are discussed in more detail in the next subsection.

Table 3 Financial risks and market risks

Financial risks	Some causes of financial risks
(1) Market risk	Market prices change in unexpected ways
(1.1) Interest rate risk	Interest rates change
(1.2) Exchange rate risk	Exchange rates change
(1.3) Equity risk	Equity prices change
(1.4) Commodity risk	Commodity prices change
(2) Credit risk	Change in credit rating or counterparty defaults
(3) Operational risk	Human errors and accidents
(4) Liquidity risk	Markets do not function perfectly
(5) Model risk	Used models and assumptions are not perfect

2.1.2 Market risks and volatility in market variables

Market risk arises from movements in the level of volatility of market prices. Generally, there are four types of financial market risks: interest rate risk, exchange rate risk, equity risk, and commodity risk, as presented in table 3 above. This classification of market risks simply represents the classification of market variables into their respective asset classes. Losses that can be attributed to market risk can occur through a combination of two factors: the *volatility* in the underlying financial variable and the *exposure* to this source of risk. Whereas firms have no control over the volatility of financial variables, they can alter their exposure to these risks, for instance, trough derivatives. (Jorion 2007, 22–23; 82–85; Hull 2007, 18–19).

Volatility in the market variables, that is market risk, is measured by the standard deviation of unexpected outcomes, or sigma (σ), also called volatility. This has been the case since Markowitz's first published his pioneering article *Portfolio Selection* (1952), which was one of the first attempts to understand the trade-off between risk and expected return. Later Sharpe (1964) and others carried the Markowitz's analysis a stage further by developing what is known as the capital asset pricing model. This is a relationship between expected return and what is termed systematic risk. In 1976 Ross developed arbitrage pricing theory – an extension of the capital asset pricing model to the situation where there are several sources of systematic risk. The key insights of these researchers have had a profound effect on the way market risks are analyzed and quantified, and ultimately all these extensions entail measures of volatility in some form.

Measurements of linear exposure to movements in underlying risk variables on the other hand appear everywhere under different disguises, depending on the market risk category in question. In the fixed income market, exposure to movements in interest rates is called *duration*, whereas in the stock market this exposure to market movements is called *systematic risk*, or beta (β). In derivatives markets the exposure to movements in the value of the underlying asset is called delta (Δ). Nonlinear exposure is measured by higher order derivatives, such as convexity and gamma (Γ) in the fixed income and derivatives markets, respectively (Jorion 2007, 76–77). The emphasis of this thesis is on measuring volatility rather than exposure, and therefore measurements of exposure are not covered extensively.

Section 1 argued that the new emphasis on risk management was partly spurred by increased volatility in the financial markets. Figures 1 and 2 plotted the price movements of some chosen financial variables. These graphs, however, described movements in the level of these financial variables, and therefore give only an indirect view of risk. Risk can be measured more precisely by short-term volatility, which in turn is plotted in Figures 4 and 5, for the same variables as in Figures 1 and 2. Figure 6 in turn plots the short-term volatility for the S&P 500 Composite index from June 1974 to October 2007. Figures 4, 5 and 6 demonstrate the standard deviation of trailing 12-month logarithmic price changes, calculated from monthly data and expressed in percent per annum.

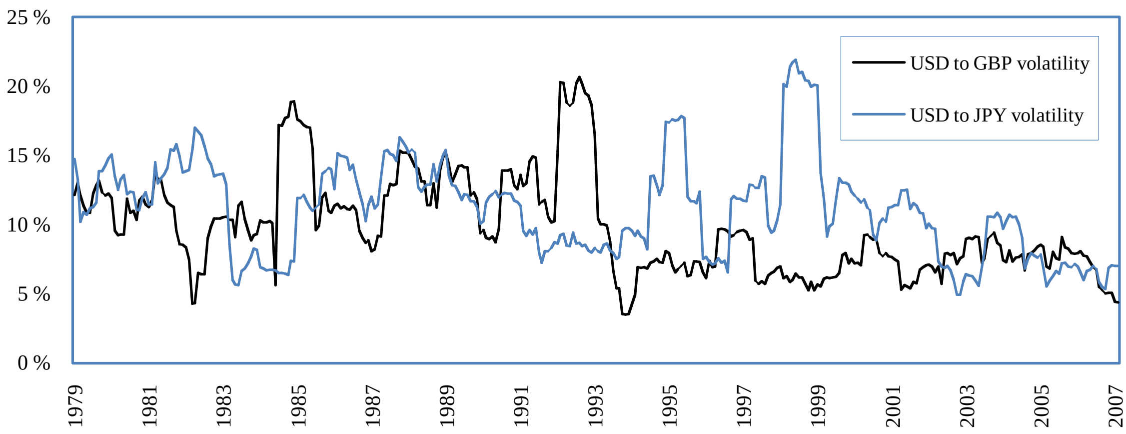


Figure 4 Volatility in exchange rates (1979-2007)

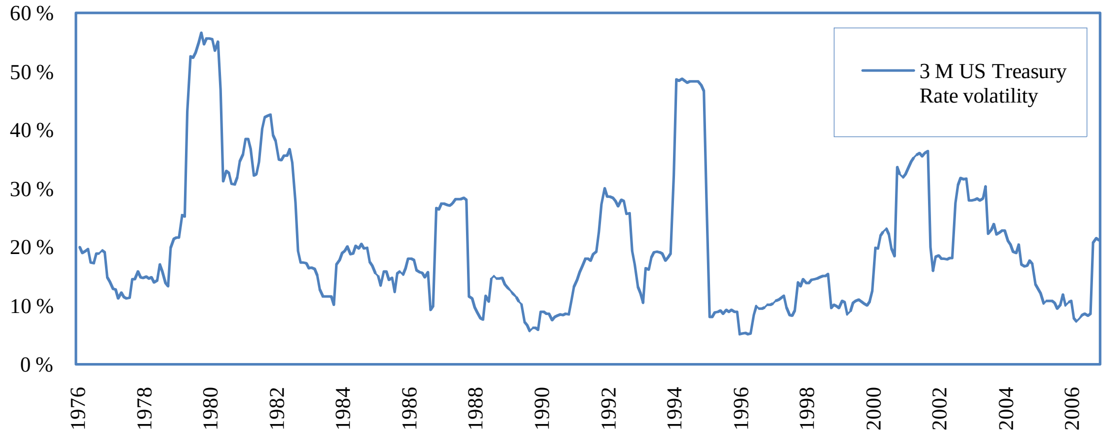


Figure 5 Volatility in 3 month U.S. Treasury rate (1976-2007)

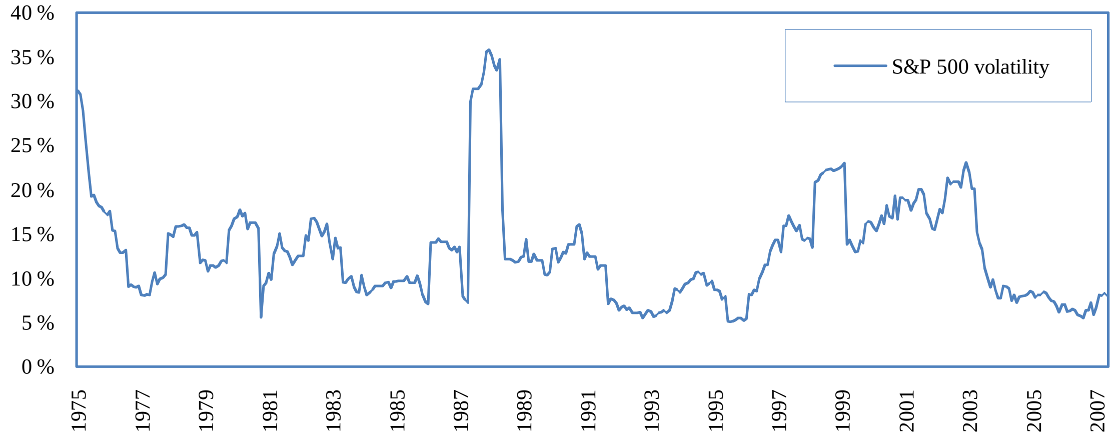


Figure 6 Volatility in S&P 500 Composite index (1975-2007)

In Figure 4 an elevated season of volatility can be observed for USD to GBP exchange rate, starting from mid 1980s and lasting until mid 1990s. For the USD to JPY exchange rate shorter periods of elevated volatility can be observed. The volatility of these exchange rates, ranging from five to twenty percent, is still rather low if compared with the volatility of 3 months U.S. Treasury rate (Figure 5), which ranges up to a maximum of 56,8 percent. Figure 5 also confirms that the volatility of the 3 month U.S. Treasury rate increased sharply after 1979. After this sudden increase the volatility also seems to remain in the elevated level for quite some time. The measure of risk seems to fluctuate over time in a way that periods of elevated volatility seem to be followed by periods of elevated volatility and periods of low volatility seem to be followed by periods of low volatility. This begs the question if risk truly is unstable over time, or whether these patterns are due to the estimation method used and just reflect “noise” in the data. This important question concerning the true nature of volatility is further discussed in upcoming parts of the thesis.

Finally, Figure 6 measures risk in the U.S. stock market from June 1974 to October 2007. Notable peaks in volatility can be observed in October 1974, when U.S. stocks went up by 17 percent after three large consecutive drops, and during the October 1987 crash, when U.S. equities lost 20 percent of their value. Therefore volatility occurs because of large unexpected price changes, whether positive or negative. This symmetrical treatment is logical because participants in these markets can be long or short, domestic or foreign, consumers or producers. Overall the volatility of financial markets creates risks and opportunities that must be measured and controlled properly (Jorion 2007, 79). The next subsection tackles the issue of measuring risk and return from observed market prices.

2.2 Measuring risk and return

2.2.1 Probability distribution function and its parameters

Value at risk, and its variance-covariance approach to be more specific, is at the end of the financial theory continuum that originates from Markowitz’s portfolio theory. Therefore, the basic concepts on which VaR is build up on were laid out as early as 1952 by Markowitz in his seminal work *Portfolio Selection*. Since then, volatility has been associated with risk and determining the level of volatility has become an important part of portfolio and risk management. This subsection introduces the basics of probability theory, which are needed to study random returns – and ultimately value at risk.

As mentioned above, risk can generally be defined as the uncertainty of outcomes. This is best measured in terms of probability distribution functions. Therefore this subsection begins by illustrating how probability distributions are created. Suppose x is a random quantity that can take on any one of a finite number of specific values, say, x_i, where $i = 1, 2,\ldots, n$. Assume further that associated with each possible x_i, there is a probability p_i that represents the relative chance of an occurrence of x_i. The p_i's satisfy $\sum p_i = 1$ and $p_i \geq 0$ for each i. Each p_i can be thought of as the relative frequency with which x_i would occur if an experiment of observing x were repeated infinitely. The quantity x, characterized in this way before its value is known, is called a random variable (Luenberger 1998, 141; Jorion 2007, 80).

It is common to display the probabilities associated with a random variable graphically as a density. The possible values of x are indicated on the horizontal axis, and the height of the line at a point represents the probability of that point. An example is shown in Figure 7a below. Figure 7a illustrates the density corresponding to the outcome of a roll of a dice, where six possibilities all have a probability of 1/6. If the outcome variable can take any real value in an interval as, for example, the rate of return on an investment, a probability density function $f(x)$ describes the probability. The probability that the variable's value will lie in any segment of the line is equal to the area of the vertical region bounded by this segment and the density function (Luenberger 1998, 141–142). Figure 7b below depicts an example of this case.

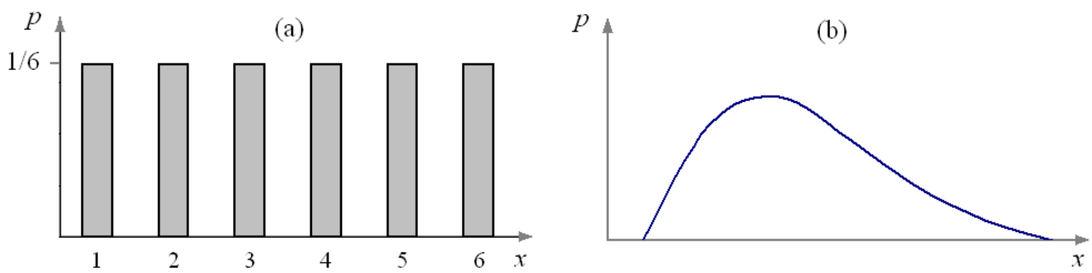


Figure 7 Some examples of probability distributions

The probability distribution can usefully be characterized by two central variables: its mean and its spread. If an investor would be immune to risk, he or she would only concentrate on maximizing the *expected value* $E(x)$ of the portfolio, often termed as *mean value*. The expected value of a random variable x is just the average value obtained by regarding the probabilities of unexpected outcomes as frequencies. In other words, it is the weighted sum of all possible values of x, each weighted by its probability of occurrence. In the case of a finite number of possibilities, expected value is defined as

$$E(x) = \mu = \sum_{i=1}^{n} p_i \cdot x_i \tag{2.1}$$

where p_i is the probability of x_i occurring. For convenience E(x) is often denoted by μ (Luenberger 1998, 142; Jorion 2007, 81). In the case of a continuous variable, such as the rate of return on an investment or the variable in Figure 7b, that has the probability density function $f(x)$, expected value is defined as (Jorion 2007, 83)

$$E(x) = \mu = \int_{-\infty}^{\infty} x \cdot f(x)dx \tag{2.2}$$

The expected value provides a useful summary of the probabilistic nature of the random variable. However, typically one wants, in addition, to have a measure of the possible dispersion of outcomes around the mean. This is in general measured by the *variance*, defined as the weighted sum of squared deviations around the mean. In general, for any random variable x, the variance of x for discrete and continuous random variables are defined respectively as

$$\mathrm{var}(x) = \sigma^2 = E\left[(x-\mu)^2\right] = \sum_{i=1}^{n} p_i (x_i - \mu)^2 \tag{2.3}$$

$$\mathrm{var}(x) = \sigma^2 = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)dx \tag{2.4}$$

In mathematical expressions, variance is represented by symbol σ^2, thus σ^2 = var(x). Variance is measured in units of x squared and thus is not directly comparable with mean. Therefore, the square root of variance is frequently used, denoted by σ and referred to as *standard deviation*. Hence, we can derive σ easily from σ^2 as $\sigma = \sqrt{\sigma^2}$. In financial context the term *volatility* often refers to standard deviation, which is also the case in this thesis. Volatility has the same units as the mean and therefore is a useful measure of how much the variable is likely to deviate from its expected value (Luenberger 1998, 143; Jorion 2007, 81–83).

Mean and variance are the first two central moments, which describe the location and spread of the distribution. For completeness, two other central moments should also be introduced. *Skewness*, which describes departures from symmetry, and *kurtosis*, which describes the degree of flatness of a distribution, can be defined respectively as

$$\gamma = \left[\int_{-\infty}^{\infty} (x-\mu)^3 f(x)dx\right] / \sigma^3 \tag{2.5}$$

$$\delta = \left[\int_{-\infty}^{\infty} (x - \mu)^4 f(x) dx \right] / \sigma^4 \tag{2.6}$$

Negative skewness indicates that the distribution has long left tail and hence entails large negative values. A kurtosis coefficient greater than 3 indicates that the tails decay less quickly than for the normal distribution. This would imply a greater likelihood of large values, positive or negative. For normal distribution, which can be fully described by its first two moments, skewness is 0 and kurtosis 3. Hence, the degree of flatness of a distribution is often also measured by *excess kurtosis*, which is defined as $\delta - 3$. These measures can be used as quick check on whether the sample distribution is close to normal (Jorion 2007, 86–87).

It should also be mentioned here, that in the case of two or more random variables, their mutual linear dependency can be measured by *covariance*. Estimating the covariance matrix of a portfolio is at the heart of producing VaR estimates when portfolios include several assets. However, this particular thesis concentrates on evaluating different VaR methods' performance, and to do this it is not necessary to estimate covariance matrixes. This is due to the fact that the VaR methods' performance can be evaluated calculating the VaR figures using a portfolio consisting only of a single asset. Introducing more assets to the underlying portfolio would not make the results derived from this thesis any more robust – it would only make the calculations heavier. It is worth noting, however, that in the upcoming sections, when estimation of variance is discussed the introduced concepts can easily be extended to cover covariance as well.

2.2.2 Estimating moments from data

When market risk is measured, to reach an estimate for market risk VaR and for other purposes as well, the *rate of return on a financial asset* is the random variable that we are interested in. It can be described by its probability distribution function, as explained above. Rates of return, however, are not directly observable from the market. Therefore we need to transform the price data into return data. Usually price data are recorded on daily frequency but databases for higher frequency data are also available as electronic storing of data has developed. Between the available high frequency databases it might be worth mentioning the Olsen and Associates database comprising all the bid and ask quotes of the foreign exchange market collected from information vendors since 1986 and the Trade and Quote (TAQ) database, which comprises all the trades and quotes related to all the securities listed in the New York Stock Exchange.

The transformation of price data into return data can be done as follows. Define *T–t* (*T*>*t*) as the time horizon, which can be of any length, and *S* as the price of a security. In this case the *arithmetic,* or *discrete,* and the *geometric,* or *continuously compounded,* rate of return can be defined respectively as

$$r_{t,T} = \frac{(S_T - S_t)}{S_t} \tag{2.7}$$

$$r_{t,T} = \ln\left(\frac{S_T}{S_t}\right) = \ln(S_T) - \ln(S_t) \tag{2.8}$$

if it is assumed that there are no interim payments, such as dividends, during time *T–t*. The use of geometric returns, sometimes also referred to as *logarithmic returns*, instead of arithmetic returns has two advantages. First, geometric returns may be economically more meaningful, since if they are assumed to be normally distributed, then the prices are lognormally distributed which leads to nonnegative prices. This is not the case if arithmetic returns were used. The second advantage of using geometric returns is that they easily allow extensions into multiple periods. For instance, the geometric return $r_{1,3}$ from period 1 to 3 can be decomposed as

$$r_{1,3} = \ln\left(\frac{S_3}{S_1}\right) = \ln\left(\frac{S_3}{S_2}\right) + \ln\left(\frac{S_2}{S_1}\right) = r_{2,3} + r_{1,2} \tag{2.9}$$

Still, in many situations the differences between the two returns are minimal when price changes are relatively small. However, significant differences may occur in markets with large moves, such as emerging markets, or when the time horizon is long, such as years (Jorion 200, 93–95; Luenberger 1998, 301; Hull 2007, 240–241). This thesis implements geometric (logarithmic) returns in its calculations.

In practice, the distribution of rates of return is usually estimated over a number of previous periods, such as days, months or years, assuming that all observations are *identically* and *independently distributed* (i.i.d.). If *n* is the number of observations, the first moment μ (mean), the second moment σ^2 (variance), the third moment γ (skewness), and the fourth moment δ (kurtosis) can be measured from return data respectively as (Luenberger 1998, 214–217; Tsay 2005, 11)

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} r_i \tag{2.10}$$

$$\hat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n} (r_i - \hat{\mu})^2 \tag{2.11}$$

$$\hat{\gamma} = \frac{1}{n-1}\sum_{i=1}^{n}(r_i - \hat{\mu})^3 / \hat{\sigma}^3 \tag{2.12}$$

$$\hat{\delta} = \frac{1}{n-1}\sum_{i=1}^{n}(r_i - \hat{\mu})^4 / \hat{\sigma}^4 \tag{2.13}$$

2.2.3 Estimating time-varying volatility

When looking more closely into the graphs presented in the beginning of this section (Figures 4, 5 and 6), it appears as risk changes in a predictable manner over time. This can be seen as volatility clusters, or groups, in the time series; high observations of volatility appear to be followed by high observations and low observations seem to be followed by low observations. This perception has important implications for risk management, since when we are calculating value at risk through parametric methods, we are most interested in the current level of volatility (and correlations as well). This is because we are estimating possible changes in the value of a portfolio over a short period of time. Hence, this subsection introduces some ways in which estimates of current level of volatility can be produced from historical data by different time series models that capture time variation in volatility. Correlations are also of great interest to risk managers when calculating VaR, but their modelling is not necessary for the research problem in question. Therefore, correlations are not discussed in this thesis per se. However, it is worth noting that the concepts on volatility modelling that are introduced here could easily be extended to cover correlations as well.

A widely employed method for estimating *unconditional volatility* is to use a moving window of fixed length and equally weight each of the observations in the window. The term unconditional refers to the fact that this estimation method assumes that the return distribution does not change over time. Hence, the variation in the volatility estimates produced by this method can only be attributed to sampling errors. For instance a typical length to be used is 20 trading days (about a calendar month) or 60 trading days (about three calendar months). This fixed size data window is then rolled through time, each day adding a new return and taking off the oldest return. Assuming that returns r_i over n days are observed, this volatility estimate is constructed from a *moving average* (MA) as

$$\sigma_t^2 = \frac{1}{n}\sum_{i=1}^{n} r_{t-i}^2 \tag{2.14}$$

The term *historic volatility* is also applied to statistical forecasts that are based on equally weighted moving averages (Alexander 2001, 49–52; Jorion 2007, 222). The focus here is on raw returns r_i rather than returns around the mean, $r_i - \mu$. This is so because for most financial time series, ignoring expected returns over very short time intervals makes little difference for volatility estimates. Hence, for analysing daily financial series it has become standard to base variances on squared returns (see e.g. Figlewski 1997; Alexander & Leigh 1997). This mean exclusion is further discussed in upcoming parts of the thesis, especially in the context of absolute and relative VaR.

This volatility model, while simple to implement, has some serious drawbacks. First, it ignores the dynamic ordering of observations: recent information receives the same weight as older observations in the window that may no longer be relevant. Further, the effect of a large return affects the MA estimate of volatility for the time that is chosen as the time window. Also, this approach leaves wholly unanswered the optimal choice of the length of moving window (Jorion 2007, 222–223).

Due to the deficiencies of the MA method, volatility estimation has moved towards *conditional volatility* models that put more weight on recent information. The first such model was the generalized *autoregressive conditional heteroskedastic* (GARCH) model suggested by Engle (1982) and Bollerslev (1986). Heteroskedastic refers to the fact that the underlying returns are not identically distributed over time, hence variance also changes as time passes by.

The GARCH model assumes that the variance of returns follows a predictable process. The conditional variance depends on the latest innovation, r_{t-1}, but also on the previous conditional variance, σ_{t-1}^2. Define σ_t^2 as the conditional variance, using information up to time t–1, and r_{t-1} as the previous day's return. The simplest such model is the GARCH(1,1) process, defined as

$$\sigma_t^2 = \gamma \cdot V + \alpha \cdot r_{t-1}^2 + \beta \cdot \sigma_{t-1}^2 \tag{2.15}$$

where V is the long-run average variance rate, γ, α and β are the weights assigned to V, r_{t-1}^2 and σ_{t-1}^2, respectively. γV is usually set as ω for the purposes of estimating the parameters. For a stable GARCH(1,1) process it is required that $\alpha + \beta < 1$. Otherwise the weight assigned to the long-term variance rate would be negative (Hull 2006, 465–466; Jorion 2007, 223–224).

The exponentially weighted moving average (EWMA) model is particular case of the MA model, where weights assigned to observations decrease exponentially as we move back trough time. It is also a particular case of the GARCH(1,1) model, where $\gamma = 0$, $\alpha = 1 - \lambda$, and $\beta = \lambda$. The formula for EWMA therefore is

$$\sigma_t^2 = \lambda \cdot \sigma_{t-1}^2 + (1 - \lambda) \cdot r_{t-1}^2 \tag{2.16}$$

where the parameter λ is called the *decay factor* and must be less than unity. This is the approach that J. P. Morgan's (1996) RiskMetrics uses for estimating volatilities and correlations. It uses decay factors of 0,94 and 0,97 for daily and monthly data, respectively (Jorion 2007, 230–232; Hull 2006, 463-465).[1]

If exponentially weighted moving averages are used correctly they can produce much more reasonable estimates of short-term volatility than equally weighted moving averages. In fact in some cases an exponentially weighted moving average model might even be preferred to a more sophisticated GARCH models. This might be the case particularly for very short-term forecasts. For longer-term forecasts, required for instance for pricing options, GARCH models are preferable due to their mean-reversal property, as EWMA models are not mean reverting (Alexander 2001, 49–116).

More complex volatility models – that account for additional factors, such as leverage – could be constructed. See e.g. Nelson (1991); Engle & Ng (1993); Zakoian (1994); and Sentana (1995). However, as mentioned earlier, this thesis concentrates on the return distribution and uses only these above mentioned simple volatility models to make one-step-ahead volatility forecasts. This choice is backed up by the findings of González-Rivera et al. (2004) and Hansen and Lunde (2005), who found that simple volatility models often perform as well as more complex ones, especially when making one-step-ahead predictions for relatively short time intervals. Furthermore, according to Alexander (2001, 64) the majority of practitioners do not use GARCH models at all, and if they do it is in a relatively basic form.

2.3 Statistical behaviour of actual asset returns

2.3.1 Pioneering research and theoretical background

The quantitative modelling of financial markets started in 1900, when a French mathematician Louis Bachelier wrote his pioneering Ph.D. Thesis with the title Théorie de la Spéculation (reprinted in Cootner 1964). In proposing the normal distribution as a model of price variations, Bachelier seemed to have been the first scientist who formulated a testable hypothesis on the statistical behaviour of financial data.

[1] In fact, RiskMetrics no longer is a part of J.P. Morgan and it now functions as an individual company. Furthermore, RiskMetrics Group has only recently upgraded the foundations of its VaR prinsiples and they no longer rely on EWMA for their volatility estimates. Their updated approach utilizes a volatility forecast derived from a long memory autoregressive conditional heteroskedastic (LM-ARCH) process, and a distribution that accurately captures fat tails for the residuals (see RiskMetrics Group 2006). This approach is beyond the scope of this thesis, and therefore when referring to the RiskMetrics approach this thesis particularly refers to the original RiskMetrics 1994 approach (see J.P. Morgan 1996).

Although it turned out later that the normal distribution can usually be overwhelmingly rejected by explicit statistical tests, due to its familiarity it is still often used in theoretical work as well as by market practitioners. Theoretically, the appeal of the Gaussian stems, of course, in finance like in other fields, from its simplicity and its stability under addition property: i.e. the Gaussian distribution is the limit distribution for sums of independent and identically distributed (i.i.d.) random variables, as proved by P. L. Laplace in *the central limit theorem* (Jorion 2007, 84). This property, in fact, comes into play quite naturally when dealing with financial prices; the most easily available data, price changes on a daily frequency, can be considered as the sums of a multitude of smaller increments on the intra-day level. If it is reasonable to assume that these high-frequency variations are i.i.d. random variables, then the central limit law should apply to price changes of a lower frequency and it, then, seems natural to postulate that their distribution approaches a Gaussian shape.

