UPTEC STS10 028
Examensarbete 30 hp
Maj 2010

Estimation periods in risk parameter computations

Johan Wickman
Rikard Tullberg

UPPSALA UNIVERSITET

Teknisk- naturvetenskaplig fakultet
UTH-enheten

Besöksadress:
Ångströmlaboratoriet
Lägerhyddsvägen 1
Hus 4, Plan 0

Postadress:
Box 536
751 21 Uppsala

Telefon:
018 – 471 30 03

Telefax:
018 – 471 30 00

Hemsida:
http://www.teknat.uu.se/student

Abstract

Estimation periods in risk parameter computations

Johan Wickman, Rikard Tullberg

Today, there exist a broad range of different methods to handle financial risk. During the last decade, Value at Risk has become a popular and widely used risk measure due to its simplicity and convenient features. Value at Risk can be estimated by the use of three different methods; Historical Simulation, Variance Covariance or Monte Carlo.

This paper is made on behalf of the risk management department at NASDAQ OMX. Today, OMX uses two years of historical data to estimate VaR with the Historical Simulation method and questioning if the parameter might be better off, if it instead is based on a different estimation period. Our issue is thus: Is two year of historical data the optimal estimating period for OMX to estimate VaR with Historical Simulation?

Our main contribution in this study is a comprehensive empirical investigation where different estimation periods have been compared. This is done by a technique called backtesting. Four different methods have been used to analyze conditional and unconditional coverage, including Kupiec´s proportion of failures test, the Basel committee´s traffic light test, Christoffersen´s interval forecast test and a test for autocorrelation.

The results indicate that the best estimation period to calculate Value at Risk with the method "Historical Simulation" is one year. Our conclusion also points out shortcomings of the assumptions behind Value at Risk. The quality of data and the strong assumption that price changes are independent and equally distributed is seriously questioned.

Handledare: Karl Klasén
Ämnesgranskare: Erik Ekström
Examinator: Elísabet Andrésdóttir
ISSN: 1650-8319, STS10028

Populärvetenskaplig sammanfattning

Idag finns det många metoder för att behandla finansiell risk. Under det senaste decenniet har riskmåttet Value at Risk (VaR) blivit alltmer förekommande tackvare sin enkla struktur och användarvänlighet. VaR kan beräknas med tre olika matematiska metoder; Historical Simulation, Variance Covariance method och Monte Carlo methond.

Den här undersökningen är gjord för riskavdelningen på Nasdaq OMX. Idag använder de sig av två års historisk data för att beräkna VaR med Historical Simulation method, och har ifrågasatt ifall parametern skulle vara bättre estimerad med en annan tidslängd. Vår problemfrågeställning är därav: Är två års historisk data den optimala tidsperioden för OMX att estimera VaR med Historical Simulation?

Huvuddelen av arbetet består av en omfattande kvantitativ studie där utfallet av olika estimeringsperioder har undersökts. Detta har genomförts genom så kallad back-testing. Fyra olika metoder har använts för att analysera villkorlig och ovillkorlig täckning, med Kupiecs test, Basel test, Christoffersens test och test för autokorrelationen.

Resultaten har indikerat att den bästa estimeringsperioden för estimering av VaR med Historical Simulation är ett års historia. Våra slutsatser pekar även mot bristerna i antagandena för VaR. Datakvaliteten och de starka antagandena om att prisförändringar är oberoende och jämnt fördelade över tiden är ifrågasatt.

Acknowledgements

During the progress of the work on this project we have received a great deal of support, guidance and feedback from a several number of people. In particular, we would like to thank our supervisor Karl Klasén at NASDAQ OMX for his guidance. Moreover, we would like to send our appreciation to the Risk Management at NASDAQ OMX, in particular Albert Lindblom and David Sjöblom, for their guidelines and helpfulness during the practical work.

Further, we would like to thank Henrik Hult, Department of Mathematics at the Royal Institute of Technology for taking time to meet with us and discus our mathematical issue.

Nevertheless, we would like to express our gratitude to our subject reviewer Erik Ekström, Department of Mathematics at Uppsala University for guidelines, feedback and support during the whole progress of our work.

Stockholm, March 2, 2010

Johan Wickman and Rikard Tullberg

Table of contents

1 Introduction

This paper is made on behalf of the risk management department at NASDAQ OMX Derivatives Markets. The risk management department is responsible for managing NASDAQ OMX Derivatives Markets counterparty risk i.e. the risk that one or several market participants will default in their obligations to the clearing organization. To monitor counterparty risk, different risk parameters have been developed with objective to function as alarm signals if the value of a participants position changes drastically over a short period of time. Today, a widely used risk measure called Value at Risk is being used. This measure can be calculated in different ways, depending on the effort of the company, time and resources, and how much information that is available about your risky assets. NASDAQ OMX Derivatives Markets uses a framework that is built upon the method called Historical simulation which is estimated by the historical data of the underlying asset. From here the risk measure Value at Risk is going to be denoted as VaR and NASDAQ OMX Derivatives Markets as OMX.

During the last five months we have studied the implications of this method and will in this study present a comprehensive empirical investigation of the method's accuracy.

This paper is divided in the following parts. First, we present the objective and put the problem in a context. Further we give an introduction to OMX as a clearinghouse and the risk management department. We then present the risk measure Value at Risk and the different kind of backtesting methods our empirical investigation is based on. Finally, results, analysis and conclusions are drawn from the empirical investigation and as a sum up a discussion of the investiagation and further studies are presented in the last section.

2 Problem formulation

Today, OMX is using the Historical Simulation (HS) method to calculate its VaR parameter. Apart from its advantages and convenient features, it brings various weaknesses and drawbacks. The weaknesses of the method are mostly related to the data, as the parameter completely depends on the particular historical data sets used. The underlying assumption is that the past, as captured in the historical data set is sufficiently similar to the future to give a reliable idea of future risks. Implicitly, it presumes that the risks we face in the future are similar to those we have faced in the past (Dowd, 1998). While this assumption is often reasonable, it can lead to seriously vague estimates of VaR.

One of the major problems is that the data of the volatility of the data may fluctuate over time. It may be below average, which will result in a too low estimate of risk or it may be unusually high, in which case our estimates of VaR are likely to be too high. The data could also include unusual events that we would not expect to return in a foreseeable future, e.g. a major stock market crash. These examples will keep the VaR estimation too high as long as the data of the event still remains in the estimation period, e.g. two years as OMX is using today (see chapter 4 for fundamentals of VaR). By the same fundamental principal, the HS approach also has difficulty of dealing with permanent changes in risk factors. The estimation period will still be represented of old non up-to-date observations, until the old estimation period has been replaced by the data consisting of the new permanent changes. For example, if the estimation period is two years of historical data, and today some major important change appear which will affect the future, it will take us two years before the old data is out of our estimation period. Therefore the HS method is sensitive to single large changes and insensitive against macroeconomic changes, depending on how long estimation period we choose to use.

There is also another problem when dealing with the length of estimation periods. As the HS method uses its empirical distribution to estimate VaR, the inference of the distribution will clearly be more accurate the longer run of data we include, especially when VaR is dealing with high confidence levels. If we base our VaR on a 95% confidence level we would have to wait on average 20 days (20*(1-0.95)=1, see chapter 4)) to expect a single loss in excess of VaR and 100 days with a 99% confidence level and so forth. The more extreme VaR we want to estimate i.e. high confidence level, the more observations are required. Assuming that the underlying distribution remains constant over the whole estimation period, we would presumably want the longest possible period to maximize the accuracy of our results.

However, it is likely that at some point, the data will exhibit some systematic changes over time, and due to the first problem mentioned above, this might lead us to prefer a somewhat shorter estimation period. The question is; do we want a well estimated VaR, based on long runs of data and implicitly not consider new changes, or do we want a VaR based on fewer data, but instead more accurate to new changes? This is the question OMX wants an answer too, and therefore is the topic of this study.

2.1 Issue and objective

Today, OMX are using two years of historical data to estimate VaR with the Historical Simulation method and questioning if the parameter might be better off if it instead would be based on a different length of historical data. Our issue is thus: Is two year of historical data the optimal estimating period for OMX to estimate VaR? If not, which would be better?

