LUND UNIVERSITY
School of Economics and Management

Master Thesis, Spring 2009
School of Economics and Management
Lund University

Star VaRs

Finding the optimal Value-at-Risk approach for the banking industry

Authors
Olle Billinger
Björn Eriksson

Supervisors
Birger Nilsson
Anders Vilhelmsson

Abstract

This paper explores the concept of Value-at-Risk (VaR) through a comparative study of non-parametric and parametric models in order to find the best risk model for banks' trading portfolios. The non-parametric methods consist of three different approaches: Simple Historical Simulation, Age Weighted Historical Simulation and Volatility Weighted Historical Simulation by means of the EWMA and GARCH models for forecasting volatility. The parametric methods comprise six different approaches: VaR based on the normal distribution, VaR with Student's t-distribution, RiskMetrics, VaR with implied volatility and VaR with GARCH volatility dynamics (both assuming normality and t-distribution). The models are estimated and tested on the S&P500 and a hypothetical bank trading portfolio. The evaluation of the models follows the Christoffersen framework of testing for correct conditional coverage together with assessment of model performance according to the regulatory requirements of the Basel Accord. The general finding is that models with leptokurtic features and time-varying volatility perform the best, while naïve models assuming normality and/or without volatility dynamics in general display poor performance. The GARCH(1,1)-t model by far outperforms its competitors as it can correctly account for both correct unconditional coverage and volatility clustering. The implications of market risk regulation are explored and it is argued that the current regulatory environment might provide incentives for low-quality risk management practices with significant room for regulatory improvements.

Keywords: Value-at-Risk, Historical Simulation, Normal Distribution, t-distribution, RiskMetrics, GARCH estimation, Volatility forecasting, Implied volatility, Backtesting, Christoffersen, Basel, Financial regulation.

“Never tell me the odds”

Han Solo, Star Wars V: Empire Strikes Back

Table of Contents

List of Tables and Figures

1. Introduction

When examining the concept of risk in the financial literature several types of risk can be distinguished: business risk, strategic risk and financial risk. Business risk and strategic risk pertains primarily to corporations and their risk exposure towards certain product markets or certain economic and political environments. Financial risk in turn, can be divided into several sub-categories: *market risk* due to changes in market prices, *credit risk* resulting from inability on behalf of the counterpart to fulfill his/her obligations, *liquidity risk* caused by insufficient market activity, *operational risk* arising from staff or systems failures and *legal risk* caused by the inability of a counterparty to engage in a certain transaction. This thesis will focus only on the area of market risk, and more specifically, how this type of risk can be captured through the risk measure Value-at-Risk (VaR). VaR can be loosely defined as the loss we can expect given a certain probability over a certain time horizon (van den Goorbergh and Vlaar 1999) and was developed as a tool of understanding and managing market risk against the background of several noteworthy trading activities and their respective losses for institutions like Barings Bank, Orange County, Metallgesellschaft and Long Term Capital Management (Kuester et al. 2006). Our sole focus on VaR can be motivated by its introduction into the regulatory environment through Basel I and Basel II where it is used as a means of determining the capital requirements for banks' trading portfolios. VaR has received much criticism over the years since being introduced and refinements of the measure such as Expected Shortfall (ES) have been suggested. But through its expediency – easy to understand and implement – VaR is here to stay and the measure has become a mainstay of academic research as well as practical risk management (Bao et al. 2006). The latter point is the primary focus of this thesis, and is examined through the construction of a portfolio resembling, as far as possible, a bank's portfolio. By estimating and testing several models on this portfolio a better evaluation of the practical implications of different VaR models and methods might be obtained. Of course, further examination of other portfolios is performed and backtested one-step ahead and out-of-sample by means of the Christoffersen (1998) methodology in order to find the model that can capture the risk the best. The thesis will also offer an outlook on the regulatory requirements in the Basel Accord and how the methods used to estimate VaR perform according to the practical implications of setting regulatory capital.

1.1 Purpose

The purpose of this paper is thus to add to the current line of VaR research by finding the method that can capture the market risk inherent in banks' portfolios the best, by comparing the performance of different methods against each other. Especially the explicit and atypical focus of a trading portfolio might add interesting results. A total of ten models are tested; four non-parametric and six parametric. The non-parametric methods used are the Historical Simulation and versions of the Historical Simulation; Age Weighted and Volatility Weighted (by means of the GARCH and EWMA). The parametric methods used are VaR with normal distribution, VaR based on the t-distribution, conditional VaR based on GARCH, the RiskMetrics approach and VaR with implied volatility. With backtesting based on the Christoffersen (1998) framework, we expect to find the best model or models for banks' risk measurement and risk management.

1.2 Delimitations

A plethora of different estimation methods for VaR exist, using different estimations procedures and different distributional assumptions. For an example, Hansen and Lunde (2004) work with 330 different GARCH-models in forecasting volatility – which all potentially could be incorporated into VaR measurement. Other models include mixture models which use several distributions, switching models or models based on extreme value theory, which explicitly model the tails of the distribution. However, due to the complexity of certain models and the quantitative knowledge needed to really grasp them, and due to time limitations we have decided to limit ourselves to the models above. They should provide an extensive examination enough for our purpose of the thesis and provide us with valuable insights into VaR estimation and forecasting.

1.2 Structure of the paper

The structure of the paper is as follows. The first section digs deeper into the VaR theory and framework in order to provide the reader with a solid foundation for understanding later concepts. The next section deals with the methods mentioned above and explain the different approaches in detail. Section 4 examines the data used and the outlines the construction of the hypothetical banking portfolio. After that, results are presented and backtested in section 5. An outlook on the regulatory implications of our findings is presented in section 6. Finally, we discuss our findings which are presented in the conclusion followed by suggestions on future research needed on the subject of VaR.

2. Theoretical background

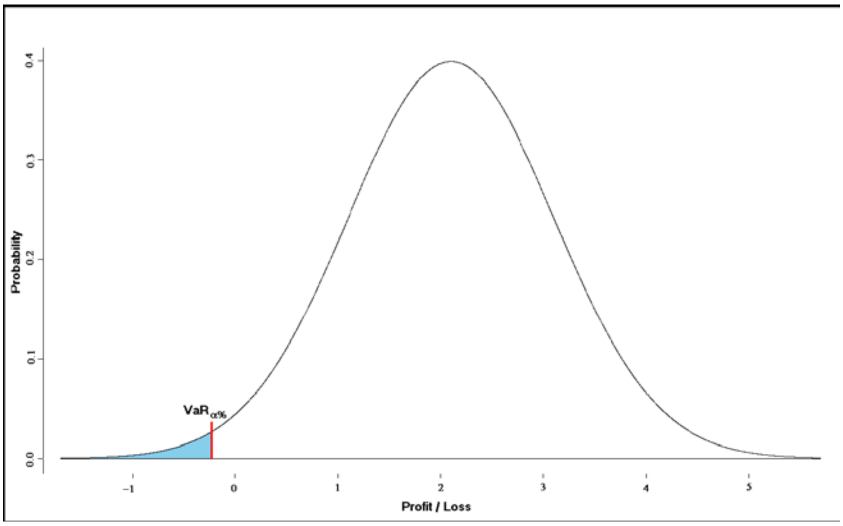
In this thesis we have decided to use Value-at-Risk (VaR) to estimate banks' risk by trying to mimic a bank's portfolio later used to determine capital requirements. The use of VaR is now widespread but was first used by the American bank JP Morgan in the early 1990s. According to the legend, the late CEO and chairman of JP Morgan demanded at the end of each trading day a report on the market risk the firm was facing during the next day. To meet this demand the staff of JP Morgan created a system to get a complete overview of the risk the firm was facing. This system was called VaR, or in other words, how much the maximum loss, given a certain probability, over the next trading day was. The firm was keen on developing VaR into the most used risk measure; hence in 1994 JP Morgan released RiskMetrics to the public free of charge. RiskMetrics was basically a simplified version of the firm's own risk model. With RiskMetrics available to anyone on the Internet, the popularity of VaR grew rapidly into the most used measure for financial risk. But what is VaR? VaR is defined as:

$$\Pr(R \leq VaR) = \alpha$$

or:

VaR summarizes the worst loss over a target horizon with a given level of confidence.

Figure 1.1 Graphical illustration of VaR. VaR is defined as a negative value in the left tail for a given level of confidence for a certain probability density function or histogram (e.g. 5% or 1%).



Hence VaR gives you a simple monetary explanation of the downside risk a firm is facing and is a natural development, or progression as one might say, from the traditional portfolio theory which was traditionally used when estimating financial risk (Dowd, 2005 p 9-11). This progression can be attributed to several arguments:

- When portfolio theory measures risk in standard deviation VaR gives an actual number of the risk the firm or agent is facing. This makes VaR more easily accessible for laypeople.
- VaR was created to deal with a large variety of distributions making it more flexible than portfolio theory which is restricted to the normal distribution which is uncommon in financial data.
- VaR can as below stated be used on a wide variety of risks such as credit risk, liquidity risk and other while portfolio theory only is able to determine market risk.

VaR has also other important features which make it superior to its predecessors (Dowd, 2005 p 12):

- VaR aggregates the risk over a number of positions thus enabling us to take an aggregated view of the risk we are facing.
- VaR is holistic which means that it takes several risk factors into account; on the contrary, its predecessors only looked at one risk factor at a time or resorted to simplifications in order to reduce multiple risk factors and analyze them one by one.
- VaR gives a probability of the potential loss to the risk manager while traditional measures leave out the probabilistic view and resort to answering the "what if" questions.

From its creation in the late 1980s VaR quickly established itself as the dominant risk measure, used not only by investment banks but also by commercial banks, pension funds and other financial institutions (Dowd, 2005 p. 10-11). But due to its widespread popularity and usage one, when using VaR, must be cautions of its drawbacks.

- VaR was motivated due to financial data not being normally distributed, however VaR works best under the assumption of normal distribution
- VaR is not coherent (it cannot account for the diversification effect in the case of non-normality)
- VaR does not give any information about the magnitude of the tail loss; losses can be far greater than what VaR anticipates. This can be contrasted with the risk measure Expected Shortfall (ES) which calculates the magnitude of the expected tail loss.

3. Methodology

This section deals with and explains the methodologies later used on the data to estimate VaR. Before examining these methods it is important to grasp the basic features of financial data since good models should take these into account in order to be successful.

3.1 Characteristics of financial data

In choosing a method to estimate VaR it is essential to acknowledge the stylized facts that financial data exhibits. Being plagued by leptokurtic characteristics, the usual workhorse of statistics – the normal distribution – is usually insufficient in order to capture the non-normal features of financial time series, despite the Central Limit Theorem. Benoit Mandelbrot, the mathematician who invented fractal theory, highlighted the shortcomings of the normal distribution when dealing with financial data. Mandelbrot assumed the stock market returns (Dow Jones Industrial Average 1916-2003) to follow a normal distribution, accordingly calculated the number of times the stock market moved by a certain magnitude and compared it to the actual movements. Thus, for an example, the Dow should have moved by more 3,4% on 58 days but in fact did so 1001 times. On six days it should have moved by more than 4,5%, but in reality did so on 366 occasions. Finally, again given the normal distribution, the index should have moved by more than 7% once every 300 000 years but in the 20th century did so on 48 occasions (The Economist January 24th 2009). Clearly the normal distribution cannot capture the non-normal properties of financial data. More specifically, in general there are three stylized facts relating to financial time series which need to be taken into account: (i) volatility clustering indicated by autocorrelation in absolute and squared returns (ii) excess kurtosis with the density return distribution more peaked around the center and fatter tails than the normal distribution (iii) skewness, possibly of time-varying nature (Kuester et al. 2006). Thus, an optimal strategy would be one where these stylized facts can be taken into account, since this would give a more accurate estimate of the actual behavior of financial data. In doing this one can separate methods as being either parametric or non-parametric. Parametric approaches explicitly model the distribution whereas the non-parametric methods work directly with the empirical distribution.

3.1.1 Modeling volatility

The important feature of financial data of time-varying volatility merits extra attention. As volatility measures the dispersion of some variable around its mean, it is a convenient measure in order to capture risk – ultimately the most significant input into models of risk measurement and risk management (Dowd 2005, p. 129). By incorporating volatility into the analysis, the two stylized facts of volatility clustering and leptokurtosis can be taken into account (see above). This stems from the basic observation that large (small) return observations tend to be followed by more large (small) returns (Mandelbrot 1963). Econometrically speaking, an obvious serial correlation can be observed for financial return series. Specifically, assume η_t is an innovation governing an asset return process. This innovation or error term is assumed to be distributed with zero mean and variance conditional on some previous information Ω_{t-1}:

$$\eta_t | \Omega_{t-1} \sim N(0, \sigma_t^2) \tag{3.1}$$

The unconditional variance of the innovation is simply the unconditional expectation of σ_t^2:

$$\sigma^2 \equiv E[\eta_t^2] = E[E_t[\eta_t^2]] = E[\sigma_t^2] \tag{3.2}$$

Thus, the variability of σ_t^2 around its mean does not affect the unconditional variance. However, it can be shown that the variability of σ_t^2 does change the higher moments of the unconditional distribution of η_t resulting in 'fat tails' and leptokurtosis (Campbell et al. 1997, p. 480). Thus by accounting for volatility by means of the η_t we can better capture the non-normal features of financial data. Engle (1982) sought to incorporate these characteristics in the Autoregressive Conditional Heteroscedasticity, ARCH(q) process where q represents the number of past squared innovations (lags):

$$\begin{gathered} r_t = \mu + \eta_t \\ \sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i \eta_{t-i}^2 \\ \eta_t \sim N(0, \sigma_t^2) \end{gathered} \tag{3.3}$$

In order to avoid estimation of higher order ARCH models with several lags it was soon shown by Bollerslev (1986) that the essence of volatility easily could be captured by means of the General Autoregressive Conditional Heteroscedasticity, GARCH(p,q) where p depicts the number of which is defined as:

$$r_t = \mu + \eta_t$$

$$\sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i \eta_{t-i}^2 + \sum_{j=1}^{p} \beta_i \sigma_{t-j}^2 \tag{3.4}$$

$$\eta_t \sim N(0, \sigma_t^2)$$

r_t depicts the mean equation with constant and residual. σ_t^2 represents the conditional variance and depends on previous innovations and itself through the autoregressive term. The above GARCH(p,q) model assumes normality but can also be implemented with other distributions such as the t.-distribution or Generalized Error Distribution (GED). The workhorse of this paper will be the simple GARCH(1,1) with one autoregressive term and one lag with past squared innovations. The estimation of the GARCH-parameters requires the use of the Maximum Likelihood function as it is applicable for the non-linearity of the model. More on the issues of Maximum Likelihood estimation and error distribution is supplied below in the (section 3.3.4).

A simpler model used in modelling time-varying volatility, yet belonging to the GARCH family (Angelidis et al 2003), is the Exponentially Weighted Moving Average (EWMA) used by RiskMetrics. It defined as:

$$\sigma_t^2 = \lambda \sigma_{t-1}^2 + (1-\lambda)\eta_{t-1}^2 \tag{3.5}$$

The EWMA methodology requires the estimation of the λ parameter. More on this issue will be discussed in section 3.3.3 below[1].

[1] In this thesis the term 'EWMA' will mainly refer to its incorporation in the non-parametric VaR method of Volatility Weighted Historical Simulation whereas the term 'RiskMetrics' will refer mainly to the parametric VaR method assuming normality. Both methods however, incorporate a λ of 0.94. Motivation for this is supplied in section 3.3.3.

3.2 Non-parametric approaches

Non-parametric models use the empirical distribution in order to model financial data. In doing this, one must possess a historical return series or historical profit/loss (P/L) series, on which the Value-at-Risk is estimated. The important feature of the non-parametric approaches is that non-normal features such as kurtosis and skewness already are inherent in the data and that, accordingly, no specific distributional assumptions modeling them are needed. Implicitly then, one also assumes the past to be a good estimate of the future when forecasting with historical simulation (Bao et al. 2006). This thesis uses three methods of dealing with historical return series: the plain or simple historical simulation (HS), the age weighted historical simulation (AWHS) and the volatility weighted historical simulation (VWHS) by means of the GARCH(1,1) and EWMA (two variants of volatility weighting). Furthermore, it has been reported that the vast majority of banks actually use the HS or some variant of it in their risk management (Perignon and Smith 2008).

3.2.1 Plain Historical Simulation (HS)

The plain or simple historical simulation simply utilizes the histogram to plot the data whereupon VaR is read off the histogram. Given a certain confidence level α (e.g. 1% or 5%) the VaR is simply estimated as (van den Goorbergh and Vlaar 1999):

$$V\hat{a}R_{t+1|t} = q(\alpha) \tag{3.5}$$

Where $q(\alpha)$ represent the quantile corresponding to the chosen confidence level α. For an example, with a sample of $n=1000$ and a confidence level of $\alpha=0,05$ the VaR is then represented by the 51^{st} largest loss. In the case where there is no single observation corresponding to the confidence level, interpolation is necessary in order to retrieve the VaR estimate. For an example, working with a window size of 250 and α of 1 percent, the VaR is taken to be the average of the second largest portfolio loss and the third largest portfolio loss. The HS use the same weight for all observations making it very sensitive to the length of the estimation window, since the VaR estimate can change swiftly as large or small observations might appear or drop out of the window (Dowd 2005, p. 94). The problem of the simple historical simulation is thus one of choosing the right window length: a small window size would give an up-to-date estimate with the cost of a highly varying VaR, whereas a long window would yield a more accurate description of the historical VaR, with the cost of this estimate having low relevance for forecasting (van den Goorbergh and Vlaar 1999). At the same time, the HS is conceptually simple, easy to implement, very widely used and has a fairly good historical track record (Dowd 2005, p. 83). That makes the method an excellent point of departure for further refined methods building on the foundations of the HS. The following methods circumvent the shortcomings of the simple HS by assigning different weights to different observations.