The economic reasoning for the i.i.d. assumption for price changes can be supported by the so-called *efficient market hypothesis* (EMH), originally formulated in the 1960s (Fama 1970). A market is said to be efficient if all the information is instantly processed when it reaches the market and it is immediately reflected in a new value of prices of the assets traded. The EMH was for a long time the only available theoretical background to the statistical behaviour of asset prices. It states that, at any point in time, asset prices should reflect the discounted expected stream of earnings from holding the underlying asset. In this case, price changes come about by arrival of some new items of information about future dividends which changes the rationally computed expectations of future prices and dividends. Of course, such information continuously hits the market; one can, for instance, think of such elementary items like weather or climatic events affecting output in certain industries, political changes as well as any kind of firm specific events. According to the EMH, the entirety of all these factors would explain the history of price changes. Therefore, the news arrival process governs the characteristics of the distribution of price increments. As news should be events that are independent of the preceding ones, the ‘independence’ part of the i.i.d. assumption can immediately be justified by the above considerations. If we are prepared to accept the second part as well (identical distributions), we end up with the central limit law for low-frequency price changes as the aggregate of high-frequency ones.

2.3.2 Mandelbrot’s stable distribution hypothesis and alternative models

During the 1960s and 1970s, the above picture underwent some changes: first, it has been found that price increments themselves are less suitable for statistical analysis than relative changes. One, therefore, routinely transforms the raw time series of prices into returns, as described in above sections, for statistical analysis. A second and more important modification was brought about by Mandelbrot’s (1963) and Fama’s (1965) finding that daily returns are decisively nonnormal. In fact, standard tests for normality usually reject their null hypothesis at extremely high levels of significance. Since then, most empirical research into the statistical properties of asset returns has found systematic deviations from normality, namely excess kurtosis and left-skewness (Simons 1996). For this reason, a number of alternative models have been proposed with the aim of explaining the empirical evidence of the ‘fat tails’ and the time fluctuations of the second moment, i.e. variance, of price changes.

Giving up normality, but maintaining stability under aggregation, Mandelbrot (1963) proposed the Lévy stable distribution as an alternative model. Under this new hypothesis on the distribution of returns, relative price changes would have no finite second and higher moments and prices would follow a Lévy stable process, which is a stochastic process obeying a generalized central limit theorem under time aggregation. Stability means that, the sum of two independent stochastic processes, both characterized by the same Lévy stable process, is itself a stochastic process characterized by a Lévy distribution. Besides their desirable stability property, the Lévy distributions are also in harmony with the usual visual appearance of the empirical distribution of returns: they posses more probability mass in their tails and center than the Gaussian (i.e. they are *leptokurtic*). Unfortunately, the stable Lévy distributions lack an analytical closed form solution and can only be characterized by their characteristic function. Furthermore, at the time of publication of Mandelbrot’s and Fama’s papers, almost no statistical method was known for estimating the parameters and testing goodness-of-fit of the Lévy distributions. Nevertheless, Mandelbrot’s argument appeared so convincing, that a number of researchers accepted the hypothesis on Lévy stable distributions without further testing. At the same time, Mandelbrot’s hypothesis was questioned by others on the basis cleverly designed tests for stability-under-addition itself (Lux & Ausloos 2002, 377).

For more than 25 years, these two strands of research coexisted in the literature and the issue of the appropriate distributional assumptions remained basically undecided. However, from a practical point of view, a number of alternative, more easily tractable statistical models have also been proposed, although they lack a strong theoretical foundation. Only a few of them are mentioned here emphasizing the broad spectrum of distributions that have been proposed: (1) a model where a geometric diffusive behaviour is superimposed to Poissonian jumps (Merton 1976; Friedman & Laibson 1989); (2) the Student’s t-distribution as a fat tailed distribution with finite higher moments (Blattberg & Gonedes 1974); (3) mixtures of normal distributions with different means and variances (Kon 1984; Tucker & Bond 1988); and (4) hyperbolic distributions (Eberlein & Keller 1995).

2.3.3 Influence of the statistical theory of extremes

This huge variety of distributional forms makes it hard to find universal laws in the statistical behaviour of financial prices. Starting in the early 1990s, a new type of analysis has appeared in empirical finance that, in fact, highlights typical features by abstracting from a specific distributional shape. This strand of literature took up a proposal by DuMouchel (1983) to concentrate on the behaviour of the tails instead of trying to fit the entire distribution. This approach is particularly of interest for VaR analysis, since modelling extreme events, i.e. the tails, is exactly what VaR is after. The theoretical background to this research program is provided by statistical extreme value theory (laid out, for example, in the following textbooks: Berlant, Teugels & Vynckier 1996; and Reiss & Thomas 1997).

The key result in extreme value theory was proved by Gnedenko in 1943, which concerns the properties of the tails of a wide range of different probability distributions. This classification of the behaviour in the distributions outer parts is derived from the classification of the extrema (minima or maxima). Namely, it can be shown that the limiting distribution of the extrema converges to the *generalized extreme value* (GEV) distribution of von Mises (1936) and Jenkinson (1955) which has the representation of

$$F_{\xi}(x) = \exp\left(-(1+\xi x/\beta)^{-1/\xi}\right) \tag{2.17}$$

where β is a *scale parameter* ($\beta > 0$) and ξ is a *shape parameter*. Further, the variable $\alpha = 1/\xi$ is referred to as the *tail index*. This limiting distribution (eq. 2.17) encompasses three types of limiting distributions of Gnedenko (1943): i.e. the Gumbel, Fréchet and Weibull families. The (left) tail of the distribution declines exponentially for the Gumbel family ($\xi \to 0$), by a power function for the Fréchet family ($\xi > 0$), and is finite for the Weibull family ($\xi < 0$). For risk management we are mainly interested in the Fréchet family that includes stable and Studen-t distributions. The Gumbel family consists of thin-tailed distributions such as the normal and the lognormal distributions.

A variety of estimators have been developed for the tail index α or ξ. Estimation of this quantity for the tail shape alone is favourable over an overall fit of the empirical distribution for a number of reasons: first, it gives an indication of the type of behaviour in the outer parts, therefore, may allow us to exclude a number of candidate processes from the outset. For example, an estimate of ξ significantly different from zero would simultaneously imply rejection of the normal, mixtures of normals as well as diffusion-jump processes, as they all lack the required behaviour of large returns. Second, in financial applications, one is often more interested in extreme realizations (crash risks) than in the exact shape of minor fluctuations. Furthermore, the choice of a particular overall model already restricts the results that can be found for the tail behaviour. For instance, the choice of the Lévy distributions always implies that tails are assumed to be extremely heavy and restricts the tail index to the region $\alpha < 2$ ($\xi > 0{,}5$) (Lux & Ausloos 2002, 380).

The majority of studies applied maximum likelihood estimators for the tail index α or ξ. Due to its simplicity, the Hill tail index estimator (Hill 1975) has become the standard work tool in most studies of tail behaviour of economic data. The Hill estimate is obtained by maximization of the likelihood of the relevant tail function conditional on the chosen size of the 'tail'. The results from a large body of research in this area has yielded astonishingly uniform behaviour for data from different markets in that one usually finds a tail index α (ξ) in the range of 2,5 (0,4) to about 4 (0,25). Examples include Jansen and de Vries (1991); Longin (1996); Loretan and Phillips (1994); Lux (1996); and Lux (1998). This conveniently allows for the rejection of the hypothesis of Lévy stable processes governing price changes, as they would require for $\alpha < 2$, as mentioned above. This has important implications for risk management, since the GEV has a property that the kth moment $E(x^k)$ of x is infinite for $k \geq \alpha$ (Hull 2007, 225). Therefore, the possibility of infinite variance – which would complicate the important task of risk estimation significantly – can be rejected. This conclusion on the finiteness of the second moment has a deep consequence on the stability of the return PDF. The finiteness of the second moment and the independence of successive returns imply that the central limit theorem asymptotically applies. Hence, the form expected for the return PDF must be Gaussian for very long time horizons. We then have two regions – at short time horizons we observe leptokurtic distributions whereas at long time horizons we expect a Gaussian distribution (Mantegna & Stanley 2002, 367). The application of the extreme value theory for VaR analysis is further discussed in section 3 of the thesis.

2.3.4 Stochastic processes governing price dynamics

Among the important areas of researchers dealing with financial and economic systems, one concerns the complete statistical characterization of the stochastic process of price changes of a financial asset. Several studies have been performed that focus on different aspects of the analyzed stochastic process, e.g. the shape of the distribution of price changes, as discussed in the section above. Therefore, the underlying research problem, when studying return distributions, is to model the stochastic process that prices follow and which essentially generates the return distribution. This is still an active area, and attempts are ongoing to develop the most satisfactory stochastic model describing all the features encountered in empirical analyses. The knowledge of the statistical properties of price dynamics in financial markets is fundamental, since it is necessary for any theoretical modelling aiming to obtain a rational price for a derivative product

(Hull 2006) and it is the starting point of any valuation of risk associated with financial position (Duffie & Pan 1997). In VaR analysis the stochastic process of price changes is the key element in performing credible Monte Carlo simulations. In spite of this importance, the modelling of such variable is still rather inconclusive.

Bachelier's original proposal of Gaussian distributed prices was soon replaced by a model in which stock prices are log-normally distributed, i.e. stock prices are performing a *geometric Brownian motion* (GBM). In a geometric Brownian motion, the differences of the logarithms of prices are Gaussian distributed. This model is known to provide only a first approximation of what is observed in real data, but it still remains, until today, the most widely assumed model for the stochastic process governing asset prices (Hull 2006, 270). Geometric Brownian motion is, for example, assumed by the most commonly used options pricing model, the Black-Scholes (or Black-Scholes-Merton) model (Black & Scholes 1973; Merton 1973). Geometric Brownian motion can be described in its continuous form for stock price S as

$$dS = \mu S dt + \sigma S dz \tag{2.18}$$

$$\frac{dS}{S} = \mu dt + \sigma dz \tag{2.19}$$

where the variable μ is the expected rate of return for the stock (i.e. the *drift* parameter), σ is its volatility, and dz is a Wiener process (i.e. increments dz are independent and normally distributed with mean zero and variance dt) (Alexander 2001, 21; Hull 2006, 265–271). The geometric Brownian motion produces a log-normal distribution for the prices, and a Gaussian distribution for the price changes. Hence, it cannot account for the leptokurtic (fat tailed) distributions observed in actual data. Theorists and practitioners are aware of this fact, and that the Black-Scholes model needs correction in its application, meaning that the problem of which stochastic process describes the changes in the logarithm of price changes in financial markets is still an open one. Nevertheless, in the absence of a better model for price dynamics, the geometric Brownian motion remains as the most popular, and it is also the one applied in this thesis, in its discrete time form.

A number of alternative models to the geometric Brownian motion have been proposed with the aim of explaining (1) the empirical evidence that the tails of measured distributions are fatter than expected for a geometric Brownian motion; and (2) the time fluctuations of the second moment of price changes. Geometric Brownian motion assumes that prices change continuously in a way that produces a lognormal distribution for the price at any future time. There are, though, many alternative processes that can be assumed. One possibility is to retain the property that the asset price changes continuously, but assume a process other than geometric Brownian motion, such as the Lévy process (Mandelbrot 1963). Another alternative is to overlay continuous asset price changes with jumps. Yet another alternative is to assume a process where all the asset price changes that take place are jumps. A model where stock prices change continuously is known as *diffusion model*, whereas a model where this process is overlaid with jumps is known as a *mixed-jump-diffusion model*. Further, a model where all stock price changes are jumps is known as a *pure jump model* (Hull 2006, 562). Examples of these models are the *constant elasticity variance* (CEV) model (Cox & Ross 1976), Merton's (1976) *mixed jump-diffusion* model, and the *variance-gamma* model (Madan, Carr & Chang 1998), respectively. Even though these suggested alternative models are (to some extent) able to produce the features observed in actual data, none of them has been convincing enough to entirely supersede geometric Brownian motion in practice.

2.4 Factors underlying value at risk

When calculating value at risk, first one needs to decide which risks one desires to measure with VaR. The most common risks measured with VaR are market risks, as is the case in this thesis. Market risk can be defined as risk arising from movements in the level of volatility of market prices, i.e. assets' prices change unpredictably. A traditional measure of market risk, since the influential work of Markowitz (1952), has been volatility. VaR, and its VC approach to be more precise, can be seen as an extension to only using volatility as a measure of market risk. To be more precise, it combines volatility with a chosen probability to arrive at a new measure for market risk, i.e. VaR. More generally, VaR represents a single point on the estimated return distribution, and volatility – or other parameters depending on the assumed distribution for that matter – can be used in estimating this particular point.

For the above given reason, in order to reach VaR estimates one needs an estimate of the current level of volatility (or other parameters) of one's portfolio (or of the assets in one's portfolio). To measure the level of volatility (parameters of distribution), in order to generate an estimate of the return distribution, one, first, needs to transform the directly observable price data into return data. Usually the transformation is executed in a way that produces logarithmic returns for observation. Logarithmic returns are generated by the differences of logarithms of market prices (see equation (2.8)). From this transformed data, volatility can be calculated using a moving average of fixed length (see equation (2.14)).

However, from several studies it appears that the second moment (volatility) fluctuates over time (see e.g. figures 4, 5 and 6). This fluctuation is generated by heteroskedastic return distributions, i.e. not identical return distributions. For this reason, econometric models, such as EWMA and GARCH, that are able to capture this time-varying volatility have been developed (see e.g. Engle 1982; Bollerslev 1986). One cannot stress enough the importance of correctly estimating the current level of volatility in VaR estimation when using the normality assumption.

The quantitative modeling of financial markets can be seen to have started in 1900 with Bachelier's seminal Ph.D. Thesis Theorié de la Spéculation. Since the 1960s, another distinctive feature of financial variables, besides the above mentioned time fluctuations in the second moment of price changes, has been identified. To be more precise, daily returns have been discovered to be decisively nonnormal. These observed deviations from normality have been excess kurtosis and negative skewness. These stylized features in financial data can be explained by two alternative viewpoints: (1) the underlying distribution in fact is normal, but time-varying volatility causes us to observe leptokurtic distributions at short time horizons, and (2) the underlying distribution is not normal and rather follows an alternative distribution. Both of these explanations might carry some truth in them, but profound answers are yet to be discovered.

A new type of analysis in empirical finance has emerged since the early 1990s that concentrates on the behavior of the tails of the return distribution instead of trying to fit the entire distribution. This extreme value theory approach has been successful in explaining the tail behavior of a multitude of financial data. Further, results from this strand of research have made it possible to exclude a number of candidate distributions proposed as alternatives to the normal distribution. It has, for example, in many cases allowed for the rejection of using Lévy distributions as proposed by Mandelbrot (1963). As the EVT appears theoretically robust method for modeling the tails of financial return distributions, it opens a new interesting avenue from which to approach value at risk.

3 VALUE AT RISK APPROACHES

3.1 Value at risk measure

3.1.1 Universal definition of value at risk and related terminology

The basic intuition behind value at risk (VaR) was laid out in section 1.2 of the thesis. Now, this section turns to a more formal definition of VaR. As explained earlier, VaR summarizes conveniently in a single, easy to understand number the downside risk of an institution due to financial market variables. To be more precise, VaR gives an estimate of the expected maximum loss over a target horizon within a given confidence interval (Jorion 2007, 105–106). Alternatively, from the viewpoint of a regulatory committee, VaR can be defined as the minimal loss under extraordinary market circumstances (Tsay 2005, 288). Both definitions will lead to the same VaR measure, even though the concepts appear somewhat different. Value at risk is mainly concerned with market risk, as is the case with this particular thesis, but the concept is applicable to credit and operational risk as well (see e.g. Basel Committee on Banking Supervision 2005). Value at risk figures can be obtained through various approaches to modelling the return distribution – variance-covariance, historical simulation, Monte Carlo simulation, and extreme value theory approaches being the most popular ones. A detailed description of each of these approaches will be given in upcoming parts of the thesis.

In what follows, VaR is defined under a probabilistic framework. This definition applies universally for any distribution, discrete or continuous, fat- or thin-tailed, and it is not affected by the choice of VaR approach. Suppose that at the time index t we are interested in the risk of a financial position for the next period. Let $\Delta S(t, t + 1)$ be the change in the value of the assets in the financial position from time t to $t + 1$. This quantity is measured in currency units and is a random variable at the time index t. Denote the cumulative distribution function (CDF) of $\Delta S(t, t + 1)$ by $F_t(x)$. In this case VaR for a long position over the time horizon $(t, t + 1)$ with probability α can be defined as (Tsay 2005, 288; Jorion 2007, 109)

$$\alpha = \Pr[\Delta S(t,t+1) \leq VaR] = F_t(VaR) \tag{3.1}$$

Since the holder of a long financial position suffers a loss when $\Delta S(t, t + 1) < 0$, the VaR defined in equation (3.1) typically assumes a negative value when α is small. The negative sign signifies a loss. From the definition, the probability that the holder would encounter a loss greater or equal to VaR over the time horizon $(t, t + 1)$ is α. Hence, 1 –

α here signifies the confidence level X used in determining VaR (see Figure 3). Alternatively, VaR can be interpreted as follows. With probability X, the potential loss encountered by the holder of the financial position over the time horizon (t, t + 1) is less than or equal to VaR.

The previous definition shows that VaR is concerned with tail behaviour of the cumulative distribution function (CDF) $F_t(x)$. For a long position, the left tail of $F_t(x)$ is important. Yet a short position focuses on the right tail of $F_t(x)$. However, the definition of VaR in equation (3.1) continues to apply to a short position if one uses the distribution of $-\Delta S(t, t + 1)$. Therefore, it suffices to discuss methods of VaR calculation using a long position (Tsay 2005, 289).

For any univariate CDF $F_t(x)$ and probability α, such that $\alpha \in$ (0, 1), the quantity $q_t(\alpha) = \inf\{x | F_t(x) \geq \alpha\}$ is called the αth quantile of $F_t(x)$, where inf denotes the smallest real number satisfying $F_t(x) \geq \alpha$. If the CDF $F_t(x)$ of equation (3.1) is known, then VaR is simply its αth quantile (i.e. $\text{VaR}_t = q_t(\alpha)$). The CDF is unknown in practice, however. The different methods for calculating VaR are essentially concerned with estimation of the CDF and its quantile, especially the tail behaviour of the CDF.

Alternatively, in equation (3.1), if $\Delta S(t, t + 1)$ is divided by the asset price S_t (or considering the differences of logarithms of prices) a percentage form for value at risk is reached. Then equation (3.1) transforms to

$$\alpha = \Pr[r_t \leq VaR\%] = F_t(VaR\%) \tag{3.2}$$

and daily returns r_t of financial asset prices can be used directly. In this theoretical context, VaR in its percentage form is more informative than the currency unit form, and therefore this thesis implements the definition in equation (3.2) for VaR. Henceforward, VaR refers to the percentage form of VaR, as defined by equation (3.2). Now suppose that r_t follows the stochastic process (see section 2 for discussion on stochastic prcesses)

$$r_t = \mu_t + \varepsilon_t = \mu_t + \sigma_t z_t \tag{3.3}$$

where $\mu_t = \mathrm{E}(r_t)$, $\sigma_t^2 = \mathrm{E}(\varepsilon_t^2)$ and z_t has the conditional distribution function $G_t(z)$. In this case VaR can be estimated by inverting the distribution function

$$VaR_t(\alpha) = q_t(\alpha) = F_t^{-1}(\alpha) = \mu_t + \sigma_t G_t^{-1}(\alpha) \tag{3.4}$$

Hence, a VaR model involves the specification of $F_t(\cdot)$, or μ_t, σ_t^2, $G_t(\cdot)$ (Bao et al. 2006). In practical applications, calculation of VaR involves several factors (Tsay 2005, 289): (1) the probability of interest α, such as α = 0,01 or α = 0,05; (2) the time horizon, which might be set by a regulatory committee, such as 1 day or 10 days; (3) the frequency of the data, which may not be the same as the time horizon; and (4) the CDF $F_t(x)$ or its quantiles. The next subsection will consider the first two of these factors, (1) and (2). After that time aggregation in value at risk calculations is discussed (3). Finally, different methods for estimating the CDF give rise to different approaches to VaR calculation (4). These different approaches are further considered in their respective upcoming subsections.

The terminology and classifications related to these approaches is rather inconclusive and many variations coexist in literature. Table 4 below offers some clarification to this terminology. The first mentioned terms are the ones applied in this thesis, and the ones inside the parentheses represent their alternatives. The denotations for the approaches are also included in this table.

Table 4 Terminology referring to different VaR approaches

Terminology referring to different VaR approaches	Denotation
Variance-covariance (delta-normal, normal, parametric)	VC
Historical simulation (nonparametric, quantile estimation, empirical density)	HS
Monte Carlo simulation (stochastic simulation)	MC
Extreme value theory (depends on the applied EVT distribution and method)	EVT

Notes:
(1) In this thesis the implemented extreme value theory approach is based on the generalized Pareto distribution and, thus, this particular model is referred to as GPD. When referring more generally to the extreme value theory the denotation EVT is implemented (compare e.g. tables 1 and 4).
(2) In some cases the EVT approach is considered as an extension to the HS approach
(3) The term conventional VaR method refers to simple VC, HS, and MC methods.
(4) VC and MC methods are sometimes collectively referred to as model-building approach

3.1.2 Uses of value at risk and choice of factors

Two central factors are to be considered when calculating VaR: the length of the *holding horizon* and the *confidence level* (Hull 2007, 202–206). In general, VaR will increase with either a longer holding horizon or a greater confidence level. VaR can potentially be used for three alternative purposes: as a *benchmark measure*, as a *potential loss measure*, or in *defining equity capital* (Jorion 2007, 116–119). The choice of these factors should depend on the use of the VaR number. The first, most general, use of VaR is simply to provide a companywide yardstick to compare risks across different markets. For this purpose, the choice of the confidence level and horizon does not matter much as long as consistency is maintained throughout divisions and over time.

Another application of VaR is to give a broad idea of the worst loss an institution can incur. If this is the case, the horizon should be determined by the nature of the portfolio. A first interpretation is that the horizon is defined by the *liquidation period*. This refers to the time it would take to liquidate the entire portfolio. For this reason commercial banks currently report their trading VaR over a daily horizon because of the liquidity and rapid turnover of their portfolios. On the contrary, investment portfolios, such as pension funds generally invest in less liquid assets and adjust their risk exposures only slowly, which is why a one month horizon is generally chosen for investment purposes. A related interpretation is that the horizon represents the time required to hedge market risks. An opposite view to the one presented above is that the horizon corresponds to the period over which the portfolio remains relatively constant. Since VaR assumes that the portfolio's composition does not change during the horizon, this measure gradually loses significance as horizon extends. However, perhaps the main reason for banks to use a daily VaR is that this is consistent with their *daily profit and loss* (P&L) *measures*. This allows an easy comparison between the daily VaR and the subsequent P&L number. Still, for this application, the choice of the confidence level is rather arbitrary. It should be recognized that VaR does not describe the worst possible loss but is rather a probabilistic measure that should be exceeded with some frequency (Jorion 2007, 116).

The third possible use of VaR is for setting up a capital cushion for the institution. This being the case the choice of the factors is crucial. If the VaR figure is used directly for this purpose, a loss exceeding the VaR would wipe out the entire equity capital, leading to bankruptcy. For this application of VaR the choice of the confidence level should reflect the degree of risk aversion of the company and the costs of a loss exceeding VaR, and at the same time, the choice of the time horizon should correspond to the time required for corrective actions as losses start to develop (Jorion 2007, 117). To illustrate, assume that an institution determines its risk profile by targeting a particular credit rating. The expected default rate can then be converted directly into a confidence level. Higher credit ratings should lead to lead to higher VaR confidence levels. For instance, according to Moody's default rate estimates a Baa rated company has a 0,20% change of defaulting within the next year. To maintain Baa rating, an institution should carry enough capital to cover its annual VaR at the 99,80 percent confidence level (Hull 2007, 258).

It should also be mentioned here that the choice of the considered factors is important for backtesting considerations as well. Model backtesting involves systematic comparisons of VaR estimates with the subsequently realized returns in an attempt to detect biases in the reported VaR figures (Hull 2007, 208) (backtesting is further described in section 4 of the thesis). Longer horizons reduce the number of independent observations and thus the power of the tests. For example, using a one day horizon translates into having 252 observations per year, but in the case of a two week horizon the number of observations reduces to 26. Likewise, the choice of the confidence level should be such that it leads to powerful tests. For instance, a 95 percent confidence level should lead to losses worst than the VaR figure in 1 day out of 20. On the other hand, choosing a 99 percent confidence level we would have to wait, on average, 100 days to confirm that the model conforms to reality. Hence, for backtesting purposes, the confidence level and the holding period should be kept within reasonable limits (Jorion 2007, 119–120).

3.1.3 Time aggregation

Computing VaR requires the definition of a holding period over which to measure unfavourable outcomes, as discussed above. This period may be set in terms of hours, days, or weeks. To be able to compare risk across different time horizons, a translation method is needed. This problem is known as *time aggregation* in econometrics (Jorion 2007, 97).

Suppose daily data, which is the most easily available alternative in most cases, are observed, from which VaR figure is obtained. Using higher frequency (daily) data is generally more efficient since it uses more available information. The investment horizon, however, may be 3 months, two weeks as set out by the Basel Committee for market risk (Basel Committee on Banking Supervision 1995), or something else depending on the application of VaR. Thus, the distribution for daily data must be transformed into a distribution over a longer time horizon. If returns are uncorrelated over time, in other words follow a Martingale process, this transformation is straightforward.

The assumption of uncorrelated returns over successive periods is consistent with the efficient market hypothesis (Fama 1970), where the current price includes all the relevant information concerning the particular asset. Thus, all price changes must be due to news that, by definition, cannot be anticipated and therefore must be uncorrelated over time. Hence prices follow a Martingale process, which means that our best guess for the price of an instrument tomorrow, is its price today (Hull 2006, 594–595). Put mathematically in terms of returns: $E(r_{t+1}) = E(r_t)$. In addition, if returns are assumed to be identically distributed over time, both the requirements for independently and identically distributed (i.i.d.) variable are fulfilled. Under these assumptions the expected return over n days is n times the expected return over one day, and likewise for the variance. Both the expected return and the variance increase linearly with time. The volatility, in contrast, grows with the square root of time (Jorion 2007, 98–99). These results can conveniently be used in transforming the VaR for a time horizon of one day into a time horizon of longer periods. This is done by the formula: N-day VaR = 1-day

VaR x $\sqrt{N}$. This formula is exactly true when the changes in the value of the portfolio on successive days have independent identical normal distributions with zero mean (Hull 2007, 203). This time aggregation method is also accepted by the Basel Committee (1995) as it explicitly states that the ten-day 99% VaR can be calculated using the one-day VaR times $\sqrt{10}$. Hence, it suffices to concentrate only on the calculation of the one-day VaR for market risk in this thesis.