Hence, the objective is to distinguish an optimal length of historical data, which implies a comparison of different lengths of estimation periods and to evaluate the outcomes.

2.2 Empirical Backtesting method

To assess the optimal length of historical period for estimating VaR given the HS approach, we define the optimal period as one which would result in a VaR estimation that best covers for future movements in underlying asset changes.

Optimal estimation period, definition 1: *The optimal length of historical period would estimate a VaR that generates the number of exceedances closest to its given hypothesis*

The hypothesis regards to the theory of Value at Risk, which incorporates the length of estimation period and the confidence level of VaR estimation. The term exceedances refer to the number of observations that would have exceeded the VaR parameter, i.e. the number of observations where the predetermined VaR parameter would have been to low estimated and hence would not take cover for future movements. This process is called backtesting. The process estimates a VaR parameter for a certain historical period and stress test the parameter over the next following period (during a quarter in this case). Thus, to perform a backtest the historical estimation period must not be closer than at least one quarter from today's date.

By this logic, you can perform the backtest during the same period, but for a range of different lengths of historical periods used for estimating the VaR parameter. In other words, you can estimate VaR parameters for different lengths of historical data, e.g. one, two or three years, and then backtest them throughout one certain period and compare the number of exceedances.

By backtesting different VaR parameters (different lengths of estimation periods), all estimation periods have to end at the same date so the backtest begins in the same point in time for each parameter, e.g. a quarter from now in the case of backtesting the last quarter.

The backtest concerns only a period of one quarter. To make the study more accurate, the backtest is performed on a multiple number of periods. Thus, the backtesting procedure is going to begin further back in time, and then step forward one quarter at a time, until it reaches today's date. For every quarter, the results from the backtest are summarized for each VaR parameter of all segments of stocks. The results are going to represent all segments' number of parameter breaks for each quarter and for every VaR parameter.

2.2.1Statistical backtesting methods

VaR models are only useful if they predict risk reasonably well and should always be accompanied by validation. Backtesting is a formal statistical framework that consists of verifying that actual losses are in line with projected losses. When the model is perfectly calibrated, the number of observations falling outside VaR should be in line with the selected confidence level. The number of exceedences is also known as the number of exceptions. With too many exceptions the model underestimates risk, and with too few it lead to inefficient allocation of capital across units. (Jorion, 2001)

Backtesting involves a systematically comparison between historical VaR measures with the subsequent returns. Since VaR models are reported only at specified confidence levels, the model is expected to be exceeded in some instances. A confidence level of 95 percent implies that five out of a hundred observations will exceed the VaR limit. In reality it is not uncommon that there are more or fewer exceptions. A greater or smaller percentage could occur because of bad luck. If the frequency of deviations becomes too large, the conclusions must be that the problem lies within the model and not bad luck. (Jorion, 2001)

Tests that statistically examine whether the frequency of exceptions is in line with the selected confidence level are known as tests of unconditional coverage. These tests are straightforward to implement and do not take into account when in time the exceptions occur and make no assumptions about the return distribution. This approach is in other words fully nonparametric. A good VaR model should however not only produce the correct number of exceptions but also exceptions that are evenly spread in time i.e. independent and identically distributed. Tests that take this into account are known as tests of conditional coverage. (Jorion, 2001)

Thus, we need to add the conditional coverage to definition 1 and reformulate the optimal estimation period as:

Optimal estimation period, definition 2:
The optimal length of an historical period would estimate a VaR that generates the number of exceedances that are closest to its given hypothesis and uniformly spread in time i.e. independent and identically distributed

Therefore the statistical study consists of tests that both take into account unconditional and conditional coverage. Tests that have been made to study unconditional coverage include Kupiec´s proportion of failures test (Kupiec, 1995) and Basel Committees; traffic light approach (Basle Committee, 1996). Tests for conditional coverage consist of Christoffersen´s interval forecast test (Jorion, 2001) and a test for autocorrelation (McNeil, Frey, Embrechts, 2005). These are the most common and well developed test methods in evaluating VaR models and hence good reason for us to use it in our investigation.

2.2.2 Choice of lengths for estimation periods
The suspicion of OMX is that a shorter estimation period than two years would generate more accurate results. Hence, we have chosen a half, one, two and three years of historical data as the different lengths of estimation periods for the VaR parameter. Thus, the study captures a range from longer to shorter periods than two years.

2.2.3 Choice of backtesting periods
When dealing with high confidence levels, it is as mentioned above, significant how much data there is available. With too low amount of data, the statistical test will not be valid. Another motive to use long history of data is to cover as many market situations as possible that have occurred in the past. Time periods with or without a financial crisis will result in very different outcome. Thus, our ambition is to choose a period not too far from today, but still a period that covers different types of market fluctuations. We have therefore chosen six years, starting quarter three, year 2003. This incorporates 24 quarters and is the length of each backtesting period. I.e. the VaR parameter is going to be backtested during the period from 2003 to 2009.

2.2.4 Choice of stocks
As mentioned in previous section we have chosen a six year backtesting period. The longest estimation period we are going to backtest is three years. Thus, it demands nine years of history for every stock to make it possible to backtest a three year estimation period during a six year period. Therefore we omit stocks with shorter than nine years of history.

The stocks registered at OMX are divided into three segments; Large, Mid and Small Cap, due to its size of market capitalization. As a consequence, we do the study separately for each stock segment.

2.2.5 Backtesting Program in Visual Basic

Our study is based on a constructed macro in Excel, programmed in Visual Basic. A list of stocks with nine years of history is the only input to the program. The outcome consists of the quarterly VaR parameter and its number of exceedances for every stock. The steps in the backtesting process are as follows:

1. Load historical data for one stock
2. Calculate the stocks daily returns, i.e. the daily change in percent
3. Calculate the VaR parameter
4. Backtest the parameter and record the number of exceptions over the present quarter
5. Step one quarter ahead and repeat point 3 and 4 until today's date has been reached
6. Continue with the next stock in line and repeat the same procedure until the last number of stock is reached
7. Change the length of estimation period, and backtest the rest of the VaR parameters, i.e. 0.5, 1,2 or 3 years
8. Now all VaR parameters are backtested for one particular stock segment. Repeat 1-7 until all stock segments are backtested

3 NASDAQ OMX Derivatives Markets

In 2008 the NASDAQ OMX Group was formed as a result of a merger between NASDAQ Stock Market Inc and OMX AB. Today, NASDAQ OMX Group is the world´s largest exchange company. It delivers trading, exchange, technology and public company services across six continents. (http://Nordic.nasdaqomxtrader.com)

One of the principal functions of a clearing organization is to guarantee that all contracts traded will be fulfilled. At OMX, clearing is integrated with derivatives trading, whereby OMX becomes the counterparty in all transactions i.e. acts as a buyer to the seller and as a seller to the buyer. As a central counterparty, the main tasks of OMX are to;

- Participate as counterparty in every transaction
- Monitor the market and market participants
- Handle all transactions
- Calculate and administrate collateral

All market participants with an agreement with OMX have a direct legal relationship with OMX. No end customers normally participate directly in trading but are represented by approved exchange and clearing members. OMX offers a broad range of products; forwards, futures and options on single Swedish, Norwegian, Finnish, Danish, Icelandic and Russian stocks as well as options and futures on the OMXS30 index (the 30 most traded Swedish stocks). (http://Nordic.nasdaqomxtrader.com)

3.1 Counterparty Risk Management

NASDAQ OMX faces several kinds of risks every day; everything from traditional business risk to risk associated specific with the derivative clearing services it provides. The most noteworthy is the risk that one or several market participants will default on their obligations to the clearing organization, i.e. counterparty default. To handle this kind of risk, NASDAQ OMX relies on several vital factors such as quality and control of the overall clearing operations, the counterparty risk management framework, the collateral that is being pledged by participants and the established rules and regulations from authorities. (http://Nordic.nasdaqomxtrader.com)

Counterparty risk is the risk that one participant in a transaction will not be able to fulfill its obligations in the future, due to the contract’s obligations. NASDAQ OMX Derivatives markets enter as counterparty for both the buyer and the seller of a binding derivative and therefore handle all the risk in a transaction, se fig 1.

