

Lund University EHL

**A comparative study of VaR
models**

Master thesis Spring 2012

Authors: Kaifeng Wang & Sebastian Recht
Supervisor: Birger Nilsson

Abstract

In this paper, we investigate the predictive performance of eight common Value at Risk (VaR) estimation methods with the daily data of OMX Stockholm 30 and FTSE 100 indexes. These methods include non-parametric, parametric as well as Extreme Value Theory approaches whose daily and 10-day estimates are then evaluated using the Christoffersen frequency test. The reason why we choose these two stock indexes is because that they respectively represent markets with small and big capitalizations, and the main goal of our study is to search for empirical evidences on whether or not the market size could have some important implications on the Value at Risk estimation process, such as if a universal approach could be applied to both types of markets.

Key words: Value at Risk, Market capitalization, Historical simulation, Volatility clustering, Volatility weighted historical simulation, EWMA, Normal distribution, t -distribution, Extreme value theory, Backtesting, Christoffersen frequency test, Square root rule

Contents

1. Introduction	- 3 -
1.1 Development of Value at Risk.....	- 3 -
1.2 Purpose of our study	- 4 -
2. Theoretical background.....	- 5 -
3. Methodology	- 7 -
3.1 Non-parametric approach	- 7 -
3.1.1 Basic historical simulation (HS)	- 7 -
3.1.2 Volatility weighted historical simulation (VWHS).....	- 7 -
3.2 Parametric approaches	- 8 -
3.2.1 Normal distribution	- 9 -
3.2.2 Student t -distribution.....	- 10 -
3.2.3 Normal and t -distribution with time-varying volatilities	- 10 -
3.3 Extreme value theory (EVT)	- 11 -
3.4 Backtesting – Christoffersen frequency test	- 13 -
4. Empirical study	- 15 -
4.1 Data.....	- 15 -
4.2 VaR estimation results.....	- 17 -
4.2.1 Non parametric approaches	- 17 -
4.2.2 Parametric approaches.....	- 17 -
4.2.3 EVT	- 18 -
4.2.4 Backtesting	- 19 -
5. Conclusions	- 22 -
References	- 24 -
Data sources	- 26 -
Appendix	- 27 -

1. Introduction

1.1 Development of Value at Risk

Value at Risk (VaR) is a widely used measure of market risk and has been developed into a standard procedure within risk management during the last two decades. VaR has gained its acceptance from academics and practitioners as well as the regulatory authorities.

Its origins can be dated back as far as to 1922 when the New York Stock Exchange imposed capital requirements on its member firms. However, VaR did not become popular until the 1990's. Thanks to numerous researchers' efforts, the theories in this field were being continuously developed and VaR has become a comprehensive tool in today's risk management. The earliest quantitative example of VaR is probably contributed by Leavens in 1945, which was further explored by, among others, Harry Markowitz in the early 1950's (Holton, 2002). Thereafter, in the 1970's and 1980's a number of financial institutions started to develop systems that could aggregate and measure risks across the entire institution. One of these systems is JP Morgan's RiskMetrics which was based on VaR and it could measure the risk related to different trading positions and aggregate these risks across the entire institution and express it in a single number (Dowd, 2002). VaR is conceptually simple, which is one of the major reasons why it has become so popular both in academia and practice (Emrah, Sayad, Levent, 2012). Another factor contributing to the popularity of VaR is that it is also accepted and applied as a proper measure of market risk in the regulatory framework such as "Basel accord", (Stelios, Dimitris, 2004) (Hendricks, 1996).

Despite its celebrity status in the risk management world, VaR is also receiving a huge amount of criticisms. According to Holton (2002) the critiques against VaR have mainly been focusing on three issues. Firstly, an extensive use of VaR would lead to systematic risks. Secondly, VaR is conceptually flawed as a measure of risk. Finally, different VaR methods could produce inconsistent results. For instance, Hendricks (1996) found in his study that twelve VaR approaches generated entirely different VaR estimates even though they were applied to the same underlying data. It is indeed a delicate process to find the appropriate VaR approach and to minimise the risk of ending up with a misspecified model (Christoffersen, Hahn, Inoue, 2001). Beder (1995) stressed that even if two firms apply the same quantitative technique, the underlying assumptions in the implementation process could differ. Due to these circumstances there is no consensus among researchers and practitioners on the 'best' method in estimating VaR, which lays the ground and provides the motivation for our research: a comparative study of some most popular approaches in this area. We will test non-

parametric, parametric approaches and the method based on Extreme Value Theory (EVT) whose performances will then be examined by the Christoffersen frequency test.

1.2 Purpose of our study

Through our study, we will examine the performance of various Value at Risk models at different-sized stock markets which, in our case, are OMX Stockholm 30 and FTSE 100 indexes. We chose these two indexes because they are likely to possess similar characteristics in all aspects except their market capitalizations. In this way we can control all other variables while focusing on the effect of the market size. There have been plenty of empirical studies concerning the implication of the size of a market, among which, for example, the paper “Trading frequency and volatility clustering” by Yi Xue and Ramazan Gençay (2012) appeared to be very interesting and relevant for our research. They found that:

“...the number of traders has an impact on the formation of volatility clustering. This relationship is a consequence of the fact that when the mean intensity of arrivals increases, the strategic competition outcome will converge to the competitive outcome. Hence, the effect of strategic interaction diminishes. Monte Carlo simulations show that in all settings, as the mean arrivals of traders increase, the variance of returns becomes less persistent. In the extreme, when the number of traders is sufficiently large, the model predicts that there is only one statistically significant first-order autocorrelation of variance of returns, while other autocorrelation coefficients are statistically indistinguishable from zero.”

Therefore, we would suspect that a market with more participated traders could display less volatility clustering, which in turn has a substantial implication on the choice of suitable VaR models, i.e. the time-constant volatility model vs. the time-varying volatility model. There are probably more findings such as this one which concerns market size and also affects the VaR estimation. Hence, by conducting this comparative study, we would like to find out whether or not two markets will suggest in using different VaR models. A positive result would provide evidence against a policy specifying a fixed VaR model for all markets, whereas a negative result would, to certain extent, support the use of a universal VaR model.

There are four sections in the remaining part of this thesis. In the second section, the theoretical background of Value at Risk and its historical development are presented. Next, in the third section the VaR models of our choice as well as the backtesting technique are described in great detail. Finally, the fourth section summarizes the empirical results and the fifth section concludes.