This assumption of i.i.d. returns is often made in risk management, but it may not be appropriate for some thinly traded markets. The worry is that when markets are trending, the risk over a longer period may be much higher than from a simple extrapolation of daily risk. This is due to autocorrelation between successive returns. In this case, the variance over *n* periods can be written as

$$\sigma^2\left(n + 2(n-1)\rho + 2(n-2)\rho^2 + ... + 2(1)\rho^{n-1}\right) \tag{3.5}$$

When markets are trending the correlation coefficient ρ is positive, which leads to higher variance than in the case of uncorrelated variable (i.i.d.). Hence, risk managers should be careful when assuming independent returns when aggregating VaR to different time horizons (Hull 2007, 204; Jorion 2007, 99–101).

3.1.4 Absolute and relative value at risk

It should be noted, that since the volatility grows with the square root of time and the mean linearly with time, the mean will dominate the volatility over long horizons. However, over short horizons, such as one day, volatility dominates. This provides a rationale for focusing on VaR measures (and volatility models) that ignore expected returns. Fixing the expected return at zero might seem an unconventional statistical procedure, but the estimation error associated with badly determined mean estimates in relatively small samples may reduce the efficiency of the estimated volatilities (Jackson & Perraudin 1998, 261).

When mean μ is fixed at zero and expected returns are normal the calculation of VaR is referred to as *absolute VaR*. Stated in other words, this assumption considers that our portfolio will not yield any returns. This approach makes perfect sense when dealing with highly liquid trading books that are revalued daily. There are many rationales behind this mean-zero framework. From a theoretical standpoint, the mean is computed by summing the logarithmic returns (i.e. the difference in logarithms of prices). Since the logarithmic returns are the differences in logarithms of prices and these returns are summed as they are sampled, they all cancel out, except for the first and last price, which is then divided by the number of samples minus one. From this, it is fairly obvious that in most cases dealing daily sampling will result in a rather small mean, compared with the volatility. From a modelling standpoint, this has various consequences: (1) the computation of volatility σ (volatility and correlations that are fed into the model are mean zero volatilities and correlations); (2) the simulation process (a relative diffusion, but with zero drift); and (3) the final VaR which does not take into account the portfolio mean returns. This produces a very neat and consistent model for analyzing risks over very short horizons, such as one day. However, if the investment horizon extends beyond ten days (two weeks) the use of *relative VaR* should be considered (Fan, Wei & Xu 2004).

Relative VaR – instead of assuming that the position's mean is zero – incorporates computation of the mean of the portfolio into volatility estimation and VaR calculation. Therefore, instead of assuming risk as pure volatility, it incorporates the expected returns into the equation. Usually, when discussing relative VaR, the emphasis is on the mean expected return of the portfolio from which scaled volatility is deducted in order to obtain VaR. However, mean inclusion affects all the same parts of the VaR computation as mean zero assumption (points (1) to (3) above). The most problematic part in using relative VaR is obtaining a reliable estimate for the expected return μ, as previously discussed. As absolute VaR makes perfect sense for traders who mark-to-market positions daily, relative VaR is ideally suited for individual investors, portfolio managers and corporations who rebalance positions weekly, monthly or quarterly (Fan et al. 2004).

With respect to this mean exclusion, this thesis follows the conventions set out by practical implementers of VaR models. That is, all the models are calculated in the absolute VaR framework as described above, except for the MC model's simulation process, where the drift parameter, i.e. mean, is estimated by $\hat{\mu}_t = \frac{1}{n-1}\sum_{i=1}^{n} r_i$. This means that all volatility estimates are mean-zero volatilities, i.e. μ is set at zero, and all the VaR estimates do not take into account mean returns.

3.1.5 Coherence of value at risk measure

Value at risk is certainly not the only measure besides volatility proposed for quantifying market risk. Early examples of alternative risk measures are e.g. Roy (1952) and Baumol (1963). More recently, a proposal on the desirable properties for risk measures for capital adequacy purposes (Artzner, Delbaen, Eber & Heath 1999) has spurred the development of a number of novel risk measures attempting to satisfy these properties. These four properties of any acceptable risk measure $\rho : X \rightarrow R$ are (Artzner et al. 1999):

(a) *Positive homogeneity*: $\rho(\lambda x) = \lambda\rho(x)$ for all random variables x and all positive real numbers λ.

(b) *Subadditivity*: $\rho(x + y) \leq \rho(x) + \rho(y)$ for all random variable x and y; It can be proved that any positively homogenous functional ρ, is convex if and only if it is subadditive.

(c) *Monotonicity*: $x \leq y$ implies $\rho(x) \leq \rho(y)$ for all random variables.

(d) *Transitional invariance*: $\rho(x + r_0) = \rho(x) - \alpha$ for all random variables x and real numbers α, and risk free rates r_0.

If ρ satisfies all these properties, then it is said that ρ is a coherent risk measure. These properties can be interpreted for a portfolio as follows. Positive homogeneity means that changing the size of a portfolio by factor λ while keeping the relative amounts of different items in the portfolio the same should result in the risk measure being multiplied by λ. Subadditivity requires that the risk measure for two portfolios after they have been merged should not be greater than the sum of the their risk measures before they were merged. Monotonicity property states that if a portfolio has lower returns than another portfolio in every state of the world, its risk measure should be greater. Finally, translation invariance means that if an amount of λ is added to a portfolio, its risk measure should go down by λ (Hull 2007, 199).

The homogeneity, monotonicity, and trnaslation invariance conditions are straightforward, given that the risk measure is the amount of cash needed to be added to the portfolio to make its risk acceptable. The subadditivity condition states that diversification helps reduce risks. If the opposite was true, in order to decrease risk, it could be convenient to split up a company into different distinct divisions. From the regulatory point of view this would allow to reduce capital requirements. Artzner et al. (1999) show that the quantile-based VaR measure does not always satisfy this subadditivity condition, and therefore cannot be regarded as coherent measure of risk. This notion has inspired a wealth of criticism towards VaR, and the development of other risk measures satisfying this and all the other conditions. For an introduction to these alternative coherent risk measures and criticism towards VaR see e.g. Granger (2002) and Szegö (2005).

An example of an alternative risk measure, satisfying all four conditions above and therefore referred to as coherent, is given by *conditional VaR* (or *expected shortfall*). Risk measures can generally be characterized by the weights they assign to quantiles of the loss distribution. Value at risk gives a 100% weighting to the αth quantile and zero to other quantiles. *Conditional VaR*, on the other hand, is determined as the expected loss conditional on exceeding the αth quantile (or VaR), when we are concentrating on the right tail of the distribution. It therefore gives equal weights to all quantiles greater than the αth quantile and zero weight to all quantiles below the αth quantile. Further, it can be shown that a risk measure is coherent if the weight assigned to the αth quantile of the loss distribution is a nondecreasing function of α. As expected shortfall satisfies this condition, it is a coherent measure of risk (see e.g. Hull 2007, 201).

Despite the presented criticism and alternative coherent risk measures, VaR has been able to hold its own. This is mainly because it has been chosen as an industry standard by the regulators, and the fact that the advantages of VaR supersede its disadvantages (Stahl 2005). Namely, the simplicity of VaR surpasses the usual arguments against VaR, i.e. financial markets' time series are not usually normally or log-normally distributed and exhibit fat tails. The modern argument against VaR is that it is not coherent, as it does not satisfy the subadditivity property. This, however, is the case only with portfolios that do not have *elliptic* probability distribution functions (Jorion 2007, 114). In these cases, VaR, of course, is not a proper risk measure and some coherent risk measure is required. Still, with most common portfolios, and return distributions, it can be said that VaR provides a simple and adequately accurate measure of risk.

3.2 Variance-covariance approach

3.2.1 Basic principles

The most simple and intuitive way to calculate value at risk is to assume that the return distribution belongs to a parametric family, such as the normal distribution, the Student-t distribution, the generalized error distribution (Nelson 1991), the exponential generalized beta distribution (Wang, Fawson, Barret & McDonald 2001), the stable Paretian distribution (Mittnik, Paolella & Rachev 2002) and the mixture of normal distributions (Venkatraman 1997). This assumption can be made if the dependence structure of the return series r_t can be fully described by the first two conditional moments (Bao et al. 2006, 102). Out of numerous parametric distributions, the normal distribution is by far the most commonly used, due to its simplicity and convenience. Hence, the term variance-covariance approach, or parametric approach, in most cases refers to using the normal distribution. When implementing this approach, the VaR figure can be derived directly from the portfolio standard deviation using a multiplicative factor that depends on the confidence level $X = 1 - \alpha$. Therefore $VaR_t(\alpha)$ is estimated by inverting the distribution function

$$VaR_t(\alpha) = \mu_t + \sigma_t G^{-1}(\alpha) \tag{3.6}$$

When normality is assumed, $G(\cdot)$ is the standard normal distribution function $\Phi(\cdot)$ so that $\Phi^{-1}(0{,}01) = -2{,}326$ and $\Phi^{-1}(0{,}05) = -1{,}645$ (Bao et al. 2006, 105). These figures can be obtained by turning to tables of the cumulative standard normal distribution function or using the inverse function of standard normal distribution found in statistical programs.

As it can be seen, this approach is solely based on the estimate of the variance (variance-covariance matrix) of asset returns, using historical time series to calculate their standard deviations (and correlations) (Simons 1996, 6). To illustrate, using a 99% confidence level (α = 0,01) and the normality assumption, VaR in equation (3.6) simply reduces to

$$VaR_t(0{,}01) = \mu_t - 2{,}326\sigma_t \tag{3.7}$$

and one only needs to estimate the parameters μ_t (mean) and σ_t (volatility). In addition, if the usual assumption of mean exclusion in risk management is made, the only required parameter is volatility. This approach is referred to as absolute VaR. Assuming other distributions than the normal for the price variations will entail different values for $G^{-1}(\alpha)$ and naturally for VaR as well. The main assumption of the parametric VaR is that the distributions of asset returns follow a certain distribution that can completely be described by its parameters. In the case of the normal distribution, this means that the variance-covariance matrix (and mean) completely describes the return distribution (Simons 1996, 7).

3.2.2 Extensions to variance-covariance approach

As can be seen above, by far the most significant advantage of the VC approach is its simplicity. However, this simplicity comes at a price. As discussed earlier, empirical studies on the statistical behaviour of asset returns have systematically found significant deviations from normality. This raises the question how realistic the normality assumption of the variance-covariance approach is and how much it distorts the estimated VaR figures. For some portfolios, that are large and well diversified, the fit is quite good, but certainly not for portfolios with heavy option components and exposures to a small number of financial risks (Jorion 2007, 111). This is the main disadvantage of the variance-covariance approach, i.e. that the normal distribution does not adequately approximate the actual distribution of profits and losses. In the case of asymmetrical return distributions – as is the case with option, loan and credit portfolios (see e.g. Szegö 2005, 7) – the approximation error due to nonlinearities can be reduced by using *delta-gamma*, *delta-gamma-delta* and the so-called *Cornish-Fisher expansion* methods (see e.g. Jorion 2007, 255–257; Hull 2007, 244–248). These extensions, however, exacerbate the computation of VaR noticeably and other approaches, such as historical simulation and Monte Carlo simulation, offer a more direct route to VaR. The most noticeable advantage of the simple variance-covariance method is its speed and simplicity, but with the above mentioned extensions this advantage is lost, and turning to other approaches makes sense.

The other reason for the inadequacy of the variance-covariance method with normal distribution – even if the underlying return distribution is sufficiently symmetrical – is that the tails of the normal distribution decay much too fast compared to what is observed in actual data, i.e. actual financial data universally seems to exhibit excess kurtosis. This much too fast decay of the tails of the normal distribution can be compensated for in two ways pertaining to the VC approach. First, another distribution, that possesses fatter tails than the normal distribution, can be assumed. Examples of these 'heavy tailed' distributions are the Student-t and the generalized error distribution (see e.g. Jorion 2007, 87–88). These alternative distributions, however, lack in strong theoretical foundation, as discussed in section 2.3.2. Hence, their use is questionable to some extent. Second, the conditional variance can be estimated by various volatility models, such as the family of GARCH models, to allow for the volatility to fluctuate over time in the VC approach (Bekiros et al. 2005, 227). This approach is based on the view that the distribution of returns, and therefore volatility, changes through time. As a result, in times of higher volatility, a stationary model could view large observations as outliers, when they are really drawn from a distribution that temporarily has a greater dispersion (Jorion 2007, 221). In practice, both of these explanations carry some truth in them.

As the focus of this thesis is on other approaches than the VC approach, some alternatives of this particular approach are still calculated to offer perspective. This thesis applies the VC approach with normal distribution and the volatility parameter estimated by equally weighted moving average MA, GARCH(1,1) and EWMA. The MA estimate is estimated as presented in equation (2.14), i.e. the mean is assumed to be zero. The EWMA volatility is estimated in a way that the VaR figures produced by this approach conform to the ones produced by the RiskMetrics 1994 approach (see J. P. Morgan 1996, 236). This RiskMetrics approach is used as a benchmark for all the other models in this thesis. A further extension to the VC approach, following the ideas of Hull and White (1998) and Barone-Adesi, Giannopoulos and Vosper (2002), referred to as 'filtering', is also implemented in this thesis. This means applying the standard VC approach to $y_t = r_t/\sigma_t$, the filtered return series using a time-varying volatility model (see e.g. Alexander 2001, 97). Time-varying volatility is estimated by GARCH(1,1) for the purposes of filtering r_t (see appendix 3 for the estimated GARCH(1,1) processes). Table 5 below lists these variations implemented in this thesis and their mnemonics.

Table 5 Variance-covariance VaR models and mnemonics

Distribution	Volatility	Return series	Mnemonic
Normal	EWMA(0,94)	r_t	RiskMetrics
Normal	MA(250)	r_t	VC
Normal	GARCH(1,1)	r_t	VC(GARCH)
Normal	GARCH(1,1)	y_t	VC*

Notes:

*(1) All the models in this thesis that are applied to the filtered return series $y_t = r_t/\sigma_t$ are denoted by asterisk *. Filtering the return series is based on the ideas of Hull and White (1998) and Barone-Adesi et al. (2002).*

(2) All these extensions to the VC approach are calculated in the mean-zero, i.e. absolute VaR, framework.

(3) MA(250) refers to the fact that MA is estimated using the past 250 days' observations, which is also the minimum length of data required by the Basel Committee (1996b).

(4) EWMA(0,94) refers to using a decay factor λ of 0,94 when estimating EWMA, which corresponds to the approach introduced by J.P. Morgan (1996).

3.3 Historical simulation approach

3.3.1 Basic principles

The simplicity and convenience of the normal distribution are powerful inducements for its use in VaR analysis, but this does not necessarily make its use appropriate for the reasons given in the previous section. Fortunately, it is possible to calculate VaR without resorting to the assumption of normality (or other distinct parametric distribution for that matter), by using a simple historical simulation method. The only assumption that this approach makes is that the empirical distribution continues to hold within the prediction period (Tsay 2005, 298; Simons 1996, 8).

This approach simply entails finding the lowest returns in the real historical data. Assuming that the distribution of return in the prediction period is the same as in the sample period, one can use the empirical quantile of the returns r_t to calculate VaR. Let $r_1, \ldots, r_n$ be the returns of a portfolio in the sample period. The *order statistics* of the sample are these observations arranged in increasing order. Here the notation

$$r_{(1)} \leq r_{(2)} \leq \cdots \leq r_{(n)}$$

is used to denote the arrangement and refer to $r_{(i)}$ as the ith order statistic of the sample. In particular, $r_{(1)}$ is the sample minimum and $r_{(n)}$ the sample maximum.

Assume that the returns are independent and identically distributed random variables that have a continuous distribution with probability density function (PDF) $f(x)$ and cumulative probability density function (CDF) $F(x)$. Then the following asymptotic result from the statistical literature (see e.g. Cox & Hinkley 1974) for the order statistic $r_{(l)}$, where $l = n\alpha$ with $0 < \alpha < 1$, can be used to determine VaR. Let $q(\alpha)$ be the αth quantile of $F(x)$, that is, $q(\alpha) = F^{-1}(\alpha)$. Assume that the PDF $f(x)$ is not zero at $q(\alpha)$. Then the order statistic $r_{(l)}$ is asymptotically normal with mean $q(\alpha)$ and variance $\alpha(1-\alpha)/[nf^2(q(\alpha))]$. That is,

$$r_{(l)} \sim N\left(q(\alpha), \frac{\alpha(1-\alpha)}{n[f(q(\alpha))]^2}\right), l = n\alpha \tag{3.8}$$

Based on this result, one can use $r_{(l)}$ to estimate the quantile $q(\alpha)$, where $l = n\alpha$. In practice, the probability of interest α may not satisfy that $n\alpha$ is a positive integer. In this case, one can use a simple interpolation to obtain quantile estimates. More specifically, for noninteger $n\alpha$, let l_1 and l_2 represent the two neighboring positive integers such that $l_1 < n\alpha < l_2$. Define $\alpha_i = l_i / n$, n representing the size of the sample. This yields $\alpha_1 < \alpha < \alpha_2$. Therefore the quantile $q(\alpha)$, i.e. VaR(α), can be estimated by (Tsay 2005, 299)

$$\hat{q}(\alpha) = \frac{\alpha_2 - \alpha}{\alpha_2 - \alpha_1} r_{(l_1)} + \frac{\alpha - \alpha_1}{\alpha_2 - \alpha_1} r_{(l_2)} \tag{3.9}$$

An extremely simple illustration of HS, using the past 100 daily observations and a 99% confidence level, would simply entail finding the lowest observation ($n\alpha$ = 0,01*100 = 1) to obtain next day's VaR. As 1 is an integer, there is no need for interpolation (eq. 3.9) in this case. A one-day time horizon and a 99% confidence level are those typically used for a market risk VaR calculation. 250 to 750 days of data is a popular choice for the number of days of data used, which is taken as a reasonable tradeoff between precision and nonstationarity (Hull 2007, 217; Jorion 2007, 265). As always, the choice of length of the sample period reflects a trade-off between using longer and shorter sample sizes. Longer intervals increase the accuracy of estimates but can use irrelevant data, thereby missing important changes in the underlying process (Jorion 2007, 263).

The HS method has a number of advantages over the VC method. First, it makes no explicit assumptions about the volatilities of returns and the covariances between them. Second, it makes no assumptions about the shape of the distributions themselves. In particular, it makes no assumption of normality (Simons 1996, 8). Furthermore, the HS method is relatively simple to implement if historical data have been collected in-house for daily marking-to-market. The same data can then be stored to be later reused in estimating VaR. The HS method also deals directly with the choice of horizon for measuring VaR: returns are simply measured over intervals that correspond to the length of the horizon. For instance, to obtain a monthly VaR, the user would reconstruct historical monthly portfolio returns over, say the last five years. The daily VaR can also be multiplied by the square root of time to obtain longer holding periods (Hull 2007, 220), but the first mentioned time aggregation method is theoretically more robust. Perhaps the most important advantage of HS is that it can account for fat tails and, because it does not rely on valuation models, it is not prone to model risk. The method is intuitive, robust and perhaps the most widely used method to compute VaR (Jorion 2001, 223).

On the other hand, the HS method has a number of drawbacks. First, it assumes that the distribution of the return r_t remains unchanged from the sample period to the prediction period. Given that VaR is mainly concerned with tail probability, this assumption implies that the predicted loss cannot be greater than that of the historical loss. This is definitely not the case in practice. Further, this assumption implies that the distribution remains stationary trough time, and therefore the simple HS method will miss situations with temporarily elevated volatility (Jorion 2007, 265). Second, for extreme quantiles (i.e. when α is close to zero or unity), the empirical quantiles are not efficient estimates of the theoretical quantiles (Tsay 2005, 300). In fact, the VC method provides much more efficient estimates in terms of sampling error than the HS method (see e.g. Jorion 2007, 124–128). This problem of inefficient estimates in conjunction with the HS method can, though, be alleviated by an econometric sampling method called bootstrapping, laid out for example in Tsay (2005).

The HS method also entails some practical problems in addition to the theoretical problems given above. First, it assumes that a sufficient history of price changes is available. Some assets may have short histories, or there may not even be a record of an asset's history (Jorion 2007, 265). The approach also leaves wholly unanswered the optimal choice of the moving window that is used as the sample period. As longer sample periods lead to more efficient quantile estimates – as can be seen for example from equation (3.8) – they might entail events that are no longer relevant or are unrepresentative of the immediate future. Further, when using long samples, the method is also unable to distinguish between high and low volatility periods and as a result it might generate inaccurate estimates (Bekiros et al. 2005, 227). Then again, short samples lead to more inefficient estimates and may omit important events, meaning that the tails will neither be well represented. A final drawback is that the method quickly becomes cumbersome for large portfolios with complicated structures (Jorion 2007, 265). Many of these problems related to the HS method can though be solved or alleviated by various extensions to the method. The next subsection considers these extensions.

3.3.2 Extensions to historical simulation approach

The basic historical simulation approach assumes that each day in the past is given equal weight. More formally, if n day-to-day changes are used, each of them receives a weight of $1/n$. Boudoukh, Richardson and Whitelaw (1998) suggest that more recent observations should be given more weight because they are more reflective of current volatilities and current macroeconomic variables. The natural weighting scheme to use is one where weights decline exponentially. Suppose we are now at the end of day n. The weight assigned to the change in the portfolio value between day $n - i$ and $n - i + 1$ is λ times that assigned to the change between day $n - i + 1$ and $n - i + 2$. In order for the weights to add up to 1, the weight given to the change between day $n - i$ and $n - i +$ 1 is

$$\frac{\lambda^{i-1}(1-\lambda)}{i-\lambda^{n}} \tag{3.10}$$

where n is the number of days. As λ approaches 1, this weighting scheme approaches to the basic HS approach, where all observations are given a weight of $1/n$. VaR is then computed by ranking the observations from the worst outcome to the best. Starting from the worst outcome, weights are summed until the required quantile of the distribution is reached. The best value for λ can be obtained by experimenting which value backtests best. An illustration of applying this hybrid (combination of the HS and RiskMetrics) method is given in Boudoukh et al. (1998). This extension solves the problems related to the optimal size of the sample period, since it is not really necessary to disregard old observations as we move forward in time because they are given very little weight. Empirical results in Boudoukh et al. (1998) show significant improvement in statistical performance of this hybrid approach to simple HS, and this improvement is most pronounced for fat-tailed data series.

The assumption of stationarity of the return distribution in HS can be relaxed by incorporating volatility updating into the HS approach. Barone-Adesi, Giannopoulos and Vosper (2002), and Hull and White (1998) suggest a way in which a volatility updating scheme, such as EWMA or GARCH(1,1) can be used in parallel with the HS approach. Suppose that the daily volatility for a particular market variable estimated at the end of day $i - 1$ is σ_i. This is an estimate of the daily volatility between the end of day $i - 1$ and the end of day i. Suppose it is now day n. The current estimate of the volatility of the market variable is σ_{n+1}. This applies to the time period between today and tomorrow, which is the time period over which VaR is being calculated.

Suppose that σ_{n+1} is twice σ_i. This means that we estimate the daily volatility of a particular market variable to be twice as great today as on day $i - 1$. This means that we expect to see changes between today and tomorrow that are twice as big as between day

i – 1 and i. When carrying out the historical simulation and creating a sample of what could happen between today and tomorrow based on what happened between day i – 1 and day i, it therefore makes sense to multiply the latter by 2. In general, when this approach is used, the value of a market variable S under the ith scenario becomes

$$S_n \frac{S_{i-1} + (S_i - S_{i-1})\sigma_{n+1} / \sigma_i}{S_{i-1}} \tag{3.11}$$

Each market variable is handled in the same way. This approach takes account of volatility changes in a natural and intuitive way and produces VaR estimates that incorporate more current information. The VaR estimates can be greater than any of the historical losses that would have occurred for our current portfolio on the days we consider. Hull and White (1998) produce evidence using exchange rates and stock indices to show that this approach is superior to traditional historical simulation and to the exponential weighting scheme of Boudoukh et al (1998). More complicated models could be developed where observations are adjusted for the latest information on correlations as well as for the latest information on volatilities (other models than EWMA or a simple GARCH). A comparison of the approach of Barone-Adesi et al. with alternative risk measurement methods is in Pritsker (2000).

The bootstrap method is yet another variation to the basic HS approach. It involves creating a set of changes in the portfolio value based on historical movements in market variables in the usual way. We then sample with replacement from these changes to create many new similar data sets. From these new data sets we then aggregate a new VaR figure, which usually contains less variance, and is therefore more accurate, than estimates produced by unbootstrapped HS methods (Tsay 2005, 192).

The implemented methods in this thesis for the HS approach are: the traditional approach to HS, using a moving window of 250 days; HS with GARCH(1,1) volatility incorporated by the method suggested by Hull and White (1998); and the traditional HS method implemented to the filtered return series $y_t = r_t / \sigma_t$ (see e.g. Jorion 2007, 265). No bootstrapping method is implemented in this thesis. Table 6 below lists these implemented variations to the HS method and their denotations.

Table 6 Historical simulation VaR models and mnemonics

Distribution	Volatility	Return series	Mnemonic
Historical distribution	N/A	r_t	HS
Historical distribution	GARCH(1,1)	r_t	HS(GARCH)
Historical distribution	GARCH(1,1)	y_t	HS*

Notes:

(1) The HS(GARCH) model uses the volatility updating scheme suggested by Hull and White (1998), whereas the HS model can be seen as a simplification of this method (see Jorion 2007, 265).*

(2) Common choises for the window length in simple HS are between 250 and 750 days.This Thesis applies a window of 250 days based on the observations of Brooks and Persand (2002, 91) and to align the HS approach with the VC approach in this thesis.

(3) No bootsrapping methods are implemented in these HS models.

(4) These HS methods are implemented in the mean-zero framework (i.e. absolute VaR).

3.4 Monte Carlo simulation approach

3.4.1 Basic principles

Monte Carlo (MC) analysis is a powerful method for calculating VaR. It can account for a wide range of exposures and risks, including nonlinear price risk and volatility risk. It is flexible enough to incorporate variation in volatility, fat tails, and extreme scenarios. Simulations generate the entire return distribution, not just one quantile, and can be used to examine, for instance, the expected loss beyond particular VaR. MC simulation can also incorporate the passage of time, which will create structural changes in the portfolio. This includes the time decay of options; the daily settlement of fixed, floating, or contractually specified cash flows; or the effect of prespecified trading or hedging strategies (Jorion 2007, 267).

The Monte Carlo simulation method for computing VaR is twofold. First, the stochastic process, and its parameters, that governs the dynamics of price changes of financial variables in question needs to be specified. Second, this process is used to generate the probability distribution for price changes ΔS, from which a certain quantile contingent upon the confidence level, i.e. VaR, can be estimated (Hull 2007, 248–249; Jorion 2007, 265–266).