Figure 3.1: *OMX Derivatives Markets as a counterparty for both the buyer and the seller of a financial contract*

A participant has to pledge collateral to NASDAQ OMX when involved in a transaction. The collateral requirement is calculated based on the particular counterparty´s trading position and determined by risk models used to calculate an individual counterparty´s daily margin requirements, which in case of a default would be the resources to liquidate the underlying portfolio. (http://Nordic.nasdaqomxtrader.com)

The margins are required to avoid the risk of a loss for the clearing organization if a participant defaults. The margin requirements should not be too low, as this would exceed the risk of a loss, but neither too high, as this could disrupt trading and clearing.

If the counterparty defaults, the time to close an account varies depending on the type of the account. Under normal circumstances it takes time to neutralize an account and during this time the value of the account can change. It is conservatively assumed that it takes two days on average to close a counterparty´s position. The margin parameters are for this reason calculated with a two day lead time. (http://Nordic.nasdaqomxtrader.com)

3.2 Risk management department

The risk management department is primarily responsible for managing NASDAQ OMX Derivate Markets counterparty risk. This is done by a framework of policies, standards, procedures and resources. The department manages changes in counterparty risk exposures against a range of risk limits on a daily and intraday basis. This work involves handling parameter breaks, intraday risk reporting and monitoring, intraday margin calls and blocking of trades and offsetting trades. Every day the risk management department is made aware of any risk parameter breaks (situations where the change in day-to-day market prices has exceeded the approved risk interval parameter level for any underlying security). When this happens the counterparty accounts are reviewed in detail and the cause of the price movement is also reviewed. A decision is then taken by the risk management department. A risk parameter break is from a statistical point of view expected. This means that a risk parameter break doesn´t have to signify a need to change the parameter. A more detailed investigation has to be done. If the desired collateral that is being pledged by a customer is not satisfied, OMX Derivatives Markets has the authority to make intraday margin calls. This is done until the desired level of collateral is reached. (http://Nordic.nasdaqomxtrader.com)

4 Value at risk

How much you can lose at most on this investment is a question all investors at some time ask themselves before an investment or during a possession in some asset. Every consideration of an investment incorporates a financial risk which therefore creates an interest among the investor as to what value that is on stake.

What determines the financial risk can briefly be explained by both human-created as well as natural elements. As for the first mentioned, we humans have given rise to and have impact on phenomenon such as business cycles, inflation, changes in government policies, wars etc., and secondly natural phenomenon can arise both predictable and unpredictable such as hurricanes and earth quakes. No less important, the financial risk also comprises today's evolution, such as long-term economic growth and technological innovations. All these factors have an impact on financial risk and as a result it has caused escalating volatility the last decades for exchange rates, interest rates and commodity prices. These three risk variables go by the common name market risk, which is one of several risks that are linked to financial risk, such as credit risk, liquidity risk, legal risk and operational risk. (Penza, Bansal 2001)

As a result of technology innovations and new economic theories, emergence of new risk management methods have been developed. Corporations today such as banks, brokerage firms and investment funds use similar methods to compute financial risk. These together have with statistical-based risk-management systems evolved a risk measurement called Value at Risk (VaR). (Penza, Bansal 2001)

The underlying mathematics in VaR can originally be deduced from prior portfolio theory by Harry Markowits. The evolution of the risk measure has its most important turning points after earlier financial crises because of the aftermath of the events. The consequence was more pressure on the banks, which induced an increase in regulatory capital requirements. In return, it followed by a demand from the banks themselves to elaborate risk measures and control devices to ensure that they met these capital requirements. After the mid 80s, VaR began to occur in corporations risk departments, but it was not before ten years after that VaR became a standard risk measurement as a consequence of the investment bank J.P Morgan officially introduced it, and in the same time it was published in the G30[1] report 1993. (Jorion, 2001)

[1] Group of thirty, also known as G30, is a consultive group composed of academics and financiers whose goal is to facilitate understanding of financial and economic issues in the private and public sector. (Investopedia.com, 3-12-09)

4.1 Fundamentals of VaR

VaR is a statistical concept that represents an estimation of the maximal loss an institution could ever lose on a portfolio of assets. (Penza, Bansal 2001)

Jorion defines VaR by:

Definition 1.1 VaR summarizes the worst loss over a target horizon that will not be exceeded with a given level of confidence. (Jorion, 2001)

This implies that losses bigger than VaR only occur with a certain small probability (1 – conf. level). (jorion, 2001)

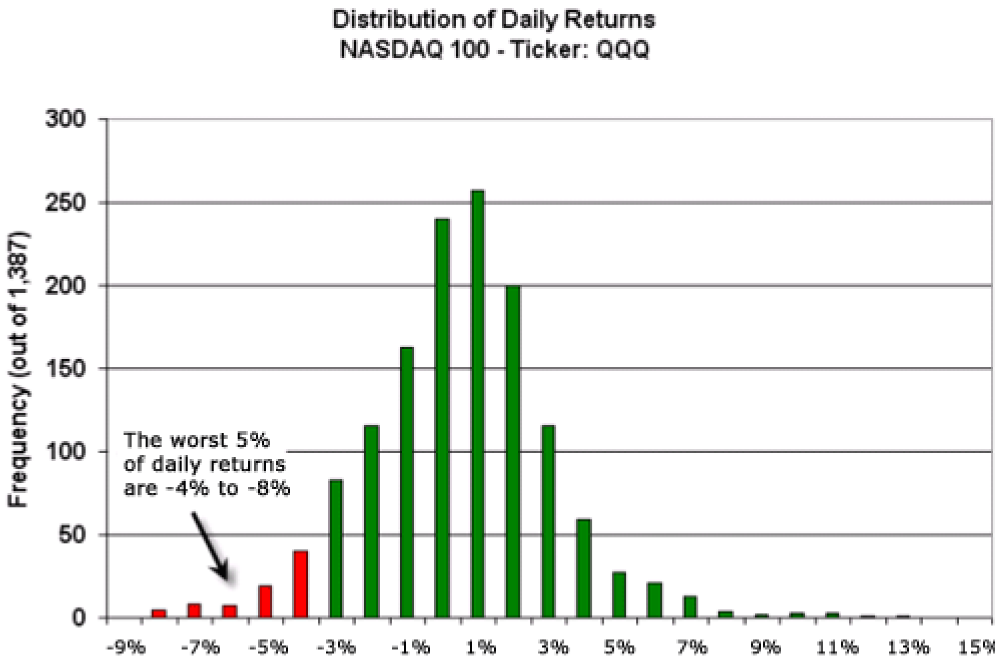

Figure 4.1: *An illustration of VaR with confidence level 95% on the NASDAQ 100 index. The samples to the left are negative and are the losses. The red marked stacks is the 5% worst samples, i.e. exceeds of VaR which in this example is -3% (Investopedia.com, 2009)*

VaR is a flexible instrument as it only consists of two parameters that are adjustable; the time horizon and the probability level. The choice of the time horizon is mainly subjective and related to the business of the financial institution and the kind of portfolio. For a bank trading portfolio invested in high liquid currencies, the choice of a one day time horizon is probably the best one, compared to a portfolio with quarterly rebalancing and reporting. Ideally, the time horizon corresponds to the longest period needed for orderly portfolio liquidation. (Penza, Bansal 2001)

The dependence of probability level has no guidance from the finance theory. Instead, it is a matter of risk aversion from the manager’s perspective. The probability level determines the number of excepted future exceptions of VaR, for example 5%, over the specified time horizon. In other words it is up to the manager to decide if a loss occurring with probability equal to 5% or 1% should be treated as extreme (Jorion, 2001).

The formal definition of VaR can then be formulated as:

Definition 1.2 Given a probability of α percent and a holding period of t days, an entity's VaR is the loss that is expected to be exceeded with a probability of only 1- α percent on the t-day holding period (Penza, Bansal 2001).

When working with market risk, the time horizon is often set to one or ten day holdings, while for credit risk it is more frequent with one day. Common confidence levels are 95%, 99% and 99.9% depending on the application (OMX has chosen 99.2%). The Bank of International Settlements proposes for market risk to compute ten-day VaR with confidence level α = 99% (Hult, 2007).