3.2.2 Age Weighted Historical Simulation (AWHS)

The age weighted historical simulation (AWHS) is based on the assumption that the weight of each observation is not equally distributed and it is to be expected that this could yield a more accurate estimate of the VaR than the plain HS (Boudoukh et al. 1998). More specifically, recent observations are assumed to contribute more than older ones to the VaR estimate. If *w(1)* is the weight of the most recent observation then $\lambda w(1) = w(2)$ is the weight given to the second most recent observation, *w(3)* should consequently be $\lambda^2 w(1)$ and so forth. λ is the decay factor and is given a value between 0 and 1 and is interpreted as the decaying importance of the observations. A λ value close to 1 indicates a slow decay of the importance of the observations while a λ value close to zero indicates a rapid decay. The weight of return observation *i* is given by:

$$w(i) = \frac{\lambda^{i-1}(1-\lambda)}{1-\lambda^n} \quad for\, i = 1,..., N \tag{3.6}$$

This replaces the uniform $1/n$ weights of the HS (Boudoukh et al. 1998). Thus, the simple HS can be seen as a special case of AWHS where $\lambda=1$. The AWHS method is superior to the HS method in four ways. (i) The AWHS method is more flexible than the HS method since it values recent observations higher than HS which values the first observation as much as the 1000th observation. (ii) The AWHS is more responsive to large losses and especially to clustered large losses. This will have a more direct impact on the next-day VaR which of course will be higher than predicted by HS. (iii) Because of the λ feature in the AWHS method large distortions not recently occurred will have a moderate impact on next-day VaR. When this distortion is no longer accounted for its effect will fall from $\lambda^n w(1)$ to zero instead of $1/n$ to zero which would be the case if the HS method was used. This reduces the so called ghost-effects because the shock it is causing on VaR will be lower. (iv) By allowing our sample to grow over time, thus letting the impact of extreme events decline over time, we can eradicate the ghost-effects which would otherwise result in jumps in our sample returns. This approach cannot be done when using HS because all old observations would weight the same no matter the sample length (Dowd, 2005 p. 93-94).

3.2.3 Volatility Weighted Historical Simulation (VWHS)

Another approach is to weight our returns by volatility, the intuition being that volatility should affect the VaR estimates. The basic idea behind the volatility weighted historical simulation is to take the recent volatility into account when estimating the return. In doing this we adjust the whole sample according to how the volatility tomorrow is expected to behave compared to how it historically has performed. For an example, if the current volatility is higher than that of the last month, simply using historical data of the last month would understate the true volatility. On the other hand, if the current volatility is lower than that of the last month, using historical volatility would overestimate the true volatility (Dowd 2005 p. 94). Thus, by taking the current volatility into account and adjusting the whole sample accordingly, a more accurate expectation of the VaR could be produced. Specifically, the weights assigned to each observation are defined as (Hull and White 1998):

$$r_{t,i}^{*} = \left(\frac{\sigma_{T,i}}{\sigma_{t,i}} \right) r_{t,i} \qquad (3.7)$$

In equation (3.7), $r_{t,i}$ is the historical return on asset i at time t, $\sigma_{t,i}$ is the GARCH or EWMA-generated historical volatility estimate of the volatility of asset i at time t, $\sigma_{T,i}$ is the last observation (or forecast) of the volatility of asset i at time t. This will yield our volatility adjusted return (Dowd 2005 p. 94-95). Equation (3.7) consequently replaces $r_{t,i}$ when calculating returns. Considering the abundance of models that estimate and forecast volatility (see above), several methods could be incorporated into the VWHS. This thesis will only, as stated, deal with volatility estimates of the EWMA and the GARCH(1,1) models following the lines of Hull and White (1998). One of the true merits of the VWHS is that by taking recent volatility into account (in periods of high volatility, historical returns are scaled upwards), one can obtain VaR figures that exceed the simple historical estimates.

3.3 Parametric approaches

Parametric approaches, as opposed to non-parametric approaches, make explicit distributional assumptions and try to model the financial time series accordingly. The VaR is estimated by using the characteristics of the fitted distribution. In doing so, it is essential that the distribution chosen can accommodate the stylized facts of volatility clustering, excess kurtosis, and possibly skewness as the distribution otherwise would not be able to capture the true behavior of the financial data. As already stated, the normal distribution cannot account for these stylized facts, but due to its intuitive appeal and simplistic features it still serves the purpose of a good benchmark which the other more refined models can be judged upon. The additional VaR models considered in this section are based on the t-distribution and conditional volatility methods. These should, theoretically, be able to accommodate the features of financial time series. Parametric methods, reportedly, are used by banks in their risk management, even though to a lesser extent than variants of the HS (Perignon and Smith 2008).

3.3.1 Normal distribution

A basic VaR model can be based on the normal distribution which requires only the mean and standard deviation in order to model the distribution. The model is defined below:

$$VaR = -\mu_{P/L} + \sigma_{P/L} q(\alpha) \tag{3.8}$$

In equation (3.8) $\mu_{P/L}$ is the mean, $\sigma_{P/L}$ is the standard deviation and $q(\alpha)$ refers to the quantile of our chosen confidence level. Because of its simplicity, it requires only two independent parameters, mean μ and standard deviation σ (Dowd, 2005, p. 154). The normal distribution is also attractive because of its additivity feature. Additivity means that the sum of variables which are normal distributed are also normal distributed. This is important when calculating multi-day VaR from one-day VaR which is required by the Basel Accord. This leads us to discuss the square root of time rule:

$$VaR^{(T)} = \sqrt{T}\, VaR^{(1)} \tag{3.9}$$

When assuming independence of normally distributed returns equation (3.8) shows that T-periods-ahead risk for a portfolio can be calculated on the basis of one-period-ahead VaR. Due to this simplistic nature this is commonly used in other VaR models although in some occasions it is not appropriate because other models do not use the normal distribution, and in these cases the 'square root of time rule' works only approximately (van den Goorberg and Vlaar 1999). However when applied on financial data the normal distribution inhibits some serious drawbacks. Financial data seldom follows the normal distribution because of its fatter tails and excess kurtosis; this calls for the use other distributions alongside the normal distribution which will be further discussed below.

3.3.2 Student's t-distribution

The normal distribution's inability to correctly describe financial data shows that the use of distributions with fatter tails and excess kurtosis are needed. Student's t-distribution can accommodate this need for a leptokurtic distribution and is defined below:

$$VaR(\alpha) = -\mu_{P/L} + \sqrt{\frac{\nu - 2}{\nu}}\,\sigma_{P/L} t(\alpha, \nu) \tag{3.10}$$

Equation (3.10) defines our VaR for a t-distribution and is somewhat different from equation (3.8). Instead of referring to the normal distribution, the confidence level term, $t_{\alpha,\nu}$, now refers to the t-distribution and is dependent not only on α but also on the degrees of freedom v.

Three parameters characterizes the generalized t-distribution: μ (the mean, or location parameter), $\sigma_{P/L}$ (the standard deviation), v (degrees of freedom). The degrees of freedom determine the fatness of the tails and kurtosis. If we want relatively high kurtosis and relatively fat tails we should choose a low ν, if we want relatively low kurtosis and relatively thin tails we should choose a high value for ν; finally, if $\nu \to \infty$, then the t-distribution equals the Normal distribution (van den Goorberg and Vlaar 1999).

As noted above, the t-distribution can capture the nature of financial data better than the Normal distribution, however it has several drawbacks. As the Normal distribution, it can produce high risk estimates due to its inability to constrain maximum possible losses. We should also be wary of using the t-distribution on extremely high confidence levels because it is not consistent with extreme value theory. Also, the t-distribution is unlike the normal distribution not additive i.e. the sum of t-distributed random variables may not be distributed as a t-variable (Dowd, 2005, p. 159-160).

3.3.3 RiskMetrics

The exponentially weighted moving average (EWMA) used by RiskMetrics is simply an extension of the historical average volatility and is defined as:

$$\sigma_t^2 = \lambda\sigma_{t-1}^2 + (1-\lambda)\eta_{t-1}^2 \tag{3.11}$$

Where σ_t^2 is the volatility for day *t* calculated at day *t-1*. λ is the decay factor showing the importance of the volatility at day *t-1* relative to η_{t-1}^2 which represents the daily percentage change at *t-1*. Assume a large daily percentage change at *t-1* meaning that η_{t-1}^2 will be very large. If λ is low, η_{t-1}^2 will affect day t volatility more than if λ was high. Thus λ decides how responsive the volatility of day t should be to the percentage in market movements rather than being responsive to volatility at day t-1. λ is in this thesis assumed to take on a value of 0.94 in line with the findings of the founders of RiskMetrics since this value gives forecasts of the variance rate that come closest to the realized variance (Hull 2006 p. 463). EWMA applies more weight to recent observations and less weight to older observations, thus giving the latest observation the largest impact on the forecast. The EWMA is in two ways superior to its predecessors for two reasons: (i) its ability to apply more weight to recent observations than old observations (ii) the impact of old observations decays at the exponential rate, λ. The historical average volatility could suffer from shocks when old extreme observations are no longer included in the sample. If these shocks are not excluded from the sample they will plague the forecast even in times of tranquility in the market. Even though EWMA outperforms its predecessor it does not (like GARCH) take mean reversion into account. In the following discussions the general EWMA will be denoted RiskMetrics as we build on their estimation of the λ parameter.

3.3.4 GARCH(1,1) and GARCH(1,1)-t

The Generalized Autoregressive Conditional Heteroscedasticity (GARCH) model introduced by Bollerslev (1986) and Taylor (1986) is an extension of the Autoregressive Conditional Heteroscedasticity (ARCH) developed by Engle (1982). ARCH was developed out of the need to model volatility clustering, which neither the Normal nor Student's t-distribution are able to accommodate by themselves. Volatility clustering was first noted by Mandelbrot (1963) and is a property often observed in financial data. A possible explanation for volatility clustering is that information which often motivates volatility in markets itself occurs during a concentrated period of time instead of being evenly spaced over time (Angelidis et. al, 2003 p.1). The important feature of the GARCH model is its ability to capture not only volatility clustering but also the thick tails of returns. In order to capture volatility clustering we assume that the conditional variance $\sigma_t^2 \equiv Var[\eta_t \mid \Omega_{t-1}]$ is dependent on past innovations. The innovation η_t is some random variable with zero mean and variance conditional on information Ω_{t-1} given at time *t* (van den Goorberg and Vlaar 1999 p.15). In this thesis we have chosen to work with the GARCH(1,1) model due to the finding that adding more lags deos not necessarily lend the model greater explanatory power. The GARCH(1,1) is defined as:

$$\sigma_t^2 = \omega + \alpha\eta_{t-1}^2 + \beta\sigma_{t-1}^2 \tag{3.12}$$

This equation shows that σ_t^2 is a function which depends on the intercept ω, information about past residuals $\alpha_1\eta_{t-1}^2$ and fitted variance from previous periods $\beta\sigma_{t-1}^2$. It is worth noting that α_0, α_1 and β must always be positive to ensure positive volatility (Brooks 2008 p. 392). When estimating the parameters in models such as GARCH(1,1) one must estimate the parameters numerically by using the maximum likelihood methodology, because of its non-linearity. The maximum likelihood estimation finds the most likely values of the vector θ of GARCH-parameters given the input data and is defined as follows (Angelidis et al 2003):

$$\ln L(\theta) = -\frac{1}{2}\left[T\ln(2\pi) + \sum_{t=1}^{T} z_t^2 + \sum_{t=1}^{T}\ln(\sigma_t^2)\right] \tag{3.13}$$

z_t^2 is the independently and identically distributed standardized innovations (innovation divided by conditional volatility). The conditional volatility estimate by the GARCH(1,1) model assuming normality is then incorporated into the formula (3.8).

The GARCH(1,1) approach can also be implemented assuming a t-distribution as the innovation distribution. The conditional volatility estimate is then implemented into the t-distributional framework as defined in (3.10) in order to retrieve the VaR estimate. It has been shown that this model might be very useful in VaR forecasting as it can account for both leptokurtosis and time-varying volatility (Perignon and Smith 2008). Specifically, the log-likelihood estimator with t-distributed innovations is defined as (Angelidis et al 2003):

$$\ln L(\theta) = T\left[\ln\Gamma\left(\frac{v+1}{2}\right) - \ln\Gamma\left(\frac{v}{2}\right) - \frac{1}{2}\ln\left[\pi(v-2)\right]\right] - \frac{1}{2}\sum_{t=1}^{T}\left[\ln(\sigma_t^2) + (1+v)\ln\left(1+\frac{z_t^2}{v-2}\right)\right] \quad (3.14)$$

z_t^2 again depicts the standardized innovation. In addition, (3.14) also estimates the degrees of freedom in order to determine the exact shape of the distribution function as well as the GARCH-parameters.

3.3.5 Implied Volatility

The *raison d'être* of volatility modeling is to produce accurate forecasts, and a natural extension of this would be to use implied volatility (IV) to predict future VaR estimates along the lines of Giot (2005). The basic idea is to invert the Black-Scholes option pricing formula and then to use the observable market variables (spot price, strike price, time to maturity, risk-free interest rate) to obtain the market's anticipation of future volatility. This estimate of implied volatility is more forward-looking than the models discussed above as these all build explicitly on historical data and it could be argued that this measure should yield superior forecasts of volatility. The research conducted on the subject has produced somewhat mixed results but the general conclusion seems to be that implied volatility produces more sophisticated forecasts compared to methods building on historical volatility (Giot 2005). A natural extension, then, would be to apply this volatility forecast to the VaR framework to investigate if IV in any way could be incorporated as a successful variable in measuring market risk.

The research on this subject tends to revolve around those indices already created with the purpose of measuring implied volatility. The most noteworthy of these might be the VIX index calculated and published by the Chicago Board Options Exchange (CBOE), whose underlying index is the S&P 500 index. The exact computation of the VIX is complicated and need not be explained in great detail here (they are available in the CBOE technical document, "The CBOE Volatility Index – VIX"). For the purpose of this paper it is enough to know that the VIX is expected volatility retrieved from a weighted average of near-term and next-term put and call options with a range of different strike prices. It is thus the market's best assessment of the volatility of the underlying stock index over the remaining life of the option, usually 22 trading days. Implied volatility has been shown to provide a meaningful and comprehensive risk measure that successfully can be incorporated into the VaR framework (Giot 2005) as it can predict the right number of exceptions over VaR and account for clustering (see below on backtesting). It is important to acknowledge that implied volatility is biased upwards compared to realized volatility (RV) and that this bias needs to be accounted for. Moreover, the VIX index is tied exclusively to the underlying S&P 500 index which must be appreciated if VIX is applied to other market data than this broad equity index. For an example, in this thesis VIX should easily be applicable on the S&P 500 but maybe not on the trading portfolio. If the VIX in any way should be applied to other market entities than the underlying, some adjustments of VIX might be needed in order to make it truly applicable. These are all issues addressed in the data and results section below.

3.4 Backtesting

When a VaR model has been estimated it is important to check its reliability and accuracy. Banks such as Bank of America usually do this every quarter and due to the recent turmoil in financial markets, making sure the VaR model is accurate might be more important than ever. For an example, the Scandinavian bank Nordea chose to revise its VaR model due to the extreme market movements seen during the financial crisis (more on this issue in the result section below). The popularity of backtesting VaR models was also advanced by the Basel Committee's decision to let banks use their own internal VaR model for estimating capital requirements.

The aim of backtesting is to estimate if the amount of losses predicted by VaR is correct. A loss that exceeds VaR is called an exception[2]. For example, with a sample of 1000 observations and a confidence level of 95% we assume 50 exceptions which must hold for a successful model. If the result of the backtest yields more exceptions than we can expect, our model underestimates the losses and is rejected. Too few exceptions mean that our model overestimates the risk meaning that firms might allocate too much capital to cover non-existing risk. In both of these cases the model used needs to be recalibrated in order to capture the risk in proper fashion and to determine the required capital. To test whether our model needs to be recalibrated or not we must first determine the failure rate. For example, if x is the amount of days VaR has been exceeded and n is number of days in our sample, x/n is our failure rate. Asymptotically the failure rate should harmonize around the p, which is the confidence level or left tail of the distribution (Jorion, 2001 p. 131-133).

To test whether the empirical result x/n is close enough to p (the predicted rate of exceptions by VaR) we use the Christoffersen (1998) approach. In the first part of the Christoffersen approach, which also is Kupiec's (1995) test, we use the likelihood ratio (LR) to test the hypothesis of correct unconditional coverage (Jorion 2001 p. 134). That is, we test if the model at hand can correctly determine, within the error margin (see appendix 2), the number of exceptions. The number of exceptions is determined by the following indicator I_{t+1} (Angelidis et al 2003):

$$I_{t+1} = \begin{cases} 1, & if\ r_{t+1} < VaR_{t+1} \\ 0, & if\ r_{t+1} \geq VaR_{t+1} \end{cases}$$

Where r is the continuously compounded return at $t+1$.

[2] The terms 'exception' and 'exceedance' will be used interchangeably in following discussions.

The test is designed to find out if the number of ones is close enough to p which must hold for a successful model. The test itself is defined as follows and is distributed as a $\chi^2(1)$ (Dowd 2005 p. 329):

$$LR_{uc} = -2\ln\left[(1-p)^{n-x} p^x\right] + 2\ln\left[(1-x/n)^{n-x}(x/n)^x\right] \tag{3.15}$$

The second part of the Christoffersen approach is to test for independence[3], or more directly: to test whether the exceptions are serially independent of each other. This latter part of the Christoffersen approach is important insofar that it detects whether the exceptions occur in clusters or not, i.e. it can account for the volatility clustering mentioned above. If one can prove the existence of clustering the model is misspecified and needs to be recalibrated. A new indicator building on the exception indicator above is set up which defines n_{ij} to be the amount of days that j (exception) occurred when it was i (no exception) the day before. π_{ij} is the probability of state j being observed given that state i was observed the previous day (Jorion 2001, p. 141). The test statistic testing independence is (Dowd 2005 p. 329):

$$LR_{ind} = -2\ln\left[(1-\hat{\pi}_2)^{n_{00}+n_{11}} \hat{\pi}_2^{n_{01}+n_{11}}\right] + 2\ln\left[(1-\hat{\pi}_{01})^{n_{00}} \hat{\pi}_{01}^{n_{01}} (1-\hat{\pi}_{11})^{n_{10}} \pi_{11}^{n_{11}}\right] \tag{3.16}$$

where $\hat{\pi}_{01} = \dfrac{n_{01}}{n_{00}+n_{01}}$, $\hat{\pi}_{11} = \dfrac{n_{11}}{n_{10}+n_{11}}$, $\hat{\pi}_2 = \dfrac{n_{01}+n_{11}}{n_{00}+n_{10}+n_{01}+n_{11}}$ (3.17)-(3.19)

$\hat{\pi}_{01}$ = The probability of a non-exception being followed by an exception.