Stochastic processes are discussed in more detail in the end of section 2 of the thesis. A popular, simple stochastic process is the geometric Brownian motion (see eq. 2.20), which is the one assumed for price changes in this thesis. The solution to this stochastic differential equation is $S_t = S_0 \exp\left(\left[\mu_t - \frac{1}{2}\sigma_t^2\right]t + \sigma_t z_t\right)$. See Brodie and Glasserman (1998). Thus simulating S_t amounts to simulating z_t. Since we are predicting one-step ahead VaR, this can be written as

$$S_t = S_{t-1} \exp\left(\left[\mu_t - \frac{1}{2}\sigma_t^2\right] + \sigma_t z_t\right) \tag{3.12}$$

where z_t is simulated from a standard normal distribution (i.e. normal distribution with mean zero and standard deviation of one). It is done N times, from which the empirical αth quantile for $r_t = \log(S_t / S_{t-1})$ is estimated. The MC method is thus similar to the HS method, except that the hypothetical changes in prices ΔS_i for a asset i are created by random draws from a prespesified stochastic process instead of sampled from historical data. In this thesis N is set at 5000, which is a sufficient number of simulations to produce satisfactory VaR estimates.

The biggest drawback of this approach is its computational time, because a large number of scenarios have to be generated. If 1000 sample paths are generated with a portfolio of 1000 assets, the total number of valuations amounts to 1 million. There are though proposed methods that limit the number of portfolio valuations, such as Jamshidian and Zhu (1996). Nevertheless, this method is most expensive to implement in terms of infrastructure and intellectual development (Jorion 2007, 267). Another evident weakness of the method is model risk. As MC relies on specific stochastic processes for the underlying risk factors as well as pricing models for instruments, such as options and mortgages, it is subject to the risk that these models are incomplete. Further, as historical correlations are used to maintain the multivariate properties of the risk factors when generating scenarios in MC. During market crises, when most correlations tend to increase rapidly, a MC system is likely to underestimate the possible losses (Barone-Adesi, Giannopoulos & Vosper 2002, 33).

3.4.2 Extensions to Monte Carlo simulation approach

When Monte Carlo simulation is used, there are ways of extending the approach so that market variables are no longer assumed to be normal. One possibility is to assume that the variables have a multivariate *t*-distribution, by simulating z_t in equation (3.12) from Student’s *t*-distribution. As this distribution entails more probability mass in its tails than the normal, this has the effect of giving higher value to the probability that extreme values for the variables occur, which ultimately leads to higher VaR estimates (Hull 2007, 249–250). Numerous alternative distributions (and stochastic processes) could also be assumed for the price changes but, as discussed in section 2 of the thesis, these alternatives lack pure theoretical foundations. Hence, their use is somewhat questionable even if they might appropriately fit in the data.

For this reason, no other process, than the geometric Brownian motion, is assumed for the price changes in this thesis. Yet, two alternative approaches to MC simulation using geometric Brownian motion (and hence making the normality assumption) are implemented in this thesis: for the first model the volatility parameter for simulation purposes is estimated in traditional methods by an MA(250) model, and for the second model volatility is estimated by GARCH(1,1). These VaR models are referred to as MC and MC(GARCH), respectively. Both of these models are calculated in the mean-zero (i.e. absolute VaR) framework, except for the simulation processes where μ_t is estimated by $\hat{\mu}_t = \frac{1}{n-1}\sum_{i=1}^{n} r_i$, where n is set at 500. For the simulation process 5000 sample paths are generated to produce stabile enough VaR forecasts.

3.5 Extreme value theory approach

3.5.1 Evolution of theory

All the previous methods estimate the quantiles using information from the whole distribution. Alternatively, since the quantiles at 1% or 5% are ‘extreme’ values for a distribution, we can focus on modelling the tails directly, which brings us to the extreme value theory (Bao et al. 2006, 107). The EVT-based methods for tail estimation are attractive because they rely on sound statistical theory that offers a parametric form for the tail of a distribution (Bekiros et al. 2005, 212).

The methods for applying the extreme value theory have undergone some changes along the years. Primarily, the EVT theory was developed by examining the limiting distribution of the sub-sample minima (or maxima). Sub-sample minima are achieved by dividing the whole sample into multiple non-overlapping sub-samples and finding their minimum values. Under the independent assumption, this limiting distribution of the normalized minimum is the *generalized extreme value* (GEV) *distribution* of Jenkinson (1955). It encompasses three types of limiting distributions of Gnedenko (1943); i.e. Gumbel, Frétchet, and Weibull families. For a more comprehensive treatment of this EVT method, see Embrechts, Klupperberg and Mikosch (1997). This method, referred to as Block Maxima (BM), however, encounters some difficulties when used for VaR calculation. First, the choice of subperiod length is not clearly defined. Second, the approach is unconditional and, hence, does not take into consideration effects of other explanatory variables (Tsay 2005, 318). To overcome these difficulties, a modern approach to EVT has been proposed in the statistical literature; see Davidson and Smith (1990) and Smith (1989). Instead of focusing on the extremes (minimum or maximum), the new approach focuses on exceedances of the measurement over some high threshold and the times at which the exceedances occur. Thus, this new approach is referred to as *peaks over threshold* (POT), and it is the EVT method applied in this thesis.

According to the POT method, we fix a low threshold u and look at all exceedances of r_i over this threshold. The choise between high and low threshold depends on which tail (right or left) of the distribution is of interest to us. To be consistent with the rest of the thesis, a low threshold is chosen (i.e. $u < 0$). Namely, supposing that $F(r)$ is the cumulative distribution for the variable r and u is a value of r in the left hand tail of the distribution. The probability that r lies between $u - y$ and u ($y > 0$) is $F(u) - F(u - y)$. The probability that r exceeds u is $F(u)$. Define $F_u(y)$ as the probability that r lies between $u - y$ and u conditional on $r < u$. This is

$$F_u(y) = \frac{F(u) - F(u-y)}{F(u)} \tag{3.13}$$

The variable $F_u(y)$ defines the left tail of the probability distribution. It is the cumulative probability distribution for the amount by which r exceeds u given that it exceeds u (Hull 2007, 224).

Balkema and de Haan (1974) and Pickands (1975) studied the asymptotic behaviour of these threshold exceedances and proved for a large class of the underlying distribution $F_u(y)$ that its limiting distribution, as the threshold is raised, is the Generalized Pareto distribution (GPD) which is given by

$$G_{\xi,\beta}(y) = 1 - \left(1 - \xi \frac{y}{\beta}\right)^{1/\xi} \tag{3.14}$$

where ξ is the *shape parameter* and β is the *scale parameter*. The shape parameter ξ defines the heaviness of the tail. These parameters can be estimated from data by maximum-likelihood methods or by estimators developed by Hill (1975) and Pickands (1975). (Maximum-likelihood methods are applied in this thesis.) Essentially all the common continuous distributions of statistics belong to this class of distributions. For instance, the case $\xi > 0$, known as hyperbolic decline, corresponds to heavy-tailed distributions, such as the Pareto, Student's, etc. The case $\xi \to 0$, known as exponential decline, corresponds to distributions like the normal or the lognormal whose tails decay exponentially. The short-tailed distributions with finite endpoints, such as the uniform beta, correspond to the case $\xi < 0$. For most financial data ξ is positive and in the range of 0,25 to 0,4 (Hull 2007, 224–225; Lux & Ausloos 2002, 379; Bekiros et al. 2005, 212).

3.5.2 Estimating value at risk via extreme value theory

This subsection now discusses how the above presented results can be used in estimating VaRs. Let r_i, $i = 1, \ldots, n_u$, be the sample of returns exceeding the threshold u, with its size being n_u. If those n_u exceedances are assumed to be identically and independently distributed (i.i.d.) with an exact GPD distribution then the maximum likelihood estimates of the GPD parameters ξ and β are consistent and asymptotically normal as $n_u \rightarrow \infty$ provided that $\xi > -1/2$ (Smith 1987). Using the maximum-likelihood estimators of these parameters, VaR with confidence level $X = 1 - \alpha$, can then be derived as (see appendix 2 for the entire derivation)

$$VaR(\alpha) = u - \frac{\hat{\beta}}{\hat{\xi}}\left[1 - \left(\frac{n_u}{n}\alpha\right)^{\hat{\xi}}\right] \tag{3.15}$$

As mentioned, this EVT method was developed under the i.i.d. assumption on series in question. The theory has, however, been extended to serially dependent observations provided that the dependence is weak. See Berman (1964) and Leadbetter, Lindgren and Rootzén (1983). Hence, EVT distributions could be directly applicable to return series that have long as well short memory (Bao et al. 2006, 103).

An important issue in implementing the POT approach is how to choose the threshold u. For example, if we are interested in the 5% quantile, then the chosen u must be greater than $q(0{,}05)$. This thesis follows Neftci (2000) and implements the empirical 10% quantile for u. The choice of the optimal threshold is a delicate issue since it is confronted with a bias-variance trade-off. If too low threshold is chosen, estimates might be biased because the limit theorems do not apply anymore while high thresholds generate estimates with high standard errors due to the limited number of observations. Alternative methods for choosing the u entail using the plot of the sample mean excess function (MEF) (see e.g. Tsay 2005, 320–321; and Embrechts et al. 1997), or alternatively one might estimate the threshold value u to decide which extremes are really extremes (see Gonzalo & Olmo 2004).

This thesis implements the more modern peaks over threshold EVT method, rather than the block maxima method, for calculating VaR estimates, which leads to using the generalized Pareto distribution. The threshold level is chosen as the 10% empirical quantile, following Neftci (2000). As the EVT was developed under the i.i.d. assumption – though later it has been extended to weakly dependent series as well –

'filtering' the return series might prove useful (see e.g. Alexander 2001, 97). Thus, besides applying the generalized Pareto distribution to the return series r_t, denoted by GPD, it is also applied to the filtered return series $y_t = r_t / \sigma_t$, denoted by GPD*.

3.6 Comparison of different approaches' assumptions and abilities

VaR models play a key role in the risk management of today's financial institutions. A number of VaR models are in use and all of them have the same aim, to measure the size of potential future losses as a predetermined level of probability. There are a variety of approaches used by VaR models to estimate the potential losses. Models differ, in fact, in the way they calculate the density function of future profits and losses of current positions, as well as on the assumptions they rely on.

Linear VaR models, basically meaning the VC approach and its extensions such as the RiskMetrics, impose strong assumptions about the underlying data. For instance, the density function of daily return follows a theoretical distribution (usually normal) and has constant mean and variance. The empirical evidence about the distributional properties of speculative price changes provides evidence against these assumptions, e.g. Kendall (1953) and Mandelbrot (1963). Risk managers have also seen their daily portfolio's profits and losses to be much larger than those predicted by the normal distribution. Embrechts et al. (1997) and Longin (2000) propose the use of extreme value theory (EVT) to overcome the last problem. Unfortunately, this approach is unsuitable in general for portfolios of derivatives, where losses may be limited by contractual terms. As an example, the maximum loss on an option spread (see e.g. Hull 2006, 225–234) is bounded by the difference between the two strike prices.

The VC VaR method has two additional major limitations. It linearises derivative positions and it does not take into account expiring contracts (Barone-Adesi et al. 2002, 33). These shortcomings may result in large biases, particularly for longer VaR horizons and for portfolios weighed with short out-of-the-money options. Therefore this approach is often used for investment portfolios. (It is after all closely related to the popular Markowitz mean-variance method of portfolio analysis.) It is less commonly used for calculating the VaR for the trading operations of a financial institution. This is because financial institutions like to maintain their deltas with respect to market variables close to zero (Hull 2007, 251). To overcome the problems of linearising derivative positions and to account for expiring contracts, risk managers look at *simulation techniques* (HS and MC). Pathways are simulated for scenarios for linear positions, interest rate factors and currency exchange rates and are then used to value all positions for each scenario. The VaR is then estimated from the distribution (e.g. the 1st percentile) of the simulated portfolio values.

Monte Carlo simulation is widely used by financial institutions around the globe. Nevertheless, this method can attract severe criticism. Namely, the generation of the scenarios is based on random numbers drawn from a theoretical distribution, often normal. Such a distribution not only does it not conform with the empirical distribution of most asset returns, but it also limits the losses to around three or four standard deviations when a very large number of simulation runs is carried out. To put this into perspective, as large as twenty standard deviations movement has been witnessed in actual financial data, when the S&P index lost 20% of its value on October 19, 1987 (Jorion 2007, 358).

Recognizing the fact that most asset returns cannot be adequately described by a theoretical distribution, an increasing number of financial institutions are using historical simulation, where each historical observation forms a possible scenario, see Butler and Schachter (1998). A number of scenarios are generated and in each of them all current positions are priced. The resulting portfolio is more realistic since it is based on the empirical distribution of risk factors.

HS has still some serious drawbacks. Long time series of data are required to include extreme market conditions. The fact that asset risks are changing all the time is ignored. Historical returns are used as if they were i.i.d. random numbers. A consequence of this usage is that during highly volatile market conditions HS underestimates risk, as documented for example by Vlaar (2000). To remedy this problem Hull and White (1998) and Barone-Adesi et al. (2002) suggest to draw random standardized returns from the portfolio's historical sample and scale these returns by the current level of volatility. This method generates the complete distribution of the current portfolio's profits and losses including the effects of volatility changes, overcoming limitations of most current VaR models. Table 7 below briefly summarizes the main benefits and disadvantages of each presented VaR approach.

Table 7 Main advantages and disadvantages of VaR approaches

Approach	Nonlinear instruments	Advantages	Disadvantages
VC	To some extent	Speed, simplicity	Unaccuracy
HS	Yes	No specific distribution	Strong i.i.d. assumption
MC	Yes	Versatility	Model risk, heaviness
EVT	No	Allows for 'fat tails'	(Strong) i.i.d. assumption

In table 7 a certain approach's ability to handle nonlinear instruments is reported in a separate column as this issue is of great importance when selecting the appropriate VaR approach for one's portfolio. Though the VC approach is simple it is not able to handle nonlinear instrumets satisfactorily. This often leads to its dismissal when choosing among VaR models. The EVT approach is theoretically more appealing than the VC approach but it also lacks suitability for portfolios of derivatives. Besides this the EVT approach also relies heavily on i.i.d. returns which can be somewhat questionable as one usually observes clustering in the second moment. However, the theory has been extended to serially dependent observations provided that the dependence is weak, hence the brackets in table 7 in EVT's disadvantages. These facts considered the HS and MC models would seem most appealing. The MC approach is versatile enough to be suitable for portfolios of derivatives but it still relies on a theoretical distribution for creating its scenarios. Besides relying on a theoretical distribution the MC approach is also rather heavy to implement and it is prone to model risk. The HS approach on the other hand does not rely on theoretical distributions as it implements the empirical distribution itself. This, however, comes at a cost as a straightforward implementation of the empirical distribution directly states that one assumes returns to be independent and identically distributed. Suggestions to remedy this problem have, however, been made and with the above considerations these suggested extensions to historical simulation would seem very appealing.

4 COMPARISON METHODS FOR VALUE AT RISK MODELS

4.1 Backtesting procedures with exception rates

4.1.1 Empirical coverage probability

Whatever the methodology used for calculating VaR, an important reality check is *backtesting*. This involves testing how well the VaR estimates would have performed in the past (Hull 2006, 450). An alternative method for testing VaR models is known as *stress testing,* which involves estimating how the portfolio and VaR model would have performed under extreme market moves (Hull 2007, 212). As backtesting provides a consistent framework for evaluating the performance of different VaR models, it is implemented in this thesis in order to discriminate amongst alternative VaR models. It is worth noting, however, that the chosen out-of-sample period in this particular thesis encompasses a market crisis period (see table 2 and appendix 1). Hence, the aspect of stress testing is also addressed in this thesis.

There are a number of theoretical and empirical studies on (backtesting) techniques for evaluating VaR models, such as Kupiec (1995), Crnkovic and Drachman (1995), Jackson, Maude and Perraudin (1997), Lopez (1998) and Christofferson (1998). The Basel Committee (1996a) has also given its own guidelines for backtesting. The majority of these methods concentrate on the days when the actual portfolio change exceeds the VaR estimate. These violations are referred to as *exceptions*.

There are a few attributes which are desirable for VaR estimates, but there is no direct way of observing whether the estimates are precise. However, a number of different indirect measurements will, together, create a picture of their precision (Boudoukh et al. 1998). The first desirable attribute is *unbiasedness*, i.e. a requirement that the estimates are accurate. This property requires that the VaR estimate is the α% tail. Verification of this property can be done by considering the number of exceptions. Suppose that the time horizon is one day and the confidence limit is X. If the VaR model under consideration is accurate (unbiased), the probability of the VaR estimate being exceeded on any given day is $\alpha = 1 - X$. Suppose that we look at a total of P days (length of the out-of-sample period) and we have defined an indicator function $I_t(\alpha)$, which receives the value of 1 if an exception occurs and zero otherwise. Now we can observe the days that VaR limit is exceeded, by $m = \sum_{t=R+1}^{T} I_t(\alpha)$, (see table 2 for definitions of R and T) from which we can derive *the empirical coverage probability* (i.e. the violation rate) as (Hull 2007, 208)

$$\hat{\alpha}_P = \frac{1}{P}\sum_{t=R+1}^{T} I_t(\alpha) \tag{4.1}$$

This empirical coverage probability $\hat{\alpha}_P$ can be compared with the chosen probability of loss α, i.e. its nominal value. The model that gives VaR forecasts with $\hat{\alpha}_P$ closest to its nominal value is the preferred model. In addition, this thesis implements a powerful statistical *likelihood ratio test* for coverage probability, suggested by Kupiec (1995) to obtain significance levels for the acceptance or rejection of $\hat{\alpha}_P$. Thus, the relevant null hypothesis is H_0: $\hat{\alpha}_P = \alpha$, against the alternative H_1: $\hat{\alpha}_P \neq \alpha$. As the number of exceptions follows a binomial distribution *the likelihood ratio test statistic for unconditional coverage* (LR_{uc}) is constructed as

$$LR_{uc} = -2\ln\left((1-\alpha)^{P-m}\alpha^m\right) + 2\ln\left((1-\hat{\alpha}_P)^{P-m}(\hat{\alpha}_P)^m\right) \tag{4.2}$$

where the maximum likelihood estimator of $\hat{\alpha}_P$ can be obtained from equation (4.1). This test statistic should follow a chi-square distribution with one degree of freedom. The null hypothesis H_0 is accepted whenever LR_{uc} is statistically insignificant as values for the test statistic are high for either very low or very high numbers of exceptions. There is exactly a 5% probability that the value of a chi-square variable with one degree of freedom will be greater than 3,84. Therefore, it follows that at 5% significance level the VaR model should be rejected, based on biasedness, whenever the expression in equation (4.2) receives a value greater than 3,84.

4.1.2 Basel Committee's guidelines for backtesting

The Basel Committee (1996a) has also announced its own framework for backtesting in the Amendment of 1996, and it requires for banks' VaR models to be backtested. Banks should use both *actual* and *hypothetical* changes in the daily profit and loss (P&L) to test a one-day VaR model that has a confidence level of 99%. The use of actual and hypothetical changes makes different assumptions on the portfolio composition during the holding period considered. First, comparing the VaR with the hypothetical change in the portfolio value is calculated on the assumption that the composition of the portfolio remains unchanged during this time period. The other method is to compare VaR to the actual change in the value of the portfolio during this time period. VaR itself is invariably calculated on the assumption that the portfolio will remain unchanged during the holding period, and ergo the former comparison based on hypothetical changes is more logical. However, it is the actual changes in the portfolio value that risk management is ultimately interested in. The backtesting procedures in this thesis conform to both of these definitions, as the composition of our portfolio does not change during the holding period (portfolios under scrutiny consist only of a single asset and therefore their composition cannot change).

Basel Committee's approach includes a requirement that banks will suffer increases in their capital requirements if, over a twelve-month period (250 trading days) using a rolling window of 60 days, their VaR models under-predict the number of losses exceeding the 1% cut-off point, i.e. under-predict the number of exceptions. If a bank's VaR model has generated zero to four exceptions, it is said to be in the Green Zone; if five to nine, it is in the Yellow Zone; and if there are more than ten exceptions, it is in the Red Zone. Capital requirements for banks whose models are in the Yellow or Red Zone will be increased by regulators by setting a higher multiplicative factor *k* for the capital requirement (see section 1.2). Table 8 below illustrates the relationship between the number of exceptions and the multiplicative factor *k*.

Table 8 Determination of the multiplicative factor

Zone	Number of exceptions	Multiplicative factor *k*	Significance level
Green	0	3	<0,05
Green	1−4	3	0,278–0,380
Yellow	5	3,4	0,162
Yellow	6	3,5	0,059
Yellow	7	3,65	0,019
Yellow	8	3,75	0,005
Yellow	9	3,85	0,001
Red	≥10	4	<0,000

From table 8 it can be observed that the multiplicative factor *k* ranges from a minimum of 3 to a maximum of 4 as the number of exceptions increases. This is a high inducement for banks to develop VaR models that produce unbiased VaR estimates. However, it is worth noting that this assessment method required by the Basel Committee does not account for bunching of the exceptions, nor size of the VaR estimates. These particular aspects will be addressed in the upcoming subsections. Further, the Basel method clearly favors models that overestimate rather than underestimate value at risk. This can be seen by applying the equation (4.2) to these Basel parameters to determine the confidence limits the Basel Committee is implicitly using for its decision to accept or reject a model. These confidence limits are reported in column four of table 8. For example, with the Basel backtesting scheme (250-day window) models exhibiting 7 or more, or zero exceptions should be rejected at 5% significance level according to equation (4.2). The Basel Committee, however, imposes no penalty on models that exhibit zero exceptions (and therefore overestimate VaR) and places them in the Green Zone.

This thesis does not per se implement the backtesing scheme proposed by the Basel Committee (1996a) but, as can be seen from the above considerations, using the empirical coverage probability and likelihhod ratio test statistic for unconditional coverage capture the same aspects of VaR precision as the Basel Committee's proposal. Ergo, the results of this thesis, based on these two criteria, are comparable to those studies on VaR models' predictive performance that rely on the Basel Committee's proposal.

4.2 Size of the estimates and quantile loss function

4.2.1 Size of value at risk estimates

The above mentioned backtesting procedures concentrate on how well a given model's empirical coverage probability conforms to its theoretical value (accuracy). This aspect does not, however, take fully into account the effects of the size of the VaR estimates. Intuitively, in a search for an optimal model we are seeking to minimize our VaR estimates – to minimize our capital charge – without jeopardizing the accuracy of our estimates – so that we are sufficiently prepared for the occurrences of extreme losses.

This aspect can simply be taken into consideration by calculating the average size of the VaR estimates during the forecast period (out-of-sample period). This criterion used alone, however, might mislead into preferring models that severely under-predict VaR. Therefore, it has to be used as a supplementary measure for model precision – accuracy being the primary comparison criterion. This is due to the fact that the lower the average VaR compared to other models, the more exceptions we expect to observe, and vice versa. However, if two models perform similarly in terms of coverage probability, we can discriminate between them by preferring the model that produces lower VaR estimates. This simple method of using average size of VaR estimates can, however, be extended to take into account additional perspectives included in VaR estimation by considering distinct loss functions, as proposed e.g. by Zielinski (2005). This desirable feature is discussed in the next subsection.

4.2.2 Predictive quantile loss function

In the context of estimating VaR, underestimating a quantile, i.e. VaR, is less desirable than overestimating it, since underestimation might lead to higher capital charges, poor capital allocation decisions and potentially even risks company survival. This suggests measuring the error of estimation by an asymmetric loss function (Zielinski 2005). Thus, this thesis implements an asymmetric quantile loss function, or check function, $Q(\alpha)$, suggested by Koenker and Basset (1978) (for introduction and examples see Koenker and Hallock 2001), in comparing the predictive ability of various VaR models. This expected loss of $q_t(\alpha)$ for a given α is given by

$$Q(\alpha) = E\big(\alpha - I_t(\alpha)\big)\cdot\big(r_t - q_t(\alpha)\big) \tag{4.3}$$

and it can provide a measure of lack-of-fit of a VaR model. The expected check function $Q(\alpha)$ can be evaluated from the VaR forecasts by

$$\hat{Q}_P(\alpha) = \frac{1}{P}\sum_{t=R+1}^{T}\big(\alpha - I_t(\alpha)\big)\cdot\big(r_t - \hat{q}_t(\alpha)\big) \tag{4.4}$$

where P is the length of the forecast period (out-of-sample period) and $\hat{q}_t(\alpha)$ is estimated by $\hat{F}_t^{-1}(\alpha) = \hat{\mu}_t + \hat{\sigma}_t\hat{G}_t^{-1}(\alpha)$ as discussed in section 3. A further inspection of equations (4.3) and (4.4) indicates that a higher weight of $(1 - \alpha)$ is given to the loss whenever an exception occurs and a lower wight of α whenever there occurs no exception. Therefore, this loss function is asymmetric, as it prefers the case of no violations, and accounts for the fact that underestimating VaR is less desirable than overestimating it.

To see how this measure works it can be decomposed into two parts: $|\alpha - I_t(\alpha)|$ and $|r_t - q_t(\alpha)|$. The first part makes the function assymtrecis as descreibed above. The second part $|r_t - q_t(\alpha)|$ can be thought to measures the forecst error of the VaR estimate as it is the difference between the actual return and the VaR estimate. Note, however, that this term cannot be defined as pure forecast error as VaR attempts to predict a ceratain quantile of the returns distribution, not actual returns. This latter part can also be seen to reflect the average size of the VaR estimates as higher VaR estimates are further away from the center of the return distribution where most of the probability mass is located, hence making this part of equation (4.3) higher for those models that generate higher VaR estimates on average.

The interpretation of this measure is similar to using the average size of the VaR estimates as a model choice criterion. More precisely, a model that gives the VaR forecasts $\hat{q}_t(\alpha)$ with the minimum value of $\hat{Q}_P(\alpha)$ is the preferred model. If two VaR models yield the same $\hat{Q}_P(\alpha)$, then the one that gives a more accurate empirical coverage probability, i.e. $\hat{\alpha}_P$ closest to its nominal value α, is the preferred model.

4.3 Bunching of exceptions

Unbiasedness and size alone are insufficient benchmarks for adequate VaR estimates. To see this, consider the case of a VaR estimate which is constant trough time, but also highly precise unconditionally (i.e. achieves an average VaR probability which is close to α). To the extent that tail probability is cyclical – due to fluctuations in the second moment – the occurrences of violations of the VaR estimate will be bunched up. This is a very undesirable property, since VaR models are required dynamic updating which is sensitive to market conditions (Boudoukh et al. 1998, 8).

Consequently, the third attribute which is controlled is that exceptions do not bunch up. If daily portfolio changes are independent, exceptions should be spread evenly through the period used for backtesting. In practice, they are often bunched together, suggesting that losses on successive days are not independent. This requires that a VaR estimate should increase as the tail of the distribution rises. If a large return is observed today, the VaR should rise to make the probability of another tail event exactly α tomorrow. In terms of the indicator variable $I_t(\alpha)$, this essentially requires that it is i.i.d. This requirement is similar to saying that the VaR estimate should provide a filter to transform a serially dependent return volatility and tail probability into a serially independent $I_t(\alpha)$ series (Boudoukh et al. 1998).