The mathematical definition of VaR is defined as:

Definition 1.3 Given a loss L and a confidence level $\alpha \in (0,1)$, $\mathrm{VaR}_\alpha(L)$ is given by the smallest number l such that the probability that the loss L exceeds l is no larger than $1 - \alpha$, i.e.

$$\begin{aligned}\mathrm{VaR}_\alpha(L) &= \inf\{l \in \mathbb{R} : P(L > l) \le 1 - \alpha\} \\ &= \inf\{l \in \mathbb{R} : 1 - F_L(l) \le 1 - \alpha\} \\ &= \inf\{l \in \mathbb{R} : F_L(l) \ge \alpha\}\end{aligned}$$

Here $\mathbb{R}$ denotes the set of real numbers, F_L the loss distribution. We might think of L as a loss resulting from holding a portfolio over some fixed time horizon. (Hult, 2007)

4.2 Historical simulation of VaR

There are three main methods to estimate VaR; Historical simulation, the Variance-Covariance method and the Monte-Carlo method. The method OMX today uses is the Historical simulation (HS). This method makes very few assumptions about the market price processes generating the portfolio's returns. It simply assumes that market price innovations in the future are drawn from the same empirical distribution as those market price innovations generated historically. I.e. it uses historical samples to construct a cumulative distribution function for future price changes. By using the empirical distribution, many of the problems inherent in modeling the evolution of market prices can be avoided. The market prices tend to have fatter tails and be slightly more skewed than predicted by the normal distribution. By applying the historical simulation, we suppose that we have observations $x_1, \ldots, x_n$ of independent and identically-distributed (iid) random variables $X_1, \ldots, X_n$ with distribution F. X_i is the percentage price change of an asset Y during the time horizon t, i.e. . $X_i =$

$Y_i/Y_{i-t} - 1$. As OMX takes cover for both upward and downward price changes due to certain derivatives, they use the absolute value of the price changes and hence does not distinguish between negative or positive returns. The empirical distribution function is then given by

$$F_n(x) = \frac{1}{n}\sum_{k=1}^{n} \mathbb{I}_{[X_k,\infty)}(x).$$

Here $\mathbb{I}_A$ is the indicator function: $\mathbb{I}_A(x) = 1$ if $x \in A$ and 0 otherwise. The empirical VaR is then given by

$$\mathrm{VaR}_\alpha(F) = \inf\{x \in \mathbb{R} : F_n(x) \geq \alpha\}$$

If we order the sample $X_1, \dots, X_n$ such that $X_{1,n} \geq \cdots \geq X_{n,n}$ (if F is continuous, then $X_j \neq X_k$ a.s. for $j \neq k$), then the empirical quantile is given by

$$\mathrm{VaR}_\alpha(F) = X_{[n(1-\alpha)],n} \qquad \alpha \in (0,1)$$

where $X_{[y],n}$ is an interpolation of $X_{y,n}$, since y is not a discrete number, which is required.

The interpolation is solved as follows:

$$R_1 = [n(1-\alpha)]$$
$$R_2 = R_1 + 1$$
$$dx = R_2 - (n-1)(1-\alpha)$$

where [w] is the integer part of w, $[w] = \sup\{n \in \mathbb{N} : n \leq w\}$ (the largest integer less or equal to w). Further,

$$V_1 = X_{R_1,n}, \quad V_2 = X_{R_2,n}$$

which represents the R_1:th and the R_2:th largest values of the n number of observations. Finally, the VaR is estimated as

$$\mathrm{VaR}_\alpha(F) = X_{[n(1-\alpha)],n} = V_1 dx + V_2(1-dx)$$

For example, if we have 250 observations with confidence level 95% we get:
$R_1 = [250(1-0.95)] = [12.5] = 12$
$R_2 = 12 + 1 = 13$,
$dx = 13 - 249(1-0.95) = 13 - 12.45 = 0.55$.

This means that V_1 is the 12th largest value of the 250 observations, $X_{12,250}$, as V_2 the 13th, $X_{13,250}$. Then VaR is calculated as: $\mathrm{VaR}_{0.95}(F) = X_{12,250} \times 0.55 + X_{13,250} \times (1 - 0.55)$

5 Statistical Backtesting methods

As what have been mentioned, an accurate VaR model needs to qualify in both conditional and unconditional coverage. The presentation of this section is hence divided into two sections, named after the above criteria.

5.1 Tests of unconditional coverage

In a good VaR model the number of exceptions should be in line with the selected confidence level. As discussed earlier this is not always the fact in reality. Too many deviations from the model should however be an alarm signal that the model is badly calibrated or even wrong.

One of the simplest methods to verify the accuracy of a VaR model is to record the failure rate, which gives the proportion of times VaR is exceeded in a given sample. Ideally, the failure rate $\frac{x}{T}$ where x is the number of exceptions and T is the number of days, should give an unbiased measure of the confidence level p i.e converge to p as the sample size increases. At a given confidence level, we want to know if x is too small or too large under the null hypothesis that:

$$H_0: p = \hat{p} = \frac{x}{T}$$

If the observed failure rate $\hat{p}$ differs significantly from the failure rate suggested by the confidence level p, the null hypothesis is rejected. (Jorion, 2001).

The setup for the test is a Bernoulli trial, which is a classical framework of success and failures. Under the null hypothesis that the model is correctly calibrated, the number of exceptions x follows a binomial probability distribution:

$$\Pr\langle X = x|T, p\rangle = \binom{T}{p} * p^x * (1-p)^{T-x}$$

When T is large, we can use the central limit theorem and approximate the binomial distribution by the normal distribution:

$$z = \frac{x - p * T}{\sqrt{p(1-p) * T}}$$

This binomial distribution can be used to test whether the number of exceptions is acceptably small. (Jorion, 2001)

When designing a verification test for unconditional coverage, there is a tradeoff between two kinds of errors:

- Type 1 error- rejection of a correct model
- Type 2 error- not rejecting an incorrect model

	Model	
Decision	**Correct**	**Incorrect**
Accept	Ok	Type 2 error
Reject	Type 1 error	Ok

Table 5.1: A summary of the two states, correct versus incorrect model and the decision.

When backtesting VaR models, the users have to balance type 1 errors against type 2 errors. (Jorion, 2001)

5.1.1 Kupiec´s POF test

Kupiec´s proportion of failures test attempts to determine whether the observed frequency of exceptions is consistent with the frequency of expected exceptions according to the VaR model and chosen confidence level. It also tries to balance between the two types of errors discussed in section 5.1. The test uses the binomial distribution to calculate the probability that a certain number of VaR breaks will occur given a selected confidence level and sample size. Under the null hypothesis that the model is "correct", the number of exceptions follows the binomial distribution.

$$\Pr\langle X = x | T, p\rangle = \binom{T}{p} * p^{x} * (1-p)^{T-x}$$

where x is the number of VaR breaks, p is the selected confidence level and T the sample size.

By using the cumulative binomial distribution, an interval can be calculated within which the number of VaR breaks must fall for the test to accept the VaR model. Kupiec defines these regions by the tail points of the log- likelihood ratio;

$$LR_{POF} = -2\ln\left(\frac{(1-p)^{T-x} * p^{x}}{\left[1-\left(\frac{x}{T}\right)\right]^{T-x} * \left(\frac{x}{T}\right)^{x}}\right)$$

This quantify is asymptotically (when T is large) chi square distributed with one degree of freedom under the null hypothesis that p is the true probability. The cutoff value for rejection of the test is given by the chi square distribution (appendix 1). If LR_{POF} exceeds the cut off value, the model is rejected, though the number of exceptions is outside the interval. A rejection implies that the confidence level that has been used when calculating VaR did not match the actual probability of VaR breaks.(Jorion, 2001)

5.1.2 Basel committee´s traffic light test

The Basel committee of banking supervision has as its objective to enhance and improve the quality of banking supervision worldwide. The committee's members come from all around the world and their work is best known for its international standards on capital adequacy. A broad framework of rules has been outlined regarding back tests of different risk models. The verification procedure consists of recording daily exceptions of the 99 percent VaR over the last year i.e. 250 trading days. The size of risk capital requirement depends of the outcome of the model backtest. If the risk becomes larger so will the capital requirement.(Basle Committee, 1996)

The test is directly derived from the failure rate test and tries to balance between the two types of errors discussed in section 5.1. The test begins with counting the number of exceptions that occurred over the selected number of days. The expected number of exceptions is given by T*(1-p), where T is the number of days and p the selected confidence level.(Jorion, 2001)

The committee has decided that over 250 trading days with 99 % VaR, up to four exceptions are acceptable. These four exceptions belong to the first of three categories and is defined as the "green zone". If the number of exceptions is in the interval from 5-9, they fall into the second category, the yellow zone. More than nine exceptions and they fall into the third category, the red zone. The green zone is defined as an accurate level of exceptions and the probability of accepting an inaccurate model is low. More exceptions are connected with a higher probability that the exceptions were produced by an inaccurate model. If the number of exceptions ends up in the yellow zone, these exceptions could be produced by both an accurate but also an inaccurate model with a higher probability than the outcomes in the green zone. The red zone generally indicates a clear problem with the model. There is only a small probability that an accurate model would generate ten or more exceptions from a sample of 250 observations. In other words, the red zone should lead to an automatic rejection of the model. (Jorion, 2001)

As illustrated in table 5.1.2 more exceptions than 4 result in a higher capital requirement.