$\hat{\pi}_{11}$ = The probability of an exception being followed by an exception.

$\hat{\pi}_2$ = The absolute probability of a non-exception or exception being followed by exception.

Equation (3.14) is distributed as a $\chi^2(2)$. These two LR tests can be combined, thereby creating a complete test for coverage and independence which also is distributed as a $\chi^2(2)$:

$$LR_{cc} = LR_{uc} + LR_{ind} \tag{3.20}$$

[3] The terms ‘independence’ and ‘no volatility clustering’ will be used interchangeably in this thesis.

This is the Christoffersen approach to check the predictive ability and accuracy of a VaR model. The upside of this test lies in the combination of two tests which can be tested separately to backtrack if the model fails due to wrong coverage or due to exception clustering. Altogether it should provide us with the tools necessary to evaluate and compare the VaR models mentioned above. A full specification of null and alternative hypotheses as well as critical values for the χ^2 distribution and no rejection regions is found in appendix 2.

3.5 Basel and the Focus on Market Risk in Trading Portfolios

The original Basel framework, the 1988 Basel Capital Accord, was introduced in order to regulate banks' credit risk inherent on their loan portfolio. However, during the 1980s and 1990s the international financial system saw great growth in the trading activities in banks and their increased exposure towards market risks. To acknowledge this, the regulators responded by developing a framework that could accommodate the risk exposure in banks' trading portfolios and was able to set specific capital requirements accordingly. The result was the 1996 Amendment to the Capital Accord to Incorporate Market Risks. Thus, credit risk pertains primarily to banks' loan and credit portfolios whereas market risk relates primarily to the trading portfolio (Jackson et al. 1997).

The drawback in the regulatory framework was its additive clause, stating separate capital requirements for each position (interest rates, equities, FX, commodities, derivatives etc.) and summing them into one capital requirement, thereby neglecting possible diversification benefits across asset classes and markets. This obvious disadvantage for multinational financial institutions possessing global portfolios caused the Basel Committee to allow the use of internal risk models within regulated banks, spurring the development of various refined VaR methods. These internal risk models would however be subject to certain standards: calculation of losses over a 10-day holding period with a 99% confidence interval. In addition, banks usually use a 95% confidence level over the next 1-day holding period to control for in-house risk (Jackson et al. 1997). In line with this, this thesis sets out on the quest of finding the optimal VaR model being able to capture the risk inherent in the banks' trading portfolio.

More specifically a banks' trading book is defined as: a portfolio consisting of positions in financial instruments and commodities held with trading intent (short-term perspective) or to hedge other components of the trading book (Basel Accord, Market Risk Amendment 2005). An accurate risk model for this portfolio must be determined by the methodology of backtesting (see above). The Basel Accord only utilizes the first step in the Christoffersen approach of correct unconditional coverage and penalizes models that fail to account for the right amount of exceptions with a backtesting window of at least 250 trading days. In these cases Basel applies an extra capital charge that should be the higher of (i) the current VaR estimate or (ii) the average VaR estimate over the previous 60 days multiplied by three or more depending on the number of VaR breaks in the model (Basel Accord, Backtesting

1996). This framework will be further discussed below in evaluating predictive performance and accuracy of the VaR models. In addition, via the second step in Christoffersen (1998) for testing conditional coverage or independence we will move beyond the simple Basel framework and be able to judge the models from the viewpoint of time-varying volatility.

3.6 Previous Research

The previous research with the explicit perspective of banks' trading portfolios is scarce in comparison to the research performed generally on the subject of VaR. To date almost all the empirical research on VaR methods have been conducted on single assets such as currencies, equity indices, fixed income portfolios etc. These studies are motivated by the non-disclosure of financial institutions of their trading positions and the fact that the exact sources of trading revenues are unknown to the public. There is no questioning that these methods are able to produce sophisticated VaR estimates, but the fact that they ignore the practical issues in risk management might make them unrealistic in dealing with the risk inherent in financial institutions.

For a full discussion of possible models beside those discussed in this thesis, the interested reader is encouraged to consult the articles of Bao et al. (2006) or Kuester et al. (2006) which provide excellent overviews over a wide range of VaR models currently used in the financial literature, or to review detailed textbooks like the ones of Dowd (2005) or Jorion (2001). As it has been shown that banks use relatively simplistic models of both parametric and non-parametric nature and that even a straightforward GARCH(1,1) model can keep up with banks' internal VaR estimates (Berkowitz and O'Brien 2002), we find that the models considered above should provide is with an extensive analysis enough for our purpose of identifying models producing reasonable VaR estimates.

Only little research has been conducted with the explicit aim of finding the best model for banks' trading portfolios. For the purpose of this thesis, the following three articles merits attention since they cast light on the relevancy of realistic data when measuring market risk. They should be enough as a background for contrasting and comparing the results achieved later on in this thesis.

Jackson et al. (1997) use data on trading revenues given to them by a bank on condition of anonymity to test parametric and non-parametric models. They stress the need for realistic portfolios as single equity indices or single FX rates are ill-advised in dealing with risk in financial institutions and conclude a slight advantage of parametric methods over non-parametric methods in forecasting VaR. It is interesting to note that they get different results for different data sets or trading portfolios which might be something to build on in this thesis.

Berkowitz and O'Brien (2002) extract daily P/L data from financial statements for six large banks with significant trading accounts together with data on internal VaR estimates and compare these estimates to a straightforward ARMA(1,1)-GARCH(1,1) estimate of VaR. The main result is that the model by Berkowitz and O'Brien better can account for volatility clustering and give at least as good or better VaR estimates as the more conservative VaR methods applied by banks.

Finally, Perignon and Smith (2008) use P/L data for five large banks extracted from P/L graphs through an estimation procedure involving image anaysis in MATLAB. They too perform a comparative study of VaR models and conclude that the best performing models are either of non-parametric nature (filtered HS) or parametric nature (GARCH-based models accounting for volatility clustering). These approaches of acquiring real trading revenue data might not be feasible for the purpose of the thesis and the next section will look more deeply into the delicate issue of constructing or replicating a banks' trading book, which should provide us with a more realistic portfolio for retrieving VaR estimates.

4. Data

This section will explain the data used in the thesis and the construction of the trading portfolio. In addition, descriptive statistics of the return time series used, S&P500 and the trading portfolio, is presented.

4.1 Sample Period

Our data has been extracted from DataStream and stretches from 2001-05-04 to 2009-01-01. This sample period is chosen because it incorporates a first period with tranquility followed by a period of extreme volatility in the market relating to the recent financial crisis (see appendix 1). This partition of the data set is very interesting from the perspective of the purpose of this paper as it allows us to test which models can actually manage the extreme change in volatility and still produce reasonable VaR estimates during the latter years of our sample. The data period chosen will give us 1000 observations to estimate our models, our in sample period, and 1000 observations for backtesting the daily forecasting performance, our out-of-sample period. This should be enough to obtain stability in the estimation procedure of the parameters, as well as giving us a backtesting period large enough for the Christoffersen test. Under this sample period we use both a rolling window and a recursive window depending on the method chosen (Brooks 2005 p. 246). The rolling window is used when calculating VaR for historical simulation and for the parametric methods when applicable. The recursive window is used for AWHS and VWHS when calculating VaR since the models are designed to circumvent the problem of data dropping out of the window: they apply less weight to older observations and more weight to recent observations.

4.2 Portfolio Composition

In this thesis we have chosen to work with two different portfolios, Standard & Poor's index of the 500 most traded stocks in United States and a weighted portfolio which aims at reflecting the asset composition in an American bank's portfolio[4]. The mimicked bank trading portfolio has been chosen by studying Bank of America's (BofA) annual reports over the years 2004-2008. Annually, Bank of America discloses the average VaR content of their trading portfolio during each year; consequently we have chosen to take the average of this as the portfolio composition over the years 2004-2008. As the bank chosen is an American bank we have selected each constituting component in the portfolio from the American financial markets.

As observed in table 4.1, our trading portfolio consists of five different kinds of assets: foreign exchange, interest rate, credit, equities and commodities. As a proxy for foreign exchange we have created a foreign exchange sub-portfolio. This portfolio is equally weighted between the most traded currencies: British pound (GBP), Japanese yen (JPY) and the Euro (EUR).

Table 4.1 Portfolio composition

The table below displays the composition of our trading portfolio. This portfolio is an average of Bank of America's trading portfolio over the years 2004-2008.

Portfolio composition	
Foreign exchange	7%
Interest rate	23%
Credit	40%
Equities	23%
Commodities	7%
Sum	100%

Our proxy for the interest rate in the portfolio is the US T-bill with a maturity of three months. The data extracted displayed the yield of the T-bill which was transformed to show a hypothetical price for the T-bill. This was a necessity due a bonds inverse relationship between yield and price. Remember that the price of a bond increases when yield decreases.

As a proxy for the credit in the portfolio we have chosen the Merrill Lynch's US corporate bond index with a rating of BBB. As a comprehensive index for companies of different grades does not exist we limit ourselves to companies with a BBB rating. This rating should yield a good proxy for measuring the credit risk inherent in the credit spread above the default free T-bill.

When trying to mimic the equity in Bank of America's trading portfolio we have selected the Standard & Poor's 500 (S&P 500). The S&P 500 is a value weighted index consisting of

[4] In the following discussion please note that we have excluded mortgages from our trading portfolio due to difficulties in quantifying this entity. We think that relevant credit parts are presented through the credit spread.

the 500 most traded stocks in the United States. The stocks included are those of large publicly held companies traded on either New York Stock Exchange (NYSE) or NASDAQ. The S&P 500 consists almost completely of US stocks making it a good proxy for the equity in Bank of America's trading portfolio.

The commodity in our trading portfolio consists of the DJ AIG commodity index. This index is constructed to be a highly liquid and diversified index for future contracts in the commodity market. To help ensure the index's diversity no group of commodity (e.g. energy, precious metals) may constitute more than 33% of the index and no single commodity (e.g. oil, pork bellies) may constitute less than 2% and no more than 15% of the index. This diversity qualifies it as a good substitute for the commodity stake in Bank of America's trading portfolio.

These four indices combined with the F/X trading portfolio creates a well diversified portfolio which is weighted according to the average of Bank of America's trading portfolio. The authors acknowledge that this constructed portfolio is a crude measure of the real trading positions of BofA. Still, without detailed information about trading profit/losses this is the second best alternative in order to analyze the performance of VaR models for banks. It adds a certain reality dimension since it contains several sources of market price risk for different asset classes and contains the diversification effect often neglected by other authors on this subject. In addition, all the models above will be tested on the S&P 500 to contrast or confirm the results found for the trading portfolio.

4.3 Descriptive Statistics

The time series for the S&P 500 and the components of the trading portfolio (TP) are all transformed into daily continuous compounded returns as follows:

$$r_t = \ln\left(\frac{P_t}{P_{t-1}}\right) \times 100 \tag{4.1}$$

Visual inspection of the return time series (Appendix 1) points to the stylized facts discussed above in section 3 as periods of tranquility are followed by periods of extreme market movements characterized by volatility clustering. This is further verified by descriptive statistics for each time series (see table 4.2). Comparing the S&P500 with the trading portfolio it is important to note the lesser volatility and the less extreme values for the trading portfolio. This is an effect of the diversification effect described previously in this thesis, and one might suspect that this effect will have an impact on the empirical results as the VaR models might be able to deal better with less extreme volatility.

For S&P 500 we note a slight negative skew and high excess kurtosis resulting in a high Jarque-Bera (JB) test statistic (Brooks 2005, p. 163). The null hypothesis of normal distribution is thus rejected. The same pattern is seen in the time series for the trading portfolio but this time with a more negative skew and slightly more excess kurtosis. Again, the null hypothesis of normality is rejected. This could potentially be a problem for the models assuming normality and it will be interesting to dig deeper into this issue. One could also expect the models based on the t-distribution to perform relatively better than methods based on the assumption of normality since this is clearly a violation of the stylized facts exhibited by the return time series. The question regarding how this affects the VaR estimates will be discussed below in the results section.

Table 4.2 Descriptive statistics for S&P 500 and TP

SP500		TP	
No obs	2000	No obs	2000
Mean	-0,02	Mean	0,00
Minimum	-9,47	Minimum	-2,72
Maximum	10,96	Maximum	2,80
Standard Deviation	1,32	Standard Deviation	0,33
Kurtosis	11,27	Kurtosis	11,96
Skewness	-0,17	Skewness	-0,74
Jarque-Bera	5713	Jarque-Bera	6872

5. Results

This section will deal with the empirical results obtained from estimating the models together with some more detailed descriptions of the actual estimating procedures used. All models are evaluated with the Christoffersen framework in order to inference about their reliability and predictive performance. In appendix 3 all estimations can be studied graphically over the backtesting window.

5.1 Non-parametric methods

In this thesis, three different non-parametric methods have been used to estimate VaR: Historical Simulation (HS), Age Weighted Historical Simulation (AWHS) and Volatility Weighted Historical Simulation (VWHS). When calculating the VWHS approach to VaR we have forecasted volatility using both GARCH(1,1) and EWMA frameworks.

Table 5.1 VaR exceedances on 5% and 1% level, expected number of exceedances are 50 and 10 for the 5% and 1% levels respectively. Backtesting using the Christoffersen methodology, critical values are found in the appendix. Values marked in bold confirm the model (null hypothesis not rejected).

	S&P 500		**Backtesting 5%**			**Backtesting 1 %**		
Model	$VaR_{5\%}$	$VaR_{1\%}$	LR_{uc}	LR_{ind}	LR_{cc}	LR_{uc}	LR_{ind}	LR_{cc}
HS (250)	77	27	13.3	**2.7**	16	19.9	**1.3**	21.2
HS (1000)	95	39	34.1	**0.7**	34.8	49	**1.5**	50.5
AWHS	80	33	16.2	**2.1**	18.2	33.3	**8.9**	42.2
VWHS EWMA	66	28	4.9	**1.1**	6	22	**4.2**	26.2
VWHS GARCH (1,1)	69	24	6.8	**0.2**	7	14.2	**1.1**	15.3

	Trading Portfolio		**Backtesting 5%**			**Backtesting 1 %**		
Model	$VaR_{5\%}$	$VaR_{1\%}$	LR_{uc}	LR_{ind}	LR_{cc}	LR_{uc}	LR_{ind}	LR_{cc}
HS (250)	73	22	9.8	**3**	12.8	20.7	**3.6**	24.3
HS (1000)	76	35	12.4	10.3	22.7	38.3	15.6	54
AWHS	76	32	12.4	10.3	22.7	30.9	18.1	49
VWHS EWMA	63	28	**3.3**	11.8	15.1	22	21.9	43.9
VWHS GARCH (1,1)	72	29	9	**4.4**	13.5	24.1	**1.6**	25.8

5.1.1 Historical Simulation

Historical simulation may not be the most sophisticated method to estimate VaR, on the other hand, historical simulation is mathematically and intuitively the simplest approach to estimate VaR which make it very widely used. As noted above, in section 3.2.1, out of 1000 profit/loss observations VaR is simply the 51st largest loss on the 95 % confidence level and the 11th largest loss when using 99 % confidence level. In our thesis we apply a rolling window using 1000 observations which corresponds to four years on both 99 % and 95 % confidence level.

Following the Scandinavian bank Nordea's last annual report we also scale down our rolling window to 250 observations which corresponds to one year. This is motivated by the large volatility seen in the market since the summer of 2007. Having two rolling windows with different sizes incorporates a more conservative VaR which will consistently have a higher VaR during tranquil periods with a more flexible VaR which adapts more quickly to the current circumstances[5].

As expected we can see (appendix 3) that HS for 250 observations is more responsive to changing volatility on both confidence levels. HS with 1000 observations incorporates more data in its estimation of VaR and consequently shows inertia when faced with changing volatility and rapidly changing market conditions; consequently it overestimates VaR during periods of tranquility and underestimates VaR during periods of higher volatility.

When backtesting the results for HS we follow the Christoffersen methodology which utilizes the χ^2 distribution. The first part of the Christoffersen methodology, also known as the Kupiec test has been detailed described in section 3.5 and has been applied to the VaR estimation.

In table 5.1 the first part of Christoffersen's test for the correct unconditional coverage of HS's VaR estimation is presented. Both the VaR forecast for the trading portfolio and SP500 give unsatisfying results and consequently fail this test which is confirmed by running the Christoffersen's likelihood ratio test. The null hypothesis of correct unconditional coverage can safely be rejected for both levels of confidence and observations.

In the second part of the Christoffersen test we determine whether the HS approach suffers from volatility clustering or not. To test for volatility clustering is from a practical point of view important. If a bank allocates capital for 50 exceedances over a period of four years it may not be able to stay liquid if a majority of the exceedances appear during the course of two

[5] To retrieve the VaR for both 5% and 1% confidence level we use the excel function PERCENTILE. This function simply extracts, with respects to chosen confidence level, the largest loss. If the desired observation is not a multiple of the sample, PERCENTILE interpolates to determine the percentile with respect to the sample.

months. HS with 1000 observations and 250 observations with a confidence level of 1% on the trading portfolio rejects the null hypothesis of no volatility clustering and is therefore disqualified. HS with 250 observations and a confidence level of 5 % on the trading portfolio as well as HS, both 5% and 1%, with 250 and 1000 observations on S&P 500 however rejects the null hypothesis of independence. It follows from the rejection of LR_{uc} that a combined hypothesis of correct unconditional coverage and independence can be safely be rejected.