One way of assessing the extent of this serial dependence (autocorrelation) is to check for autocorrelation at any lag in the $I_t(\alpha)$ time series (see e.g. Ljung & Box 1978). This method, however, quickly becomes cumbersome when lots of data are used as multiple time series and lags should be checked. Alternatively, one might use specific test statistics designed for capturing this serial dependence in violations to ease ones effort. Probably the most widely adopted such statistics are Christoffersen’s (1998) *likelihood ratio test statistics for independence* and *conditional coverage* (LR_{ind} and LR_{cc}, respectively). These tests are based on the null hypothesis that the probability of an exception occurring is independent on what happened the previous day. More precisely, the likelihood ratio test statistic for independence is defined as

$$LR_{ind} = -2\ln\left((1-\hat{\alpha}_P)^{(T_{00}T_{10})}(\hat{\alpha}_P)^{(T_{01}T_{11})}\right) + 2\ln\left((1-\alpha_0)^{T_{00}}\alpha_0^{T_{01}}(1-\alpha_1)^{T_{10}}\alpha_1^{T_{11}}\right) \quad (4.5)$$

T_{ij} measures the number of days in which state j occurred while it was at i the previous day, and α_j denotes the probability of observing an exception conditional on state j the previous day. If indicators (i, j) are assigned to 0 if VaR is not exceeded and 1 otherwise, then the maximum likelihood estimates of α_0 and α_1 are given by (T_{01} / (T_{00} + T_{01})) and (T_{11} / (T_{10} + T_{11})), respectively. If an exception is independent of previous day's conditions then LR_{ind} should not be statistically significant. LR_{ind} follows a chi-square distribution with one degree of freedom.

Further, an additional test of conditional coverage can be constructed by combining the tests for unconditional coverage and independence, i.e. equations (4.2) and (4.5) (Christofferson 1998) which allows for the simultaneous examination of biasedness and bunching. Thus, this likelihood ratio for conditional coverage follows a chi-square distribution with two degrees of freedom and is constructed as

$$LR_{cc} = LR_{uc} + LR_{ind} \tag{4.6}$$

However, as can be seen from the definition and null hypotheses of these *LR* statistics (for independence and conditional coverage) it is clear that they are only designed to secure that not too many exceptions occur on consecutive days. That is, though they are relatively powerful statistical tests, they still only capture seraial dependence at only one single lag which clearly is not adequate for the purposes of this thesis.

Therefore, in this case neither of these methods is preferable and an alternative approach is implemented. That is, clustering of the exceptions is addressed by observing the out-of-sample quantile loss plots, based on equation (4.3), from which bunching (at multiple lags) is relatively easy to detect. To see this, let us consider, for example, the case of α = 0,05 with a given time index *t*. In this case the loss value in equation (4.3) corresponds to

$$\begin{cases} 0{,}95 \cdot \left| r_t - \hat{q}_t(\alpha) \right| & \text{if} \quad r_t < \hat{q}_t(\alpha) \\ 0{,}05 \cdot \left| r_t - \hat{q}_t(\alpha) \right| & \text{if} \quad r_t > \hat{q}_t(\alpha) \end{cases}$$

Hence, a larger weight 0,95 is given to the forecast error (i.e. the difference between the estimate and actual outcome) when a violation occurs and a smaller weight 0,05 if there occurs no violation. This causes spikes in the plots whenever a violation occurs which makes it relatively easy to discover possible bunching of the exceptions. In the plots higher spikes indicate larger violations. This method, however, does not offer a direct answer to what is acceptable degree of bunching but it still allows for a sensible comparison of clustering across models.

4.4 Implemented comparison methods

This thesis evaluates one-step-ahead (one day) predictive ability of various VaR models during three out-of-sample periods (see table 2) based on their accuracy, size of estimates and bunching of exceptions. The models' predictions are generated using two of the most commonly implemented confidence levels in market risk VaR calculation, i.e. X = 95% and X = 99%. The corresponding probabilities of loss are therefore α = 0,05 and α = 0,01, respectively. Most VaR models are expected to perform better the closer the estimated quantile is to the mean of the distribution. Hence, divergent results are expected when using different confidence levels.

In this thesis models' accuracy is evaluated via empirical coverage probability, for which confidence limits are generated by a relatively powerful likelihood ratio test statistic for unconditional coverage (LR_{uc}) proposed by Kupiec (1995). Size of the VaR estimatess is evaluated by a predictive quantile loss function of Koenker and Basset (1978) which is able to capture the important property that underestimating VaR is less desireable than overestimating it. Finally, possible bunching of the exceptions is addressed by observing the plots produced by the quantile loss function from which clustering is relatively easy to detect. These evaluation methods are distinctively implemented for each of the three out-of-sample periods (before, during and after market crisis) to derive conclusions on the performance of the tested VaR models in different market circumstances. The results for these comparison criteria are reported in the next section of the thesis.

5 DATA AND RESULTS

5.1 Description of the data

This thesis applies different approaches to value at risk modelling, summarized in table 9 below, to the stock markets of four different economies. The data for the stock indices are retrieved from Thomson Financial database. Indices under examination are the S&P 500 Composite (United States), CDAX General 'Kurs' (Germany), HEX General (Finland), and India BSE National (India) (see appendix 1 for the indices' time series).[2] The return series of each index is given by the log difference of these price series. Henceforth, these indices are referred to by the names of their respective countries, i.e. United States, Germany, Finland and India. The data comprises daily closing prices of these indices from 1.1.1988 to 31.12.2004. This whole sample consisting of T observations is split into an in-sample period of size R and an out-of-sample period of size P so that $T = R + P$. For these samples a rolling window scheme is implemented. That is, the $(t - R)$th one-step-ahead VaR prediction is based on observations $t - R$ through $t - 1$, where $t = R + 1, . . . , T$.

Table 9 Summary of implemented VaR models

Distribution	Normal	Historical	Monte Carlo	Pareto
VaR Model	VC	HS	MC	GPD
	VC*	HS*	N/A	GPD*
	VC(GARCH)	HS(GARCH)	MC(GARCH)	N/A
Benchmark	RiskMetrics			

Notes:
(1) See section 3 for a complete description of these VaR models.
(2) Variance-covariance (VC), historical simulation (HS), Monte Carlo simulation (MC) and an extreme value theory approach that utilizes the generalized Pareto distribution (GPD).
(3) All the models in this thesis that are applied to the filtered return series $y_t = r_t/\sigma_t$ *are denoted by asterisk *.*

The entire out-of-sample period is further divided into three two-year out-of-sample evaluation periods according to market circumstances. These periods are referred to as before, during and after market crisis periods (see table 2 for an exact definition of these periods). More precisely, these periods represent market conditions before the bursting of the so called technology bubble in 2000 (representing normal market conditions),

---

[2] Datastream mnemonics (entity keys) for these particular indices are S&PCOMP (I0003), CDAXGNI (I900001545), HEXINDX (I000004651, and IBOMBSE (I000004256), respectively.

during the market crisis period caused by this speculative bubble bursting (representing market crisis period), and after the markets had settled from this shock (representing normal market conditions). The indices for this thesis were chosen in a way that they were all impacted by this crisis relatively simultaneously but with different intensity, allowing for the division of the time series into these distinct periods. Thus, insights into the performance of the tested VaR models under different market conditions can be gained. While two of these periods represent normal market conditions, VaR models are expected to perform divergently during these particular periods. This is due to the fact that the predictions of VaR models in the after crisis period are to some extent affected by the crisis period, whereas the before crisis period was preceded by a less turbulent period which should lead to more conservative VaR predictions.

The choice of these indices was further driven, besides the requirement that they all were simultaneously impacted by the market crisis of 2000, by another factor. That is, the United States and Germany were chosen to represent developed stock markets, whereas Finland and India represent emerging markets.[3] This choice grants the opportunity of observing the performance of different VaR models in different market contexts. To be more precise, it is expected that especially the emerging market time series do not behave normally and exhibit some degree of skewness and kurtosis. To revise whether these expectations carry some thruth in them, summary statistics for the whole sample period of each index are provided in table 10 below.

Table 10 Summary statistics, whole sample period (1.1.1988–31.12.2004)

Index	Mean	St. dev	Skewness	Excess kurtosis	J B statistic
United States	0,00036	0,010	***-0,232	***4,335	***3426,20
Germany	0,00024	0,012	***-0,401	***5,205	***4939,20
Finland	0,00037	0,018	***-0,425	***8,504	***12968,20
India	0,00072	0,018	***-0,110	***6,226	***6015,10

Notes: (1) The J B (Jarque-Bera) statistic is asymptotically distributed as a $\chi^2(2)$ under the null of normality (Jarque & Bera 1987).
(2) See Snedecor and Cochran (1980, 78) for statistical tests of skewness and kurtosis.
*(3) *** Significant at the 1 percent level.*

As anticipated, the results in table 10 suggest that the series are strongly nonnormal. All are significantly leptokurtic (i.e. exhibit positive excess kurtosis, wich indicates

[3] In general, markets that are perceived as most developed and therefore less risky are referred to as developed markets, whereas markets where "*politics matters at least as much as economics to the to the markets*" (Bremmer, 2005) can be characterized as emerging markets. In this thesis the division between these two concepts is, above all, based on stock market volatility. Hence, the United States and Germany fall under the category of developed markets, and Finland and India under the category of emerging markets.

heavier tails and a higher peak than for the normal distribution) and skewed to the left (i.e. exhibit more negative than positive returns). The Jarque-Bera (1987) statistical test of normality also indicates that all the series are significantly nonnormal. Furthermore, table 10 supports the view that United States and Germany can be classified as developed stock markets, and Finland and India as emerging stock markets, as the former group exhibits less deviations from normality than the latter based on the Jarque-Bera statistic. This view is also supported by the fact that standard deviation and kurtosis for the stock markets defined as developed are less than for the markets defined as emerging. To further study the behaviour of these series during the three out-of-sample periods (before, during and after market crisis) summary statistics covering each of these periods is provided in table 11 below.

Table 11 Summary statistics, three evaluation periods

Index (Period)	Mean	St. dev	Skewness	Excess kurtosis	J B statistic
United States (1)	0,00097	0,012	***-0,653	***5,870	***786,55
United States (2)	-0,00048	0,014	0,009	***1,534	***50,03
United States (3)	0,00063	0,009	***0,059	***1,254	***33,36
Germany (1)	0,00092	0,014	***-0,846	***2,932	***238,37
Germany (2)	-0,00066	0,015	***-0,175	***0,930	***20,81
Germany (3)	0,00069	0,014	0,030	***2,411	***123,55
Finland (1)	0,00161	0,018	***-0,432	***3,891	***330,31
Finland (2)	-0,00101	0,034	***-0,344	***2,809	***174,22
Finland (3)	0,00015	0,015	***-0,794	***5,536	***696,48
India (1)	-0,00001	0,017	***-0,075	***2,529	***130,81
India (2)	-0,00105	0,022	***-0,377	***1,117	***37,67
India (3)	0,00153	0,015	***-1,280	***10,582	***2479,19

Notes:

(1) The J B (Jarque-Bera) statistic is asymptotically distributed as a $\chi^2(2)$ under the null of normality (Jarque & Bera 1978).

(2) See Snedecor and Cochran (1980, 78) for statistical tests of skewness and kurtosis.

*(3) *** Significant at the 1 percent level.*

(4) Periods 1, 2 and 3 refer to before, during and after market crisis periods, respectively. See table 2 for specific period definitions.

The summary statistics in table 11 support the splitting of these indices into the periods presented in table 2 as all the series have higher standard deviations during the defined crisis period (period 2) than before (period 1) or after (period 3) this period. This difference is most prononunced for the emerging markets’ time series. The division is also supported by the visual appearance of these time series (see appendix 1). Furthermore, as this thesis assumes a long position in the chosen indices, and they all trend downwards during the crisis period (see means during period 2 in table 11 and appendix 1 for visual confirmation) it seems natural to postulate the period spanning 1.1.2000–31.12.2001 as the crisis period. In this subperiod examination, presented in table 11, all the indices still exhibit significant deviations from normality except for the fact that United States during the crisis period and Germany during the after crisis period do not appear to be statistically significantly skewed. These examinations of the data indicate that these indices can plausibly be divided into three out-of-sample evaluation periods – and that VaR models relying on normality might severely underpredict the actual quantiles as all the time series considered exhibit significant excess kurtosis.

5.2 Presentation of the research results

5.2.1 Tail coverage probability

Accuracy of the VaR models is measured with empirical coverage probability in this thesis, as described in section 4. Accuracy is evaluated during three out-of-sample periods (tables 12–14) in order to compare the potential changes in risk forecast precision across different market conditions.

Table 12 below presents the results for empirical coverage probability for the pre-crisis period (period 1). As shown in table 12 with α = 0,05 the conventional RiskMetrics model produces satisfactory forecasts. In fact, it is the only model that produces satisfactory forecasts across all countries according to the likelihood ratio test statistic for unconditional coverage (denoted by ** in table 12). For most countries the predicted coverage probability is close to the nominal coverage, and for the United States and Finland RiskMetrics also produces the most accurate forecasts (presented in bold font). All the models perform rather well across all countries, except for the HS model that systematically seems to overpredict coverage probability (and therefore underpredicts VaR) and the GPD* model that, on the contrary, underpredicts tail probability (and therefore overpredicts VaR). Bear in mind that overpredicting VaR is more desireable than underpredicting it as too low VaR predictions may lead to higher capital charges imposed by regulators or even risk company survival. Most models seem to overpredict tail probability for United States, Germany and Finland, except for the EVT based models (GPD and GPD*). On the contrary, most models underpredict tail probability for India. This underprediction is most severe for the GPD* model that causes no single violation to occur during period 1.

Table 12 Empirical coverage probability (%), period 1 (1.1.1997–31.12.1998)

Probability	Model	United States	Germany	Finland	India
5 %	RiskMetrics	****4,79**	**6,21	****5,41**	**5,11
	VC	**6,32	8,42	8,22	**4,70
	VC(GARCH)	**5,36	7,62	**6,81	**4,70
	VC*	****4,79**	**6,21	**6,21	2,66
	HS	7,85	9,22	8,82	**5,73
	HS(GARCH)	**6,13	**7,01	**6,01	2,86
	HS*	**6,13	**7,01	**6,01	2,86
	MC	**6,51	9,62	8,82	**4,70
	MC(GARCH)	**6,32	8,42	7,41	****4,91**
	GPD	**3,64	****5,01**	3,21	**5,32
	GPD*	2,49	**3,41	3,21	0,00
1 %	RiskMetrics	2,49	2,81	**1,80	2,25
	VC	2,30	4,21	3,41	2,04
	VC(GARCH)	2,87	3,01	2,61	**1,64
	VC*	2,30	2,40	2,20	****1,02**
	HS	2,11	2,00	3,01	2,25
	HS(GARCH)	**1,53	**1,60	2,00	****1,02**
	HS*	**1,53	**1,60	2,00	****1,02**
	MC	2,68	4,21	3,61	2,25
	MC(GARCH)	3,26	3,81	2,81	**1,43
	GPD	****1,34**	****0,60**	**0,60	****1,02**
	GPD*	**0,57	0,20	****0,80**	0,00

Notes:

(1) See equation (4.1) for empirical coverage probability.

(2) The best model in terms of empirical coverage probability is presented in bold font for each index.

*(3) ** denotes insignificance at 5% level according to Kupiec's (1995) likelihood ratio test statistic for unconditional coverage (equation 4.2). A satisfactory model is insignificant.*

During the pre-crisis period with α = 0,01, the risk forecasts of RiskMetrics are less satisfactory than with α = 0,05, but only the GPD model beats RiskMetrics across all countries. HS(GARCH) and HS* also perform well. Only the EVT based models are able to improve their performance as the confidence level is raised from 95% to 99% which is in accordance with expectations as the purpose of these particular VaR models is to provide accurate forecasts for extreme quantiles. With α = 0,01 underprediction of VaR, that was also observed with α = 0,05 for United States, Germany and Finland, is rather obvious across all countries. Only GPD and GPD* occasionally overpredict VaR at this probability level. Overall, no model appears to be particularly superior to RiskMetrics, which works reasonably well for both α = 0,05 and α = 0,01 during the pre-crisis period.

Table 13 below presents corresponding results for the crisis period (period 2). The observed findings are surprisingly similar to those during the before crisis period. With α = 0,05 all models perform rather similarly as in the pre-crisis period and it can be said that most models' performance has even improved to some extent. There still exists some overprediction of coverage probability for Germany but these results are not notably different from those obtained during period 1. RiskMetrics again produces rather satisfactory coverage and it is only beaten by the HS, HS(GARCH), HS* and GPD models. All models besides the GPD* work rather suitably during the crisis period with α = 0,05.

Table 13 Empirical coverage probability (%), period 2 (1.1.2000–31.12.2001)

Probability	Model	United States	Germany	Finland	India
5 %	RiskMetrics	**6,67	7,71	**4,60	**6,22
	VC	**5,29	7,31	**4,60	**6,83
	VC(GARCH)	**6,27	8,89	**4,80	**5,82
	VC*	****5,10**	7,51	**3,40	**5,62
	HS	**5,49	**6,92	**5,80	**6,02
	HS(GARCH)	****5,10**	****5,93**	**4,20	**5,82
	HS*	****5,10**	****5,93**	**4,20	**5,82
	MC	**5,69	7,31	**5,20	7,03
	MC(GARCH)	**6,08	8,70	****5,00**	**6,02
	GPD	**4,12	**3,75	**3,60	**6,63
	GPD*	1,96	7,31	1,80	****4,82**
1 %	RiskMetrics	**1,96	**1,19	2,00	2,81
	VC	**1,57	2,57	2,20	3,01
	VC(GARCH)	**1,76	****0,99**	2,40	2,61
	VC*	**1,37	****0,99**	****1,20**	2,41
	HS	**1,76	2,17	2,20	2,21
	HS(GARCH)	**1,57	****0,99**	****1,20**	**1,41
	HS*	**1,57	****0,99**	****1,20**	**1,41
	MC	**1,76	2,57	2,20	3,01
	MC(GARCH)	**1,57	**1,19	2,40	3,01
	GPD	****1,18**	**0,79	2,00	**1,61
	GPD*	**0,59	**1,78	**0,80	****1,20**

Notes:

(1) See equation (4.1) for empirical coverage probability.

(2) The best model in terms of empirical coverage probability is presented in bold font for each index.

*(3) ** denotes insignificance at 5% level according to Kupiec's (1995) likelihood ratio test statistic for unconditional coverage (equation 4.2). A satisfactory model is insignificant.*

With $\alpha = 0{,}01$ the results are as well rather similar to those in the before crisis period. Some differences can though be noticed. As most models significantly overpredicted coverage probability during the before crisis period, this overprediction, though it still rather universally exists, no longer is as significant during the crisis period. Therefore, it appears that the VaR models have rather adequately been able to adjust to the changes in market conditions. Best performers during period 2 with $\alpha = 0{,}01$ are the HS(GARCH), HS* and GPD* models, but it cannot be said that RiskMetrics' performance is insufficient. Overall, during the crisis period RiskMetrics works rather well with $\alpha = 0{,}05$ and $\alpha = 0{,}01$ and its performance is best challenged by the VC*, HS(GARCH), HS* and GPD models.

Corresponding results for the post-crisis period are presented in table 14 below. With $\alpha = 0{,}05$ the RiskMetrics model again produces satisfactory risk predictions across all countries. Only the HS(GARCH) and HS* models are able to match this performance. The VC* model also produces rather satisfactory estimates during the after crisis period with $\alpha = 0{,}05$. Furthermore, the remainder of the models appear to significantly underpredict tail probability (i.e. overpredict VaR) during period 3. Exactly the opposite was observed for periods 1 and 2. Therefore, it seems that the predictions of these VaR models are to some extent influensed by the preceeding turbulent period. That is, these models adjust their VaR predictions upwards during the crisis period but as a more tranquil risk regime is reached their reactions are not as dynamic downwards as they were upwards.

With $\alpha = 0{,}01$ the RiskMetrics model is also able to produce satisfactory results across all countries, and no other model is able to match this. Most models still perform well but not just as well as the RiskMetrics model. Closest challengers of RiskMetrics are VC(GARCH), VC*, HS(GARCH), HS*, MC(GARCH), GPD and GPD*. The underprediction of coverage probability, that was observed with $\alpha = 0{,}05$, can also be observed with $\alpha = 0{,}01$. As mentioned, this probably is due to the fact that the crisis period causes the VaR predictions to be somewhat conservative during the after crisis period. This overprediction of VaR is, however, not as apparent with $\alpha = 0{,}01$ as it was with $\alpha = 0{,}05$. Overall, with $\alpha = 0{,}05$ and $\alpha = 0{,}01$ the RiskMetrics performs best during period 3 in terms of coverage probability and its performance is best challenged by the HS(GARCH) and HS* models. The VC* model also performs rather well on this period with both probability levels. Before drawing more comprehensive conclusions on the overall performance of these various VaR models, i.e. the aggregate performance covering all out-of-sample periods, attention is first paid to other comparison criteria, as described in section 4.

Table 14 Empirical coverage probability (%), period 3 (1.1.2003–31.12.2004)

Probability	Model	United States	Germany	Finland	India
5 %	RiskMetrics	**4,16	****4,90**	**4,37	**3,78
	VC	2,18	2,75	1,59	**6,18
	VC(GARCH)	2,38	**4,12	2,38	2,79
	VC*	**4,16	**4,51	2,78	****4,58**
	HS	2,77	1,96	2,38	7,17
	HS(GARCH)	**4,55	**4,71	****4,56**	**6,37
	HS*	**4,55	**4,51	****4,56**	**6,37
	MC	2,38	1,96	1,39	**6,37
	MC(GARCH)	2,38	**3,92	2,38	2,79
	GPD	0,99	1,96	0,60	1,39
	GPD*	****4,75**	9,22	1,19	1,99
1 %	RiskMetrics	****0,79**	**1,76	**0,79	**1,79
	VC	0,00	**0,39	**0,79	2,39
	VC(GARCH)	0,20	**1,18	**0,60	****1,39**
	VC*	0,20	**1,57	**0,79	**1,79
	HS	**0,59	0,20	**0,79	2,19
	HS(GARCH)	**0,59	**1,37	****0,99**	1,99
	HS*	**0,59	**1,37	****0,99**	1,99
	MC	0,20	0,20	**0,79	2,59
	MC(GARCH)	0,00	****0,98**	**0,60	**1,59
	GPD	**0,40	0,20	**0,40	**0,40
	GPD*	**0,40	5,49	**0,60	**1,59

Notes:

(1) See equation (4.1) for empirical coverage probability.

(2) The best model in terms of empirical coverage probability is presented in boldface for each index.

*(3) ** denotes insignificance at 5% level according to Kupiec's (1995) likelihood ratio test statistic for unconditional coverage (equation 4.2). A satisfactory model is insignificant.*

5.2.2 Predictive quantile loss

In addition to measuring the accuracy of value at risk estimates by empirical coverage probability this thesis also considers additional perspectives of VaR models' precision. To account for the asymmetric nature of VaR forecasting and average size of the VaR estimates, a quintile loss function, as described in section 4, is implemented. Tables 15–17 present the out-of-sample average quantile loss values $\hat{Q}_P(0{,}05)$ and $\hat{Q}_P(0{,}01)$ as defined in equation (4.4) for the before, during and after crisis periods. The best model in terms of $\hat{Q}_P(\alpha)$ (i.e. the model with the smallest $\hat{Q}_P(\alpha)$) is indicated in bold font in the tables for each country and for each α.

Table 15 below presents the results for the average quantile loss values for the pre-crisis period (period 1). As shown in table 15 with α = 0,05 RiskMetrics produces the smallest (i.e. the best) quantile loss values for Germany and India. In the United States and Finland RiskMetrics' performance is beaten by several models but the differences between models are only marginal. The only notable differences on period 1 with α = 0,05 can be observed in India where the average quantile loss values of GPD* are much higher than for most models. As GPD* produced no violations during this period (see table 12) it can be deduced that its VaR predictions are much too high. VC*, HS(GARCH) and HS* models' performance is also rather poor in this particular case.

Table 15 Predictive quantile loss (%), period 1 (1.1.1997–31.12.1998)

Probability	Model	United States	Germany	Finland	India
5 %	RiskMetrics	0,148	**0,170**	0,206	**0,192**
	VC	0,143	0,184	0,222	0,196
	VC(GARCH)	0,145	0,173	0,205	0,193
	VC*	0,149	0,171	**0,202**	0,288
	HS	0,147	0,188	0,224	0,196
	HS(GARCH)	0,149	0,174	0,203	0,285
	HS*	0,149	0,174	0,203	0,286
	MC	0,144	0,187	0,226	0,196
	MC(GARCH)	0,146	0,176	0,208	0,194
	GPD	**0,142**	0,180	0,217	0,193
	GPD*	0,165	0,175	0,207	0,563
1 %	RiskMetrics	0,054	0,053	0,064	0,063
	VC	0,054	0,070	0,079	0,070
	VC(GARCH)	0,054	0,060	0,067	0,061
	VC*	0,051	0,052	0,061	0,084
	HS	0,053	0,059	0,073	0,069
	HS(GARCH)	0,050	0,051	0,061	0,087
	HS*	0,050	0,051	0,061	0,087
	MC	0,056	0,073	0,081	0,069
	MC(GARCH)	0,056	0,063	0,070	**0,060**
	GPD	**0,047**	**0,050**	0,066	0,062
	GPD*	0,051	0,053	**0,060**	0,174

Notes:
(1) See equation (4.4) for predictive quantile loss.
(2) The best model in terms of predictive quantile loss is presented in bold font for each index and for each α (i.e. the model with the smallest out-of-sample average quantile loss).

During the pre-crisis period with α = 0,01 the results are very similar to those with α = 0,05. RiskMetrics no longer produces the best results in terms of average quantile loss in any country but it is not far behind from the best models' performance. Differences in model performance are still rather marginal with the same exceptions as with α = 0,05, i.e. the GPD* model is clearly the worst model and VC*, HS(GARCH) and HS* do somewhat worse than the rest of the models. Particularly the GPD model seems to perform rather well with α = 0,01 as it produces the best estimates in two countries out of four (United States and Germany). Overall, in the pre-crisis period with α = 0,05 and α = 0,01 RiskMetrics is among the best models when measured by average quantile loss and no considerable differnces between models' average quantile losses are observed. The only remarkable difference is for India where GPD* performs much worse than the remaining models.