Zone	Number of exceptions	Increase in scaling factor	Cumulative probability
	0	0	8,11
	1	0	28,58
	2	0	54,32
Green	3	0	75,81
	4	0	89,22
	5	0,4	95,88
	6	0,5	98,63
Yellow	7	0,65	99,6
	8	0,75	99,89
	9	0,85	99,97
Red	10 and more	1	99,9

Table 5.1.2: *Illustrates the boundaries for the Basel test over 250 days. If the number of exceptions is five or more, the financial institution incurs a progressive penalty (the scaling factor increases). The cumulative probability is the probability of obtaining a given number or fewer exceptions when the model is correct (true 99 % coverage). The yellow zone begins at the point where the cumulative probability exceeds 95 % and the red zone where the cumulative probability exceeds 99.99%.*

5.2 Tests of conditional coverage

A good VaR model should not only produce the "correct" number of exceptions but also produce exceptions that are evenly spread in time i.e. independent. Clustering of exceptions is an indicator that the model is inaccurate. The market could experience increased volatility that is not captured by VaR. The investigation of a good VaR model should therefore consist of tests for conditional and unconditional coverage.

5.2.1 Autocorrelation

This measure is part of our investigation of conditional coverage. Clustering of exceptions is a signal that the model does not capture the volatility in the market (Jorion, 2001). The autocorrelation is represented as a number between$[-1,1]$, where 1 defines perfect correlation and -1 perfect negative correlation. In other words, this is an indicator how two time periods correlates to each other.

Autocorrelation is defined as;

$$R(\tau) = \frac{E[(X_t - \mu)(X_{t+\tau} - \mu)]}{\sigma^2}$$

Where "E" is the expected value operator, X_t is a time series with mean μ and variance σ^2, $X_{t+\tau}$is the same time series with a lead time of τ with the same mean and standard deviation. (McNeil, Frey, Embrechts, 2005).

5.2.2 Christoffersen´s interval forecast test

Christoffersen´s interval forecast test examines whether the probability of an exception today depends on the outcome the previous day. The same log-likelihood framework as Kuipec POF test is used but extended to include a statistic for independence of exceptions. (Jorion, 2001)

The test takes its beginning by first defining an indicator variable. If VaR is exceeded any day during the specific time period, the indicator gets a value of 1. If there´s no exception it gets the value 0.

$$I_t = \begin{cases} 1 & if\ violation\ occurs \\ 0 & if\ no\ violation\ occurs \end{cases}$$

The next step is to define n_{ij} as the number of days in which state j occurred in one day while it was as i the previous day and π_i as the probability of observing an exception conditional on state i the previous day. If an exception is independent of what happened the previous day π_0 and π_1 should be equal. (Jorion, 2001)

$$\pi_0 = \frac{n_{01}}{n_{00} + n_{01}} \qquad \pi_1 = \frac{n_{11}}{n_{10} + n_{11}} \qquad \pi = \frac{n_{01} + n_{11}}{n_{00} + n_{01} + n_{10} + n_{11}}$$

The test statistic for independence of exceptions has the form

$$LR_{ind} = -2\ln\left(\frac{(1-\pi)^{n_{00}+n_{10}} * \pi^{n_{01}+n_{11}}}{(1-\pi_0)^{n_{00}} * {\pi_0}^{n_{01}} * (1-\pi_1)^{n_{10}} * {\pi_1}^{n_{11}}}\right)$$

This framework allows an investigation whether the reason for not passing the test depends on clustering of exceptions, by calculating the statistic LR_{ind} and compare the value against the chi square distribution with one degree of freedom (see appendix 1) as the cutoff value. The value has to be lower than the critical value of the chi square distribution to pass the test. (Jorion, 2001)

For example, if you of 252 days have 20 exceptions this is equivalent with π = 7.9 percent. Of these, 6 exceptions occurred following an exception the previous day and 14 exceptions occurred when there was none the previous day. The conditional probabilities are;

π_0 = 14/232=6 %
π_1 = 6/20 = 30 %

This indicates that we have a much higher probability of having an exception following another one. We find LR_{ind} = 9.53. This is a higher value than the cutoff value of 3.84 at the confidence level 95% (see appendix 1) and therefore we reject independence.

6 Empirical results

In our study, we observe stocks over a period of six years. During this time, the volatility of the stocks fluctuates between the start and end point of the investigation period. Obviously, there had been at least one major financial crisis during this time. However, that will neither be taken into account nor be noticeable in the concluding analysis, as we look at the result as one final outcome, not what has occurred one quarter to another. As an orientation we therefore show the histogram for the average parameter breaks 99.2% VaR, with a two year estimation period for each of the three segments of stocks.

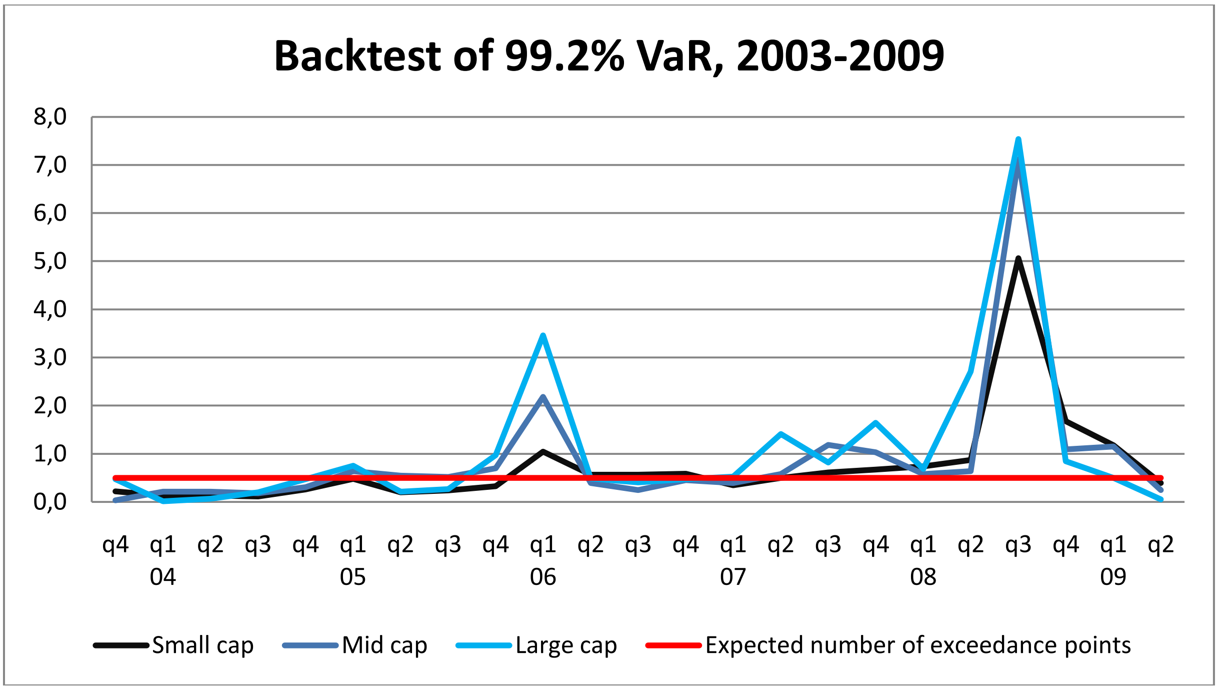

Figure 6.1: *Histogram of the average parameter breaks 99.2% VaR for the three stock segments; Small Cap, Mid Cap and Large Cap and the theoretical expected number of exceedance points given the 99.2% confidence level for a two year estimation period.*

As table 6.1 illustrates, there is two major increases in parameter breaks around quarter one 2006 and quarter three 2008 and some larger fluctuations during the whole year of 2007. The years with no parameter breaks, the parameter seems to be too high. We also notice that Large Cap generally has a higher level of fluctuations than the other segments.