2. Theoretical background

From its development through RiscMetrics system at JP Morgan up until today, VaR has evolved into an acknowledged risk measure by both academics and practitioners and has become a standard tool in risk management. But what is actually Value at Risk? As previously mentioned VaR measures market risk and it does so by:

“...determining how much the value of a portfolio could decline over a given period of time with a given probability as a result of changes in market prices or rates. For example, if the given period of time is one day and the given probability is 1 percent, the value-at-risk measure would be an estimate of the decline in the portfolio value that could occur with a 1 percent probability over the next trading day. In other words, if the value-at-risk measure is accurate, losses greater than the value-at-risk measure should occur less than 1 percent of the time.”(Hendricks, 1996)

Further Linsmeier and Pearson (1996) described VaR as a: *“single, summary, statistical measure of possible portfolio losses. Specifically, value at risk is a measure of losses due to “normal” market movements. Losses greater than the value at risk are suffered only with a specified small probability.”*

VaR is then simply the maximum loss we could expect over a certain time period at a certain confidence level (Dowd, 2002). It is defined mathematically as:

$$\text{VaR}_\alpha(L) = \min\{l: \Pr(L > l) \leq 1 - \alpha\} \quad (1)$$

and under a continuous loss distribution it takes the form as:

$$\text{VaR}_\alpha(L) = \Pr(L > \text{VaR}_\alpha(L)) = 1 - \alpha \quad (2)$$

There are some important aspects regarding VaR. The first one is the time period of measurement which is also called the holding period and can be, for example, daily, weekly or monthly. Secondly, the composition of the portfolio during the given holding period is assumed to be constant, meaning that the assets included in the portfolio do not change over the given holding period. The choice of short holding periods is therefore favored so that the VaR models are more realistic. Thirdly, the confidence level chosen for the VaR measurement could be defined as the likelihood that the loss outcome is no worse than the VaR. The confidence level could be for example 60 %, 95 %, 99 % or 99,9 %. If we for example decide to work with a confidence level of 99 % this means that the future loss is expected to exceed the VaR estimate only by 1% of the time. The 99 % confidence level is sometimes also referred to as the 99th percentile VaR because the 99th percentile of the loss distribution gives the loss implied (Dowd, 2002) (Hendricks, 1996). Under the 1996 amendment to the Basel Accord institutions, for example banks, should derive their VaR at the 99% confidence level

and the holding period used should be 10 days. Banks are allowed to use 1-day VaR but then it should be multiplied by the square root of 10 (Dowd, 2002).

As mentioned in the introduction VaR partially has its theoretical roots in portfolio theory. As a measure of market risk it of course has some similarities with portfolio theory, but the two theoretical frameworks also differs in some aspects. Dowd highlights the following differences as important:

- The two frameworks differ in how they interpret risk. VaR interprets risk as the maximum likely loss, while portfolio theory uses the standard deviation of the return
- VaR can be modified to work with a wide range of different distributions and makes it possible to adapt the VaR-framework to the distribution that fits the data best. Portfolio theory on the other hand assumes that, for example, returns are normally distributed.
- Portfolio theory is limited to work with market risk, while VaR can be applied to a broader spectrum of risks.
- In general it is only the variance-covariance approach to estimate VaR which shares the same theoretical basis with Portfolio theory, while other approaches do not.

As mentioned earlier VaR is widely used in both academics and in practice because of its attractive features. However, VaR also has some weaknesses or drawbacks. First of all VaR is silent about the size of a loss if a tail event occurs. If we for example work with a 99% confidence level VaR tells us what we can lose 99% of the time, but we get no information about what we can lose the remaining 1% of the time. We should expect to lose more than VaR but we don't know how much. Second, because of the lack of information about the size of the tail loss VaR does not always encourage diversification. In other words, VaR is not sub-additive, which is a required property for a coherent risk measure. This means that we could end up with a higher VaR for a diversified portfolio than for a undiversified one. This is not a reasonable result for a realistic measure of market risk (Dowd, 2002). Third, VaR is sensitive to what kind of assumptions we make regarding the distribution of the losses. These assumptions has a "critical bearing" on how VaR should be estimated (Dowd, 2002). Forth, If everyone uses VaR as measure of market risk, it could be destabilizing during a financial crisis. If everyone is using VaR there is a potential risk that uncorrelated risks become highly correlated and firms, like banks, will be exposed to much larger risks than the VaR measure suggests (Dowd, 2002).

3. Methodology

3.1 Non-parametric approach

The most intuitive way to estimate VaR would be the non-parametric approaches. The fundamental pre-requisite for it to work is to assume that the recent past will be repeated in the near future (Dowd, 2002). The non-parametric approach is appealing because it does not require any distribution assumption for the underlying data, and therefore it is free from the errors due to the false distribution assumptions. Furthermore, it solves the dimensionality problem by directly using the empirical data of the portfolio losses. However, its straightforward property does also bring some severe drawbacks. For example, the VaR estimate is likely to be affected by some extreme losses in the very past and it is constrained by the largest losses in the sample. It is also difficult to decide on the appropriate length of the sample data to be used for the estimation, i.e. we need to filter out the irrelevant information while keeping a sufficient amount of the relevant information.

Two non-parametric methods will be addressed in our study: the basic historical simulation (HS) and one of its extensions, the volatility weighted historical simulation (VWHS).

3.1.1 Basic historical simulation (HS)

Recall that the definition of Value at Risk is $\text{VaR}_\alpha(L) = \min\{l: \Pr(L > l) \leq 1 - \alpha\}$ as in the equation (1) where L is the future losses and α is the chosen confidence level.

If we assume the sample size is N , then there will be simply $(1 - \alpha)N$ losses that are larger than $\text{VaR}_\alpha(L)$ and the $(1 - \alpha)N + 1$ largest loss will be naturally the estimate of $\text{VaR}_\alpha(L)$ when using the HS approach.

3.1.2 Volatility weighted historical simulation (VWHS)

Compared with HS, VWHS improves its predicting precision by taking into account of volatility clustering in the data. Volatility clustering is an empirically well-known phenomenon in financial returns. As the name implies volatility tend to cluster in a way that a period of high volatility is likely to be followed by another period of high volatility and period of low volatility is likely to be followed by another period of low volatility. In principle,

VWHS equals to applying the basic historical simulation method to the rescaled losses. We denote the original in-sample losses observations¹ as $l_1, l_2 \dots l_T$ which are then rescaled as:

$$\begin{aligned} l_1^* &= \frac{\sigma_{T+1}}{\sigma_1} l_1 \\ l_2^* &= \frac{\sigma_{T+1}}{\sigma_2} l_2 \\ &\vdots \\ l_T^* &= \frac{\sigma_{T+1}}{\sigma_T} l_T \end{aligned} \quad (3)$$

where $\sigma_1, \dots, \sigma_T$ are the volatilities of the corresponding in-sample loss observations, and σ_{T+1} is the volatility of the first loss in the out-of-sample period.

In practice, there are usually two ways to estimate these volatilities, the GARCH(1, 1) model or the exponentially weighted moving average volatility model (EWMA). Due to the fact that EWMA will be used in our estimations, we will concentrate on its methodology here. One issue when applying the GARCH(1, 1) model is that it is problematic to specify how often the parameters in the model should be updated in order to truly reflect the underlying data. This issue is, however, not encountered by the EWMA model:

$$\sigma_{T+1}^2 = \frac{1 - \lambda}{1 - \lambda^T} \sum_{t=1}^T \lambda^{T-t} \varepsilon_t^2 \quad (4)$$

where the constant $\lambda = 0,94$ is a general practice used in RiskMetrics. When the sample size T is reasonably large, this form reduces and becomes even more appealing:

$$\sigma_{T+1}^2 \approx (1 - \lambda) \varepsilon_T^2 + \lambda \sigma_T^2 \quad (5)$$

Empirically, we can approximate the initial value ε_0^2 with zero and σ_0^2 with the sample variance of the loss data.

After the original losses have been rescaled, the $(1 - \alpha)T + 1$ largest rescaled loss is the VWHS-EWMA estimate of $\text{VaR}_\alpha(L)$ according to the same principle used in the basic historical simulation approach,

3.2 Parametric approaches

In order for the parametric approaches to work, we need to explicitly specify the statistical distribution of the underlying loss data, which in turn will produce the estimates needed based on their statistical properties. Compared to the non-parametric approaches, the parametric

¹ This method is often considered as a time-series method. Therefore, T is used instead of N here to represent the total number of the in-sample observations.

methods have both advantages and disadvantages. It is relatively easy and straightforward to implement in practice, which unfortunately also comes with a price: the estimates are very vulnerable to the errors caused by the inaccurate distribution assumptions (Dowd, 2002).