Table 16 below presents corresponding results for the crisis period (period 2). Once again, no substantial disparities can be observed between models with both α = 0,05 and α = 0,01. RiskMetrics is chosen as the best model only once, compared to twice in the pre-crisis period, but its average quantile losses are very close to those of the best performers. The only rather poor performance compared to other models in terms of average quantile loss can be observed in Finland for GPD* where it produces a 0,06% higher average quantile loss than the penultimate performer (HS) with α = 0,05. This difference is not as obvious with α = 0,01. Furthermore, GPD* produces very similar results as the other models in the remaining countries. In addition it is chosen as the best model in India with α = 0,05. When looking at table 16 with α = 0,01 the GPD model clearly stands out as it is chosen as the best model three times out of four. Nevertheless, other models are not far behind GPD.

A comparinson of the results for the pre-crisis (table 15) and during crisis (table 16) periods shows that the average quantile loss values have risen in the crisis period for the emerging markets, whereas no remarkable difference can be observed for the developed markets. This indicates that the impacts of the market crisis have been more severe for Finland and India, making it even more difficult to forecast risks in these countries. Then again, the market crisis appears to have had no remarkable impact on the risk forecasts in United States and Germany measured in terms of predictive quantile loss. This indicates that the market crisis period was less turbulent in the developed countries than in the emerging countries. The summary statistics in table 11 verify this observation.

Table 16 Predictive quantile loss (%), period 2 (1.1.2000–31.12.2001)

Probability	Model	United States	Germany	Finland	India
5 %	RiskMetrics	0,154	0,159	**0,395**	0,256
	VC	**0,150**	0,169	0,398	0,271
	VC(GARCH)	0,153	0,150	0,405	0,253
	VC*	0,154	0,148	0,422	0,252
	HS	0,150	0,167	0,403	0,279
	HS(GARCH)	0,156	**0,148**	0,406	0,256
	HS*	0,156	0,148	0,406	0,256
	MC	0,150	0,169	0,403	0,273
	MC(GARCH)	0,153	0,151	0,408	0,254
	GPD	0,150	0,164	0,396	0,270
	GPD*	0,167	0,157	0,486	**0,246**
1 %	RiskMetrics	0,050	0,040	0,134	0,082
	VC	0,048	0,052	0,160	0,082
	VC(GARCH)	0,050	0,038	0,150	0,076
	VC*	0,050	0,039	0,150	0,075
	HS	0,049	0,053	0,144	0,078
	HS(GARCH)	0,051	0,039	0,150	0,073
	HS*	0,051	0,039	0,150	0,073
	MC	0,048	0,053	0,162	0,083
	MC(GARCH)	0,050	**0,038**	0,153	0,076
	GPD	**0,047**	0,050	**0,132**	**0,067**
	GPD*	0,055	0,044	0,169	0,071

Notes:

(1) See equation (4.4) for predictive quantile loss.

(2) The best model in terms of predictive quantile loss is presented in bold font for each index and for each α (i.e. the model with the smallest out-of-sample average quantile loss).

Table 17 in turn presents the average quantile loss values for the after crisis period. With both probability levels (α = 0,05 and α = 0,01) RiskMetrics is chosen twice as the best model and its values are also very close to the best performers in the other countries in terms of average quantile loss. Once again, as in the previous periods, differences among models are small. Only GPD appears to produce somewhat larger values than the other models, especially with α = 0,05. With α = 0,01 this difference, however, diminishes. Overall, RiskMetrics' performance is good in the after crisis period and no single model can be said to be superior to it in terms of average quantile loss.

Table 17 Predictive quantile loss (%), period 3 (1.1.2003–31.12.2004)

Probability	Model	United States	Germany	Finland	India
5 %	RiskMetrics	0,093	0,146	**0,175**	**0,173**
	VC	0,103	0,161	0,189	0,194
	VC(GARCH)	0,095	0,147	0,182	0,180
	VC*	0,094	0,146	0,178	0,178
	HS	0,100	0,163	0,184	0,195
	HS(GARCH)	0,094	0,147	0,175	0,179
	HS*	0,094	0,147	0,175	0,179
	MC	0,103	0,162	0,191	0,196
	MC(GARCH)	0,095	**0,146**	0,184	0,179
	GPD	0,119	0,187	0,251	0,238
	GPD*	**0,092**	0,156	0,197	0,211
1 %	RiskMetrics	**0,022**	0,040	**0,065**	0,065
	VC	0,027	0,042	0,066	0,073
	VC(GARCH)	0,024	**0,038**	0,066	**0,064**
	VC*	0,023	0,039	0,065	0,067
	HS	0,026	0,042	0,069	0,074
	HS(GARCH)	0,023	0,039	0,065	0,066
	HS*	0,023	0,039	0,065	0,066
	MC	0,028	0,042	0,067	0,074
	MC(GARCH)	0,024	0,038	0,066	0,064
	GPD	0,032	0,051	0,078	0,078
	GPD*	0,025	0,064	0,076	0,070

Notes:

(1) See equation (4.4) for predictive quantile loss.

(2) The best model in terms of predictive quantile loss is presented in bold font for each index and for each α (i.e. the model with the smallest out-of-sample average quantile loss).

In the post-crisis period the average quantile loss values are once again higher for the emerging markets (Finland and India) than for the developed markets (United States and Germany). The same observation was made during the two previous periods as well. Therefore, it appears that defining Finland and India as emerging markets and United States and Germany as developed markets is well justified. Further, this indicates that risk forecasting is more difficult in the emerging market context. As there appeared no significant differences among different models' performance in terms of average quantile loss values, despite few minor exceptions, these results still provide important information for the analysis as a whole. Comparing the quantile loss values across the three periods unveils interesting information about the overall performance of these VaR models.

For the developed markets the average quantile loss values are of the same magnitude for periods 1 and 2, but the after crisis values are considerably lower. For the emerging markets the lowest values are also produced in the after crisis period, but periods 1 and 2 also differ in their results. That is, the crisis period produces the highest values, whereas the pre-crisis values lie inbetween the values of the crisis and after crisis periods. For the developed markets this indicates that the VaR models adapt to the high risk regime rather well (or alternatively the crisis period is not turbulent enough to cause difficulties in risk forecasting in the developed markets). On the contrary, in the emerging markets the crisis period causes poor performance in terms of average quantile loss, indicating that VaR models are not fully able to adapt to this high risk regime. Recall that (poor) performance in terms of average quantile loss is a constituent of two separate factors: (high) number of exceptions and (high) forecast errors (see equation (4.4)). Therefore, by combining the results of tables 12–14 and 15–17 it can be seen whether models' poor performance in terms of average quantile loss is due to high number of exceptions, high forecast errors, or both. For Finland this dissection reveals that models' poor performance during the crisis period is mostly due to high forecast errors, whereas for India the both factors (high number of exceptions and high forecast errors) appear to be contributing to this poor performance (tables 13 and 16). Further, this dissection also reveals that the good performance of the models in the post-crisis period may be due to low number of exceptions rather than low forecast errors. Before drawing final conclusions on the overall performance of VaR models yet another important factor in value at risk models' performance evaluation is considered: serial dependence of violations.

5.2.3 Serial dependence of violations

The third desireable feature of VaR forecasts under inspection in this thesis is serial dependence of exceptions (i.e. violations). As described in section 4, there are multiple ways of examining this possible bunching of the exceptions. One might check for autocorrelation at multiple lags in the indicator functions, $I(\alpha)$, time series or alternatively use specific test statistics designed to capture this time dependence, such as the likelihood ratio test statistics for independence and conditional coverage. However, as mentioned in the previous section, checking for autocorrelation at multiple lags quickly becomes cumbersome when lots of data are used and the *LR* statistics are only able to capture one lag in their calculations. Therefore, in this case, neither of these methods is preferable and an alternative approach is implemented. That is, bunching of the exceptions is addressed by observing the out-of-sample quantile loss plots, from which bunching (at any lag) is relatively easy to detect.

Appendices 5–7 plot the out-of-sample loss values with α = 0,05 for the United States every time $t = R + 1, . . . , T$. This is the summand of the out-of-sample average loss $\hat{Q}_P(\alpha)$ in equation (4.4). The figures with α = 0,01 deliver largely the same results and are therefore suppressed for the sake of space. Furthermore, the figures for the other countries (Germany, Finland and India) exhibit similar features as for the United States and are also suppressed. Note that a larger weight 0,95 is given to the loss when a violation occurs and thus when there is a violation the loss value increases simultaneously generating a spike in the plots. Therefore, the spikes in the plots in appendices 5–7 represent violations making it relatively easy to observe possible clustering of exceptions.

An overall observation of these plots reveals that in most cases the VaR models generate large loss values (i.e. large spikes in the plots) at the same times, whereas moderate loss values (i.e. a violation occurs but the forecast error is not that large, perceived as small spikes in the plots) are not that universal across all models. Consider, for example, the risk forecasts for the United States on period 1 (appendix 5). All VaR models exhibit two large spikes during this period at the same times, whereas the smaller spikes are not observed for every model. A closer look also reveals that even if a spike occurs it is not of the same magnitude for each model. Naturally, a smaller spike is preferable as it indicates that even though a violation occurred the forecast error was not that large. Overall, the plots in appendices 5–7 indicate that the VaR models behave rather similarly in terms of when violations actually occur.

To address the issue of bunching observing these plots, episodes where many spikes occur within a small amount of time are identified. Even though conclusions cannot be made on what is an acceptable degree of bunching this still allows for a sensible comparison across models (i.e. is this bunching observed across all models or are some models able filter possible autocorrelation in the second moment?). As α = 0,05 for all the plots in appendices 5–16 one would expect to observe around 25 evenly spread violations per plot. Episodes of significant bunching can be observed at least in appendices 5 and 6. In most cases, when bunching is identified for one VaR model it is also present for the rest of the models (see e.g. the second half of 1998 in appendix 5). Only the magnitude of this bunching varies across models (compare e.g. HS and GPD* models at the end of 1998 in appendix 5). However, occasions where some models exhibit bunching and others do not are rather scarce. In appendices 5–7 such incidents can be identified in the favour of EVT models in appendix 6 (at the beginning and end of 2000). As mentioned these differences, however, are rare and do not allow for systematic discrimination among the studied VaR models based on their ability to filter serial dependence from the second moment of the return series. Next, given the above findings, overall performance of the tested VaR models across all three out-of-sample periods is considered.

5.2.4 Overall performance

The above presented findings suggest using empirical coverage probability as the primary comparison criteria for model choice (as it showed most variation in models' performance, not to mention it measures the same aspect of model performance as the Basel Committee's framework) and average quantile loss (as it measures a slightly different aspect of model performance and showed sufficient differences between models to allow for their sensible ordering). No significant differences were observed in models' performance in terms of ability to filter the serial dependence in the second moment. Therefore, bunching of exceptions as a comparison criteria is dropped out from the remaining analysis. Hence, concluding remarks are given considering the combined performance of models during all three out-of-sample periods in terms of tail coverage and quantile loss. First, each approach (and the models pertaining to this particular approach) is bencmarked against the RiskMetrics model. After this, the overall best VaR models are identified and thoroughly examined, separately for α = 0,05 and α = 0,01.

First, the models within the variance-covariance approach (VC, VC(GARCH) and VC*) are benchmarked against the RiskMetrics model. Note that RiskMetrics also pertains to the variance-covariance approach, and hence differences between these models are simply due to differences in volatility estimation (as all the VC models in this thesis are built upon the normality assumption). With α = 0,05 the RiskMetrics model performs satisfactorily 11 times out of 12 during all three periods according to Kupiec's (1995) likelihood ratio test statistic for unconditional coverage (tables 12–14). The corresponding results are 6, 7 and 9 times for the VC, VC(GARCH) and VC*, respectively. The overall average quantile loss for RiskMetrics is 2,27% compared to 2,38%, 2,28% and 2,38% for VC, VC(GARCH) and VC*, respectively (tables 15–17). Therefore, it seems that none of the considered VC models is able to match RiskMetrics' performance with α = 0,05 as it produces the most satisfactory tail coverage and lowest average quantile loss. Out of the considered VC models (RiskMetrics excluded) VC* appears most appealing. Further, this indicates that filtering the retuns series might prove to be a useful within the VC approach as VC(GARCH) and VC* implement the same GARCH(1,1) volatility model but in different manner. With α = 0,01 RiskMetrics produces satisfactory tail estimates 7 times out of 12 (tables 12–14) and 0,73% average quantile loss (tables 15–17). VC, VC(GARCH) and VC* produce 3, 6 and 7 satisfactory tail estimates and 0,82%, 0,75% and 0,76% average quantile losses, respectively. Therefore, with α = 0,01 it seems that RiskMetrics still dominates but it is closely challenged by VC(GARCH) and VC*. Overall, it is obvious that the simple VC model performs the poorest proving that extensions to this simple model are useful. Out of the two considered extensions filtering (VC*) appears somewhat more useful than a straightforward implementation of a volatility model (VC(GARCH)). Still, neither of these extensions performs as well as RiskMetrics.

Next, the models within the historical simulation approach (HS, HS(GARCH) and HS*) are benchmarked against RiskMetrics. With α = 0,05 the HS, HS(GARCH) and HS* models produce 5, 11 and 11 satisfactory tail estimates out of 12 and average quantile losses of 2,40%, 2,37% and 2,37%, respectively. The corresponding results for RiskMetrics were 11 and 2,27%. Hence, HS(GARCH) and HS* match RiskMetrics' performance in terms of tail coverage but are somewhat inferior in terms of average quantile loss. This difference, however, is very small. With α = 0,01 the HS, HS(GARCH) and HS* models generate 3, 10 and 10 satisfactory tail estimates out of 12, compared to 7 by RiskMetrics, and 0,79%, 0,76% and 0,76% average quantile losses, compared tro 0,73% by RiskMetrics. Therefore, it seems that with α = 0,01 the HS(GARCH) and HS* models beat RiskMetrics in terms of tail coverage but not in terms of average quantile loss. Differences in average quantile losses are only marginal and, hence, it can be concluded that HS(GARCH) and HS* are superior to RiskMetrics. Overall, the simple HS model performs the worst in terms of both tail coverage and average quantile loss and seems inferior to RiskMetrics. HS's performance, however, can be improved by incorporating volatility updating into the model. This thesis considers two distinct approaches, HS(GARCH) and HS*, to allow for volatility updating within the HS approach. These extension appear useful and perform as well as the RiskMetrics model with α = 0,05 and even beat its performance with α = 0,01. Between these two methods allowing for volatility updating within the HS approach no winner can be identified as the results produced by these two models (HS(GARCH) and HS*) are almost identical. The HS* model, however, is more simple in terms of required programming, hence tilting the scale towards its benefit.

This thesis considered two models within the Monte Carlo simulation approach: the simple Monte Carlo simulation (with normal distribution and historic volatility) and an extension which utilizes GARCH(1,1) in its volatililty estimates. Both of these models were estimated in mean-zero framework, except for the drift parameter in the stochastic simulation process for which mean was estimated by historic average. Therefore, it would be expected that the estimates of MC and MC(GARCH) would not differ much from those of the VC and VC(GARCH) as they both rely heavily on the normality assumption. With α = 0,05 MC and MC(GARCH) produce 5 and 6 times satisfactory tail coverage out of 12, according to Kupiec's (1995) likelihhod ratio for unconditional coverage, and 2,40% and 2,30% average quantile losses, respectively. As expected, these results are very close to those of VC and VC(GARCH). Corresponding results for RiskMetrics were 11 and 2,27%. Therefore, it seems that MC(GARCH) beats MC, but it is no match for RiskMetrics. With α = 0,01 the MC and MC(GARCH) models produce 2 and 6 satisfactory tail estimates and 0,84% and 0,76% average quantile losses, respectively. These results, once again, are very close to those of the VC and VC(GARCH) models and no match for RiskMetrics (7 and 0,73%). Therefore, overall, it seems that the Monte Carlo simulation VaR method (MC) can be enhanced by implementing GARCH(1,1) instead of equally weighted moving average (i.e. historic) volatility. This extension, however, cannot match the performance of RiskMetrics.

Finally, the models within the extreme value theory approach (GPD and GPD*) are benchmarked against RiskMetrics. While both of these models implement the generalized Pareto distribution to account for the heavy tails of the retutn distribution the GPD* is implemented on the filtered return series and GPD directly on the return series. With α = 0,05 GPD and GPD* produce 7 and 3 times out of 12 satisfactory tail coverage and 2,51% and 2,82% averge quantile losses, respectively. Analogous results for RiskMetrics were 11 and 2,27%. Therefore, neither of these models is able to challenge RiskMetrics with α = 0,05. With α = 0,01 GPD and GPD* produce 10 and 9 satisfactory tail estimates and average quantile losses of 0,76% and 0,91%, respectively, compared to 7 and 0,73% by RiskMetrics. Thus, both GPD and GPD* beat RiskMetrics in terms of tail coverage but not in terms of average quantile loss. The difference in average quantile loss between GPD and RiskMetrics is only marginal but GPD*'s average quantile loss is noticeably higher than GPD's and RiskMetrics'. Therefore, though GPD* is able to beat RiskMetrics in terms of tail coverage with α = 0,01 it does not prove to be a useful extension as it does not perform better than simple GPD. Overall, filtering does not appear useful in the context of EVT based VaR models as GPD* is outperformed by GPD. Further, GPD appears to seriously challenge RiskMetrics as the VaR confidence level is raised: with α = 0,05 RiskMetrics still performs better but with α = 0,01 GPD already dominates. Next, attention is turned to those models that were perceived as best in the above considerations.

5.2.5 Further inspection of superlative models

Most satisfactory models with α = 0,05 were identified among RiskMetrics (11 satisfactory tail forecasts out of 12), HS(GARCH) (11), HS* (11) and VC* (9). While RiskMetrics produced the lowest average quantile loss, no significant differences could be observed in these models' average quantile losses. To further examine the behaviour of these models, figure 8 below plots these models' out-of-sample VaR forecasts. The results of HS(GARCH) and HS* are so indistinguishable that HS(GARCH) is suppressed from figure 8 and HS* can be seen to represent both HS* and HS(GARCH). Further, as all periods and countries deliver largely similar results only forecasts for the United States during period 1 are presented in figure 8 for the sake of space.

Figure 8 Superlative VaR models' out-of-sample forecasts, α = 0,05, period 1

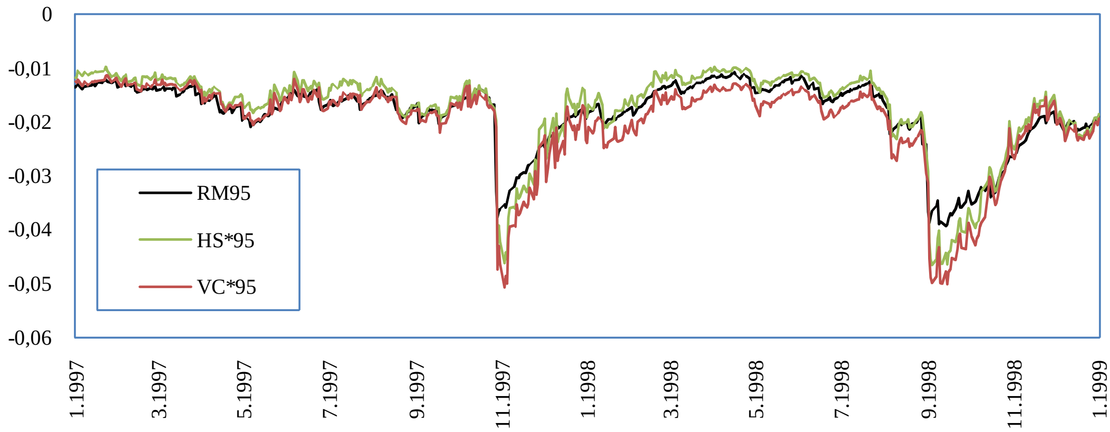


Figure 8 shows that the best models with α = 0,05 (i.e. RiskMetrics, HS(GARCH), HS* and VC*) behave in a rather similar manner. On a more tranquil period they all give quite low tail estimates and when a spike in the losses occurs they all react simultaneously to this. There are though some minor differences in how strong the models react to large negative returns: RiskMetrics appears to be somewhat more conservative in its reactions than HS(GARCH), HS* and VC*. This difference, however, is barely noticeable. Overall, it appears that there are no fundamental disparities in the way that these models react to returns and create their forecasts with α = 0,05. As RiskMetrics and VC* are built upon the normality assumption with time-varying volatility (EWMA and GARCH(1,1) it seems that this combination and its assumptions are sufficient for estimating the loss that is not expected to be exceeded 95% of the time, i.e. the 0,05 quantile. The empirical distribution with time-varying volatility (HS(GARCH) and HS*) also appears very appealing for VaR forecasts at this confidence level.

To see whether the best models with α = 0,01 also exhibit similar features, the superlative models with this particular confidence level are examined next. Most satisfactory models with α = 0,01 were identified among HS(GARCH) (10 satisfactory tail forecasts out of 12), HS* (10) and GPD (10). These three models also generated the same average quantile loss values (0,76%). While GPD* also performed well (9 satisfactory tail forecasts out of 12) it did not perform better than GPD and therefore was not considered as a prominent extension. Figure 9 below plots the out-of-sample forecasts of these VaR models for the United States with α = 0,01 during period 1, while periods 2 and 3 deliver largely similar results and are thus suppressed. The results of HS(GARCH) and HS* are once again so indistinguishable that HS* represents both of these models. Further, though RiskMetrics was not among the best performers with α = 0,01 (7 satisfactory tail forecasts out of 12) it is stil plotted in figure 9 as it represents the benchmark model in this thesis.

Figure 9 Superlative VaR models' out-of-sample forecasts, α = 0,01, period 1

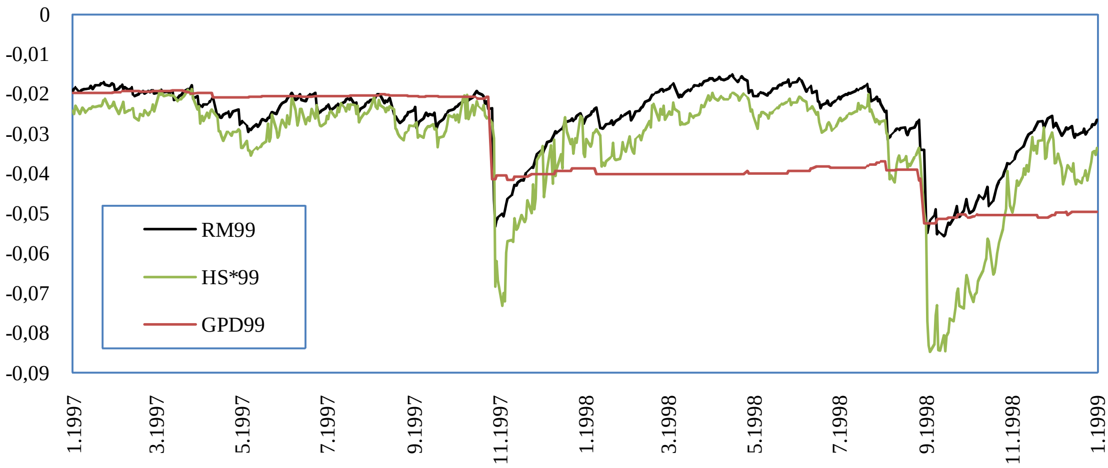


Figure 9 reveals that RiskMetrics probably does not perform as well as the best models with α = 0,01 because it generates too low VaR estimates: HS* (and HS(GARCH)) and GPD almost universally produce higher VaR estimates than RiskMetrics. Therefore, it appears that the normality assumption, even if coupled with time-varying volatility, is not reasonable when estimating the loss that is expected to be exceeded only 99% of the time (i.e. the 0,01 quantile of the return distribution). Other distributions and methods appear to be more accurate with such extreme quantiles. In this case the generalized Pareto distribution (GPD) and empirical distribution coupled with volatility updating (HS(GARCH) and HS*) produce most satisfactory estimates of the 0,01 quantile. Another interesting feature revealed by figure 9 is the divergent behaviour of GPD compared to HS* and RiskMetrics: it produces much more stable and generally higher VaR forecasts than its competing models. This, of course, is merely a reflection of the assumptions underlying GPD VaR approach, as it basically assumes a stable distribution that does not change over time.

5.2.6 Robustness of the results

Robustness checks with different length out-of-sample periods were conducted to verify the robustness of the above presented results to varying (before, during and after crisis) period explications. Different two- and three-year periods defined as the before, during and after crisis periods delivered largely the same results as the above presented results with period definitions according to table 2. Results from one-year out-of-sample periods were too volatile to allow for a sensible division of the data into periods reflecting prevailing market conditions.

5.3 Conclusions

With respect to periods of turmoil, represented in this thesis by the technology boom and its collapse at the turn of the millennium, and risk measurement by Value at Risk this thsesis provides evidence that in the precence of turbulence most VaR models tend to react rather satisfactorily to this elevated risk regime. However, when the turbulence is extensive enough some VaR models experinse difficulties in risk forecasting. This can be seen in differing average quantile loss values for emerging markets (Finland and India) between the before and after crisis periods. Similar deductions could not, however, be made when utilizing empirical coverage probability as performance measure. Results concerning VaR models performance on a tranquil period that is preceded by a turbulent period indicate that the preceeding turbulence causes significant lag effects on risk forecasts, therefore causing most VaR models to overpredict risks. This suggests that while most VaR models react swiftly to regimes of higher risk, reactions to low risk regimes are not as expeditious. This, of course, is not a severe problem as overpredicting risk in VaR setting is not as fatal as underpredicting risk. Nevertheless, testing VaR model predictive performance under even more severe market crisis periods might offer more insightful information on VaR models ability to perform satisfactorily under serious turbulence.

With respect to predictive performance of different VaR approaches and their extensions the results can be described as follows. The considered extensions – filtering the returns series or possibly a more direct utilization of GARCH(1,1) volatility process – were discovered to be advantageous within all approaches with one exception. More precisely, the extreme value theory where the generalized Pareto distribution is applied directly on the return series outperformed applying the generalized Pareto distribution on the filtered return series. Within the variance-covariance VaR approach filtering appeared somewhat more useful than a straightforward implementation of GARCH(1,1). Still, neither of these extensions to the variance-covariance approach outperformed RiskMetrics, the benchmark model in this thesis. Within the historical simulation VaR approach filtering the return series and an alternative method that allows for volatility updating in the historical simulation approach were found to be equally advantageous. Further, both of these extensions matched RiskMetrics' performance with $\alpha = 0{,}05$ and even outperformed it with $\alpha = 0{,}01$. Within the Monte Carlo simulation VaR approach a straightforward utilization of GARCH(1,1) proved advantageous but not even with this extension could the Monte Carlo simulation approach challenge RiskMetrics. Finally, the extreme value theory approach to VaR was implemented by the generalized Pareto distribution. As mentioned, filtering did not prove advantageous for this particular approach, but the simple approach was found to outperform RiskMetrics with $\alpha = 0{,}01$.