6.1 Data quality

As mentioned in chapter 2, the underlying assumption for the HS method is that the past, as captured in the historical data set, is sufficiently like the future. As a measure of data validity to presume future outcomes, we regard the distribution of the historical trade frequency of each segment of stocks. The amount of absent trading days is thus of great importance. With a great lack of trading days, the final conclusions would be vague. Thus, the greater frequency of trades the more accurate assumptions we presume.

Data quality

Small Cap	Mid Cap	Large Cap
8,58%	9,03%	5,25%

Table 6.1: *The amounts of days in each stock segments that has a price change equal to zero, i.e. the amount of days that the stock wasn't traded*

The result shows that Large Cap has significantly lower amount of absent trading days. Small Cap and Mid Cap has similar results.

6.2 Parameter breaks

In this section the results from the empirical backtest are presented. Each of the segments is illustrated separately. Results from all estimation periods are captured in the same table, including total amount of parameter breaks, average and theoretical number of parameter breaks and the correlation factor.

The total amount of exceedance points is a sum of all exceedances during a six year period. We generalize the number of trading days to 62 per quarter, i.e.248 per year. Hence the total number of exceedance points is a summation over 1488 days, for all stocks in each segment. The average number of exceedance points is the total number of exceedance points divided by the current number of stocks. It represents the average number of exceedance points of the current estimation period, for one stock in that segment. The theoretical number of exceedance points is the expected number of exceedances given a six year period with VaR at confidence level 99.2%.

The correlation indicates how well the parameter is correlated between each quarter. Zero indicates no correlation.

Small Cap

Parameter	0,5	1	2	3
Total	978	837	793	745
Average	21,26	18,20	17,24	16,20
Theoretical	11,90	11,90	11,90	11,90
Correlation	-7%	0%	12%	15%

Table 6.2: *Empirical results for the segment Small Cap for all four estimation periods; 0.5, 1, 2 and 3 years during a six year period of backtest*

The Small Cap study engages 46 stocks. As illustrated in table 6.2, the three year parameter shows the lowest average of exceedance points. Though, all the four estimation periods exced the theoretical expected number of parameter breaks.

Besides, the three year parameter is the one with the highest dependency from historical data, i.e. has greatest correlation. In contrast, the one year parameter has no dependency and the six months parameter is negative correlated.

Mid Cap

Parameter	0,5	1	2	3
Total	655	627	681	664
Average	19,85	19,00	20,64	20,12
Theoretical	11,90	11,90	11,90	11,90
Correlation	-18%	1%	5%	6%

Table 6.3: *Empirical results for the segment Mid Cap, for all four estimation periods; 0.5, 1, 2 and 3 years during a six year period of backtest*

The Mid Cap study engages 33 stocks. As what can be distinguished from table 6.3, the one year parameter has both the lowest amounts of parameter breaks and correlation. Unlike the result from Small Cap, the two year parameter has the largest amount of parameter breaks. The six months parameter shows a strong negative correlation.

Large Cap

Parameter	0,5	1	2	3
Total	1620	1569	1853	1989
Average	22,50	21,79	25,74	27,63
Theoretical	11,90	11,90	11,90	11,90
Correlation	-2%	6%	11%	15%

Table 6.4: *Empirical results for the segment Large Cap, for all four estimation periods; 0.5, 1, 2 and 3 years during a six year period of backtest*

The Large Cap study incorporates 72 stocks. The one year parameter shows the closest average of parameter breaks to the theoretical number. Both the two and the three year parameter have a relative higher average number of exceedances and correlation factor, in comparison with the other two parameters.

6.3 Kupiec´s POF test

As discussed earlier, Kupiec's test is conducted as a log likelihood- ratio test. Under the null hypothesis that the model is correct, LR_{POF} has a chi-square distribution with one degree of freedom. The test is performed with 95% confidence level and the corresponding cut off value is 3.84 (see appendix 1). If the observed failure rate $\hat{p}$ differs significantly from the failure rate suggested by the confidence level p, the null hypothesis is rejected and the model is considered inaccurate.

The results from the three segments are separately represented in tables 6.5, 6.6, and 6.7. The tables contain the number of stocks in the segment that pass the test. An average value for the whole segment of the POF ratio is also computed, which is presented against the theoretical cut-off value from the Chi-square distribution.

Small Cap

Parameter	0,5	1	2	3
#Pass the test	24	32	33	30
LR-POF	8,15	4,94	4,30	4,59
Cut Off value	3,84	3,84	3,84	3,84

Table 6.5: *The results of Kupiec's test for each estimation period/parameter, for the Small Cap segment*

For Small Cap, the two year parameter shows best results in comparison to the other parameters, in terms of number of stocks that pass the test and average closest POF ratio. The one year parameter shows second best results.

Mid Cap

Parameter	0,5	1	2	3
#Pass the test	22	18	17	19
LR-POF	6,80	5,59	8,21	8,33
Cut Off value	3,84	3,84	3,84	3,84

Table 6.6: *The results of Kupiec's test for each estimation period/parameter, for the Mid Cap segment*

The six month parameter shows best results for Mid Cap in terms of number of stocks that pass the test. Nevertheless, the one year parameter has the lowest POF ratio on average, and hence closest to the theoretical value.

Large Cap

Parameter	0,5	1	2	3
#Pass the test	26	28	16	17
LR-POF	9,45	8,49	15,24	19,03
Cut Off value	3,84	3,84	3,84	3,84

Table 6.7: *The results of Kupiec's test for each estimation period/parameter, for the Large Cap segment*

For Large Cap, the one year parameter shows best results for both number of stocks that pass the test and closest average POF ratio to the theoretical value. Not far behind comes the six month parameter. The two and three year parameters show significantly worse results.

6.4 Basel test

The traffic light approach is built upon the same theoretical framework as Kuipec's POF- test. The binomial distribution is used to estimate if the number of exceptions is in line with the theoretical number. To pass the test, the total number of exceptions for every stock have to fall into the first of three categories, the "green". Table 6.8 shows the limits of exceptions for every stock at 99.2 % confidence level, and tables 6.9, 6.10 and 6.11 show the results for Small, Mid and Large Cap for the different estimation periods.

Limits for Basel test

#Exceptions	Outcome
<19	Green
19-25	Yellow
>25	Red

Table 6.8: *Limits of parameter breaks for the three different categories in the Basel test*

Small Cap

Parameter	0,5	1	2	3
Green	24	32	33	32
Yellow	9	8	7	8
Red	13	6	6	6

Table 6.9: *The results of the Basel test for all estimation periods of the stock segment Small Cap*

For Small Cap, the one, two and three year parameters show similar results, with around a quarter of the stocks equally distributed over the yellow and red category. The six month parameter stands out, with a significantly higher amount of stocks in the red category.

Mid Cap

Parameter	0,5	1	2	3
Green	22	18	17	19
Yellow	5	10	5	5
Red	6	5	11	9

Table 6.10: *The results of the Basel test for all estimation periods of the stock segment Mid Cap*

The results from Mid Cap are well dispersed, with the six month parameter of significantly highest amount of stocks in the green category.

Large Cap

Parameter	0,5	1	2	3
Green	26	28	16	17
Yellow	21	27	18	13
Red	25	17	38	42

Table 6.11: *The results of the Basel test for all estimation periods of the stock segment Large Cap*

For Large Cap, the one year parameter shows both the highest number of stocks in the green category and the lowest number of stocks in the red category.