In our study, we will examine two most used distributions: the normal distribution and the student t -distribution. Even though the normal distribution cannot accommodate the fact that the financial data are often fat-tailed, it can serve as a good benchmark due to its simplicity. Another well-known feature of financial returns is that volatilities tend to cluster in some time periods, which cannot be accounted for by the models in their original forms. Therefore, we will also test the two distributions by incorporating time-varying volatilities estimated by the exponentially weighted moving average volatility model (EWMA).

3.2.1 Normal distribution

There are some characteristics that make the normal distribution a very intuitive and simple model to use. The first one is the central limit theorem which states that the mean of an independent random variable will be approximately normally distributed as the number of observations goes towards infinity. This is an attractive feature when we are working with sample means. The second one is that the normal distribution only has two independent parameters, namely the mean μ and the standard deviation σ . The skewness of the normal distribution is always zero, i.e. symmetric, and the kurtosis is always equal to 3. This means that we only need to estimate the mean and the standard deviation in order to derive VaR estimates (Dowd, 2002):

$$\text{VaR}_\alpha(L) = \mu + \sigma z_\alpha \quad (6)$$

where z_α is the quantile at the chosen confidence level α for the standard normal distribution.

It happens that sometimes the holding period of our desired VaR estimate is not the same as the frequency of the available samples. Ideally, we should try to match the two. However, it could result in too few observations and consequently distort the VaR estimate. As an alternative, the Basel accord allows financial institution to scale up 1 -day VaR estimates to h -day VaR estimates by using the ‘square root rule’. For the normal distribution this works as follows²:

$$\text{VaR}_\alpha^h(L) = h\mu + \sqrt{h}\sigma z_\alpha \quad (7)$$

As mentioned earlier, the normal distribution also has some drawbacks when it is used as the basis for estimating VaR. First of all, we can only motivate the use of normality through

² For non-parametric approach, $\text{VaR}_\alpha^h = \sqrt{h}\text{VaR}_\alpha$

the central limit theorem when we work with more centrally located quantiles. Its validity becomes questionable when we analyse ‘tail events’. Secondly, the normal distribution fails to capture high kurtosis which is often present in the financial data.

3.2.2 Student t -distribution

As stated earlier the student t -distribution can accommodate the positive excess kurtosis by incorporating ν , the degrees of freedom. Now the estimate of VaR takes the form as:

$$\text{VaR}_\alpha(L) = \mu + \sqrt{\frac{\nu - 2}{\nu}} \sigma t_{\alpha, \nu} \quad (8)$$

$$\text{VaR}_\alpha^h(L) = h\mu + \sqrt{h \frac{\nu - 2}{\nu}} \sigma t_{\alpha, \nu} \quad (9)$$

where $t_{\alpha, \nu}$ is the α -quantile with ν as the degrees of freedom for t -distribution.

It requires ν to be larger than 4 for kurtosis to exist and we can use the following relation to derive ν when conducting empirical studies:

$$\nu = \frac{4k - 6}{k - 3} \quad (10)$$

where k is the sample kurtosis³.

Just like the normal distribution, the student t -distribution is either very reliable when we analyse extreme values or quantiles at very high confidence levels.

3.2.3 Normal and t -distribution with time-varying volatilities

As stated earlier neither the normal distribution nor the student t -distribution can account for volatility clustering in their original forms. In order to capture this, we need to replace the constant volatility parameter σ with its time-dependent counterpart σ_{T+1} which is the same as defined before, i.e. it is the volatility of the first loss in the out-of-sample period. The methodology to estimate σ_{T+1} under a EWMA model has been described in the non-parametric approach part. And now we have normal distribution and t -distribution with EWMA estimated volatilities as:

$$\text{VaR}_\alpha(L) = \mu + \sigma_{T+1} z_\alpha \quad (11)$$

³ When k is smaller than 3, it means the sample data have a thinner tail than the normal distribution. Then neither normal distribution nor t -distribution would be suitable for the data.

$$\text{VaR}_\alpha(L) = \mu + \sqrt{\frac{v-2}{v}} \sigma_{T+1} t_{\alpha,v} \quad (12)$$

3.3 Extreme value theory (EVT)

The basic idea of applying Extreme value theory when estimating VaR is that the ‘extreme losses’ will follow certain known distribution under some given assumptions, regardless of the distribution of the original losses. The VaR can be then estimated from this known distribution of the extreme values. It is very appealing to use EVT to estimate VaR in the sense that the theory focuses on the behavior of tail values, something that Risk management is actually interested in.

EVT consists of two branches which can be distinguished by the way how the ‘extremes’ are defined. The traditional EVT, also known as the block maxima method, divides the loss observations into same-sized blocks and then selects the largest loss from each block to compose the extreme values. The second one is the modern EVT, also known as Peaks over threshold (POT). This method pre-defines a threshold at first and then chooses the losses that are larger than this threshold to be the extreme values. One significant drawback of the traditional EVT is that some important information is lost during the selection of extreme values: it ignores very large losses if they are clustered in the same block. Therefore, its application is not suitable for the analysis of data with volatility clustering such as financial losses in our study. We will hence concentrate only on POT extreme value theory in the following sections.

The limit theory proposed by Pickands-Balkema-deHaan states that the limiting distribution of $F_u(l-u)$ is a generalized Pareto distribution (GPD) when the threshold u goes to infinity:

$$F_u(l-u) \approx G(l-u) = \begin{cases} 1 - \left(1 + \xi \frac{l-u}{\beta}\right)^{-\frac{1}{\xi}} & \text{if } \xi \neq 0 \\ 1 - \exp\left(-\frac{l-u}{\beta}\right) & \text{if } \xi = 0 \end{cases} \quad (13)$$

where ξ is a shape parameter and β is a scale parameter, which can be estimated by applying Maximum likelihood (ML) estimation to the extreme losses. Assume that there are M losses that are larger than the chosen threshold u (the extreme values) and let us denote them as $l_u^1, l_u^2, \dots, l_u^M$. Then in order to obtain the ML estimates we should find the parameters that maximize the log-likelihood functions below:

$$\log L(\beta) = \left[-M \ln \beta - \frac{1}{\beta} \sum_{i=1}^M (l_u^i - u) \right] \text{ if } \xi = 0 \quad (14)$$

$$\log L(\beta, \xi) = -M \ln \beta - \left(1 + \frac{1}{\xi} \right) \sum_{i=1}^M \ln \left(1 + \xi \frac{l_u^i - u}{\beta} \right) \text{ if } \xi \neq 0 \quad (15)$$

Then the unconditional POT estimates of VaR can be empirically calculated as follows:

$$\text{VaR}_\alpha = u - \beta \ln \left(\frac{N}{N_u} (1 - \alpha) \right) \text{ if } \xi = 0 \quad (16)$$

$$\text{VaR}_\alpha = u + \frac{\beta}{\xi} \left[\left(\frac{N}{N_u} (1 - \alpha) \right)^{-\xi} - 1 \right] \text{ if } \xi > 0 \quad (17)$$

where N is the number of all loss observations and N_u is the number of the extreme values, i.e. the losses that are larger than the pre-defined threshold. The reason why ξ is always non-negative here is because that when $\xi > 0$, the underlying distribution of loss observations would correspond to a fat-tailed distribution; and when $\xi = 0$, it would correspond to a semi-heavy tailed distribution. Therefore, it is only these two cases that are relevant for analyses of financial data.