The superlative models with α = 0,05 were identified among RiskMetrics, HS(GARCH), HS* and VC*, and no single model appeared to outperform the benchmark model (RiskMetrics). Thus, it appears that with moderate VaR confidence levels (i.e. e.g. 0,05) RiskMetrics is a respectable choice to base ones VaR framework on. Alternatilvely one could consider building one's VaR models according to historical simulation with filtering or volatility updating (HS* or HS(GARCH), or filtered variance covariance approach (VC*). When considering more extreme confidence levels (i.e. e.g. 0,01 as required e.g. by the Basel Committee for measuring market risk) the above picture undergoes some changes: RiskMetrics no longer appears as a conceivable foundation for one's VaR model. Alternative models appear to produce more accurate risk predictions. Specifically, with α = 0,01 the HS(GARCH), HS* and GPD models perform superiorly to this benchmark. This implies that the normality assumption gradually loses its accuracy and usability as VaR confidence level is raised, thus making other approaches more appealing.

These findings are in line with extant studies that have showed that extensions to the HS approach (filtering and volatility updating) outperform the simple HS approach and approaches within the VC family (Hull & White 1998; Barone-Adesi et al. 2002). This thesis contributes to these extant studies by validating this outperformace of extended HS against a wider variety of VaR models with two distinct confidence levels (α = 0,05 and α = 0,01) and data that include a crisis period. Further, the finding on the prominence of the extreme value theory approach with high confidence levels is also in line with extant studies. This thesis contributes by considering a wide variety of competing models and data that include a crisis period. Also, it has been very common in previous studies to implement the Basel backtesting scheme (see e.g. Kalyvas & Sfetsos 2006) and, thus, this thesis contributes by considering alternative approaches in verifying VaR models predictive performance. Especially the good performance of the EVT approach in the Basel backtesting framework can be driven by on average higher VaR forecasts, as this framework only penalizes for underpredictions of VaR. Thus, verifying the EVT approaches performance with alternative performance measures offers important confirming information.

The prominence of the GPD model still has to be interpreted with some caution. As this thesis and previous studies have shown, the forecasts of the extreme value theory VaR approach are on average more conservative than those of alternative approaches. Thus, before implementing it for market risk measurement, one has to thoroughly examine the effectiveness of this method by determining whether the lower multiplicative factor for capital requirements is counter-balanced by this conservatism exhibited in its VaR forecasts. In fact, it seems that this aspect has been considered by financial institutions and deters them from implementing approaches based on the extreme value theory (see e.g. Rogachev 2007).

The impacts of the market crisis period on risk forecasting performance of the considered VaR models were not as strong as might have been anticipated. Thus, further studies covering periods of even more extensive turbulence – such as the Asian financial crisis of 1997–1998 or the Sub-prime crisis of 2007–2008 – might provide interesting (and deviating) results from those of this thesis. Nevertheless, this thesis showed some evidence on deviating performance between VaR models contingent upon market conditions. Thus, it might prove useful to consider regime-switching VaR models that are based on regime-switching Markov processes (Tsay 2005, 588–594; Guidolin & Timmermann, 2004; Li & Lin, 2004). Further, the validity of the results gained from this thesis could be tested on portfolios containing nonlinear instruments, thus emphasisng the adaptability of different VaR approaches. Also, extending the study into a multivariate setting might unveil some important issues that might complicate the adaptation of the VaR models tested in this thesis into real life portfolios.

6 SUMMARY

As the financial markets have continuously become more volatile and complex, reliable risk measurement techniques are required to properly address the increased risks. Value at Risk (VaR) provides an easy-to-understand method for quantifting market risk and, thus, it has become the industry standard for the financial sector. The triumphal march of VaR can, to some extent, be attributed J. P. Morgan, that was one of the first exponents of VaR as it released its RiskMetrics database in 1994, and the Basel Committee, that explicitly allowed for VaR to be used in determining banks' minimum capital requirements since 1996. Despite the criticism that VaR has received, mainly based on the fact that VaR is not a coherent measure of risk, it still continues to gain ground on other sectors besides the financial as well.

During a short span of time a number of papers have studied various aspects of VaR methodology. The recent research in this field has progressed so rapidly that comparing the relative predictive performance of various traditional and novel VaR models has not yet been fully matched. This thesis contributes to this particular problem by bechmarking the predictive performance of various VaR models within the variance-covariance, historical simulation, Monte Carlo simulation and extreme value theory approaches to the RiskMetrics model, thus providing valuable informations for risk practitioners and regulators. The out-of-sample period is chosen in a way that it encompasses a market crisis period – the bursting of the technology bubble at the turn of the millennium. Hence, this thesis also contributes by examining the effects of market turbulence on risk forecasting precision by VaR models. Furthermore, the comparison methods for discriminating among various VaR models in this thesis (empirical coverage probability, predictive quantile loss and clustering of violations) are more thorough than the backtesting framework suggested by the Basel Committee, thus completing the results of extant studies that solely rely on the Basel framework.

Within the tested VaR models this thesis considers three different distributions – the normal, empirical and generalized Pareto distributions – as their use can robustly be rationalized. Thus, distributions such as the Student's-t that lack strong theoretical foundations are excluded from this thesis. With respect to volatility modelling, an extremely important component in VaR calculation, this thesis also applies rather strict restrictions: to be included in this thesis a volatility model is required to be simple and tractable enough to be available and within the grasp of risk practitioners. Hence, this thesis only considers equally and exponentially weighted moving averages (EQMA and EWMA) and the simplest model within the GARCH family: GARCH(1,1).

Given these restrictions alternative VaR models are constructed and out-of-sample tests are conducted on three periods between 1997 and 2004, more specifically on before, during and after crisis periods, utilizing data from four national stock indices that were all impacted by the chosen market crisis relatively simultaneously. The selected stock indices are S&P 500 Composite (United States), CDAX General (Germany), HEX Geneeral (Finland) and BSE National (India). Furthermore, the United States and Germany were chosen to represent developed markets, whereas Finland and India proxy for emerging markets. This conveniently allows for observing the performance of VaR models in different market contexts.

The results of this thesis with respect to periods of turmoil and risk measurement via Value at Risk indicate that while most VaR models react swiftly to high risk regimes, reactions to low risk regimes are not as expeditious. This, of course, is not a severe problem as overpredicting risk in VaR setting is not as fatal as underpredicting risk. With respect to predictive performance of different VaR models the results can be described as follows. With 95% VaR confidence level no single model appeared to outperform the benchmark model, RiskMetrics. However, as the VaR confidence level was raised to 99% three models were found to clearly outperform RiskMetrics. These particular models were historical simulation with volatility updating or filtered return series and an extreme value theory based model utilizing the generalized Pareto distribution. This indicates that the normality assumption, made for example by the RiskMetrics model, gradually loses its accuracy and usability as VaR confidence level is raised, thus making other VaR approaches, such as extensions to the historical simulation or extreme value distributions, more appealing. In general, the considered extensions in this thesis were found to be efficient compared against the basic versions of these VaR approaches. However, with the three above mentioned exceptions, these extensions hardly ever were found to outperform the RiskMetrics model.

The significance of the information provided by this thesis is most pronounced for financial institutions that are contemplating between different VaR approaches and models. In addition, companies on other industries that are interested in market risk measurement, for example companies whose performance is highly dependent on the price development of certain raw materials, might also find the information provided by this thesis very useful. Whether it is an entirely new risk measurement system or an update to an existing system, the results of this thesis offer guidelines for choosing among various competing VaR models. However, the results of this thesis have to be interpreted with some caution: the results were derived in a univariate setting (single asset portfolios), which hardly ever is the case in real life. Therefore, trying to implement some of the models tested in this thesis into real life portfolios, consisting possibly of hundreds or even thousands of assets, might unveil some unanticipated problems. Thus, it would be useful to test the validity and applicability of the gained results in a multivariate setting (multiple asset portfolios). Furthermore, it might also prove useful to try to reproduce the results of this thesis with portfolios containing derivatives, thus emphasizing the ability of different VaR approaches to handle nonlinear instruments. Nevertheless, these restrictions considered, the results of this thesis still offer valuable information for risk practitioners and regulators concerning the predictive performance of various competing VaR models.

REFERENCES

Alexander, Carol (2001) *Market models*. John Wiley & Sons: Chichster.

Alexander, Carol – Leigh, C. (1997) On the covariance matrices used in VaR models. *Journal of Derivatives*, Vol. 4, No. 3, 50–62.

Artzner, Philippe – Delbaen, Freddy – Eber, Jeanne-Marc – Heath, David (1999) Coherent measures of risk. *Mathematical Finance*, Vol. 9, No. 3, 203–228.

Baillie, Richard – Bollerslev, Tim (1992) Prediction in dynamic models with time-dependent conditional variances. *Journal of Econometrics*, Vol. 52, No. 1–2, 91–113.

Balkema, A. A. – de Haan, Laurens (1974) Residual lifetime at great age. *Annals of Probability*, Vol. 2, No. 5, 792–804.

Bank for International Settlements (1994) *Public disclosure of market and credit risks by financial intermediaries*. Euro-currency standing committee of the central banks of the Group ten countries (Fisher report).

Bao, Yong – Lee, Tae-Hwy – Saltoglu, Burak (2006) Evaluating predictive performance of Value-at-Risk models in emerging markets: A reality check. *Journal of Forecasting*, Vol. 25, No. 2, 101–128.

Barone-Adesi, G – Giannopoulos, K – Vosper, L (2002) Backtesting derivative portfolios with filtered historical simulation (FHS). *European Financial Management* Vol. 8, No. 1, 31–58.

Basel Committee on Banking Supervision (1994) *Risk management guidelines for derivatives*. BIS, Basel, Switzerland.

Basel Committee on Banking Supervision (1995) *An internal model-based approach to market risk capital requirements*. BIS, Basel, Switzerland.

Basel Committee on Banking Supervision (1996a) *Supervisory framework for the use of "backtesting" in conjunction with the internal models approach to market risk capital requirements*. BIS, Basel, Switzerland.

Basel Committee on Banking Supervision (1996b) *Amendment to the Basel capital accord to incorporate market risk*. BIS, Basel, Switzerland

Basel Committee on Banking Supervision (2005) *Basel II: International convergence of capital measurement and capital standards; A revised framework*. BIS, Basel, Switzerland.

Baumol, William (1963) An expected gain-confidence limit criterion for portfolio selection. *Management Science*, Vol. 10, No. 1, 174–182.

Bekiros, Stelios – Georgoutsos, Dimitris (2005) Estimation of value-at-risk by extreme value and conventional methods: a comparative evaluation of their predictive performance. *International Financial Markets, Institutions and Money*, Vol. 15, No. 3, 209–228.

Berkowitz, Jeremy – O'Brien, James (2002) How accurate are Value-at-Risk models at commercial banks? *The Journal of Finance*. Vol. LVII, No. 3, 1093–1111.

Berlant, J. – Teugels, J. L. – Vynckier, P. (1996) *Practical analysis of extreme values*. Leuven University Press: Leuven.

Berman, Simeon (1964) Limiting theorems for the maximum term in stationary sequences. *Annals of Mathematical Statistics*, Vol. 35, No. 2, 502–516.

Black, Fisher – Scholes, Myron (1973) The pricing of options and corporate liabilities. *Journal of Political Economy*, Vol. 81, No. 3, 637–659.

Blattberg, Robert C. –Gonedes, Nicolas J. (1974) A comparison of the stable and student distributions as statistical models for stock prices. *Journal of Business*, Vol. 47, No. 2, 244–280.

Bollerslev, Tim (1986) Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, Vol. 31, No. 3, 307–327.

Brodie, Mark – Glasserman, Paul (1998) Simulation for option pricing and risk management. In: *Risk management and analysis*, Vol. 1, ed. by Alexander, Carol. John Wiley & Sons: New York.

Brooks, Chris (1998) Predicting stock index volatility: Can market volume help? *Journal of Forecasting*, Vol. 17, No. 1, 59–80.

Brooks, Chris – Persand, Gita (2002) Model choice and Value-at-Risk performance. *Financial Analysts Journal*, Vol. 58, No. 5, 87–97.

Boudoukh, Jeremy – Richardson, Matthew – Whitelaw, Robert (1998) The best of both worlds: A hybrid approach to calculating Value at Risk. *Risk*, 11 (May), 64 – 67.

Bremmer, Ian (2005) Managing risk in an unstable world. *Harward Business Review*, Vol. 83, No. 6, 51–62.

Butler, J. S. – Schachter, Barry (1998) Estimating value at risk with a precision measure by combining kernel estimation with historical simulation. *Review of Derivatives Research*, Vol. 1, No. 4, 371–390.

Cootner, P. H. (1964) *The random character of stock market prices*. MIT: Cambridge.

Cox, D. R. – Hinkley, D. V. (1974) *Theoretical statistics*. Chapman and Hall, London.

Cox, John C. – Ross, Stephen A. (1976) The valuation of options for alternative stochastic processes. *Journal of Financial Economics*, Vol. 3, No. 1–2, 145–166.

Christoffersen, Peter F. (1998) Evaluating interval forecasts. *International Economic Review*, Vol. 39, No. 4, 841–862.

Crnkovic, C. – Drachman, J. (1995) A universal tool to discriminate among risk measurement techniques. Mimeo, Corporate Risk Management Group, JP Morgan.

Davidson, A. C. – Smith, R. L. (1990) Models for exceedances over high thresholds (with discussion). *Journal of the Royal Statistical Society Series B*, Vol. 52, No. 3, 393–442.

Derivatives Policy Group (1995) *Framework for voluntary oversight*. Derivatives Policy Group, New York.

Doherty, Neil A. (2000) *Integrated risk management: Techniques and strategies for managing corporate risk*. McGraw-Hill: New York.

Duffie, Darrell – Pan, Jun (1997) An overview of Value at Risk. *Journal of Derivatives*, Vol. 4, No. 3, 7–49.

DuMouchel, William H. (1983) Estimating the stable index α in order to measure tail thickness: A critique. *The Annals of Statistics*, Vol. 11, No. 4, 1019–1031.

Eberlein, Ernst – Keller, Ulrich (1995) Hyperbolic distributions in finance. *Bernoulli*, Vol. 1, No. 3, 281–299.

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

Engle, Robert (1982) Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, Vol. 50, No. 4, 987–1007.

Engle, Robert – Ng, Victor (1993) Measuring and testing the impact of news on volatility. *Journal of Finance*, Vol. 48, No. 5, 1749–1778.

European Union (2006) Directive 2006/48/EC and Directive 2006/49/EC (Capital requirements directive). *Official Journal of the European Union*. Vol. 49, L 177, 1–255.

Fama, Eugene (1965) The behaviour of stock market prices. *Journal of Business*, Vol. 38, No. 1, 34–105.

Fama, Eugene (1970) Efficient capital markets: A review of theory and empirical work. *Journal of Finance*, Vol. 25, No. 2, 383–417.

Fan, Ying – Wei, Yi-Ming – Xu, Wei-Xuan (2004) Application of VaR methodology to risk management in the stock market in China. *Computers & Industrial Engineering*, Vol. 46, No. 2, 383–388.

Figlewski, Stephen (1997) Forecasting volatility. *Financial Markets, Institutions and Instruments*, Vol. 6, No. 2, 1–88.

Friedman, Benjamin M. – Laibson, David I. (1989) Economic implications of extraordinary movements in stock prices. *Brookings Paper on Economic Activity*, Vol. 1989, No. 2, 137–189.

Gnedenko, D. V. (1943) Sur la distribution limité du terme d’une série aléatoire. *The Annals of Mathematics*, Vol. 44, No. 3, 423–453.

González, G – Lee, T-H – Mishra, S (2004) Forecasting volatility: a reality check based on option pricing, utility function, value-at-risk, and predictive likelihood. *International Journal of Forecasting*, Vol. 20, No. 4, 629–645.

Gonzalo, J. – Olmo, J. (2004) Which extreme values are really extremes. *Journal of Financial Econometrics*, Vol. 2, No. 3, 349–369.

Granger, Clive W. J. (2002) Some comments on risk. *Journal of Applied Econometrics*, Vol. 17, No. 5, 447–456.

Group of Thirty Global Derivatives Study Group (1993) *Derivatives: Practices and principles*. Washington, D.C. (G-30 report)

Guidolin, Massimo – Timmermann, Alan (2004) Value at risk and expected shortfall under regime switching. EFA 2004 Maastricht Meetings Paper No. 2983. Available at SSRN: http://ssrn.com/abstract=557091.

Hansen, Peter R. – Lunde, Asger (2005) A forecast comparison of volatility models: Does anything beat a GARCH(1,1)? *Journal of Applied Econometrics*, Vol. 20, No. 7, 873–889.

Hendricks, Darryl (1996) Evaluating Value-at-Risk models using historical data. *Economic Policy Review*, April.

Hill, Bruce M. (1975) A simple general approach to inference about the tail of a distribution. *Annals of Statistics*, Vol. 3, No. 5, 1163–1173).

Hull, John C. (2007) *Risk management and financial institutions*. Pearson Prentice Hall: New Jersey.

Hull, John C. (2006) *Options, futures, & other derivatives*. 6th ed. Pearson Prentice Hall: New Jersey.

Hull, John – White, Alan (1998) Incorporating volatility updating into the historical simulation method for Value at Risk. *Journal of Risk*, Vol. 1, No. 1, 5–19.

Jackson, Patricia – Maude, David – Perraudin, William (1997) Bank capital and value-at-risk. *Journal of Derivatives*, Vol. 4, No. 3, 73–89.

Jackson, Patricia – Maude, David – Perraudin, William (1998) Testing value-at-risk approaches to capital adequacy. *Bank of England Quarterly Bulletin*, Vol. 38 (August), 256–266.

Jamshidian, Farshid – Zhu, Yu (1996) Scenario simulation: theory and methodology. *Finance and Stochastics*, Vol. 1, No. 1, 43–67.

Jansen, Dennis W. – de Vries, Casper G. (1991) On the frequency of large stock returns: Putting booms and busts into perspective. *The Review of Economics and Statistics*, Vol. 73, No. 1, 18–24.

Jarque, Carlos M. – Bera, Anil K. (1987) A test of normality of observations and regression residuals. *International Statistical Review*, Vol. 55, No. 2, 163–172.

Jenkinson, A. F. (1955) The frequency distribution of the annual maximum (or minimum) of meteorological elements. *Quarterly Journal of the Royal Meteorological Society*, Vol. 81, No. 348, 158–171.

Jorion, Philippe (2001) *Value at risk: The new benchmark for managing financial risk*. 2nd ed. McGraw-Hill: New York.

Jorion, Philippe (2007) *Value at risk: The new benchmark for managing financial risk*. 3rd ed. McGraw-Hill: New York.

J.P. Morgan (1996) *RiskMetrics technical document*. 4th ed. J.P. Morgan: New York.

Kahra, Hannu (2007) Risk and uncertainty [Riski ja epävarmuus]. *Turun Sanomat* 25.9.2007.

Kalyvas, Lampros – Sfetsos, Athanasios (2006) Does the application of innovative internal models diminish regulatory capital? *International Journal of Theoretical and Applied Finance*, Vol. 9, No. 2, 217–226.

Kasanen, Eero – Lukka – Kari – Siitonen, Arto (1993) The constructive approach in management accounting research. *Journal of Management Accounting Research*, Vol. 5, Fall 1993, 243–264.

Kasanen, Eero – Lundström, Thomas – Puttonen, Vesa – Veijola, Risto (1996) *Corporations' financial risks* [*Rahoitusriskit yrityksissä*]. Tammisaaren Kirjapaino Oy: Tammisaari.

Kendall, Maurice (1953) The analysis of economic time-series. *Journal of the Royal Statistical Society*, Vol. 116, No. 1, 11–34.

Knüpfer, Samuli – Puttonen, Vesa (2004) *Modern finance* [*Moderni rahoitus*]. Dark Oy: Vantaa

Koenker, Roger – Basset, Gilbert (1978) Regression quantiles. *Econometrica*, Vol. 46, No. 1, 33–50.

Koenker, Roger – Hallock, Kevin (2001) Quantile regression. *Journal of Economic Perspectives*, Vol. 15, No. 4, 143–156.

Kon, Stanley J. (1984) Models of stock returns – A comparison. *Journal of Finance*, Vol. 39, No. 1, 147–165.

Kupiec, Paul (1995) Techniques for verifying the accuracy of risk management models. *Journal of Derivatives*, Vol. 3, No. 2, 73–84.

Leadbetter, M. R. – Lindgren, G. – Rootzén, H. (1983) *Extremes and related properties of random sequences and processes*. Springer-Verlag: New York.

Li, Ming Yuan Leon – Lin, Hsiou-wei William (2004) Estimating value-at-risk via Markov switching ARCH models – an empirical study on stock index returns. *Applied Economics Letters*, Vol. 11, No. 11, 679–691.

Ljung, G. M. – G. E. P. Box (1978) On a measure of lack of fit in time series models. *Biometrica*, Vol. 65, No. 2, 297–303.

Longin, Francois (1996) The asymptotic distribution of extreme stock market returns. *The Journal of Business*, Vol. 69, No. 3, 383–408.

Longin, Francois (2000) From value at risk to stress testing: the extreme value approach. *Journal of Banking and Finance,* Vol. 24, No. 7, 1097–1130.

Lopez, Jose A. (1998) Methods for evaluating value-at-risk estimates. *FRBNY Economic Policy Review*, Vol. 4, 119–124.

Loretan, Mico – Phillips, Peter C.B. (1994) Testing the covariance stationarity of heavy-tailed time series. *Journal of Empirical Finance*, Vol. 1, No. 2, 211–248.

Luenberger, David (1998) *Investment science*. Oxford University Press: New York.

Lux, Thomas (1996) The stable Paretian hypothesis and the frequency of large returns: An examination of major German stocks. *Applied Financial Economics*, Vol. 6, No. 6, 463–475.

Lux, Thomas (1998) The socio-economic dynamics of speculative markets: Interacting agents, chaos, and the fat tails of return distributions. *Journal of Economic Behaviour & Organization*, Vol. 33, No. 2, 143–165.

Lux, Thomas – Ausloos, Marcel (2002) Market fluctuations I: Scaling, multiscaling, and their possible origins. In: *The science of disasters: Climate disruptions, heart attacks, and market crashes*, ed. by Bunde, A. – Kropp, J. – Schellnhuber, H, 352–371. Springer-Verlaug: Berlin.

Madan, Dilip B. – Carr, Peter P. – Chang, Eric C. (1998) The variance-gamma process and option pricing. *European Finance Review*, Vol. 2, No. 1, 79–105.

Mandelbrot, Benoit (1963) The variation of certain speculative prices. *Journal of Business,* Vol. 36, No. 4, 394–419.

Mantegna, Rosario – Stanley, Eugene (2002) Investigations of financial markets using statistical physics methods. In: *The science of disasters: Climate disruptions, heart attacks, and market crashes*, ed. by Bunde, A. – Kropp, J. – Schellnhuber, H, 352–371. Springer-Verlaug: Berlin.

Markowitz, Harry (1952) Portfolio selection. *Journal of Finance,* Vol. 7, No. 1, 77–91.

Merton, Robert C. (1973) Theory of rational option pricing. *Bell Journal of Economics and Management Science,* Vol. 4, No. 1, 141–183.

Merton, Robert C. (1976) Option pricing when underlying stock prices are discontinuous. *Journal of Financial Economics*, Vol. 3, No. 1–2, 125–144.

Mises, Richard, von (1936) La distribution de la plus grande de n valeurs. *Rev Math Union Interbalcanique*, Vol. 1, 141–160. (reproduced: (1964) Selected papers of Richard von Mises, Vol. II, *American Mathematical Society*, 271–294.)

Mittnik, Stefan – Paolella, Marc S. – Rachev, Svetlozar T. (2002) Stationarity of stable power-GARCH processes. *Journal of Econometrics*, Vol. 106, No. 1, 97–107.

Neftci, Salih (2000) Value at risk calculations, extreme events, and tail estimation. *Journal of Derivatives*, Vol. 7, No. 3, 23–38.

Nelson, Daniel B. (1991) Conditional heteroskedasticity in asset returns: a new approach. *Econometrica*, Vol. 59, No. 2, 347–370.

Partnoy, Frank (1999) *Fiasco: The inside story of a Wall Street trader*. Penguin: New York.

Pickands III, James (1975) Statistical inference using extreme order statistics. *Annals of Statistics*, Vol. 3, No. 1, 119–131.

Pritsker, M. (2000) *The hidden risk of historical simulation*. Federal Reserve Board: Washington DC.

Reiss, R.-D. – Thomas, M. (1997) *Statistical analysis of extreme values with applications in finance, hydrology and other fields*. Birkhäuser: Basel.

RiskMetrics Group (2006) *The RiskMetrics 2006 methodology*. RiskMetrics Group: New York.

Rogachev, Andrey (2007) Value-at-risk concept by Swiss private banks. *The Journal of Risk Finance*, Vol. 8, No. 1, 73–78.

Ross, Stephen (1976) The arbitrage theory of capital asset pricing. *Journal of Economic Theory*, Vol. 13, No. 3, 343–362.

Roy, Andrew (1952) Safety first and the holding of assets. *Econometrica*, Vol. 20, No. 3, 431–449.

Sentana, Enrique (1995) Quadratic ARCH models. *Review of Economic Studies*, Vol. 62, No. 4, 639–661.

Sharpe, William (1964) Capital asset prices: A theory of market equilibrium. *Journal of Finance*, Vol. 19, No. 3, 425–442.

Simons, Katherina (1996) Value at risk – New approaches to risk management. *New England Economic Review*, Sep/Oct, 3–14.

Smith, Richard L. (1989) Extreme value analysis of environmental time series: An application to trend detection in ground-level ozone (with discussion). *Statistical Science*, Vol. 4, No. 4, 367–393.

Snedecor, G. W. – Cochran, W. G. (1980) *Statistical Methods*, 7th ed. Iowa State University Press: Ames.

Stahl, Gerhard (2005) Why Value at Risk holds its own. *Euromoney*, Vol. 36, No. 430, 96–100.

Sullivan, Joe – Stoumbos, Zachary – Brooks, Robert (2007) Real-time assessment of Value-at-Risk and volatility accuracy. *Nonlinear Analysis: Real World Applications*, Vol.8, No. 1, 323–336.

Szegö, Giorgio (2005) Measures of risk. *European Journal of Operational Research*, Vol. 163, No. 1, 5–19.

Taylor, James (1999) Evaluating volatility and interval forecasts. *Journal of Forecasting*, Vol. 18, No. 2, 111–128.

Tsay, Ruey S. (2005) *Analysis of financial time series*. 2nd ed. John Wiley & Sons: New Jersey.

Tucker, Alan L. – Pond, Lallon (1988) The probability distribution of foreign exchange price changes: Tests of candidate processes. *Review of Economics and Statistics*, Vol. 70, No. 4, 638–647.

Venkatraman, S. (1997) Value at risk for a mixture of normal distributions: the use of quasi-Bayesian estimation techniques. *Economic Perspectives*, March 12–13, Federal Reserve Bank of Chigago.