6.5 Christoffersen´s interval forecast test

The objective of the interval forecast test is to test conditional coverage i.e. the probability that an exception today depends on the outcome the day before. The same log-likelihood framework as Kuipecs POF test is used, but extended to include a statistic for independence of exceptions (see chapter 5.2.2). The following three tables 6.12, 6.13, 6.14 illustrate the results for Small, Mid and Large Cap. The tables contain the number of stocks in the segment that pass the test. An average value for the whole segment of the IND ratio is also computed which is presented against the theoretical cut-off value from the Chi-square distribution.

Small Cap

Parameter	0,5	1	2	3
#Pass the test	28	30	31	29
LR-IND	6,61	6,21	5,65	6,61
Cut Off value	3,84	3,84	3,84	3,84

Table 6.12: *The results of Christophersen's test for each estimation period for the Small Cap segment*

For Small Cap, the two year parameter shows both the highest amount of stocks that pass the test as well as closest IND ratio to the theoretical value.

Mid Cap

Parameter	0,5	1	2	3
#Pass the test	26	15	15	17
LR-IND	4,33	6,19	7,72	8,37
Cut Off value	3,84	3,84	3,84	3,84

Table 6.13: *The results of Christophersen's test for each estimation period for the Mid Cap segment*

The six month parameter shows a significantly higher amount of stocks that pass the test compared to the other three parameters. Consequently, the average IND ratio is also relatively much lower than the others. The three year parameter has the second highest amount of stocks that pass the test, but contradictory the relative highest IND ratio.

Large Cap

Parameter	0,5	1	2	3
#Pass the test	42	36	30	28
LR-IND	5,5	5,83	8,29	9,66
Cut Off value	3,84	3,84	3,84	3,84

Table 6.14: *The results of Christophersen's test for each estimation period for the Large Cap segment*

For Large Cap, the six month parameter continues to show significantly highest amount of stocks that pass the test. Similarly, the parameter also has the average IND ratio closest to the theoretical value. The one year parameter has a clear second place in terms of stocks that pass the test and average IND ratio to the theoretical value.

7 Consolidated analysis

Due to the different segments and its different features, we summarize the results of each segment separately and conclude which estimation period that best makes the different tests, given the theory and our objective.

Given the results from chapter 6, the outcome for Small, Mid and Large Cap is presented in Table 7.1, 7.2 , 7.3, graded from 1-4 depending on how well the parameters have passed the test relatively to each other.

For breaks, number one represents the estimation period which resulted in the average number of exceedance points closest to the theoretical value and number four the furthest away.

For correlation, number one represents the estimation period which had a correlation closest to zero and number four the estimation period furthest from zero.

For the statistical tests, some considerations have been made due to both the total number of test passed for the current parameter and its average statistical test result, as for LR_{POF} and LR_{ind}, for the tests proven similar results. For the Basel test, final considerations have been made upon a balance of all categories incorporated to the test i.e. green, yellow and red category.

For the final conclusion, all test results have the same weight and meaning in our analysis. Thus, we sum up the total score of each estimation period to distinguish them and conclude which one that has proven best results in total.

7.1 Small cap

The three year estimation period performed 745 parameter breaks out of 1488 days and the average number of 16.2 parameter breaks, compared to the two year parameter with 17.2, is the closest to the theoretical expected number of exceedance points of 11.9. The one year estimation period resulted in zero correlation and therefore labeled with a number one.

From table 6.5, the results from Kupiec's test show that the two year estimation period has both the highest amounts of stocks that pass the test (33 against 32), and on average closest LR_{ind} result to the theoretical value (4.30 against 3.84). The outcome from the Basel test, for the one, two and three year parameter is hard to distinguish. Clearly the six month parameter has the worst results, with the highest amounts of stocks that ended up in the red and yellow category (see table 6.9). For Christophersen's test, the two year parameter shows best results. The six month and the three year parameter show similar results. Thus, we call it a tie.

Small Cap

Parameter	0,5	1	2	3
Breaks	4	3	2	1
Correlation	2	1	3	4
Kupiec	4	2	1	3
Basel	4	1	1	1
Christophersen	3	2	1	3
Conclusion	4	2	1	3

Table 7.1: *Overall results for Small Cap, labeled in a scale from 1-4, where one represents the estimation period resulting relatively best in the group*

Given the results for Small Cap in table 7.1, the overall outcome is hard to distinguish, due to the uncertainty of the mixed results and close fallout between the one and the two year parameter. Correlation for the one year parameter is relatively much better than for the two year parameter and for the Basel test there is a tie. What determine our conclusion are the results from Kupiec's and Christophersen's tests, which show that the two year parameter generally provides best results.

7.2 Mid Cap

From table 6.3, it's clear that the one year estimation period shows best results in both number of breaks and correlation, with an average number of parameter breaks of 19.0 against the six month parameter of 19.8, and correlation 1.2% against the two year parameter with 4.6%.

For Kupiec's test in table 6.6, the one year parameter shows the lowest LR_{POF}, but has at the same time a low number of stocks that pass the test. The six month parameter has the second lowest average LR_{POF} , but the highest number of stocks that pass the test. Therefore it is rewarded as the best estimation period for this test. The equivocation of these results is a feature of this particular segment, containing a great variety of different stock markets and thus skewed distribution of the results.

The Basel test also proves that the six month parameter is the best, with the highest number of stocks in the green category. As for Christophersen's test, the three year parameter shows the second best result in number of stocks that pass the test in table 6.13, but contradictive the highest average LR_{ind}. The one and two year parameters both come in third place, with the same amount of stocks that pass the test.

Mid Cap

Parameter	0,5	1	2	3
Breaks	2	1	4	3
Correlation	4	1	2	3
Kupiec	1	3	4	2

Basel	1	2	4	3
Christophersson	1	3	3	2
Conclusion	1	2	4	3

Table 7.2: *Overall results for Small cap, labeled in a scale from 1-4, where one represents the estimation period resulting relatively best in the group*

Results from the Mid Cap segment presented in table 7.2, are ambiguous and makes it difficult to draw conclusions. The results of the six month and the one year parameter are close. But still, the six month parameter shows best result in total and therefore is considered as number one overall.

7.3 Large Cap

The one year parameter shows the best results in terms of parameter breaks, with an average of 21.8 in table 6.4. The six month parameter has a correlation of -2.4% and is closest to zero.

Due to the statistical results, the one and two year estimation periods shows the best results. For Kupiec's and Basel test, the one year parameter shows best results while Christophersen's test holds the six month parameter as number one (see tables 6.7, 6.11 and 6.14).

Large Cap

Parameter	0,5	1	2	3
Breaks	2	1	3	4
Correlation	1	2	3	4
Kupiec	2	1	4	3
Basel	2	1	3	3
Christophersen	1	2	3	4
Conclusion	2	1	3	4

Table 7.3: *Overall results for Small cap, labeled in a scale from 1-4, there one represents the estimation period resulting relatively best in the group*

As stated in table 7.3, the one and two year parameters show the best results throughout the whole investigation for Large Cap. After summarizing the results, the conclusion is that the one year parameter shows best results in total and the six month parameter the second best. The three year parameter shows unambiguous results for being the third best, which omits the two year parameter to the fourth place.

8 Conclusion

In the empirical part of our investigation, we summarized the number of parameter breaks that occurred over a six years period. If our investigation had ended here, we could just conclude that the parameter with the number of exceptions closest to the theoretical expected number is the most accurate one. This however is not the case. An accurate parameter, as discussed in previous chapters, involves two dimensions;

- Unconditional coverage: In a good VaR model the number of exceptions should be in line with the selected confidence level (Jorion, 2001)
- Conditional coverage: A good VaR model should not only produce the correct number of exceptions but also produce exceptions that are uniformly spread in time, i.e. independent and identically distributed. (Jorion, 2001)

In other words, to give a definitive answer, consideration must be taken to the both above dimensions.

To make a conclusion generalized for the whole stock market, we summarized the results from all three stock segments; Small Cap, Mid Cap and Large Cap. By doing so, we also incorporate the data quality and the number of stocks engaged in each segment for this study.