The same as non-parametric or parametric approaches, there is also a twist to the POT so that it could capture the market dynamics by taking into account of volatility clustering. This modified POT is called conditional POT and is in fact applying POT to the standard residuals instead of original losses. These standardized residuals are defined as:

$$\begin{aligned} \varepsilon_1^* &= \frac{l_1 - \bar{l}}{\sigma_1} \\ \varepsilon_2^* &= \frac{l_2 - \bar{l}}{\sigma_2} \\ &\vdots \\ \varepsilon_N^* &= \frac{l_N - \bar{l}}{\sigma_N} \end{aligned} \quad (18)$$

where l_1, l_2, \dots, l_N are the loss observations and \bar{l} is the sample mean of all these losses. $\sigma_1, \sigma_2, \dots, \sigma_N$ are the volatilities of the corresponding losses and can be estimated by using EWMA model described earlier.

Next, the α - quantiles can be derived as:

$$q_\alpha = u^* - \beta^* \ln \left(\frac{N}{N_{u^*}} (1 - \alpha) \right) \text{ if } \xi^* = 0 \quad (19)$$

$$q_\alpha = u^* + \frac{\beta^*}{\xi^*} \left[\left(\frac{N}{N_{u^*}} (1 - \alpha) \right)^{-\xi^*} - 1 \right] \text{ if } \xi^* > 0 \quad (20)$$

The added sign of “*” here indicates that these parameters are GDP parameters when POT is applied to the standardized residuals. u^* is the pre-defined threshold which will be used on the standardized residuals and its choice is independent of the threshold u used in the unconditional POT.

The procedure to estimate β^* and ξ^* is very similar to the one used in the unconditional POT, i.e. we need to find the parameters that will maximize the log-likelihood functions as follows:

$$\log L(\beta^*) = \left[-M^* \ln \beta^* - \frac{1}{\beta^*} \sum_{i=1}^{M^*} (\varepsilon_{u^*}^i - u^*) \right] \text{ if } \xi^* = 0 \quad (21)$$

$$\log L(\beta^*, \xi^*) = -M^* \ln \beta^* - \left(1 + \frac{1}{\xi^*} \right) \sum_{i=1}^{M^*} \ln \left(1 + \xi^* \frac{\varepsilon_{u^*}^i - u^*}{\beta^*} \right) \text{ if } \xi^* \neq 0 \quad (22)$$

where $\varepsilon_{u^*}^1, \varepsilon_{u^*}^2, \dots, \varepsilon_{u^*}^{M^*}$ are the standardized residuals which are larger than the chosen threshold u^* and there are M^* of them in total.

Finally, the conditional POT estimates VaR as:

$$\text{VaR}_\alpha = \mu + \sigma_{N+1} q_\alpha \quad (23)$$

μ is the sample mean of the in-sample losses and σ_{N+1} is the volatility of the first loss in the out-sample period and can be estimated by EWMA model.

3.4 Backtesting – Christoffersen frequency test

We will use Christoffersen test to examine the accuracy of different estimators of VaR in our study. This test consists of three steps: firstly the observed probability of violations⁴ from the underlying model (estimator of VaR) is compared with the expected probability of violations in order to see if the former is close enough to the latter in a statistical sense. This step is also called ‘unconditional converge test’.

To illustrate this procedure in a mathematical fashion, let us denote the sample size of the testing period, i.e. the out-of-sample period as N , the observed number of violations from the underlying model as x , hence the observed probability of violations is $\pi = \frac{x}{N}$. The expected probability of violations according to the selected confidence level α is $p = 1 - \alpha$. Then a likelihood ratio test can be applied:

⁴ A violation occurs when the actual loss is larger than the corresponding VaR estimate.

$$LR_{uc} = -2\ln\left(\frac{L_0}{L_1}\right) = -2[\ln(p^x(1-p)^{N-x}) - \ln(\pi^x(1-\pi)^{N-x})] \sim \chi^2(1) \quad (24)$$

where L_0 is the log-likelihood under the null hypothesis that the underlying model provides a statistically correct probability of violations and L_1 is the log-likelihood under the alternative hypothesis. If the derived test statistics LR_{uc} from the test period is larger than the Chi-square critical value with one degree of freedom, the null hypothesis is rejected, i.e. we do not consider this estimator of VaR to be statistically sufficient.

The objective of the second step in Christoffersen test is to examine if the observed violations resulted by the underlying model clusters. In other words, we want to see if a violation today would also cause a violation tomorrow, or they are statistically independent. A failure to confirm this independence indicates that the underlying model being tested is misspecified and should thus be rejected.

Denote the probability of observing no violation as π_0 and observing violations as π_1 :

$$\pi_0 = \frac{n_0}{N} \quad \pi_1 = \frac{n_1}{N} \quad (25)$$

where n_0 and n_1 are the number of non-violation and observing violations in the test period respectively. Furthermore, we define n_{00} , n_{01} , n_{10} and n_{11} as:

$n_{00} \equiv$ the number of observations when there is violation neither today nor tomorrow

$n_{01} \equiv$ the number of observations when there is no violation today but there is one tomorrow

$n_{10} \equiv$ the number of observations when there is a violation today but not tomorrow

$n_{11} \equiv$ the number of observations when there are violations both today and tomorrow

Then we can calculate the probability of the four situations described above as:

$$\begin{aligned} \pi_{00} &= \frac{n_{00}}{n_{00} + n_{01}} & \pi_{01} &= \frac{n_{01}}{n_{00} + n_{01}} \\ \pi_{10} &= \frac{n_{10}}{n_{10} + n_{11}} & \pi_{11} &= \frac{n_{11}}{n_{10} + n_{11}} \end{aligned} \quad (26)$$

Next the likelihood ratio test is applied as:

$$LR_{ind} = -2\ln\left(\frac{L_0}{L_1}\right) = -2[\ln(\pi_0^{n_0}\pi_1^{n_1}) - \ln(\pi_{00}^{n_{00}}\pi_{01}^{n_{01}}\pi_{10}^{n_{10}}\pi_{11}^{n_{11}})] \sim \chi^2(1) \quad (27)$$

where this time the null hypothesis is non-violation and violations are independent over time. If the derived test statistics is larger than the corresponding critical value of Chi-square with one degree of freedom, we should reject the null hypothesis and abandon the underlying model. In practice, it could happen that we will observe no clustering, i.e. $n_{11} = 0$ and $\pi_{11} = 0$, which will make it impossible to conduct the likelihood ratio test presented above. One suggested solution is that we can calculate the likelihood under the alternative hypothesis as

$\pi_{00}^{n_{00}}\pi_{01}^{n_{01}}$. However, this is not a remedy for all cases and we should actually reconsider the feasibility of the model used to generate those ‘trouble-making’ VaR estimates instead.

Finally, the last step of Christoffersen test which is named as ‘conditional coverage test’ can be carried out as:

$$LR_{cc} = LR_{uc} + LR_{ind} \sim \chi^2(2) \quad (28)$$

The same as before, if the derived test statistics is larger than the corresponding critical value of Chi-square with two degree of freedom, we should reject the underlying model.