5.1.2 Age Weighted Historical Simulation

Moving up the ladder of complexity we evaluate our VaR estimation using the Age Weighted Historical Simulation which has been presented in section 3.2.2. Using this method, we leave the rolling window method in favor for the recursive window because of the declining importance of each observation, which is a property of the AWHS method. As in HS we estimate 1000 VaR's where the estimation window expands from 1000 observations for the first VaR to 2000 for the last VaR. For every new VaR we need to re-estimate the weight given to each observation because of the declining importance of each observation. For this purpose we have written a VBA macro which can be viewed in the appendix.

Table 5.1 presents the reliability of the VaR estimations calculated with AWHS. Compared with HS AWHS performance is quite poor with, in some cases, more exceedances than HS (it should be noted that we use a very conservative decay factor λ of 0.9999). Subsequently; for the trading portfolio as for S&P 500 the null hypothesis of correct unconditional coverage can be rejected on all confidence levels. Most of the exceedances occur late in the sample period which shows that the AWHS is slow to adjust to changing market conditions. The second part of the Christoffersen regarding volatility clustering or independence has been carried out and disqualifies AWHS on both confidence levels for the trading portfolio, that is, the null hypothesis for no volatility clustering is rejected. However AWHS manages to account for volatility clustering, or in other words, independence on both levels of confidence when applied on S&P 500. For the combined test, AWHS, is deemed unsuitable or both the trading portfolio and S&P 500 for both levels of confidence.

5.1.3 Volatility Weighted Historical Simulation

We use two approaches to estimate the volatility when calculating VaR with VWHS; EWMA and GARCH(1,1) assuming normal distribution. As with the AWHS model, the VWHS model uses a recursive estimation window, meaning that the first VaR estimation uses 1000 observations to yield a forecast, from then on, the estimation window expands as the number of observations within the window increases. For example, the second VaR forecast uses 1001 observations and the third one uses 1002 observations, and as before we forecast 1000 VaR's .

In the EWMA approach to finding conditional volatility we apply the $\lambda=0,94$ RiskMetrics decay factor. The volatility of day *t-1* is multiplied with this decay factor and is subsequently weighted as described in section 3.3.3. When volatility has been forecasted in line with the EWMA approach it is included in the AWHS model described in section 3.2.3. When backtesting VWHS using EWMA to estimate volatility, the results are somewhat more satisfying than previous methods. Although failing to yield a satisfying result for the first part of the Christoffersen in three out of four times, VWHS using EWMA passes the test for correct unconditional coverage for the trading portfolio on the 5 % level. When testing for independence of exceedances, VHWS using EWMA rejects the null hypothesis and is therefore deemed unsuitable to be used as a forecasting model for estimating VaR for S&P 500. However, the test does not reject this null hypothesis for the trading portfolio. When running the combined test, LR_{cc}, the EWMA approach fails to reject the null hypothesis for both the S&P 500 and our trading portfolio on 5% and 1% level.

Faced with the failure of the EWMA volatility weighting method, we try to incorporate another volatility measure in the VWHS model, namely the GARCH(1,1) model assuming normal distribution (that is, a normal distribution for the innovations is assumed and normality is assumed in the maximum likelihood estimation). The GARCH(1,1) parameters and conditional volatility forecast are both estimated in Eviews where a loop has been programmed to estimate GARCH(1,1) parameters and conditional volatility the necessary 1000 times. This volatility forecast works as an input in VWHS in the same manner as EWMA. When studying table 4.2 displaying descriptive statistics we reject the null hypothesis of normality for both the trading portfolio and S&P 500. A careful guess before testing this approach with the Christoffersen methodology is that results will be negative, that is, the GARCH(1,1) assuming normal distribution will not be deemed suitable. This guess is confirmed by the Christoffersen test for correct unconditional coverage where the VWHS using GARCH(1,1) reject the null hypothesis for both the trading portfolio as well as for S&P

500 on all levels. When running the second part of the Christoffersen test, the GARCH(1,1) manages to respond to volatility clustering and does not rejects the null hypothesis of no volatility clustering, that is, the test for independence. Because of the inability of GARCH(1,1) to reject the null hypothesis of correct coverage, GARCH(1,1) also fails to rejects the combined test, LR_{cc}.

To sum up the Christoffersen tests for the non-parametric models, none of the models manages to qualify as suitable for forecasting VaR, neither for the trading portfolio nor for S&P 500. The perplexing observation when looking at the test for independence is that all models manages not to reject the null hypothesis of independence (no volatility clustering), while the same test disqualifies three out of five non-parametric models for the trading portfolio.

5.2 Parametric methods

The parametric methods used in this thesis include VaR based on the normal and the t-distribution both with and without explicitly modeling time-varying volatility. For the cases where we use time-varying volatility we use the RiskMetrics model, GARCH based on the normal and the t-distribution, and implied volatility incorporated into a normal distribution.

Table 5.2 VaR exceedances on 5% and 1% level, expected number of exceedances are 50 and 10 for the 5% and 1% levels respectively. Backtesting using the Christoffersen methodology, critical values are found in the appendix. Values marked in bold confirm the model (null hypothesis not rejected), and values marked with star passes the joint test of correct conditional coverage.

	S&P 500		**Backtesting 5%**			**Backtesting 1%**		
Model	$VaR_{5\%}$	$VaR_{1\%}$	LR_{uc}	LR_{ind}	LR_{cc}	LR_{uc}	LR_{ind}	LR_{cc}
Normal	89	53	26.3	**1.3**	27.6	92.7	**3.1**	95.8
t-distribution	62	25	**3.0**	**4.0**	7.0	16.0	**5.0**	21.0
Implied volatility	96	33	35.5	**0.2**	35.7	33.3	**2.3**	35.6
RiskMetrics	66	30	**4.9**	**0.6**	**5.6**	26.3	**1.9**	28.2
GARCH(1.1)	71	29	8.3	**0.0**	8.3	24.1	**1.7**	25.9
GARCH(1.1)t	43	11	**1.1**	**0.0**	**1.1***	**0.1**	**0.2**	**0.3***

	Trading Portfolio		**Backtesting 5%**			**Backtesting 1%**		
Model	$VaR_{5\%}$	$VaR_{1\%}$	LR_{uc}	LR_{ind}	LR_{cc}	LR_{uc}	LR_{ind}	LR_{cc}
Normal	75	40	11.5	8.7	20.2	51.8	15.2	67.1
t-distribution	51	26	**0**	18	18	18	13	31
Implied volatility	90	34	27.5	**0.2**	27.7	35.8	**2.4**	38.2
RiskMetrics	60	21	**2.0**	**2.9**	**4.9**	9.3	**0.9**	10.2
GARCH(1.1)	63	25	**3.3**	**0.0**	**3.3**	16.0	**0.2**	16.2
GARCH(1.1)t	44	7	**0.8**	**0.0**	**0.8***	**1.0**	**0.1**	**1.1***

5.2.1 Normal VaR

VaR based on the normal distribution has its limitations but nevertheless serves well as a benchmark for the rest of the parametric models. Our estimations for mean and volatility are carried out over a rolling window with 1000 observations. The choice of a rolling instead of a recursive window is that we want to allow for certain flexibility over time, which a rolling window can give us. VaR estimates are retrieved from formula 3.8 above. As expected, the normal VaR performs poorly since it cannot capture the correct number of exceptions. The normal VaR for S&P 500 passes the second part of the Christoffersen test as the exceedances are non-clustered in the model (see table 5.2). This does not hold for the trading portfolio. All told, the Normal VaR does not predict the correct conditional coverage of exceedances. Thus, the Normal VaR behaves exactly as what can be expected from theoretical standpoints and the descriptive statistics above since the normal distribution proves to be too restrictive in its assumptions in order to model financial data correctly. This should disqualify the model as a tool for practical risk measurement and management.

5.2.2 t-VaR

VaR based on Student's t-distribution should yield more successful results than the normal model above since it should better account for the fat tails and leptokurtosis of financial time series. Again, our estimations are carried out over a rolling window of 1000 observations. Working with the generalized t-distribution requires estimation of the mean and volatility as well as the degrees of freedom, since these determine the shape of the distribution (the lower the degrees of freedom, the fatter tails the distribution has and vice versa). The degrees of freedom *v* are estimated from the following probability density function. Taking the natural logarithm of this function and maximizing the sum of the log-likelihood function with respect to degrees of freedom *v* we can estimate the shape of the distribution function together with the mean and the volatility. Exact calculations for this purpose with VBA macro is supplied in appendix 4.

$$\ln L(\theta) = -\frac{1}{2}\ln\left[\frac{\pi(v-2)\Gamma(v/2)^2}{\Gamma((v+1)/2)^2}\right] - \frac{1}{2}\ln\sigma_t^2 - \frac{(v+1)}{2}\ln\left[1+\frac{(r_t-\mu)}{\sigma_t^2(v-2)}\right] \quad (5.1)$$

This is done by numerical maximization in Excel 1000 times for each new backtesting window. That is, the first estimation is carried out for a backtesting window from 1 to 1001, the second estimation over observations 2 to 1002 etc. We supply the VBA macro used for this purpose in appendix 4. The estimated degrees of freedom are used as an input in formula (3.10) above to retrieve VaR estimates.

The t-VaR works surprisingly well on the 5% level where it succeeds in predicting the right number of exceedances for both the S&P 500 and the trading portfolio (table 5.2). However, it fails in capturing the behavior of the more extreme quantile on the 1% level. In addition, it also fails the test of independence. As a result, the use of the simple t-distribution is not recommended in risk management but the outcome shows that the t-distribution might be a feasible workhorse when dealing with the non-normality of financial data.

5.2.3 Implied volatility

The use of implied volatility (IV) in VaR estimation is primarily motivated by the fact that it has been shown to produce accurate forecasts compared even to sophisticated models based on historical data. This thesis poses the question if IV also can be incorporated successfully into a VaR framework. It can be expected that IV, represented through the VIX index, should produce accurate VaR estimates for the S&P 500 since this is the underlying asset. However, when measuring the IV a certain upward bias versus the realized volatility (RV) can be observed, possibly of time-varying nature and correlated with the current level of volatility (Chernov 2002). Accounting for this bias is hot question in academic research at the moment and could include complicated statistical methods, modeling the difference between IV and RV as a time-varying process. However, for simplistic reasons this thesis takes a simpler approach to the problem where the volatility from the VIX index is reduced by the mean of the difference between the two series (IV and RV). This makes the volatility measure less biased and useful together with formula (3.8) assuming normality.

If IV in any way were to be incorporated for other data than the S&P 500, e.g. the trading portfolio, further statistical manipulations are necessary. The basic problem is that the volatility is higher for the S&P 500 than for the trading portfolio since the latter exhibits a notable diversification effect. The approach of this thesis is again a simplistic one where the IV from the VIX is scaled down to fit the trading portfolio better. Since the relationship of the volatility for the S&P 500 to the volatility of the trading portfolio, on average, is 4:1 the IV is simply scaled down by this ratio.

As table 5.2 shows, VaR with IV represents a poor measure of downside risk since it fails to account both for the correct number of exceptions and independence for both S&P 500 and the trading portfolio. This might be quite a striking result since IV has proved to be such as successful parameter in forecasting volatility. The reason for the poor performance could probably be found in the statistical transformations of IV that are needed to incorporate it into the analysis. The methods concerning IV and its implementation into the VaR framework should be more thorough than those we have applied and potentially include specific time-varying modeling of the difference between IV and RV. There is also a general difficulty in applying a measure of IV that is tied to a specific index such as VIX on other market entities such as our constructed trading portfolio, where further more refined methods should be applied than our simple 1:4 ratio. The idea behind this reasoning was that the VIX possibly could be interpreted as market wide volatility forecast which could be scaled to the focal entity of interest, which however seems to be highly doubtful. Altogether, we reject implied volatility incorporated into a normal distribution framework as it cannot account for the correct number of exceedances even though the exceedances themselves are independent.

5.2.4 RiskMetrics

The RiskMetrics (RM) model is based on the normal distribution with time-varying volatility according to the EWMA model in formula (3.10) above with $\lambda=0.94$. Mean and conditional volatility are estimated over a rolling window of 1000 observations and incorporated into the formula (3.8) above to obtain VaR estimates. In the financial literature the RiskMetrics model usually shows some success when estimating financial risk, which also holds true over our sample period. On the 5% level the RM succeeds in predicting the right number of exceptions as well as keeping the assumption of independent exceedances upright – both for the S&P 500 and the trading portfolio (table 5.2). However, on the more extreme level of 1% the RM model fails and the null hypothesis of correct unconditional coverage is rejected. Even though the RM passes the test of independence the model is still rejected according to the joint hypothesis of correct conditional coverage at the 1% level (table 5.2). As such then, the RM method is rejected since a reliable model should be able to capture the behavior of financial time series data, both for more and less extreme quantiles. Since the RM works well for the 5% level but not on the 1% level, it must be rejected as tool recommendable for practical risk management. Yet, it still points towards the feasibility of normal models with time-varying volatility.

5.2.5 GARCH(1,1)

The first GARCH model we examine is incorporated into a normal distribution and thus builds on the finding above that conditional time-varying volatility is viable in VaR forecasting, even though a normal distribution framework in itself need not be. The choice of the GARCH(1,1)[6] model is motivated primarily by the fact that it is parsimonious in its assumptions and easy to work with, which is also something to consider in practical risk management. We forecast 1000 conditional variance estimates used in formula (3.8) since in this case, we assume normality (programming script in Eviews supplied in appendix 4). Parameters are estimated 1000 times with the Maximum Likelihood procedure assuming normality (see above in section 3.3.4)

The GARCH(1,1) model shows to be too restrictive for the S&P 500 on both the 5% and the 1% level and the null of correct conditional coverage is rejected. Nevertheless, the GARCH(1,1) model works better for the trading portfolio where it, like the RiskMetrics model above, works well on the 5% level but not on the 1% level. This is probably an effect of the diversification effect, where the GARCH(1,1) model proves to be workable for the smaller volatility of the trading portfolio. However, a reliable model must be able to work well on both levels in order to meet the demands of internal models as well as regulatory requirements and the GARCH(1,1) model building on a normal distribution is therefore rejected since it does not produce the correct conditional coverage. It is reasonable to assume that these results are effects of the normality assumption underlying both the calculation of VaR estimates (formula 3.8) and the maximum likelihood estimation, which depends on normality in the estimation of the GARCH parameters. As the descriptive statistics has shown, the normality assumption might be too restrictive when dealing with real financial time data.

[6] We choose to model the GARCH process without an AR(1) term in the mean equation. When working with daily financial data it has been showed that including an AR-term might be unnecessary or even destructive, which is exactly what we experienced during our estimations.

5.2.6 GARCH(1,1)-t

The second GARCH model we use in obtaining VaR estimates is based on the t-distribution[7]. This simply means that the t-distribution is used as the innovation distribution when estimating the conditional volatility. By using the Maximum Likelihood estimation procedure (see above) we are able to determine the mean, the standard deviation and the degrees of freedom. Using these as inputs into formula (3.10) we can determine VaR. Conditional variance forecasts are generated by estimating the GARCH parameters for each new rolling estimation window (i.e. 1000 parameter estimations) and using these as inputs into the GARCH(1,1)-t model used for forecasting (1000 forecasts). Programming script for this rolling estimation procedure and forecasting in Eviews is supplied in appendix 4. This time the volatility is conditional according to $\sigma_t = \sqrt{\sigma^2_{GARCH,t}} \times Z_t$ where Z_t is the estimated critical value from the standardized t-distribution. As hinted previously in this thesis, the t-distribution might be a more reasonable assumption since the distribution can deal with the leptokurtosis of financial data. As can be seen in table 5.2 the GARCH(1,1)-t performs excellently, both for the S&P 500 and the trading portfolio and on both confidence levels. The null hypotheses of correct unconditional coverage and independence cannot be rejected, which means that the joint null of correct conditional coverage also cannot be rejected. The GARCH(1,1)-t is the only model that is able to account for both the clustering phenomenon as well as predicting the right number of exceedances for both confidence levels and for both portfolios. Naturally, the authors deem this model to be the most suitable as a tool for practical risk measurement and risk management.

[7] Again, estimations are carried out without an AR-term in the mean equation.

6. Basel Evaluation and Capital Requirements

We have now evaluated ten models according to the Christoffersen framework and move on to assess the capital requirement for our trading portfolio. We hope that this regulatory perspective will allow us some extra insight into the concept of VaR and our chosen models as it allows for a certain reality dimension. This will all be done according to the Basel Accord (Basel Amendment on Market Risk 1996) which stipulates a range of criteria to assess the quantity of risk as well as capital requirements. Financial regulators determine a banks capital requirement according to the following equation (Danielsson and de Vries, 1997):

$$CR = (3 + s) \times VaR \tag{6.1}$$

The choice of 3 as a scaling factor has been subject to a debate and is often criticized by financial institutions, however, it has been deemed justified in other studies (see Stahl 1997). Depending on the number of exceedances this scaling factor can be increased by s which can reach a maximum value of one with the implication that the multiplicative factor take on values in the range of [3,4]. The VaR in equation (6.1) is defined as the average $VaR_{1\%}$ over the last 60 days. Table 6.1 displays the framework for evaluating VaR used by financial institutions according to regulatory standards and is based on a backtesting window of 250 observations.

The green zone highlights the models that are deemed suitable for risk management purposes. That is, a model which produces no more than 4 exceedances based on a sample of 250 observations is assumed to be viable and the bank's internal risk management needs no revision. An accurate model yields no penalty as seen in column three in table 6.1.

If the VaR model returns between 5-9 exceedances the model ends up in the yellow zone. In the lower range of the model the yellow zone models are assumed to be somewhat accurate and are therefore given a relatively low penalty. Models in the upper end of the yellow zone are more likely to be inaccurate and are thus given a higher penalty which is interpreted as a higher capital requirement. The penalty, or scaling factor, is according to the regulators constructed so that the model returns to the 99^{th} percentile standard. For example, 5 exceedances yield only 98% coverage, the increase in scaling factor would be sufficient to bring the model back to the 99^{th} percentile standard which is required by the Basel accord.