The quality of the data, as discussed earlier is a problem. The quality of data for stocks in Small and Mid Cap has after investigation poorer quality than stocks in Large Cap, i.e. stocks are not as frequently traded as the stocks in Large Cap. Their history is generally also shorter than stocks in Large Cap. To qualify in our study, all stocks must have a minimum of nine years of history. This criterion disqualifies a large number of stocks, especially in Small Cap but also in Mid Cap and has consequently left us with almost as many stocks of Large Cap as Small and Mid Cap together. The both features of different data qualities and different number of stocks in the segments makes it necessary to consider particularly the outcome from Large Cap to make generalize conclusions of the whole stock market. The table below shows the relative overall performance of the three stock segments for all tests, in addition with its amount of stocks and lack of trading days.

Summary

Parameter	0,5	1	2	3	Stocks	No trades
Small Cap	4	2	1	3	46	8,58%
Mid Cap	1	2	4	3	33	9,03%
Large Cap	2	1	3	4	72	5,25%

Table 8.1: *A summary of the total result of the different tests, labeled by the segments*

To summarize the overall results, there is no definite "winner" of best estimation period, i.e. no parameter shows best results in every segment.

For the segments with the fewest stocks and lowest accuracy due to data quality, i.e. for Small Cap and Mid Cap, the six month and two year parameter has each a 1 and a 4, i.e. an average of 2.5. The one year parameter has two 2s, and the three year parameter two 3s. Thus, the best parameter for these segments would be the one year parameter, second best the six months and the two year parameter and the three year parameter as the least accurate.

Summary

Parameter	0,5	1	2	3	Stocks	No trades
Small Cap	4	2	1	3	46	8,58%
Mid Cap	1	2	4	3	33	9,03%
Summation	2	1	2	3	77	
Large Cap	2	1	3	4	72	5,25%

Table 8.2: *A summary of the total result of the different tests, labeled by the segments*

The results from Large Cap indicate similar assumptions as for the other segments. The best performed parameter is the one year estimation period, second best the six month, third best the two year parameter, and the three year parameter as worst. Due to the numerical results we can therefore conclude that the parameter that overall makes the tests best i.e. take cover for both dimensions of unconditional and conditional coverage and that also performs best in Large Cap is the one year parameter. Second best would be the six month, and third best the two year parameter.

9 Discussion and suggestion to further studies

The overall results have shown that the one year parameter gives the best results in total. What should be noticed is that there is a diffuse pattern over the results between the different stock segments. No segment shows exactly the same results one to another and should in some sense thus be regarded apart from each other. If there would be a possibility and resources to expand the parameter framework we therefore recommend not to generalize one optimal estimation period but instead treat the segments separately.

As what have been mentioned before, the data quality is a general issue when dealing with historical data. In our study it has limited us to both the number of stocks and to the length of historical data, which has unfortunately resulted in uneven numbers of stocks in the different segments. As a consequence the comparison between the segments is doubtful. A solution to this problem could be to choose equal numbers of stocks in each segment. However, this would had left us with less number of stocks and hence make general conclusion less valid. Since our objective was to make generalized conclusions of the whole stock market we chose not to involve fewer stocks in our study.

One strong assumption when estimating VaR with the historical simulation method is that the daily price changes is independent and equally distributed. This assumption should however be carefully handled and can be seen as an explanation to why the model is not completely accurate. In reality, the volatility in price changes is strongly dependent of the financial environment and hence varies from different time periods. As what have been observed from the previous financial crises, the declines in asset values have not just been very large but also appeared close to each other. This fact makes it obvious that price changes is not independent and neither identical distributed and thus by the theory not makeable that the risk framework is not accurate. Therefore it makes it unrealistic to use this risk methodology to take cover for future risks in an unstable financial environment. Another doubtful feature with the historical simulation method in financial crises is that the data used in estimating the VaR parameter can never show larger losses than previous historical events and thus unable to predict eventual losses larger than earlier appeared.

Due to the statistical investigation of our study, the different backtesting approaches of VaR is hampered by different kinds of shortcomings. Kuipecs POF test gives a good overview if the number of exceptions is in line with the theoretical expected number of exceptions. On the other hand, the test lacks the ability to examine when in time the exceptions occur. It also does not capture the magnitude of the exceptions. It is therefore of great importance that this test is followed up by other tests that examines when in time the exceptions occur i.e. conditional coverage. The Basel Committee’s test is built upon the same statistical framework as the POF test and therefore contains the same kinds of shortcomings.

These two tests give a good first overview of the number of exceptions that are line with the theoretical number, but lack the ability to alone draw conclusions of the models accuracy. As a complement we used Christoffersens interval forecast test that takes conditional coverage into account. The test examines the likelihood that two exceptions follow each other. The choice of confidence level also incurs different kind of problems and should be selected with regard to backtesting. For small values of p such as 0.01, it becomes increasingly difficult to confirm deviations and to conclude if the model overestimates risk. A lower VaR confidence level is therefore from a backtesting perspective more appropriate in order to observe sufficient number of deviations to validate the model.

Our own reflection and critique about the Historical Simulation method is mainly due to the fact, that the method does not take into account that historical data used is less accurate the further back in time you go, due to the daily financial environment. Our belief is that the closer the historical period is to today, the more accurate the estimation is due to future predictions. But as mentioned, to approximate an accurate estimation large amount of data is necessary. Therefore our suggestion to further studies is to construct a weighted historical simulation method to estimate VaR. One idea is to give more weight to recent data and at the same time a lower weight to older data.

References

1.Beyond Value at risk, Kevin Dowd, 1998, John Wiley & Sons Inc

2.Measuring Market Risk With Value At Risk, Pietro Penza, Vipul K. Bansal, 2001, John Wiley and Sons Inc

3.Quantitative risk management, J. McNeil, R.Frey, P.Embrechts, 2005, Princeton University Press

4.Mathematical Modeling and Statistical Methods for Risk Management, Henrik Hult and Filip Lindskog, 2007, KTH

5.Value at Risk, The New Benchmark for Managing Financial Risk, Jorion, 2001, 2nd Edition, McGraw-Hill

6.Basle Committee of Banking Supervision, 2006, International Convergence of Capital Measurement and Capital Standards – A Revised Framework, Comprehensive Version, www.bis.org, October 2009,

7.Basle Committee of Banking Supervision, 1996, Supervisory Framework For The Use of "Backtesting" in Conjunction With The Internal Models Approach to Market Risk Capital Requirements, www.bis.org, October 2009

8.NASDAQ OMX, http://Nordic.nasdaqomxtrader.com, October 2009

APPENDIX 1 - Critical values for the chi square distribution

	P Value												
f	**0.995**	**0.99**	**0.975**	**0.95**	**0.9**	**0.75**	**0.5**	**0.25**	**0.1**	**0.05**	**0.025**	**0.01**	**0.05**
1	0.00	0.00	0.00	0.00	0.02	0.1	0.45	1.32	2.71	3.84	5.02	6.63	7.88
2	0.01	0.02	0.05	0.1	0.21	0.58	1.39	2.77	4.61	5.99	7.38	9.21	10.60
3	0.07	0.11	0.22	0.35	0.58	1.21	2.37	4.11	6.25	7.81	9.35	11.34	12.84
4	0.21	0.30	0.48	0.71	1.06	1.92	3.36	5.39	7.78	9.49	11.14	13.28	14.86
5	0.41	0.55	0.83	1.15	1.61	2.67	4.35	6.63	9.24	11.07	12.83	15.09	16.75
6	0.68	0.87	1.24	1.64	2.20	3.45	5.35	7.84	10.64	12.59	14.45	16.81	18.55
7	0.99	1.24	1.69	2.17	2.83	4.25	6.35	9.04	12.02	14.07	16.01	18.48	20.28
8	1.34	1.65	2.18	2.73	3.49	5.07	7.34	10.22	13.36	15.51	17.53	20.09	21.95
9	1.73	2.09	2.70	3.33	4.17	5.90	8.34	11.39	14.68	16.92	19.02	21.67	23.59
10	2.16	2.56	3.25	3.94	4.87	6.74	9.34	12.55	15.99	18.31	20.48	23.21	25.19