4. Empirical study

4.1 Data

The daily data of OMX Stockholm 30 and FTSE 100 indexes from April 2nd 2002 to April 4th 2012, downloaded from the websites of NASDAQ OMX Nordic and Yahoo Finance respectively, are used for our study (see Figure 3 in the appendix). Furthermore, the whole 10-year period is divided into an in-sample period and an out-of-sample period. The former provides the input for the VaR estimation while the latter lays the ground for the backtesting. Before starting with the VaR estimation, the original indexes data are transformed into losses data as:

$$l_t = -\text{LN}\left(\frac{r_t}{r_{t-1}}\right) \times 100 \quad (29)$$

where r_t and r_{t-1} indicate the index level of the current trading day and the previous trading day respectively.

The descriptive statistics of the two indexes are summarized in Table 1. Not surprisingly both indexes have a kurtosis that is significantly larger than 3 and are therefore fat-tailed. Furthermore, the OMX loss distribution is negatively skewed while the FTSE is positively skewed.

Table 1: Descriptive statistics of the daily losses

	Mean	Std.Dev	Variance	Skewness	Kurtosis	Min	Max
OMX	-,010941	1,5902421	2,529	-0,124	6,740	-9,8650	7,5127
FTSE	-,003544	1,3360929	1,785	0,118	9,409	-9,3842	9,2646

These observations suggest that we could conclude that the losses are not normally distributed. Two formal normality tests, Kolmogorov-Smirnov test and the Shapiro-Wilk test,

were also applied to the losses data in order for us to further examine the non-normality assumption (see the results in Table 2).

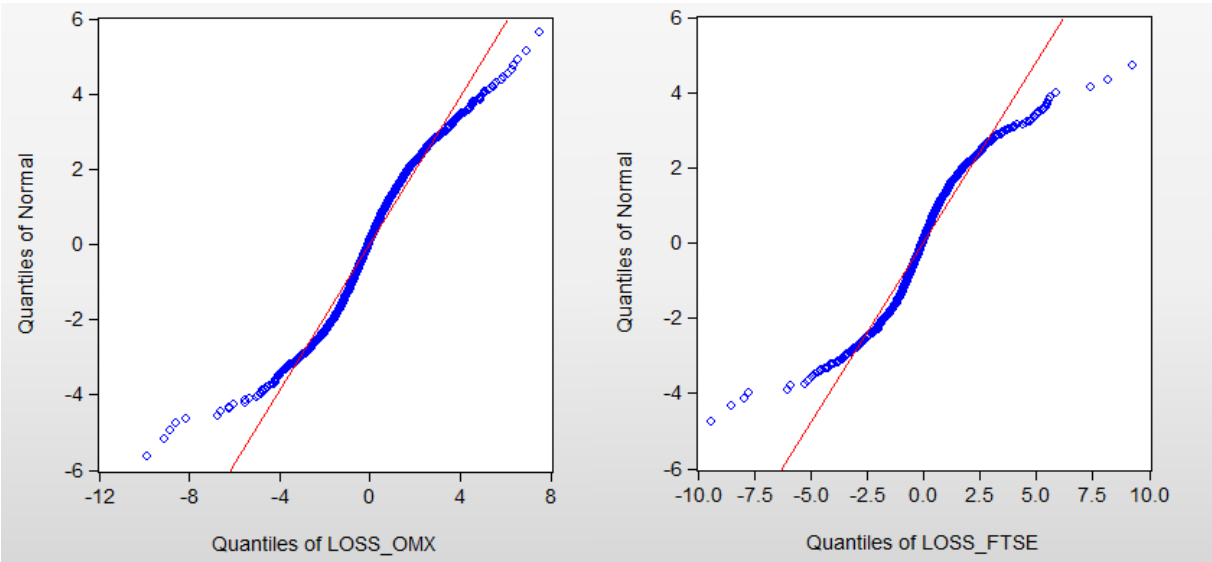
Table 2: Normality tests of the daily losses

	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	<i>p</i> -value	Statistic	df	<i>p</i> -value
OMX	0,067	2520	0,000	0,953	2520	0,000
FTSE	0,084	2520	0,000	0,922	2520	0,000

a. Lilliefors Significance Correction

The tests compare the empirical distribution at hand with a normal distribution. The null hypothesis is that the two distributions do not differ significantly. If the *p*-value is smaller than the significance level, in this case 5%, the null hypothesis should be rejected. From the test output above, we can conclude that neither the OMX nor the FTSE loss distributions was normally distributed. A graphical test, Q-Q plot, was also confirming this conclusion (see Figure 1). If an empirical distribution coincides with a normal distribution, the Q-Q plot will be a straight line. And if an empirical distribution has a fatter tail the Q-Q plot will display an ‘S’ shape such as seen in our case.

Figure 1: Q-Q plot of the daily losses



4.2 VaR estimation results

4.2.1 Non parametric approaches

- HS

The $\text{VaR}_{0,99}$ estimate is given by taking the 99th percentile of the rolling window data containing as many samples as that in the in-sample period. I.e. the 99th percentile of losses from April 3rd 2002 to March 31st 2011 approximates the first $\text{VaR}_{0,99}$ in the out-of-sample period. Then in order to estimate the second $\text{VaR}_{0,99}$ estimate in the out-of-sample period, we roll the window by dropping the first loss data in the in-sample period and adding the first loss data in the out-of-sample period. The Excel command **PERCENTILE.INC** is used to estimate the 99th percentile.

- VWHS

As described in the methodology part, we use the EWMA model $\sigma_{T+1}^2 = (1 - 0,94)\varepsilon_T^2 + 0,94\sigma_T^2$ in our study to approximate the volatility of losses. The error term ε_t is approximated by taking the difference between the actual daily loss and the sample mean of the in-sample daily losses. The initial value of ε_t is specified as zero and the initial value of σ_T^2 is specified as the sample variance of the in-sample daily losses. In addition, we calculate the σ_{T+1} for each month instead of for each day. Therefore, the $\text{VaR}_{0,99}$ estimates of April 2011 is derived by taking the 99th percentile of the corresponding rescaled losses which are rescaled by the EWMA-estimated volatility of the first trading day in April 2011. By ‘corresponding’ we mean that the same rolling window approach described earlier is applied.

4.2.2 Parametric approaches

In our study we applied four different parametric models on the OMX and FTSE indexes. They are the normal distribution, with time-constant volatility or with time-varying volatility, and the t -distribution with also these two types of volatility. To construct time-varying volatility we use the Exponentially Weighted Moving Average model.

- Normal distribution

Aligned with the non-parametric approaches we use a rolling window of the in-sample observations to estimate the mean and standard deviation which are used to estimate VaR of the out-of-sample period. Our expectation is that this model will perform worse than the other

models because the lack of normality in financial data. Still the normal VaR will be included in our tests as benchmark case. When the mean and standard deviation are derived, the $VaR_{0,99}$ is calculated as:

$$VaR_{0,99}(L) = \mu + \sigma z_{0,99} \quad (30)$$

To obtain the 99% quantile of the standard normal distribution we use the formula **NORM.S.INV(0,99)** in Excel.

- *t*-distribution

The VaR estimate for the out-of-sample period was derived by using the following model.

$$VaR_{0,99}(L) = \mu + \sqrt{\frac{v-2}{v}} \sigma t_{0,99,v} \quad (31)$$

where the mean and the standard deviation are calculated in the same way as for the normal distribution, but now we also need to calculate the degrees of freedom, v . To estimate v we inserted the following relation between kurtosis and the degrees of freedom in Excel:

$v = (4*(kurtosis)+3)-6/((kurtosis)+3)-3$. The quantile at the 99% level with v degrees of freedom $t_{\alpha,v}$ is given by **T.INV(0,99; v)** in excel.