Vlaar, Peter (2000) Value at risk models for Dutch bond portfolios. *Journal of Banking and Finance*, Vol. 24, No. 7, 1131–1154.

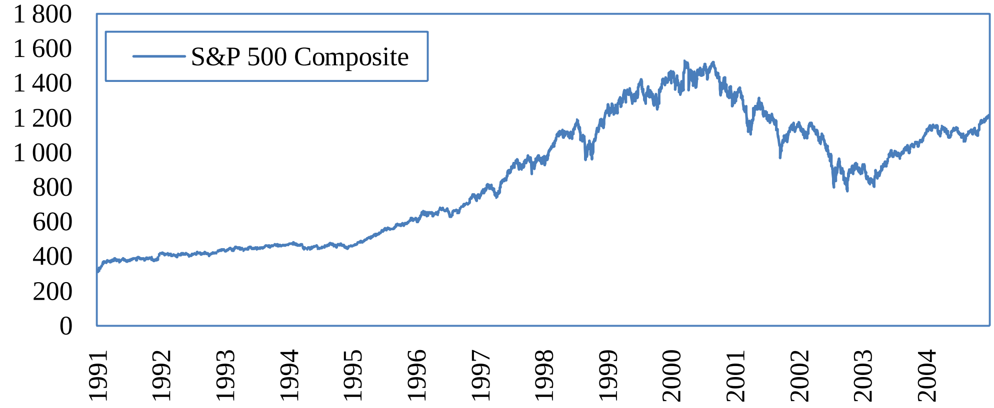
Wang, Kai-Li – Fawson, Chris – Barret, Christopher – McDonald, James (2001) A flexible parametric GARCH model with an application to exchange rates. *Journal of Applied Econometrics*, Vol. 16, No. 4, 521–536.

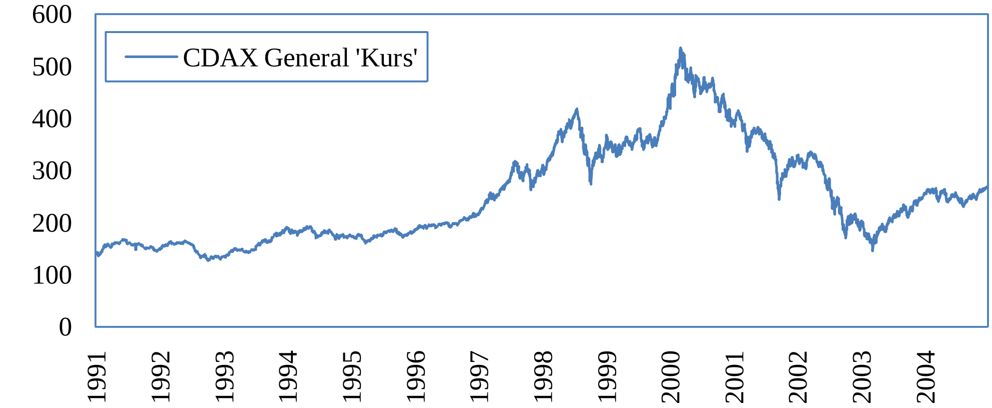
Wong, Michael – Cheng, Wai – Wong, Clement (2003) Market risk management of banks: Implications from the accuracy of Value-at-Risk forecasts. *Journal of Forecasting*, Vol. 22, No. 1,23–33.

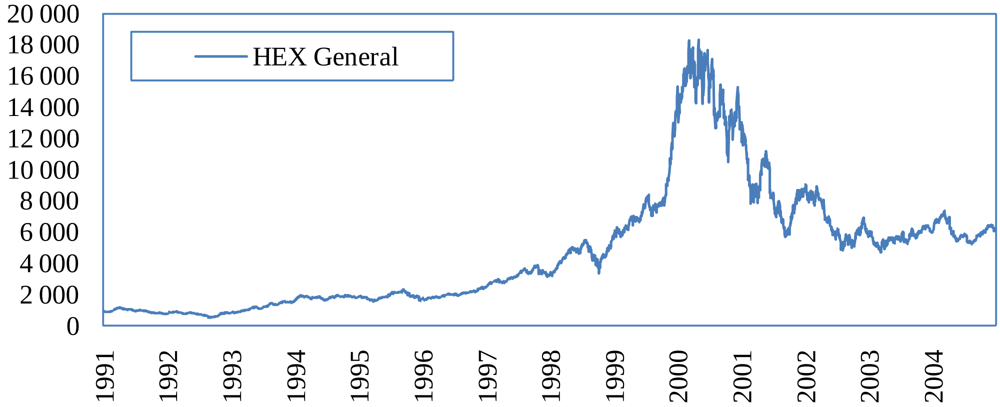
Zakoian, Jean-Michel (1994) Threshold heteroskedastic models. *Journal of Economic Dynamics and Control*, Vol. 18, No. 5, 931–955.

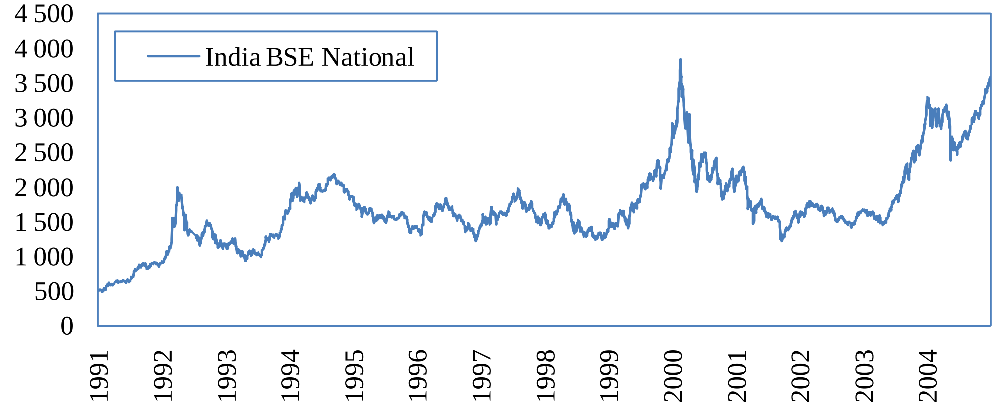
Zielinski, Ryszard (2005) Estimating quantiles with linex loss function. Applications to VaR estimation. *Applicationes Mathematicae*, Vol. 32, No. 4, 367–373.

APPENDIX 1 MARKET INDICES' TIME SERIES

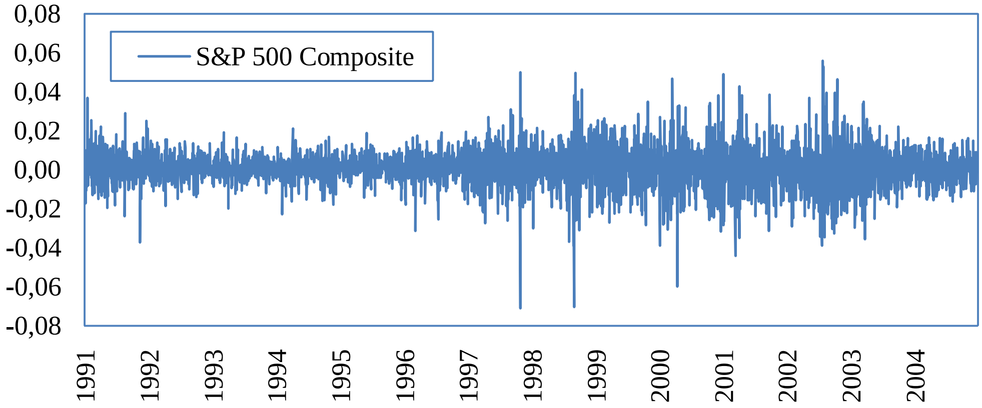


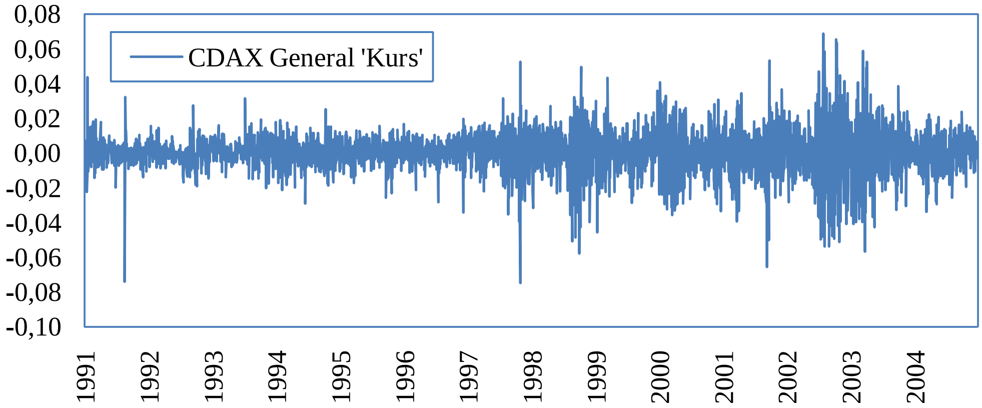


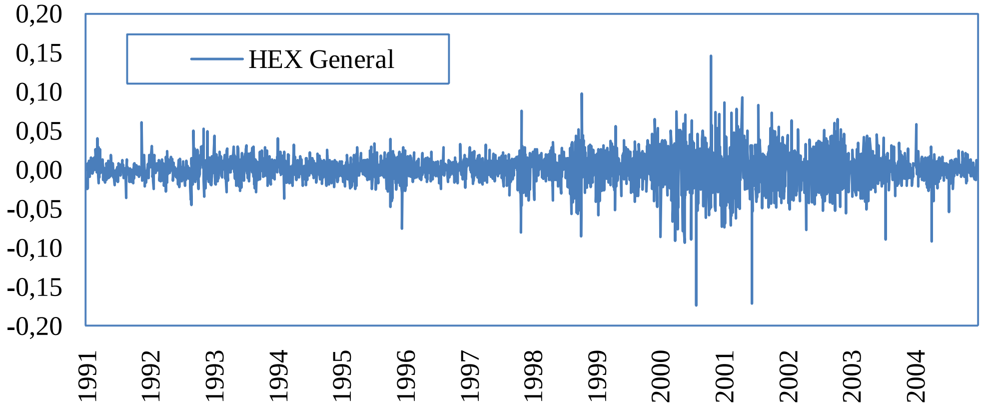


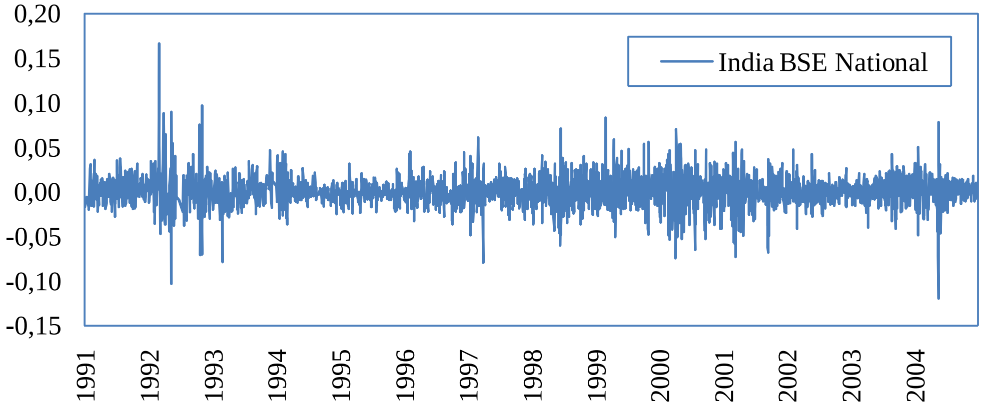


APPENDIX 2 MARKET INDICES' RETURN SERIES

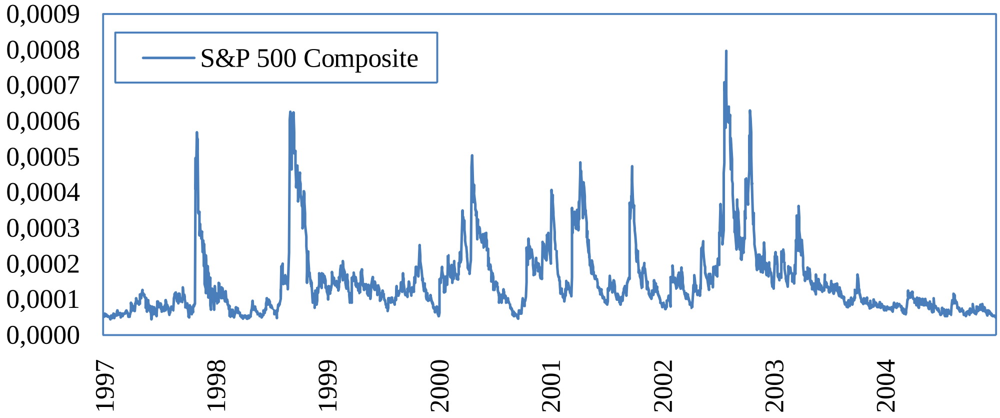


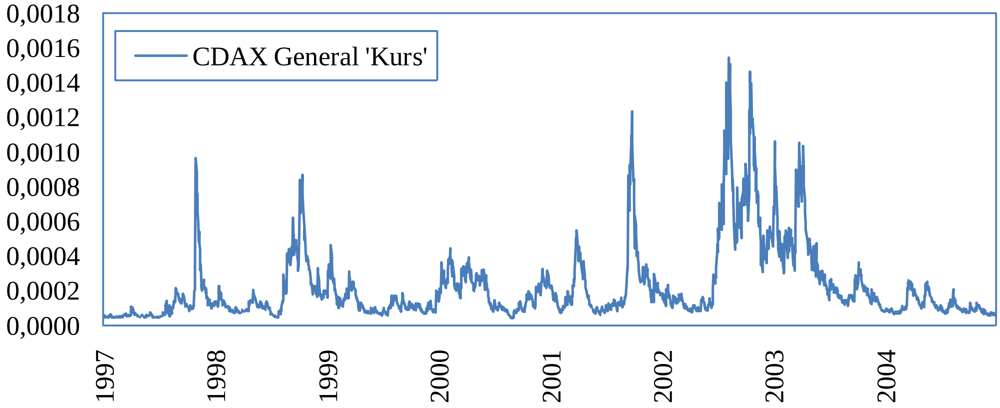


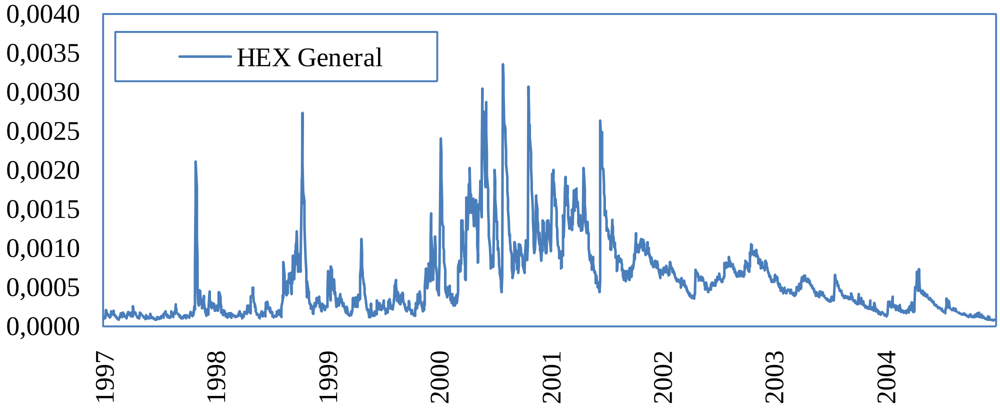


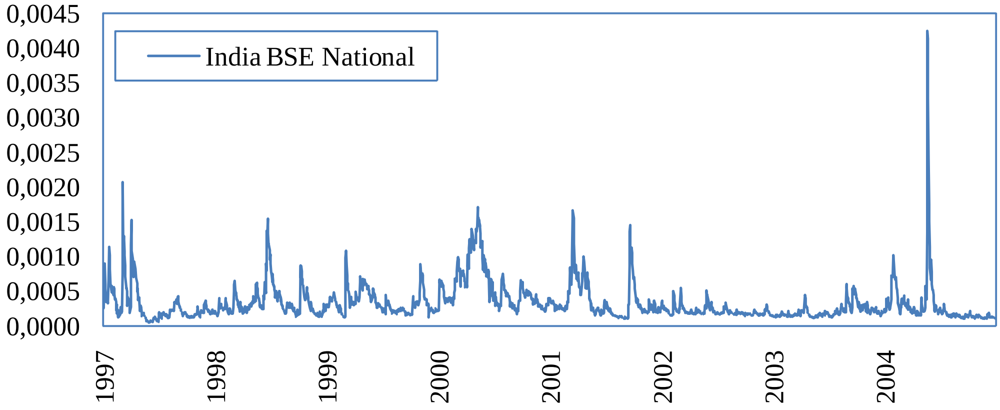


APPENDIX 3 ESTIMATED GARCH(1,1) SERIES









APPENDIX 4 DERIVATION OF GPD VAR

To be consistent with the rest of the thesis, the GPD VaR is derived concentrating on the left tail of the return distribution. Accordingly, a low threshold u ($u < 0$) is fixed and all exceedances y over u are of interest in this approach (i.e. $r_i = u - y_i$). The distribution of excess values is given by

$$\Pr\left(r > u - y \middle| r < u\right) = \frac{F(u) - F(u-y)}{F(u)}, y > 0 \tag{A.1}$$

Balkema and de Haan (1974) and Pickands (1975) showed that the asymptotic form of $\Pr(r > u - y \mid r < u)$ is

$$G(y) = 1 - \left(1 - \frac{\xi y}{\beta}\right)^{1/\xi} \tag{A.2}$$

where β is the scale parameter and ξ is the shape parameter, and $\beta > 0$, $1 - \xi y/\beta > 0$. This is known as the generalized Pareto distribution (CDF) with density (PDF)

$$g(y) = \frac{1}{\beta}\left(1 - \frac{\xi y}{\beta}\right)^{1/\xi - 1} \tag{A.3}$$

For $\{r_t\}_{t=1}^{n}$, ξ and β can be estimated by maximizing $\prod_{i=1}^{n_u} g(y_i)$, where $\{y_i\}_{i=1}^{n_u}$ is the sample of exceedances over the threshold u. Denote the maximum likelihood estimators of ξ and β by $\hat{\xi}$ and $\hat{\beta}$, respectively. Then combining (A.1) and (A.2)

$$\frac{F(u-y)}{F(u)} = \left(1 - \frac{\xi y}{\beta}\right)^{1/\xi} \tag{A.4}$$

which gives, if $F(u)$ is estimated by n_u/n,

$$\hat{F}(u-y) = \frac{n_u}{n}\left(1 - \frac{\hat{\xi} y}{\hat{\beta}}\right)^{1/\hat{\xi}} \tag{A.5}$$

or equivalently for $r < u$,

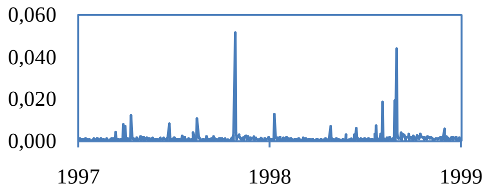
$$\hat{F}(r) = \frac{n_u}{n}\left(1 - \frac{\hat{\xi}(u-r)}{\hat{\beta}}\right)^{1/\hat{\xi}} \tag{A.6}$$

Immediately the αth quantile (i.e. $\hat{q}(\alpha)$ = VaR) can be estimated by setting $\hat{F}(r) = \alpha$ and hence

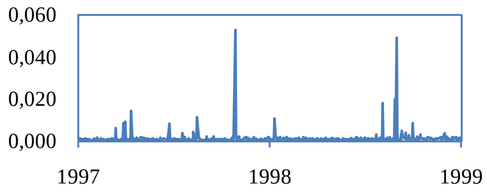
$$\hat{q}(\alpha) = u - \frac{\hat{\beta}}{\hat{\xi}}\left(1 - \left(\frac{n_u}{n}\alpha\right)^{\hat{\xi}}\right) \tag{A.7}$$

APPENDIX 5 Q LOSS PLOTS, USA, PERIOD 1, α = 0,05

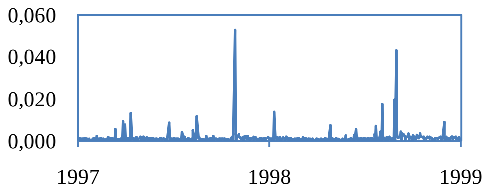
RM95



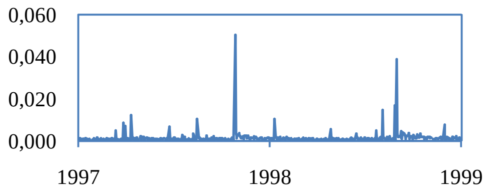
VC95



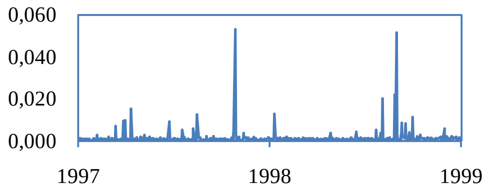
VC(GARCH)95



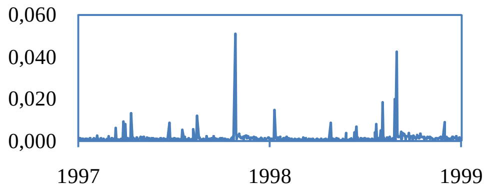
VC*95



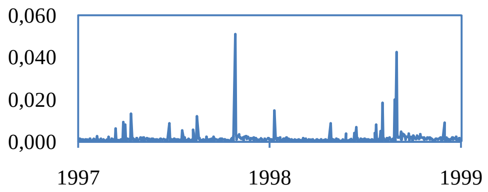
HS95



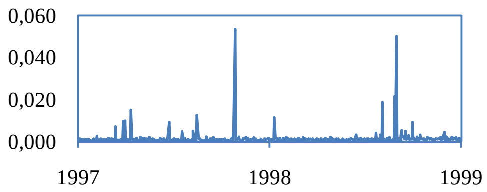
HS(GARCH)95



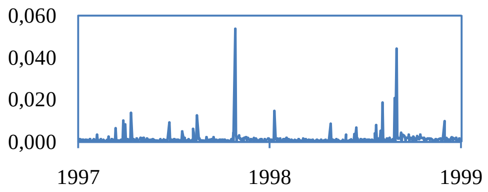
HS*95



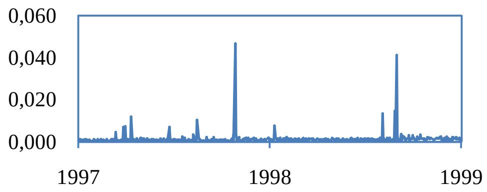
MC95



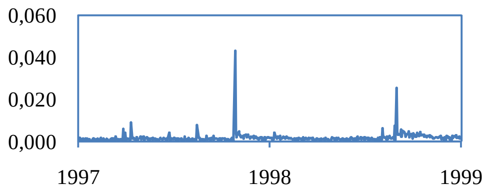
MC(GARCH)95



GPD95

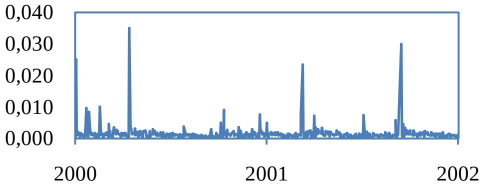


GPD*95

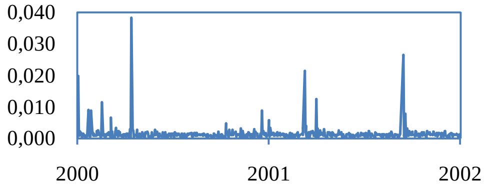


APPENDIX 6 Q LOSS PLOTS, USA, PERIOD 2, α = 0,05

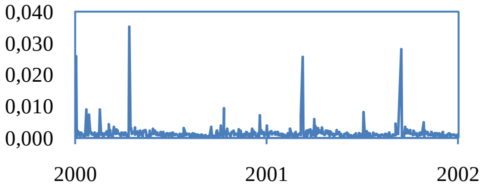
RM95



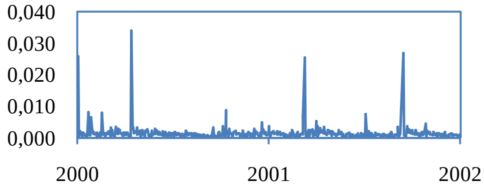
VC95



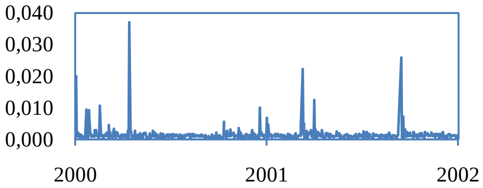
VC(GARCH)95



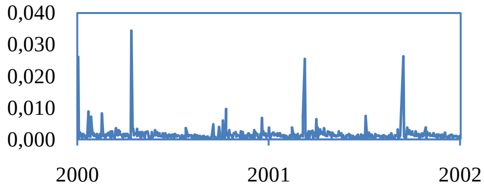
VC*95



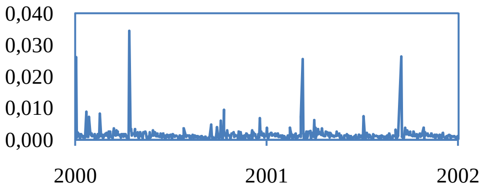
HS95



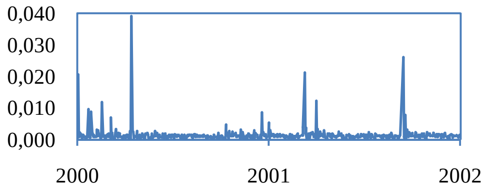
HS(GARCH)95



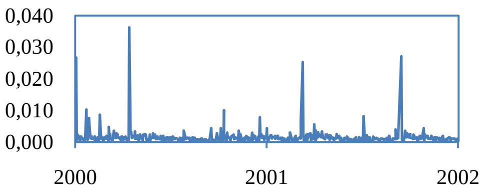
HS*95



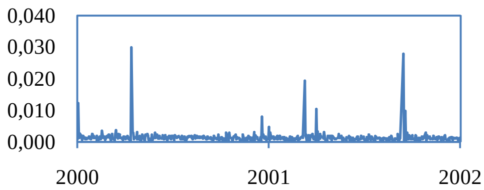
MC95



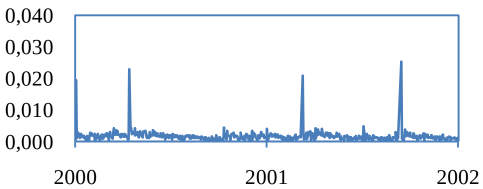
MC(GARCH)95



GPD95



GPD*95



APPENDIX 7 Q LOSS PLOTS, USA, PERIOD 3, α = 0,05