10 exceedances or more constitutes the red zone. If a financial institution's VaR model end up in the red zone financial regulators automatically presume that the model needs recalibration. In contrast with the yellow zone, capital requirements are required to increase immediately and the multiplicative factor increases to a maximum 4.

Table 6.1 The table defines the green zone, yellow zone and the red zone to assess the results retrieved from backtesting the VaR models. The boundaries set in the table are based on a sample of 250 observations.

Zone	Number of exceptions	Increase in scaling factor
Green Zone	0-4	0.00
	5	0.40
	6	0.50
Yellow Zone	7	0.65
	8	0.75
	9	0.85
Red Zone	10 or more	1.00

Table 6.2 displays the capital requirement for each of the models created. We have taken the average VaR over a period of 60 days which is required by the Basel Amendment of 1996. Not surprisingly the lowest VaR estimations are given by the most naïve (and worst) models such as the normal distribution, t-distribution and historical simulation. This shows that banks might have an incentive to choose naïve models that might understate the true risk in order to keep their capital requirement on a low level instead of a more complex model which more accurately model the risk. With the constant pressure to maximize profit, accurate risk management is not a top priority because a more complex model equals a higher capital requirement which ultimately impairs the earnings ability of the financial institution. We feel that Danielsson and de Vries (1997) have captured this tendency in a neat way:

> "It is like using a protective sunblock, because one has to, but choosing the one with lowest protection factor because its cheapest, with the result that one still gets burned."

A closer look at table 6.2 reveals that the vast majority ends up in the red zone, one model (implied volatility) in the yellow zone and one model (GARCH(1,1)-t) in the green zone. Note that these calculations are carried out over the last 250 trading days characterized by extreme volatility clustering following the financial crisis and the collapse of the investment bank Lehman Brothers. Another result could probably have been achieved over a more tranquil period where instead the vast majority of models would have been deemed suitable. This shows that most models experience difficulties when capturing extreme market movements and it also points to justification of our estimation window with inclusion of extreme market movements. Only in these extreme periods do the models reveal their true selves, making it easier to reach robust conclusions on them. The results in table 6.2 also point to the deficiencies in banks' own risk measurement methods. As mentioned above, looking at the capital requirement it is straightforward to see the incentive to use simple models. For an example, the best model according to the Christoffersen test is the GARCH(1,1)-t, which also produces a relatively high capital requirement. The variability of the VaR estimates of the GARCH(1,1)-t might imply problems in the capital allocation for markets risks as the allocation cannot change rapidly which generally is a counter-argument of banks against variable VaR methods. However, rapid adjustment can be achieved by changing market instrument exposure instead of changing capital allocation with changing risk profile which means that the GARCH(1,1)-t still could be successfully incorporated into practical risk management. In comparison the HS methods or the naïve parametric methods generate a relatively low capital requirement creating incentives for bad risk management practices. The case of how to respond to this tendency is a delicate issue which regulators somehow need to respond to (more on this in section 7 below).

Table 6.2 The table displays the average VaR for each of the models investigating the trading portfolio: the number of exceedances, the scaling factor applied to each model and the capital requirement (CR) calculated with equation 6.1. Again, VaR 1% is calculated as an average over 60 days and the backtesting window constitutes 250 days.

Model	Average VaR 60 obs	No of exceedances	Scaling factor	CR
Non-parametric				
HS (250)	-1.21	12	1.0	4.85
HS (1000)	-2.02	25	1.0	8.07
VWHS with GARCH(1,1)	-2.96	11	1.0	11.85
VWHS with EWMA	-0.93	24	1.0	3.70
AWHS	-0.82	26	1.0	3.29
Parametric				
Normal distribution	-0.78	29	1.0	3.11
t-distribution	-0.89	22	1.0	3.55
RiskMetrics	-2.52	9	1.0	10.06
Implied volatility	-1.96	5	0.4	6.68
GARCH(1,1)	-2.28	11	1.0	9.12
GARCH(1,1)-t	-3.00	3	0.0	9.01

7. Discussion

Before concluding we would like to add a few comments which could cast light on the results achieved in this paper. As with all empirical investigations, certain assumptions and methods have been presumed in order to reach tangible results. The purpose of this section is thus to set all of these into a wider perspective.

We have chosen to estimate VaR with a wide range of models with the purpose of finding the model that performs the best, both for the S&P 500 and the trading portfolio, and over the interesting time period during the financial crisis. Of course, other models and approaches could have been estimated. The current line of academic research has been focused on sophisticated methods including mixture distributions, extreme value theory, skewed distributions, switching models, and a wide range of other parametric and non-parametric methods (more on these in the future research section below). We feel that the methods we have chosen should be enough for the purpose of this thesis, especially faced with the tight time constraint.

Our trading portfolio, built from information retrieved from Bank of America is an average over their trading portfolio over the past five years. The assumption that the composition of a bank's trading portfolio is left untouched for five consecutive years is highly unlikely. This assumption was however necessary due to lack of time and to minimize complexity. Rebalancing the trading portfolio would perhaps yield another result which would be interesting to dig further into. As before, we acknowledge the shortcomings of the trading portfolio but still feel that it has provided us with some extra insight into the methods of risk management, especially regarding the diversification effect across asset classes.

In the second part of the Christoffersen test concerning independence we note a few drawbacks. When testing for independence we get a positive result if returns exceed forecasted VaR two days in a row. If exceedences for example occurs every other day, say Monday, Wednesday and Friday this would not appear as volatility clustering in the Christoffersen test even though exceedances seems to be clustered. A test which could be able to detect even this kind of unusual clustering would refine the test even more, making it even more reliable. On the matter of volatility clustering, the fact that the Basel accord does not account for this is a major drawback. An even tougher requirement demanding that financial institutions take volatility clustering into account when evaluating their internal risk management would create an incentive for adopting more refined and sophisticated models.

Indeed, the case of how to respond to the clear inefficiency of bank risk management models by regulatory requirements is a very tricky issue. On the one hand, the research performed by Berkowitz and O'Brien (2002) show that banks' risk models are trivial and that even a simple ARMA(1,1)-GARCH(1,1) model can outperform them. On the other hand, responding to this deficiency in practical risk management by regulation can prove to be counterproductive. In addition to the argument of volatility clustering above, using the internal model approach requires banks to add sub-portfolio VaRs without accounting for the diversification benefits across asset classes. Altogether, banks' risk models entail many approximations and implementation issues stemming from regulatory requirements. Clearly, there exist several arguments in favor of better regulation encouraging more reliable risk management in banks, but, faced with the difficulty of implementing such a framework with broad industry consensus the current regulation is better than none.

A final point concerns deficiencies in the models used to capture risk themselves. It is important to realize that all models are simplifications of the real world and that they all rely on different assumptions. This simply means that measurement problems might arise stemming from the specification of the model used. For an example, the underlying process or distribution might be misspecified, relationships or correlations might be unstable over time (even though they are assumed constant by the model) or the model might ignore several important real world factors such as market liquidity or transaction costs. In this thesis, model risk might originate mainly from distributional assumptions, but also from parameter estimations or from other simplifications (e.g. implied volatility). Awareness of possible shortcomings reduces the blind faith in the superiority of financial modeling and should also make room for informed judgment on the model used, which is a fundamental part of practical risk management. Altogether, though, a model characterized by the features of reliability, accuracy, simplicity and ease of implementation should be deemed suitable for risk measurement and management purposes. In the words of statistician George Box: "All models are wrong, but some are useful".

8. Conclusion

The purpose of this paper is to evaluate which model yields the best result in estimating VaR for both a bank's hypothetical trading portfolio and for S&P500. The study is based on return series during the period of 2001-05-04 to 2009-01-01 with a backtesting window from 2005-03-04 to 2009-01-01 which yields 1000 daily VaR estimations. We have in total tested ten different models; the non-parametric (HS, AWHS, VWHS using EWMA and VWHS using GARCH(1,1)) and the parametric (Normal distribution, t-distribution, RiskMetrics, Implied volatility, GARCH(1,1) and GARCH(1,1)-t). We have succeeded in forecasting VaR for all ten models and evaluating them using the backtesting framework set up by Christoffersen. The ranking in table 8.1 is based on the Basel Accord or the first part of the Christoffersen test and we can observe that the VaR models assuming non-normality and/or time-varying volatility performs the best. Indeed it is also worth noting that HS with an estimation window of 250 performs reasonably well. The normal distribution is the worst choice when forecasting the number of exceedances beyond VaR which comes as no surprise when observing table 5.1 where the null hypothesis of normality is rejected. The GARCH(1,1)-t is easily the best model for forecasting the number of exceedances, both for S&P 500 and for the trading portfolio. Beyond this conclusion and the bottom four models the model ranking differ somewhat.

Table 8.1 The ranking in table 7.1 is based on the number of exceedances. The ranking follows only the Basel accord or the first part of the Christoffersen test but for the original backtesting window of 1000 observations.

Model	S&P 500	Model	Trading Portfolio
GARCH(1,1)-t	11	GARCH(1,1)-t	7
AWHS	24	RiskMetrics	21
HS (250)	27	HS (250)	22
VWHS with EWMA	28	GARCH(1,1)	25
t-distribution	28	t-distribution	22
GARCH(1,1)	29	VWHS with EWMA	28
RiskMetrics	30	AWHS	29
VWHS with GARCH(1,1)	33	VWHS with GARCH(1,1)	32
Implied volatility	33	Implied volatility	34
HS (1000)	39	HS (1000)	35
Normal distribution	53	Normal distribution	40

By observing table 8.2 where we also account for the complete Christoffersen test we can conclude that naïve models such as the normal distribution and t-distribution are ranked in the middle or the lower end of the table. As noted above, Nordea's choice to decrease their observation sample to 250 observations for the HS is justified. This model approach outperforms several other more sophisticated models. However, one might wonder on which theoretical grounds Nordea has chosen to decrease their sample window. Bank of America for example, has decided not to decrease their window and continue to estimate VaR with 750 observations. All told, the choice of the length of the observation window is arbitrary and we feel that a successful model should be based more on sound theoretical arguments. This can be achieved by the use of parametric models with non-normality and/or time-varying volatility since these are the most successful in estimating VaR. Again, the GARCH(1,1)-t model shows its accuracy and reliability as it by far outperforms its competitors when it comes to VaR prediction.

Table 8.2 We provide an overall ranking of the calculated models according to the LR_{CC} test. The ranking is based on the premise that the better the model the closer to zero the test statistics is.

Model	S&P 500	Model	Trading Portfolio
GARCH(1.1)-t	0.3	GARCH(1.1)-t	1.1
AWHS	15.3	RiskMetrics	10.2
HS (250)	21.2	GARCH(1.1)	16.2
GARCH(1.1)	25.9	HS (250)	24.3
t-distribution	26.0	AWHS	25.8
VWHS with EWMA	26.2	t-distribution	31.0
RiskMetrics	28.2	Implied volatility	38.2
Implied volatility	35.6	VWHS with EWMA	43.9
VWHS with GARCH(1.1)	42.2	VWHS with GARCH(1.1)	49.0
HS (1000)	50.5	HS (1000)	54.0
Normal distribution	95.8	Normal distribution	67.1

A final point concerns the differences, or lack thereof, when comparing the results for the trading portfolio and the S&P 500. As we can see, the ranking of the VaR models is quite similar for both datasets, which points to the fact that even simple time series such as broad stock indices might contain valuable information for practical risk measurement and risk management.

9. Future research

After concluding this thesis a number of suggestions on further research have occurred to us. Above all, we acknowledge the shortcomings of our trading portfolio and encourage future academic research with the explicit purpose of modeling and estimating VaR models on real or constructed trading portfolios. This would allow for a certain reality dimension and also have wider implications for practical risk management, which ultimately should be the goal of VaR modeling. Especially the use of Extreme Value Theory on trading portfolios should be very interesting. Furthermore, more frequent balancing of the trading portfolio and inclusion of non-linear assets such as derivatives could be performed in order to reach more refined conclusions. Implied volatility incorporated into a VaR framework could be more successful than what our paper has shown if the right statistical methods are applied in order to counter the effect of risk neutral valuation, particularly since the forward looking measure of implied volatility has proven successful in forecasting volatility. Finally, we encourage research conducted on the exact workings of the regulatory requirements and how they create incentives within banks, as this could create a foundation for improving existing regulation.

10. References

Angelidis T., Benos A., Degiannakis S. (2003) "The Use of GARCH models in VaR Estimation" *Statistical Methodology*, Vol. 1 pp. 105-128

Bank of America Financial Statements 2004-2008 www.bankofamerica.com

Bao Y., Tae-Hwy L., Saltoglu B. (2006) "Evaluating Predictive Performance of Value-at-Risk Models in Emerging Markets: A Reality Check." *Journal of Forecasting* pp. 101-128

Basel Committee on Banking Supervision (2005) *Amendment to the Capital Accord to Incorporate Market Risks*

Basel Committee on Banking Supervision (1996) *Supervisory Framework for the Use of Backtesting in Conjunction with the Internal Models Approach to Market Risk Capital Requirements*

Berkowitz J. and O'Brien J. (2002) "How Accurate Are Value-at-Risk Models at Commercial Banks?" *The Journal of Finance* Vol. LVII No. 3 pp. 1093- 1111

Bollerslev T. (1986) "Generalized Autoregressive Conditional Heteroscedasticity" *Journal of Econometrics* 31: 307-327

Boudoukh J., Richardson M., Whitelaw R., (1998) "The best of both worlds: a hybrid approach to calculating Value at Risk" *Risk* 11 (May) pp. 64-67

Brooks C., *Introductory Econometrics for Finance* 2nd Edition Cambridge University Press

Campbell J.Y, Lo A.W. MacKinlay A.C. (1997) *The Econometrics of Financial Markets*, New Jersey: Princeton University Press

Chernov M., (2002) "On the role of volatility risk premia in implied volatilities based forecasting regressions", Columbia Business School

Chicago Board Options Exchange (2009) "The CBOE Volatility Index – VIX"

Christoffersen P.F. (1998) "Evaluating Interval Forecasts" *International Economic Review* Vol. 39, No. 4, pp. 841-862

Danielsson J., de Vries C. G. (1997) "Value-at-Risk and Extreme Returns" FMG-Discussion Paper NO 273, Financial Markets Group, London School of Economics.

Dowd K. (2005) *Measuring Market Risk* 2nd Edition Chichester: John Wiley & Sons

Engle R.F. (1982) "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of UK Inflation" *Econometrica* 50: 987-1008

Giot, P. (2005) "Implied Volatility Indexes and Daily Value at Risk Models", The Journal of Derivatives, Summer 2005 pp. 54-64

Goorbergh van den R., Vlaar P. (1999) "Value-at-Risk Analysis of Stock Returns; Historical Simulation, Variance Techniques, Tail Index Estimation" *Netherlands Central Bank Staff Reports* 40:1-37

Hull J.C. (2006) *Options, Futures and Other Derivatives* 6th Edition New Jersey: Pearson Prentice Hall

Hull J.C., White, A. (1998) " Incorporating volatility updating into the historical simulation method for Value-at-Risk" *Journal of Risk* 1 (Fall) pp. 5-19

Jackson P., Maude D.J. and Perraudin W., "Bank Capital and Value at Risk", *The Journal of Derivatives*, Spring 1997 pp. 73-89

Jorion P. (2001) *Value at Risk - The New Benchmark for Managing Financial Risk* 2nd Edition New York: McGraw Hill

Kuester K., Mittnik S., Paolella M.S. (2006) "Value-at-Risk Prediction: A Comparison of Alternative Strategies" *Journal of Financial Econometrics* Vol. 4 No. 1 pp. 53-89

Mandelbrot B., (1963) "The Variation of Certain Speculative Prices" Journal of Business 36 pp. 394-419

Nordea Financial Statement 2008 www.nordea.com

Perignon C. and Smith D.R. (2008) "A New Approach to Comparing VaR Estimation Methods", *The Journal of Derivatives* Winter 2008 pp. 54-66

The Economist January 24th, Section: *The Future of Finance* pp. 1-24

Appendix

Appendix 1: Time Series of S&P500 and Trading Portfolio

Figure A.1.1 S&P 500 daily observations over the sample period 2001-05-04 – 2009-01-01

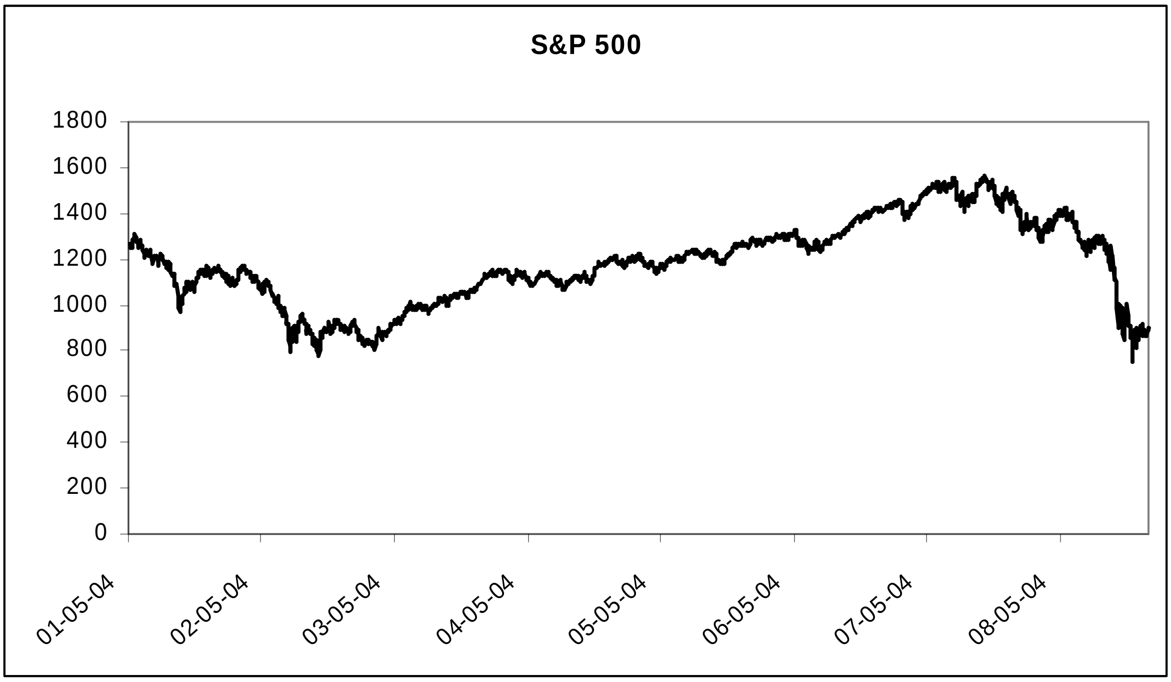


Figure A.1.2 S&P 500 daily returns over the sample period 2001-05-04 – 2009-01-01. Notice the tranquil period in the middle of the sample and the period with volatility clustering towards the end of the sample relating to the fall in S&P 500.