- Normal distribution and *t*-distribution with time-varying volatility

To produce the VaR estimates based on the normal distribution and *t*-distribution with time-varying volatility we used the same equations as those for the models with time-constant volatility except that we replaced the sample standard deviation, σ , with σ_{T+1} defined in the same way as before: $\sigma_{T+1}^2 = (1 - 0,94)\varepsilon_T^2 + 0,94\sigma_T^2$. The details of this process can be found in section 4.2.1, VWHS.

4.2.3 EVT

The threshold $u = 2$ is arbitrarily chosen in our study for both unconditional and conditional POT methods. A value should be selected as the threshold if it is high enough so that only the ‘real’ extreme values following the generalized Pareto distribution remain, and it is also low enough so that there are sufficient amount of extreme values left to estimate the distribution parameters. There is no standard procedure to identify the optimal threshold, which in fact is the biggest drawback of the POT approach. We decided to use $u = 2$ because it has been proven empirically that it is a reasonable level both practically and economically. The ML

estimates of the generalized Pareto distribution parameters are displayed in Table 3 and Table 4 below together with their corresponding log-likelihoods:

Table 3: Unconditional POT parameter estimation ($u = 2$)

	OMX		FTSE	
β	1,16505	1,28634	1,24106	1,18469
ξ	0	-0,10318	0	0,04555
$\log L$	-204,039	-203,306	-144,700	-144,588

Table 4: Conditional POT parameter estimation ($u = 2$)

	OMX		FTSE	
β^*	0,62633	0,58683	0,61895	0,71321
ξ^*	0	0,06379	0	-0,15090
$\log L$	-38,845	-38,747	-38,501	-37,823

A comparison of the log-likelihood values indicates that neither ξ nor ξ^* is significantly different from zero at 5% level for either index. Thus, we conclude that the loss data at hand follow a Gumbel distribution, i.e. ξ is zero, and VaR only needs to be estimated under this scenario. It obeys the same principle when it comes to the usage of EWMA model in the conditional POT method.

4.2.4 Backtesting

- **Daily VaR**

Christoffersen frequency test results in Table 5 indicate that, for the OMX index, all models are accepted except VWHS, Normal and Normal EWMA which failed the unconditional coverage test. All the models that passed the unconditional coverage test also passed the independence test and the conditional coverage test. For the FTSE index, both HS and the unconditional POT are accepted whereas their extensions considering time-varying volatilities, i.e. VWHS and the conditional POT, failed the independence test. When we consider the parametric approaches, for FTSE, we see that all the models are accepted in the unconditional coverage test except Normal EWMA. All the parametric models that are accepted in the unconditional coverage test are also accepted in the independence test and the conditional coverage test.

By observing these outcomes, we have some interesting findings. Firstly, the normal distribution model, either with unconditional volatility or conditional volatility, has the worst performance for both indexes. This is not a surprising result since we know that financial data

usually have fat tails and a kurtosis which is larger than the value assumed under a normal distribution. Therefore, it is obviously not plausible to assume that the losses are normally distributed while estimating the VaR.

Table 5: Backtesting on Daily VaR_{0,99} estimates at 5% critical level

OMX	Methods	LR_{uc}	LR_{ind}	LR_{cc}	Violations
Non-parametric	HS	3,36	0,34*	3,70	6
	VWHS	5,25	n.a.	n.a.	7
Parametric	Normal	15,41	n.a.	n.a.	11
	Normal EWMA	5,25	n.a.	n.a.	7
	<i>t</i> -distribution	3,36	0,34*	3,70	6
	<i>t</i> -dist. EWMA	3,36	0,29*	3,65	6
EVT	Uncondi POT	1,82	0,24*	2,06	5
	Condi POT	3,36	0,29*	3,65	6

FTSE	Methods	LR_{uc}	LR_{ind}	LR_{cc}	Violations
Non-parametric	HS	0,14	0,05	0,19	2
	VWHS	1,82	10,04	n.a.	5
Parametric	Normal	3,36	0,34*	3,70	6
	Normal EWMA	5,25	n.a.	n.a.	7
	<i>t</i> -distribution	0,14	0,05*	0,19	2
	<i>t</i> -dist. EWMA	1,26	0,02*	1,28	1
EVT	Uncondi POT	0,14	0,05*	0,19	2
	Condi POT	0,14	7,56	n.a.	2

Note:

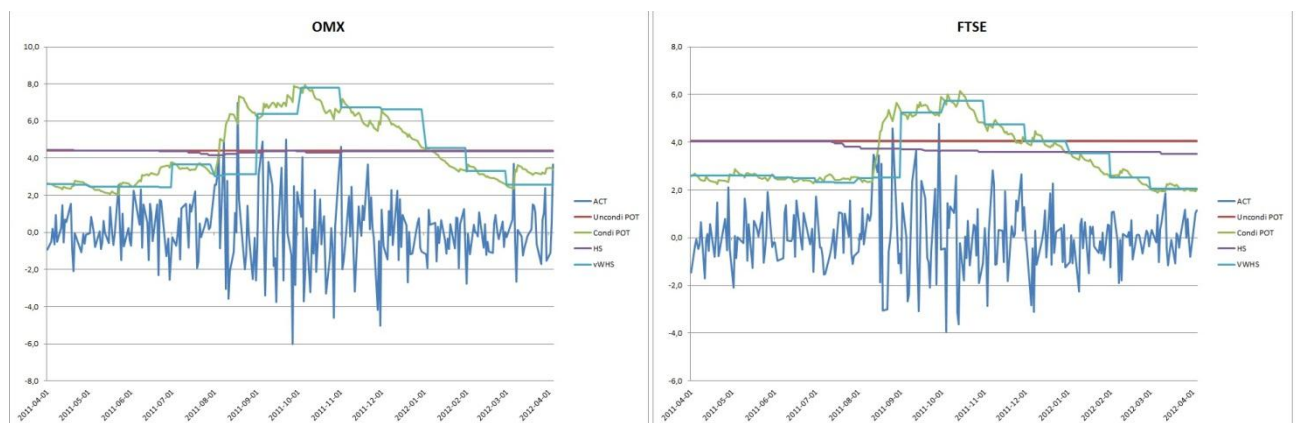
1. “*” indicates that there is no violation clustering. Therefore, the alternative hypothesis $\pi_{00}^{n_{00}}\pi_{01}^{n_{01}}$ is used in the independence test.
2. ‘n.a.’ is displayed if the underlying model has been rejected during the previous step.

Secondly, the approaches which take into account of volatility clustering did not, in general, perform significantly better than their constant-variance counterparts. For the FTSE index they even performed worse than their counterparts. At the first sight, this observation seems surprising: normally we expect that the models which take into account of volatility clustering would outperform the ones without this consideration. However, the “irregularity” in our results may be explained by the fact that the two indexes in our study have very stable volatilities during the out-of-sample period when the conditional methods do not possess advantages over the unconditional ones. In order to examine this assumption, the actual losses and the VaR estimates are plotted together in Figure 2. The plots confirm that either OMX or FTSE displays significant volatility clustering. FTSE has, in addition, less volatility clustering during the out-of-sample period than OMX, which explains why the conditional methods

perform more poorly for FTSE than for OMX. This is also confirming the finding by Yi Xue and Ramazan Gençay (2011): there is a negative relation between the number of traders in the market and the present magnitude of volatility clustering effect.

At last, if we assume that the model producing fewer violations is better at estimating VaR of the underlying data, we then have another general result independent of VaR approach. It showed that models perform better when being applied to the FTSE index than to the OMX index.