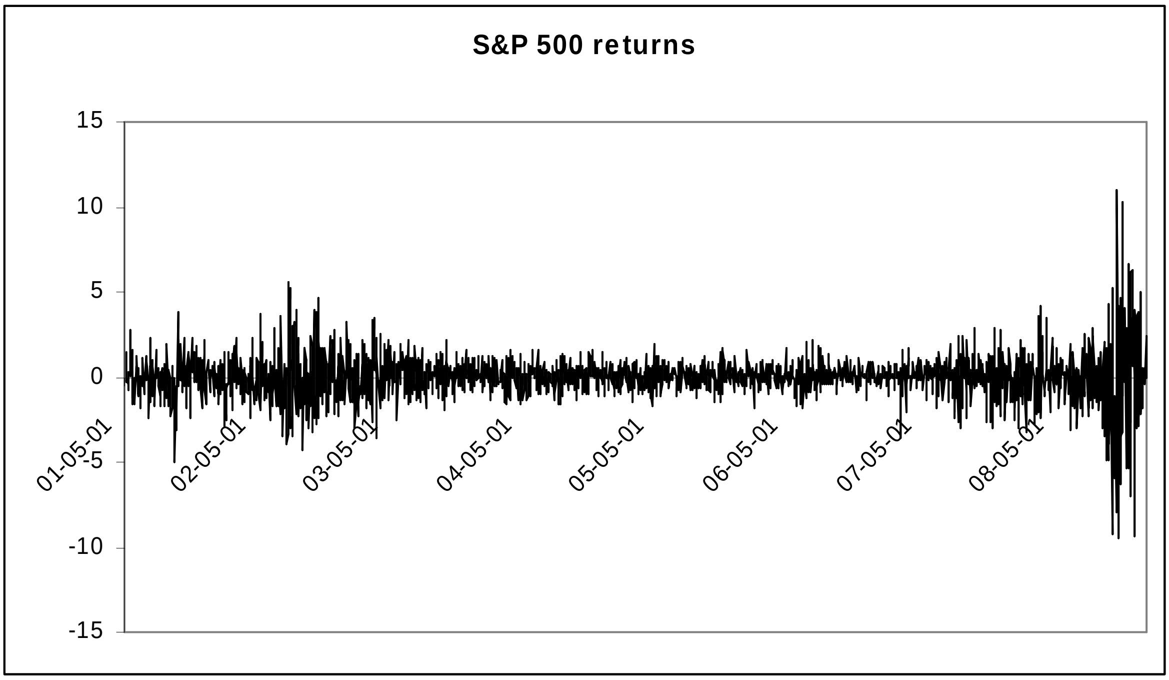
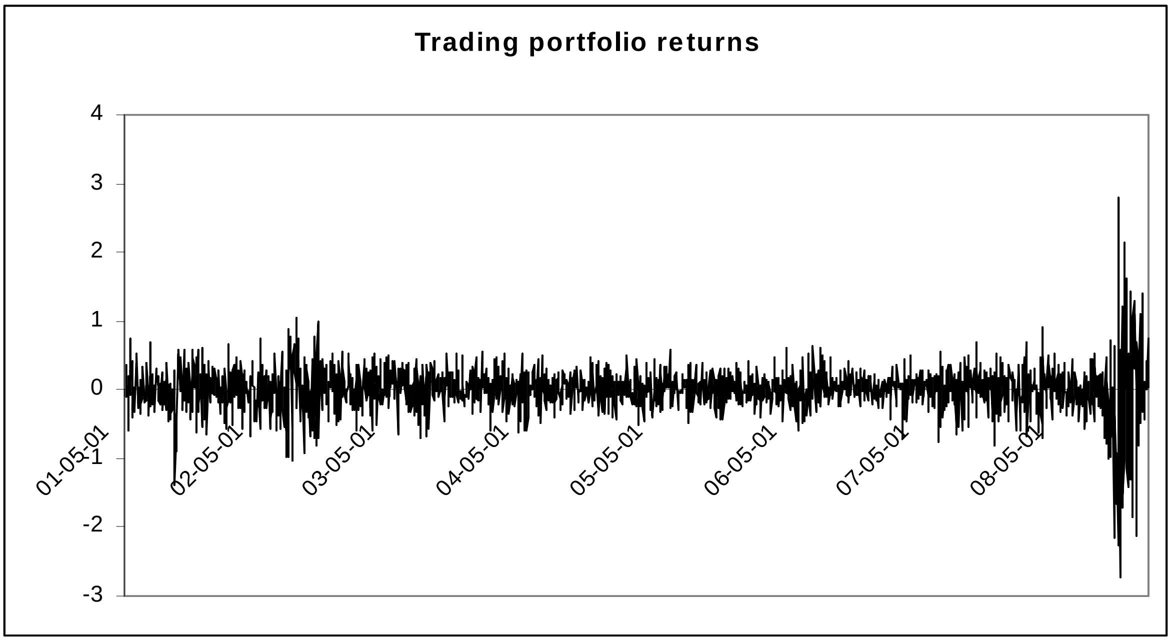


Figure A.1.3 Trading portfolio daily returns over the sample period 2001-05-04 – 2009-01-01. Notice the tranquil period in the middle of the sample and the period with volatility clustering towards the end of the sample relating to the turbulence in the financial markets.



Appendix 2: Christoffersen test

Table A.2.1 Critical values using the χ^2 distribution

DF	χ^2 Critical values	
	Critical value 5%	Critical value 1%
1	3,84	6,63
2	5,99	9,21

Table A.2.2 Null and alternative hypotheses for the three-step Christoffersen test

	LR_{UC}	LR_{IND}	LRcc
H0	Correct unconditional coverage	Exceedances are independent	Correct conditional coverage
H1	Incorrect unconditional coverage	Exceedances are not independent	Incorrect conditional coverage

Table A.2.3 Non-rejection regions for the different confidence levels and sample sizes used for the first part of the Christoffersen test.

Confidence level	Backtesting sample size	
	250	1000
5%	$7 \leq N \leq 19$	$38 \leq N \leq 64$
1%	$1 \leq N \leq 6$	$5 \leq N \leq 16$

Appendix 3: VaR over backtesting period, non-parametric and parametric methods

Figure A.3.1 The table shows VaR for Historical Simulation on both 99% and 95% confidence level with 1000 and 250 observations derived from SP 500 over the period 2005-03-04 to 2009-01-01

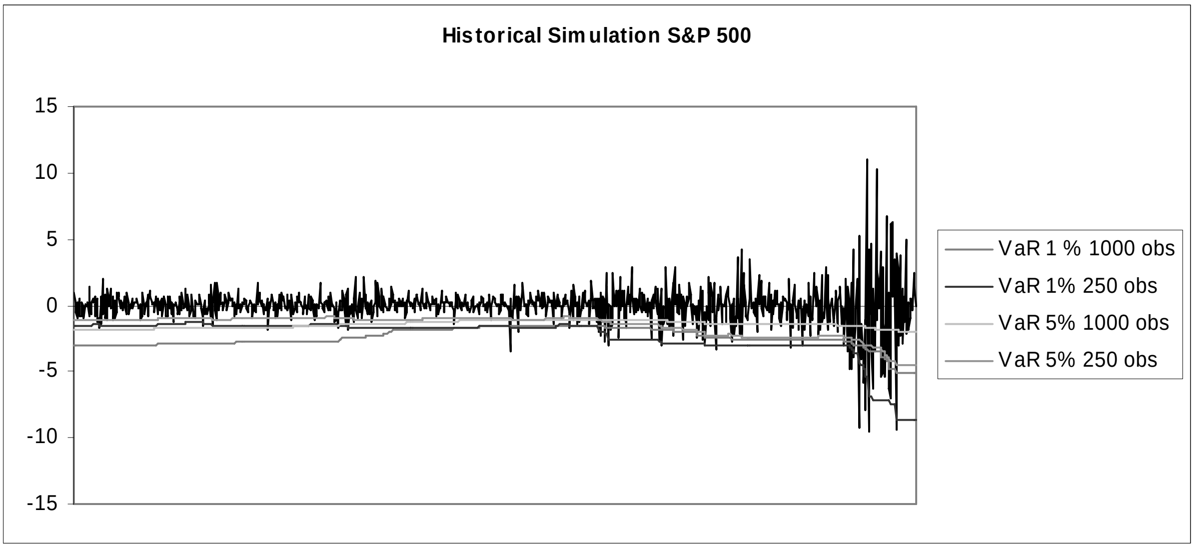


Figure A.3.2 Age Weighted Historical Simulation approach to estimating VaR over the period 2005-03-04 to 2009-01-01 for S&P 500

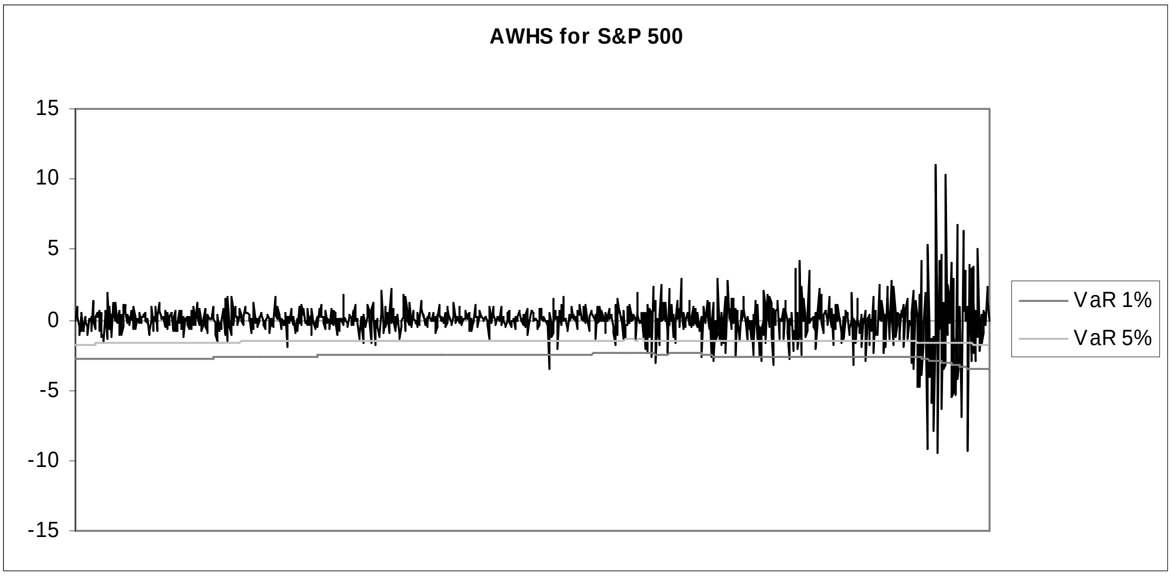


Figure A.3.3 Volatility Weighted Historical Simulation approach using EWMA to estimate VaR over the period 2005-03-04 to 2009-01-01 for S&P 500

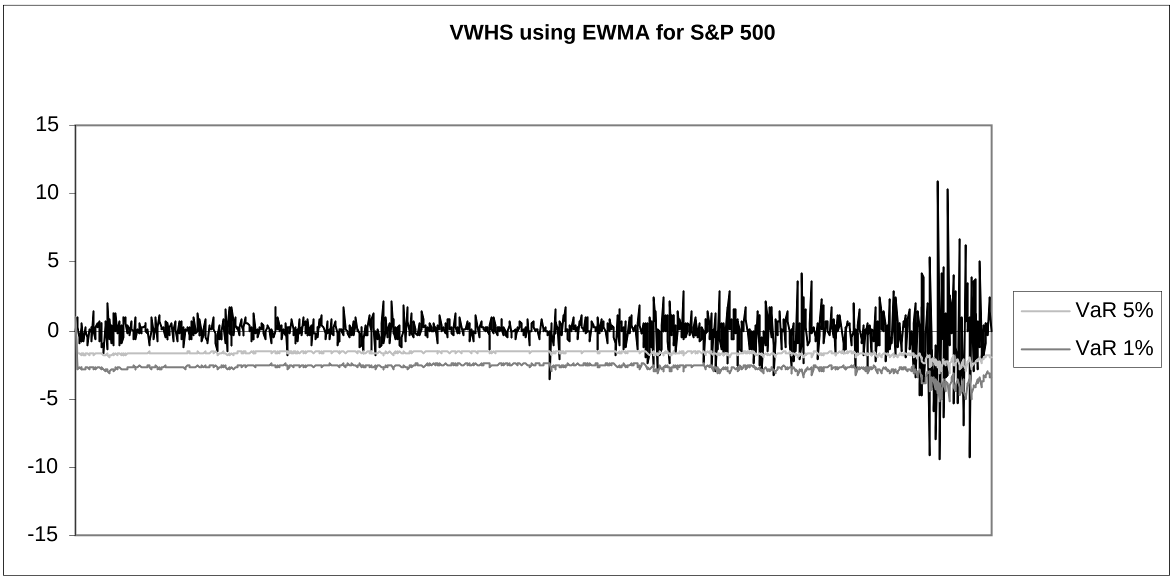


Figure A.3.4 Volatility Weighted Historical Simulation approach using GARCH(1,1) to estimate VaR over the period 2005-03-04 to 2009-01-01 for the S&P 500

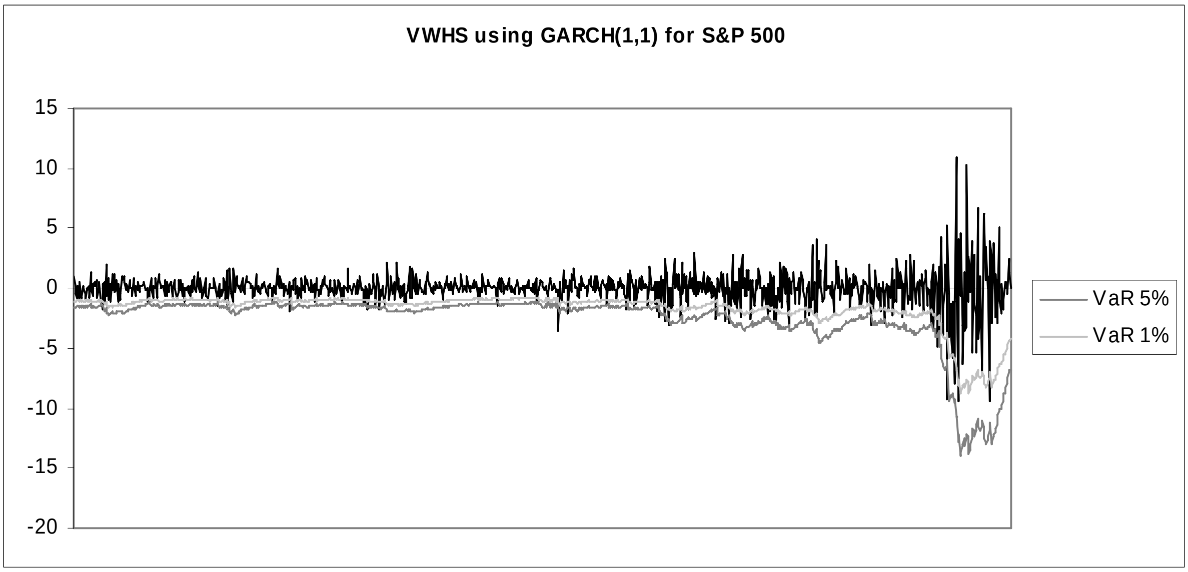


Figure A.3.5 The table presents VaR for Historical Simulation on both 99 % and 95 % confidence level with 1000 and 250 observations derived from our trading portfolio over the period 2005-03-04 to 2009-01-01

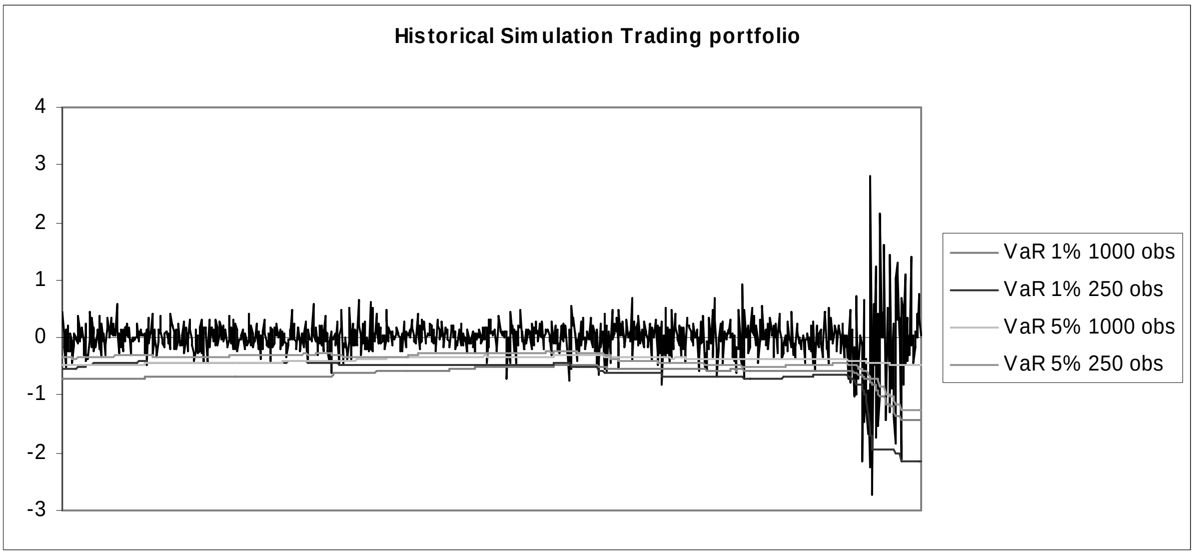


Figure A.3.6 Age Weighted Historical Simulation approach to estimating VaR over the period 2005-03-04 to 2009-01-01 for the trading portfolio

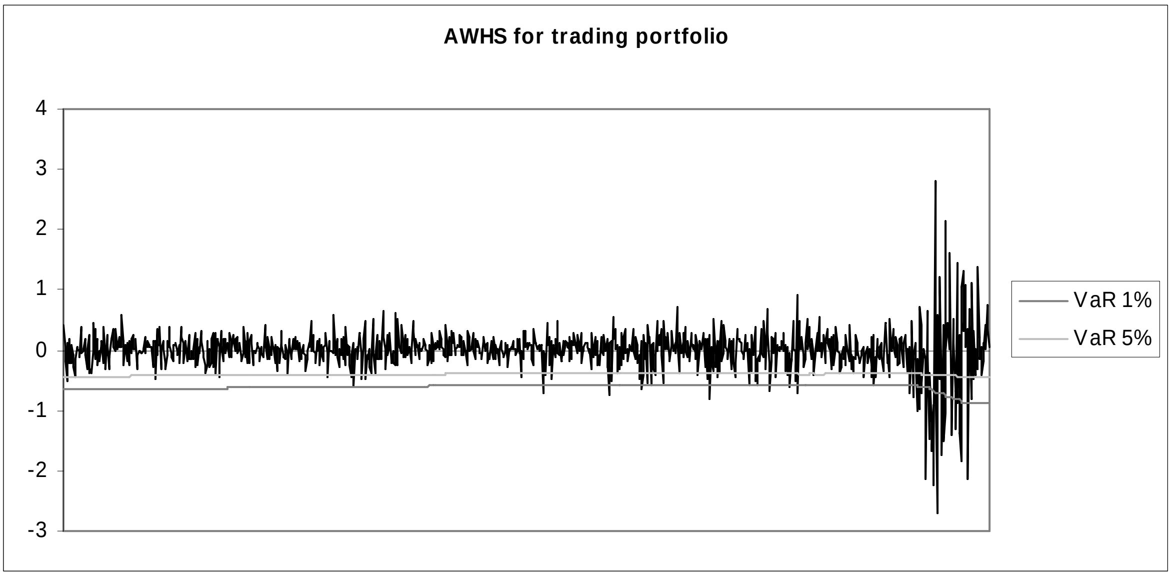


Figure A.3.7 Volatility Weighted Historical Simulation approach using EWMA to estimate VaR over the period 2005-03-04 to 2009-01-01 for the trading portfolio

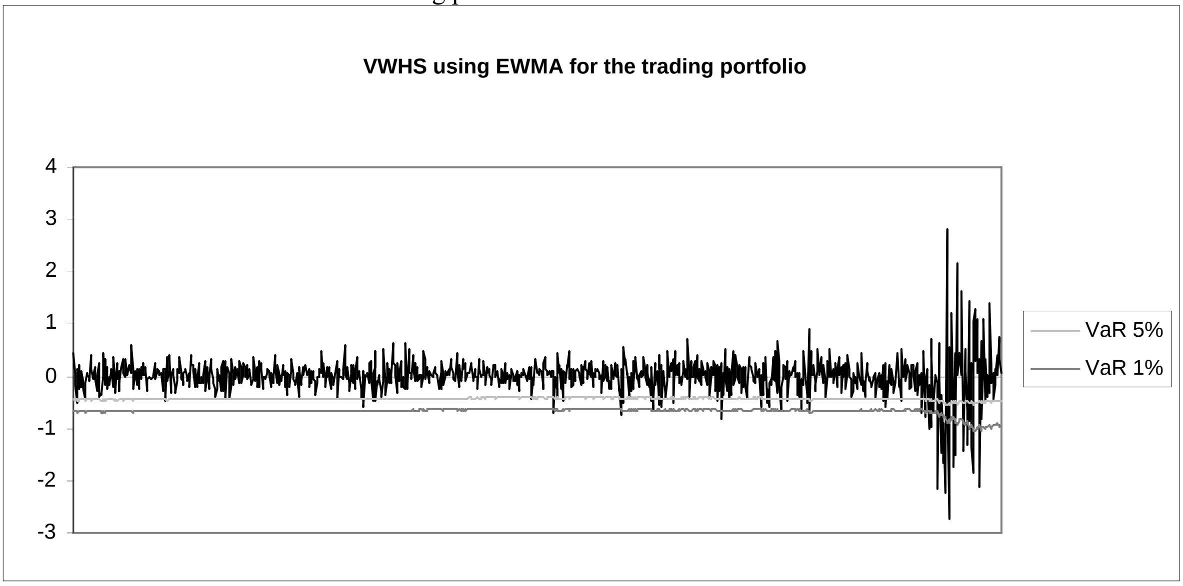


Figure A.3.8 Volatility Weighted Historical Simulation approach using GARCH(1,1) to estimate VaR over the period 2005-03-04 to 2009-01-01 for the trading portfolio

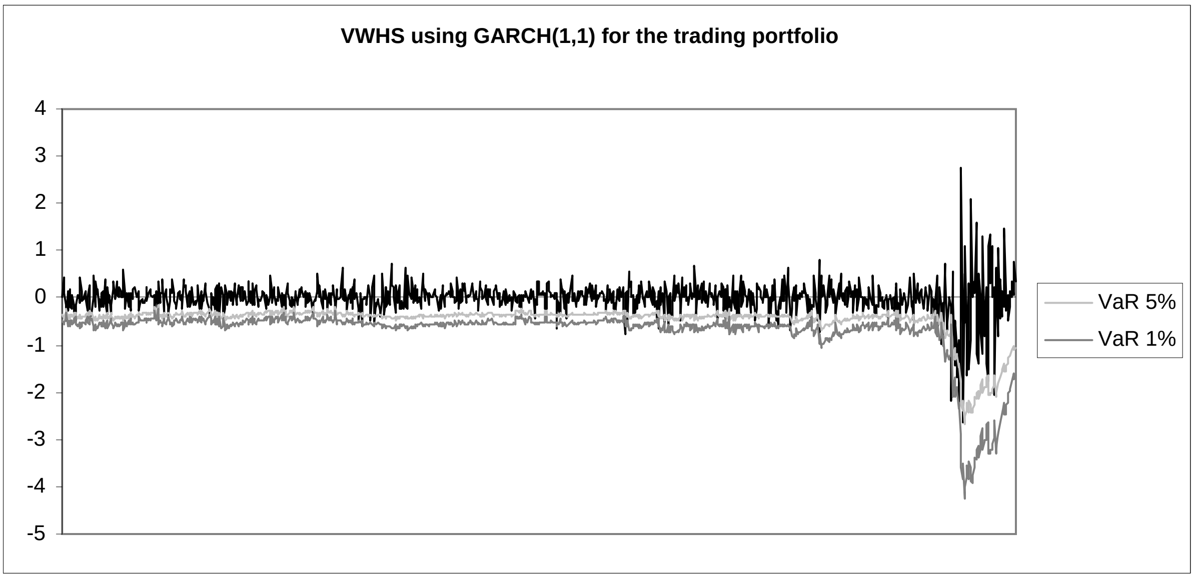


Figure A.3.9 Normal distribution approach to estimate VaR over the period 2005-03-04 to 2009-01-01 for the S&P 500

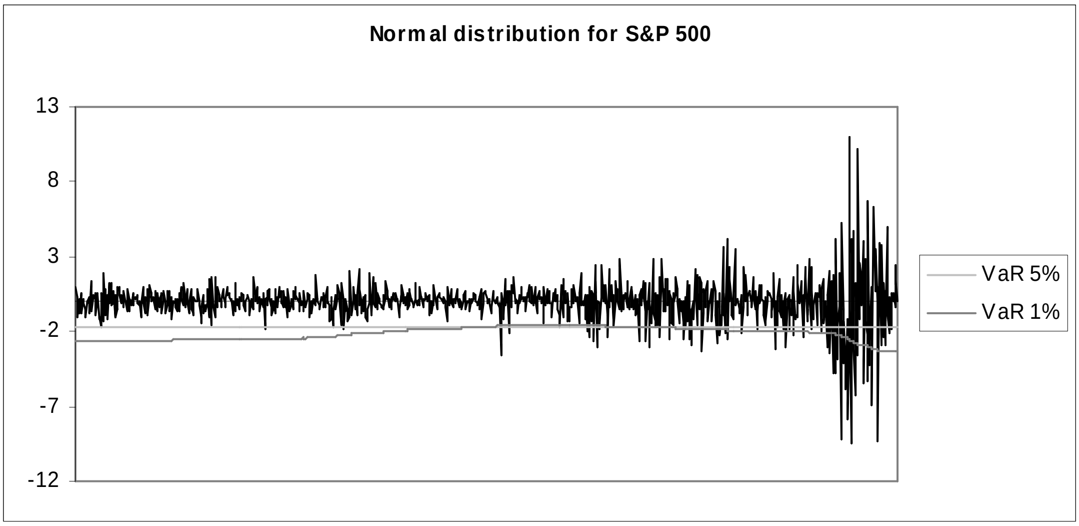


Figure A.3.10 t-distribution approach to estimate VaR over the period 2005-03-04 to 2009-01-01 for the S&P 500

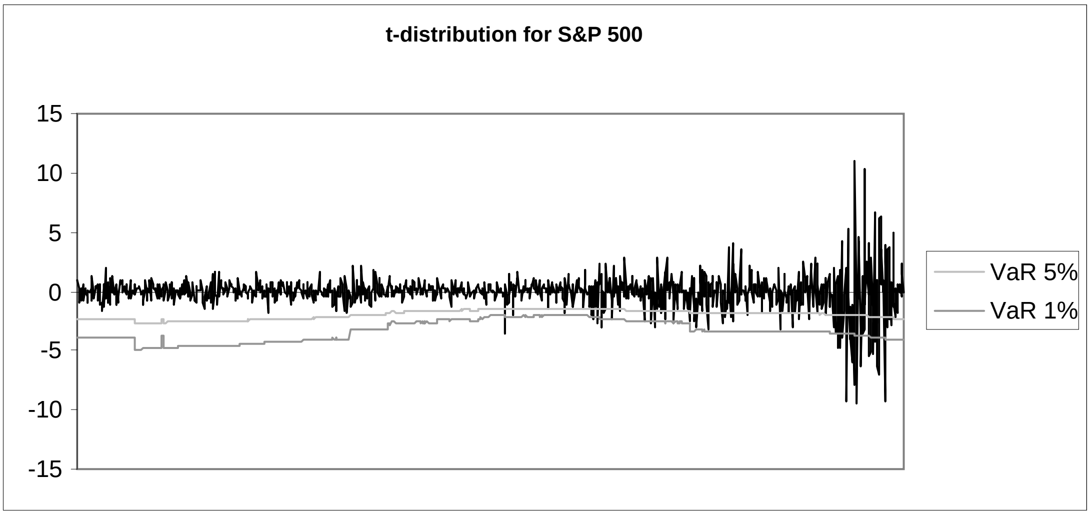


Figure A.3.11 RiskMetrics approach to estimate VaR over the period 2005-03-04 to 2009-01-01 for the S&P 500

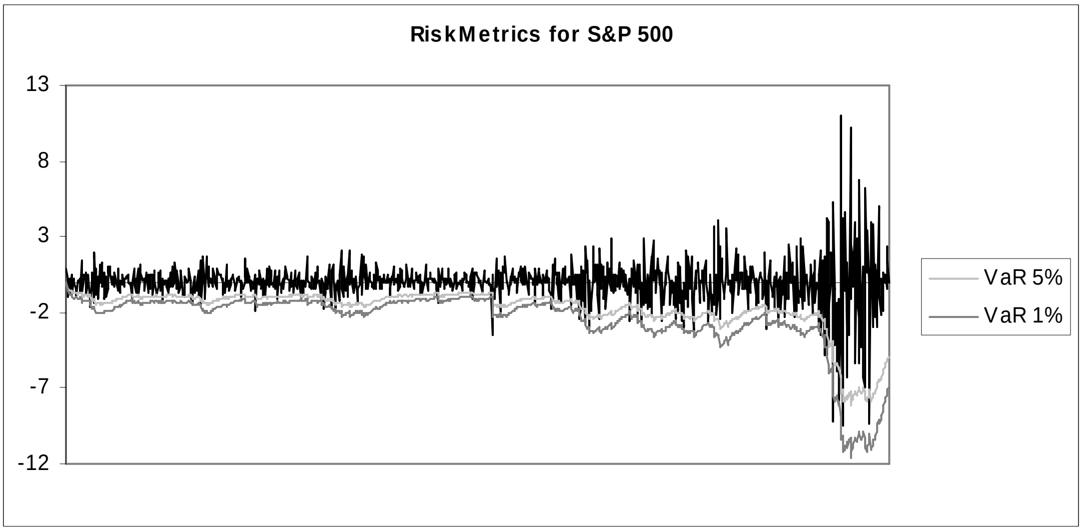


Figure A.3.12 Implied volatility approach to estimate VaR over the period 2005-03-04 to 2009-01-01 for S&P 500

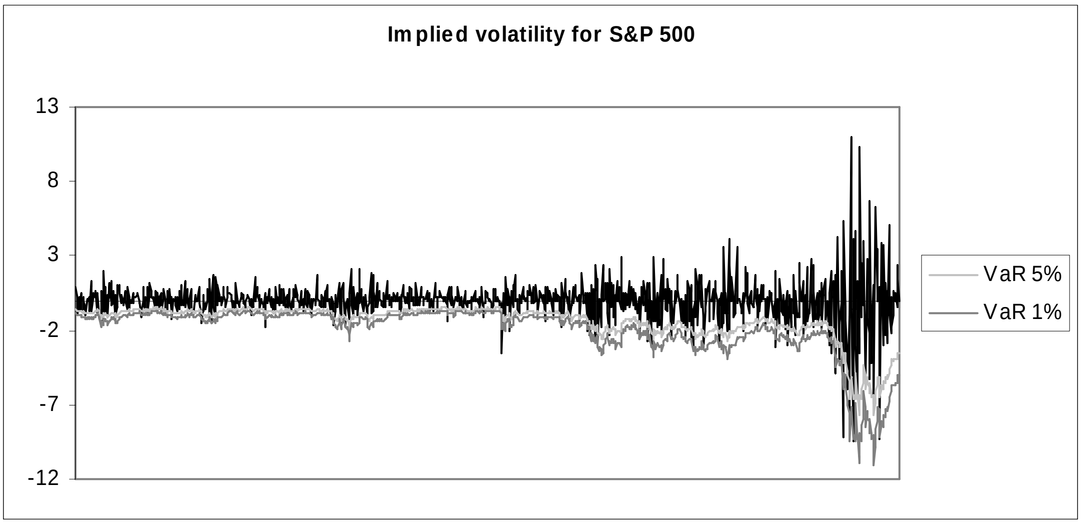


Figure A.3.13 GARCH (1,1) approach to estimate VaR over the period 2005-03-04 to 2009-01-01 for S&P 500

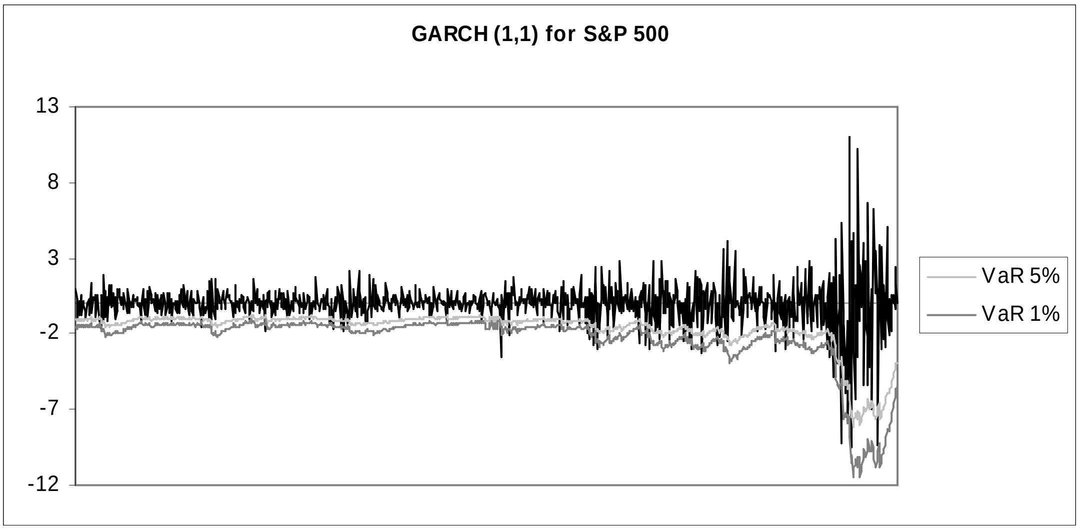


Figure A.3.14 GARCH (1,1)-t approach to estimate VaR over the period 2005-03-04 to 2009-01-01 for S&P 500