Figure 2: Plots of actual losses and VaR_{0,99} estimates



- **10-day VaR**

Recall that the “Square root rule” can be applied to scale up the daily VaR estimates to the 10-day VaR estimates (see Figure 4 in the appendix). The backtesting results for the 10-day VaR estimates are summarized in Table 6 below.

Table 6: Backtesting on 10-day VaR_{0,99} estimates at 5% critical level

OMX	Methods	LR_{uc}	LR_{ind}	LR_{cc}	Violations
Non-parametric	HS	2,00	30,96	n.a.	5
	VWHS	5,57	n.a.	n.a.	7
Parametric	Normal	3,61	38,14	n.a.	6
	Normal EWMA	13,08	n.a.	n.a.	10
	<i>t</i> -distribution	3,61	38,14	n.a.	6
	<i>t</i> -dist. EWMA	5,57	n.a.	n.a.	7
EVT	Uncondi POT	2,00	30,96	n.a.	5
	Condi POT	5,57	n.a.	n.a.	7

FTSE	Methods	LR_{uc}	LR_{ind}	LR_{cc}	Violations
Non-parametric	HS	2,00	30,96	n.a.	5
	VWHS	5,57	n.a.	n.a.	7

Parametric	Normal	2,00	30,96	n.a.	5
	Normal EWMA	10,34	n.a.	n.a.	9
	<i>t</i> -distribution	2,00	30,96	n.a.	5
	<i>t</i> -dist. EWMA	7,82	n.a.	n.a.	8
EVT	Uncondi POT	0,79	23,47	n.a.	4
	Condi POT	7,82	n.a.	n.a.	8

Note: ‘n.a.’ is displayed if the underlying model has been rejected during the previous step.

None of the models could pass all the tests. The performances were rather similar this time for both indexes. All of the models considering volatility clustering failed already the unconditional test, whereas the models without this consideration passed the unconditional test but failed the independence test miserably. The intuition behind this outcome is that if there is a huge daily loss causing violation, this daily loss will also drag up several 10-day losses around this date and could thus cause a chain of violations, which is exactly in conflict with the violation independence hypothesis.

5. Conclusions

As mentioned earlier, VaR today has a well-established position as a measure of market risk both in academia and in practice. Banks use it every day to measure their exposure to market risk and researchers constantly provides new insights concerning the validity of different VaR models that are available at the moment. The popularity of the models in combination with the huge variety of VaR approaches available has helped the subject gain a lot of attention within the research community. With the huge amount of studies done on the subject it is not our intention to give a full research overview, but to provide some valuable and, more or less, representative insights within the area of Value at Risk.

In our study, we examined and compared the performance of eight common approaches to estimate Value at Risk at the 99% level. Two European stock indexes, the OMX Stockholm 30 and FTSE 100 were chosen as the underlying data mainly due to the fact that they have similar properties in most aspects expect for their respective market sizes. Through our research we intend to find out if difference in sizes would have some important implications on the different VaR estimation models.

Regarding the daily VaR estimates, five out of eight models survived the Christoffersen frequency test for both stock indexes. According to the number of violations (see Table 7), HS and *t*-distribution model with time-constant volatility seemed to have the most consistent performance, as they are both ranked as the second best for both indexes. However, as

concluded by Baran and Witzany (2011), we should note that the viability of HS approach depends heavily on the fact whether or not the arbitrarily chosen sample window contains sufficient information to predict the VaR for the next period. Therefore, in this sense, t -distribution model would be easier to apply in a bigger scope since it is based on theoretical assumptions and has a parametric form. Furthermore, the rather unsatisfactory performance of the normal distribution model in our study, especially in the case of the FTSE 100 index, also confirms the finding by Baran and Witzany in their research.

Table 7: Ranking of the models for daily VaR estimates

OMX	Violations	FTSE	Violations
Uncondi POT	5	t-dist. EWMA	1
HS	6	HS	2
t-distribution	6	t-distribution	2
t-dist. EWMA	6	Uncondi POT	2
Condi POT	6	Normal	6

The EVT approach, in general, gave us a conflicting picture. This result could be due to among others the fitness of the selected threshold, i.e. the performance of the EVT approach depends on whether the chosen threshold fits well the underlying data. Unfortunately, there is no defined method to find the “best” threshold which is exactly the biggest drawback associated with the EVT approach. In fact, the controversy of the EVT approach has been discussed by many previous studies. For example, Bekiros and Georgoutsos (2005) claimed that even though the traditional VaR models perform satisfactorily on lower confidence levels, the EVT models clearly outperform when the level goes above 99 %. On the other hand, in a study of five Asian stock market indexes Lee and Saltoğlu (2001) questioned the superiority of EVT models. They concluded that the EVT models underperformed while models combining GARCH with t -distribution or normal distribution generated more consistent results. They also showed that some EVT models performed better during a time of crises.

To our surprise, models considering volatility clustering did not outperform their counterparts without this consideration. As far as we are concerned, there were two causes which led to this outcome. Firstly, the out-of-sample period is a relatively ‘calm’ period. Models considering volatility clustering cannot provide extra benefit when there is very few significant volatility clustering presented in the underlying data. Secondly, a market with more participated traders could display fewer volatility clustering. This explains why we saw

that models considering volatility clustering performed worse for FTSE 100 than for OMX Stockholm 30 index.

When looking at the difference between the two indexes, it seemed that all models performed better for FTSE than for OMX. Could it be that these conventional approaches are better at capturing the characteristics of the markets with a larger size, in our case, FTSE since most empirical studies were using the data from the big markets? We believe that further studies are needed in the future to answer this question.

Regarding the 10-day VaR estimates, it is very disappointing that all the models failed in the backtesting. It seems that the dependence of the violations was the major problem for the VaR models in our study. Therefore, either a better model is called for or we can question whether the “square root rule” is appropriate in all cases⁵.

All in all, we did find some differences between the two stock indexes when it comes to the estimation of Value at Risk. However, we do not think these differences are significant enough so that different VaR models are definitely required for big and small markets respectively. At last, our study also echoes the statement repeated by many other researchers such as Hendricks (1996) and Lechner and Ovaert (2010). They stressed that there is no single VaR method that is to be considered “the best” and it is important to let the circumstances at hand determine which VaR model that should be applied.

References

Baran, J., and Witzany, J (2011) “A Comparison of EVT and Standard VaR estimations” *[online] Social Science Research Network*

Bao Y. et al. (2004) “Evaluating Predictive Performance of Value-at-Risk Models in Emerging Markets: A Reality Check” *Department of Economics, University of Texas*

Bekiros S. and Georgoutsos D. (2005) “Estimation of Value-at-Risk by extreme value and conventional methods: a comparative evaluation of their predictive performance” *Int. Fin. Markets, Inst. and Money* 15 pp. 209-228

Berkowitz J. and O'Brien J. (2002) *The Journal of Finance* VOL LVII No. 3

⁵ See a relevant discussion in Dowd K. et al. (2002) “Long-term Value at Risk” *CRIS Discussion Paper Series – 2003.1*

Christoffersen P. (2002) “Elements of Financial Risk Management, Chapter 8 Backtesting and Stress Testing” *McGill University and CIRANO*

Dowd K. (2002) “Measuring Market Risk” *Wiley Finance*

Dowd K. et al. (2002) “Long-term Value at Risk” *CRIS Discussion Paper Series – 2003.I*