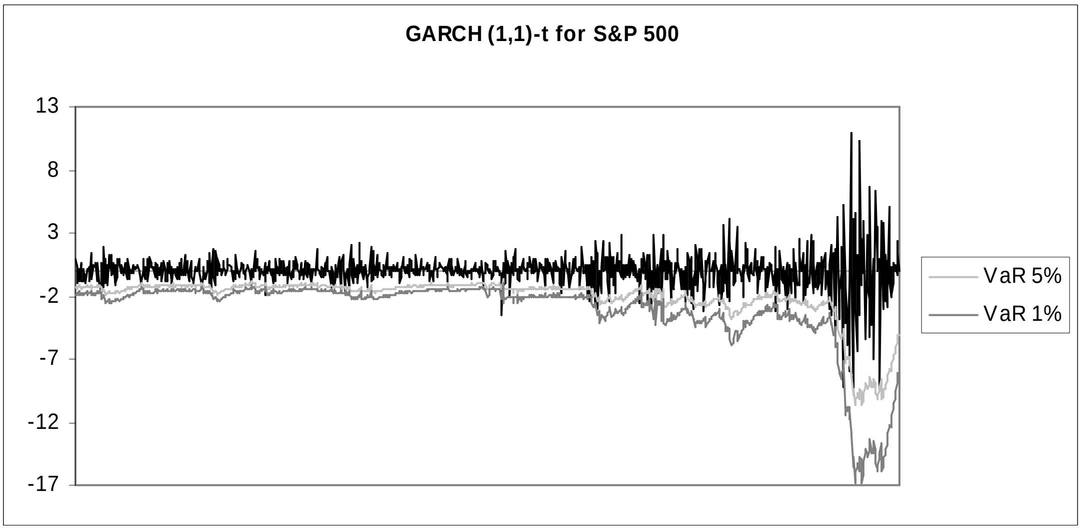


Figure A.3.15 Normal distribution approach to estimate VaR over the period 2005-03-04 to 2009-01-01 for the trading portfolio

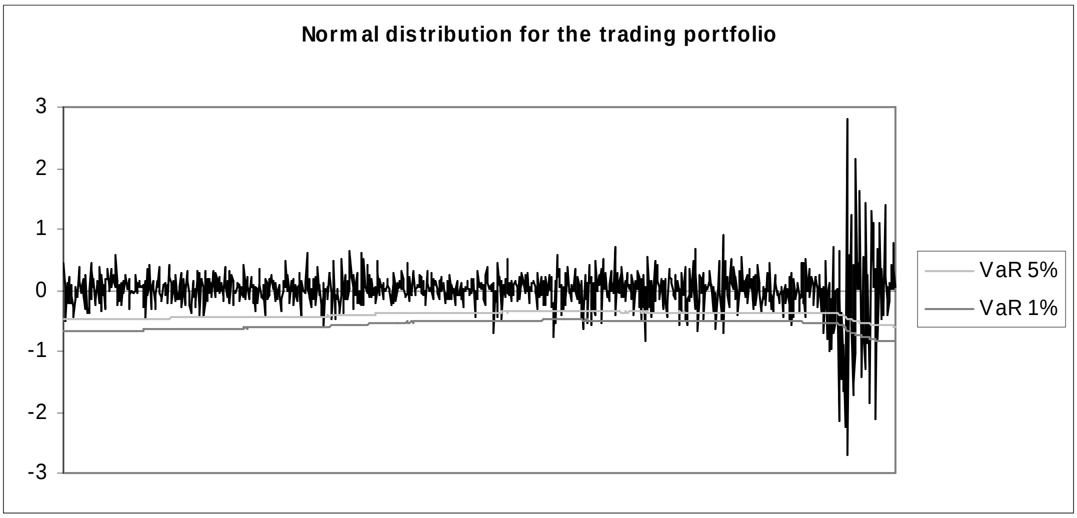


Figure A.3.16 t-distribution approach to estimate VaR over the period 2005-03-04 to 2009-01-01 for the trading portfolio

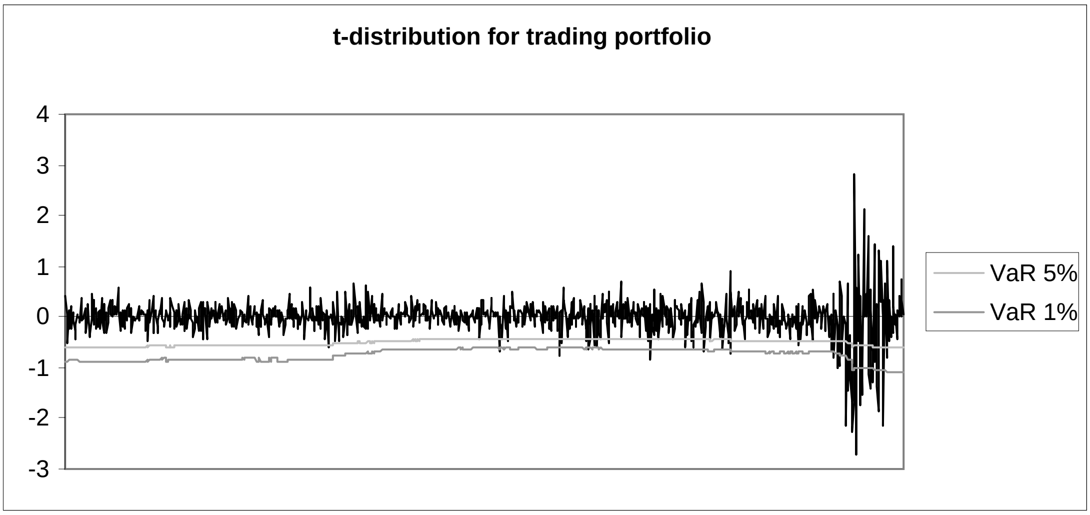


Figure A.3.17 RiskMetrics approach to estimate VaR the period 2005-03-04 to 2009-01-01 for the trading portfolio

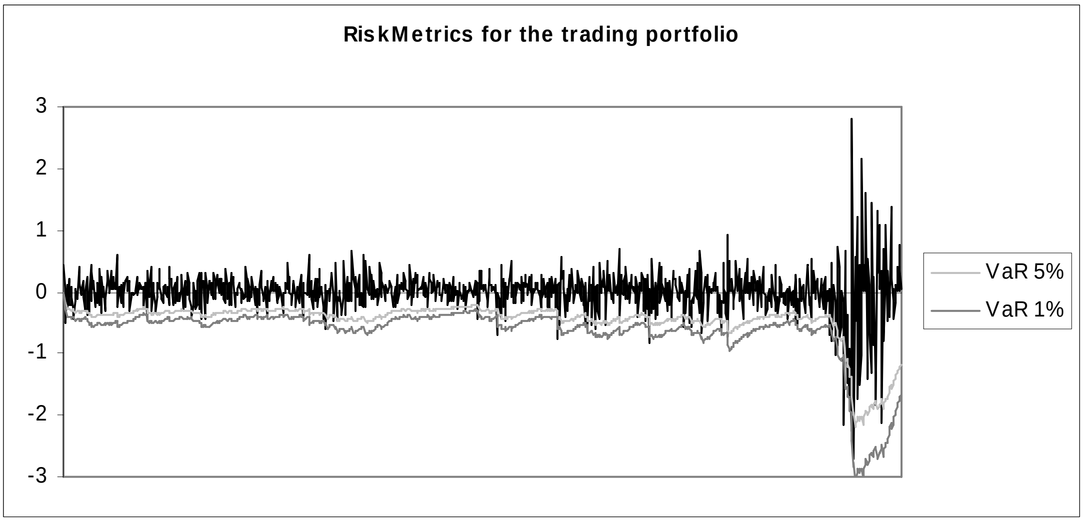


Figure A.3.18 Implied volatility approach to estimate VaR over the period 2005-03-04 to 2009-01-01 for the trading portfolio

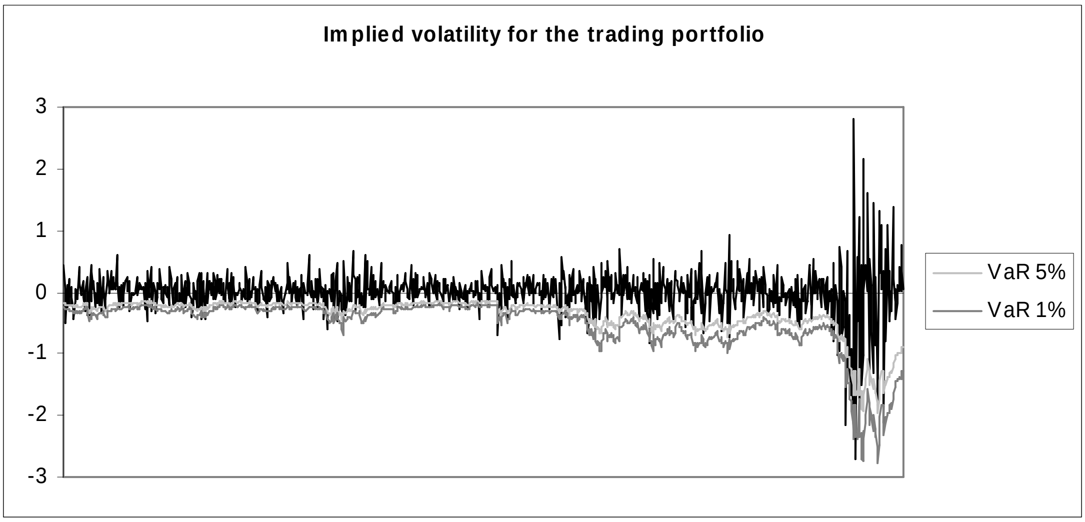


Figure A.3.19 GARCH(1,1) approach to estimate VaR over the period 2005-03-04 to 2009-01-01 for the trading portfolio

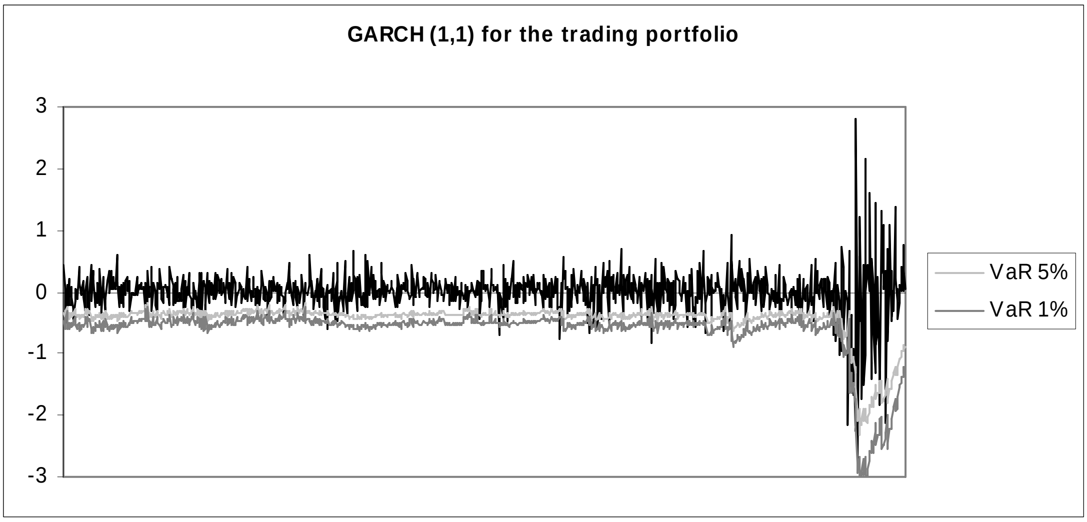
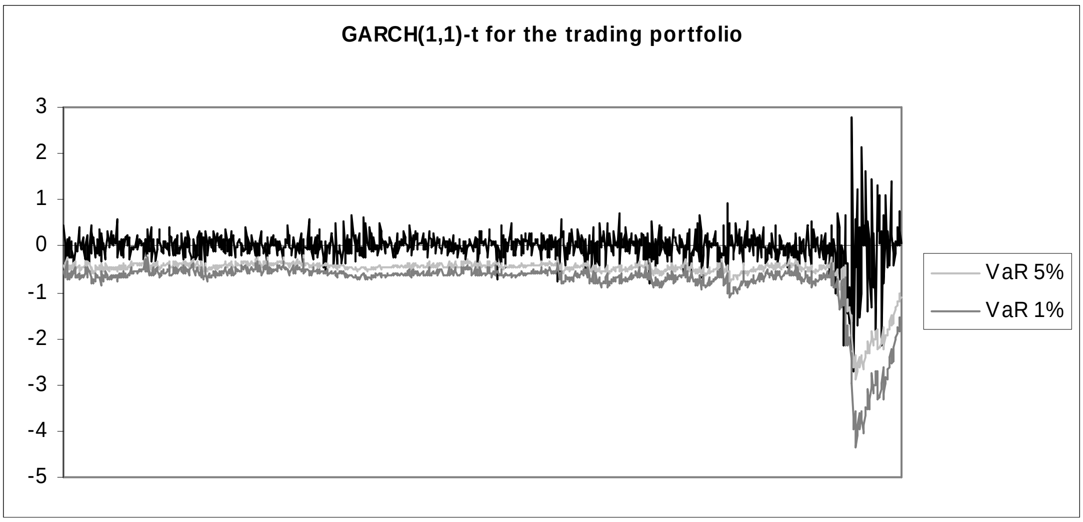


Figure A.3.20 GARCH(1,1)-t approach to estimate VaR over the period 2005-03-04 to 2009-01-01 for the trading portfolio



Appendix 4: Programming scripts

VBA macro for estimating Age Weighted Historical Simulation

```
Sub AWHS()
    Sheet9.Select
    Dim data(2000) As Variant
    Dim rescaling As Variant
    Dim weighta As Variant
    Dim weightb(2000) As Variant
    Dim i As Integer
    Dim j As Integer
    Dim percentilen As Double

    For i = 1 To 2000
        data(i) = Cells(i + 2, 2)
        'rstar(i) = Cells(i + 2, 5)
    Next i

    For i = 1001 To 2000
        ReDim returns(i - 1) As Variant
        weighta = (1 - 0.9999) / (1 - 0.9999 ^ (i - 1))
        For j = 1 To i - 1
            weightb(j) = weighta * 0.9999 ^ (i - j - 1)
            returns(j) = weightb(j) * data(j)
        Next j
        percentilen = Application.WorksheetFunction.Percentile(returns, 0.05)
        rescaling = ((1 - 0.9999 ^ (i - 1)) / (1 - 0.9999)) * percentilen
        Cells(i + 2, 3) = rescaling
    Next i

End Sub
```

VBA macro for Volatility Weighted Historical Simulation

```
Sub VWHS()
    Sheet7.Select
    Dim condvol(2000) As Variant
    Dim rstar(2000) As Variant
    Dim i As Integer
    Dim j As Integer
    Dim percentilen As Double

    For i = 1 To 2000
        condvol(i) = Cells(i + 2, 4)
        rstar(i) = Cells(i + 2, 5)
    Next i

    For i = 1001 To 2000
        ReDim rr(i - 1) As Variant
        For j = 1 To i - 1
            rr(j) = rstar(j) * condvol(i)
        Next j
        percentilen = Application.WorksheetFunction.Percentile(rr, 0.01)
        Cells(i + 2, 6) = percentilen
    Next i

End Sub
```

VBA macro for estimating degrees of freedom for the t-distribution

```
Sub solverDFnonstandtradingportfolio()
Sheet1.Select
Dim i As Integer
Dim j As Integer
Dim r1 As Integer
Dim r2 As Integer
Dim r3 As Integer

Application.ScreenUpdating = False
For i = 1904 To 2003
   For j = i - 1000 To i - 1
      r1 = i
      Cells(j, 24).FormulaR1C1 = "=-(1/2)*LN((PI()*(R" & CStr(r1) & "C23-2)*EXP(GAMMALN(R" & CStr(r1) & "C23/2))^2)/(EXP(GAMMALN((R" & CStr(r1) & "C23+1)/2))^2))-(1/2)*LN(R" & CStr(r1) & "C22^2)-(((R" & CStr(r1) & "C23+1)/2)*LN(1+((R" & CStr(j) & "C3-R" & CStr(r1) & "C21)^2/(R" & CStr(r1) & "C22^2*(R" & CStr(r1) & "C23-2)))))"
   Next j
   r2 = i - 1000
   r3 = i - 1
   Cells(i, 25).FormulaR1C1 = "=SUM(R" & CStr(r2) & "C24:R" & CStr(r3) & "C24)"
   SolverReset
   SolverOptions precision:=0.1, iterations:=32767
   SolverOk SetCell:=Range("Y" & i), MaxMinVal:=1, ValueOf:="0", ByChange:=Range("U" & i & ":W" & i)
   SolverAdd CellRef:=Cells(i, 21), Relation:=1, FormulaText:="100"
   SolverAdd CellRef:=Cells(i, 23), Relation:=3, FormulaText:="3"
   SolverSolve userFinish:=True
   SolverFinish keepFinal:=1
   Cells(i, 25).FormulaR1C1 = ""
Next i
Application.ScreenUpdating = True
End Sub
```

Programming script for estimating GARCH(1,1)-t in Eviews

```
smpl @all
scalar noobs=2000
scalar nw=1000
matrix(2000,5) results
for !i=nw to noobs
  smpl !i-nw+1 !i
  equation garch_n
  garch_n.ARCH(1,1, tdist) trading_portfolio c
  garch_n.makegarch sp500rgarch
  rowplace(results,@transpose(garch_n.@coefs),!i)
next
smpl @all
```

Programming script for forecasting GARCH(1,1) in Eviews

```
smpl @all

scalar noobs=2000
scalar nw=1000
matrix(2001,1) garchforecast
vector(1) forecastresult

for !i=nw to noobs
  smpl !i-nw+1 !i
  equation garch_n
  garch_n.arch(1,1) series02 c
  garch_n.makegarch garchcondvar

forecastresult(1)=garch_n.@coefs(2)+garch_n.@coefs(3)*resid(!i)+garch_n.@coefs(4)*garchcondvar(!i)

  rowplace(garchforecast,@transpose(forecastresult),!i+1)
next
```