Gençay R. and Selçuk F. (2004) “Extreme value theory and Value-at-Risk: Relative performance in emerging markets” *International Journal of Forecasting* 20 pp. 287-303

Gençay R. et al. (2003) “High volatility, thick tails and extreme value theory in value-at-risk estimation” *Insurance: Mathematics and Economics* 33 pp. 337-356

Hendricks, D. (1996) “Evaluation of Value-at-Risk models using historical data” *Federal Reserve Bank of New York Economic Policy Review* No. 2 pp. 39-70

Kollias C. et al. (2011) “Stock markets and terrorist attacks: Comparative evidence from a large and a small capitalization market” *European Journal of Political Economy* 27 pp. 64-77

Lee, T.-H., Saltoğlu, B (2001) “Evaluating Predictive Performance of Value-at-Risk Models in Emerging Markets: A Reality Check” *Mimeo, Marmara University*

Lechner A. L. and Ovaert C. T (2010) “Value-at-risk: Techniques to account for leptokurtosis and asymmetric behavior in returns distributions” *The Journal of Risk Finance* Vol. 11 No. 5 pp. 464-480

Marimoutou, V. et al. (2009) “Extreme value theory and value at risk: Application to oil market” *Energy Economics* 31 pp.519-530

Pérignon C. et al. (2008) “Do banks overstate their Value-at-Risk?” *Journal of Banking & Finance* 32 pp. 783-794

Şener, E., et al. (2012) “Ranking the predictive performances of value-at-risk estimation methods” *International Journal of Forecasting* doi:10.1016/j.ijforecast.2011.10.002

Xue Y. and Gençay R. (2012) “Trading frequency and volatility clustering” *Journal of Banking & Finance* 36 pp. 760-773

Data sources

<http://www.nasdaqomxnordic.com/nordic/Nordic.aspx>

<http://finance.yahoo.com>

Appendix

Figure 3: Graphs of the daily losses

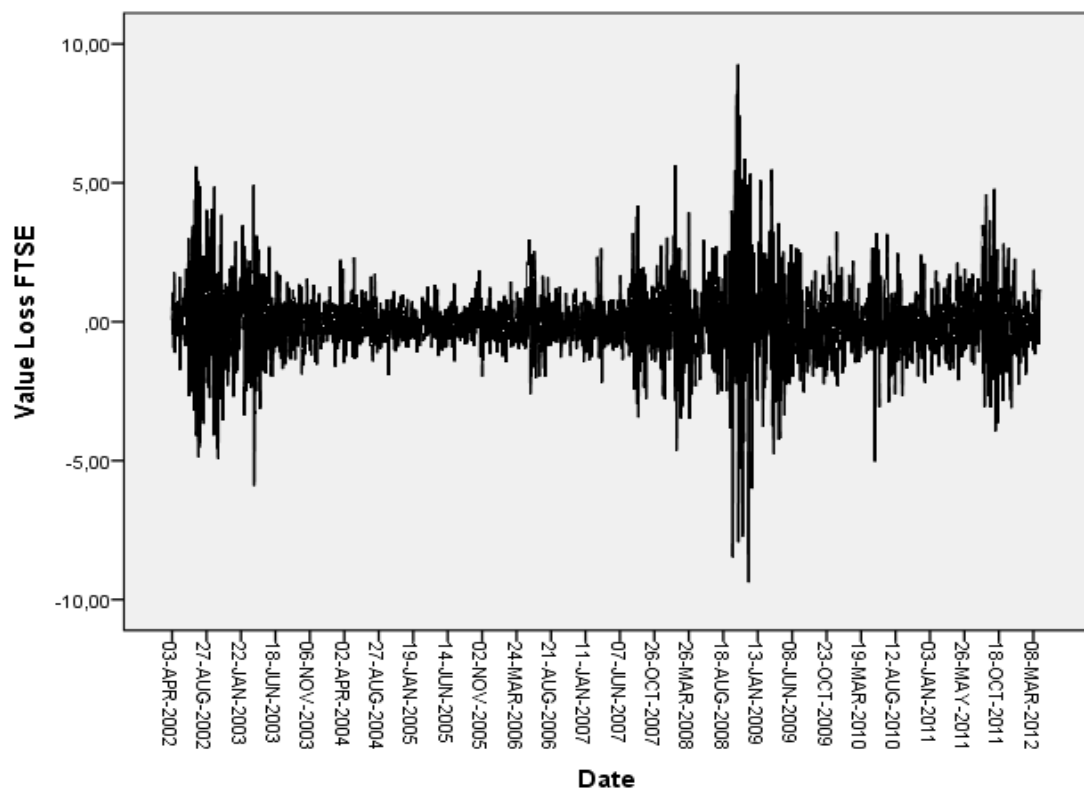
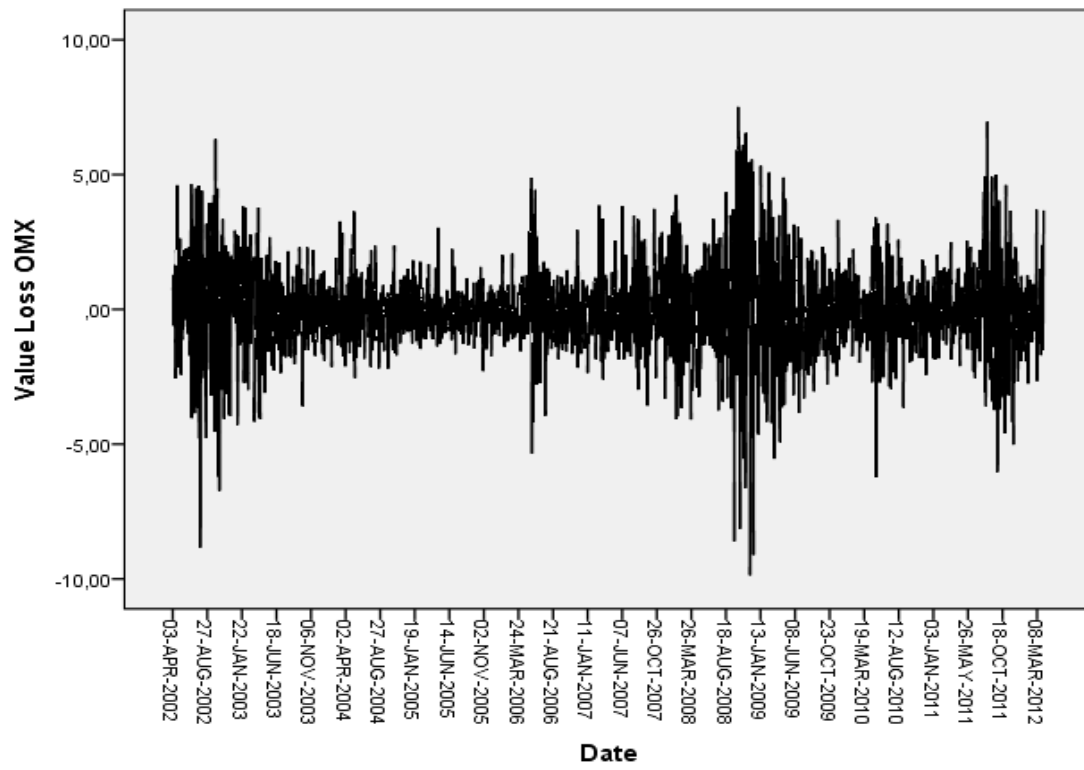


Figure 4: Graphs of the 10-day losses

