UNIVERSITY OF PIRAEUS

DEPARTMENT OF BANKING & FINANCIAL MANAGEMENT

POSTGRADUATE PROGRAM IN BANKING & FINANCIAL MANAGEMENT

VAR FOR U.S. AND EUROPEAN STOCK INDICES

MIRJANA SAZDANOSKA

A.M. MXRH1501

Committee:

Professor Skiadopoulos George (Supervisor)

Professor Antzoulatos Angelos

Lecturer Botsari Antonia

January, 2017

Abstract

The objective of this study is to determine the best Value-at-Risk (VaR) model for the biggest stock exchange indexes in the U.S. and Europe. We use the Historic Simulation and Variance-Covariance approach with estimated volatility from Moving Average, Exponentially Weighted Moving Average, GARCH (1,1) with Normal Distribution, GARCH (1,1) with Student-t Distribution, EGARCH (1,1) with Normal Distribution and EGARCH (1,1) with Student-t Distribution. We use these methods in order to obtain the VaR forecasts for the period 01/01/2007 to 26/09/2016 for the following indices: S&P500, NASDAQ, EUROSTOXX, FTSE100, DAX, CAC and ATHEX. For the backtesting we apply the three tests proposed by Christoffersen (2012). Our results show that the most accurate results are achieved when using the Student t distribution together with a volatility forecasting model that takes into account the leverage effect, as the EGARCH (1,1).

Key words: Value-at-Risk, Historical Simulation, Variance-Covariance, EWMA, GARCH, EGARCH, Backtesting

Table of Contents

List of figures:

List of Tables:

1. Introduction

Over the past years, there has been a significant change in the way that financial institutions assess their risks. Risk is the likelihood of losses resulting from events such as changes in market prices. In current times of financial uncertainty, where every day there are new financial products, more complex than ever, the need for measuring and controlling risk is essential (Crouhy et al., 2006). Moreover, because of the increased volatility in markets around the globe, and the financial crisis in 2008, the focus has been on market risk analysis and measurement. The crisis has shown that the risk assessment that was put in use by various financial institutions failed and that risk was essentially underestimated.

The Bank for International Settlement (BIS) defines market risk as “the risk of losses in on and off – balance sheet positions arising from movements in market prices” (BIS, 1996). It includes interest rate risk, equity risk, foreign exchange risk, commodity risk etc. The main source of market risk in stock indices is equity risk since they are linear assets and the only risk factor is the price of the indices. In other words, the market risk is the risk that the value of the security or the portfolio will decline.

There are many different methodologies for risk management today that were created several decades ago. Since then they have evolved and adjusted accordingly for different securities.

The first risk management regulation was provided by the Basel Committee. The Basel Committee is an international committee of banking supervisory authorities. The main issue that this committee deals with is capital adequacy. The first important agreement of the committee was the 1988 Basel accord which set the principles for capital requirement in all international banks. Within the next years, a standard measuring method of market risk was added to these principles. This method was used to compute the potential loss that a portfolio would suffer from the changes of each risk factor. In 1995, the Basel Committee replaced the standardized method with the internal model approach (BIS 1995). This approach gave the financial institutions the opportunity to develop their own measuring method of market risk and capital requirement, adjusting it to their needs. However, in order to use this kind of approach, the institutions had to be regularly reviewed and regulated by the management and other external officials. These principles were not exclusively for the financial institutions, with the result of expanding to other companies as well.

These developments, facilitated by the growing technological resources, brought on some more sophisticated risk management systems, one of which is the Value at Risk (VaR).

According to Christoffersen (2012), VaR is a simple risk measure that answers the following question: What loss is such that it will only be exceeded $p \times 100\%$ of the time in the next K trading days?

One incentive for the development of VaR is the mark to market approach in the evaluation of profits and losses. The daily revaluation of a position that the company holds has also brought the need for a more frequent estimation of the market risk on this position. (Gallati, 2003)

The popularity of VaR skyrocketed after the development of a VaR system by JP Morgan. In 1994, they published a technical document that provided a detailed description of RiskMetrics, a set of techniques and data to measure market risks in portfolios of fixed income instruments, equities, foreign exchange, commodities, and their derivatives issued in over 30 countries. They presented the framework and how to calculate VaR, the necessary statistical measures needed, existing models for calculation, data structure and how to model the exposures to different risk factors. An important advantage of the RiskMetrics methodology is the presentation of the measurement of the market risk as a single number which has been very useful and easy to understand by investors. Since then, the Basel II accord has chosen it as a recommended measure for market risk.

Despite the popularity that has gained among professionals in the risk management area, VaR has faced a lot of criticism for its lack of forecasting accuracy. (Danielsson et al., 2001 Zikovic and Filer, 2009). The main reasons behind this are two: firstly, in the computing of VaR, one needs to use the historic data of the security under question, and historic data does not always correspond to the present data. Another reason why VaR does not always give the most reliable results is that there are many choices involved in the process of calculating it. For example, one has to decide on confidence interval, estimate horizon, forecast horizon etc., making VaR subjective to some extent. As a result, the backtesting was developed with the purpose of verifying the results from the VaR calculations.

The threats of market risk became primary concern even more so with the financial crisis in 2008. Following the failure of many financial institutions, such as Lehman Brothers, one of the largest investment bank in the US, the measurement of market risk was once again put to the test. Now more than ever, methods such as VaR have to prove their value so that excess risk taken by individuals can be stopped and another crisis prevented from happening. These are the reasons behind the continuing research in the hope of finding more efficient tools and less disruption in the global economy and consequently in people's lives.

The purpose of this thesis is to find the Value-at-Risk (VaR) model for different U.S. and European stock indices that can forecast the market risk most accurately. We compare the performance of these models and evaluate which of them is the best fit for each index. We use different estimation windows and confidence intervals for a one-day forecast horizon VaR.

The U.S. stock indices taken under consideration are S&P 500 and Nasdaq Composite. The European stock indices are Euro Stoxx 50, FTSE 100, DAX 30, CAC 40 and Athens Stock Exchange (ATHEX) Composite Index. For each of these we compute the daily logarithmic returns and apply the VaR methods.

A total of five models are tested, one is parametric with 4 different ways of estimated volatility and one non-parametric. The non-parametric method is the Historical Simulation (HS). The parametric model used is the Variance/Covariance matrix. In order to compute the Variance/Covariance Matrix we need to estimate the volatility, and do this by using the Simple Moving Average and Exponentially Moving Average as well as GARCH and EGARCH (using both normal and Student's t-distribution). Finally, we use the three criteria suggested by Christoffersen in order to backtest VaR statistically and find the best model for each stock index.

The rest of the thesis is structured as follows: Section 1 gives the background of VaR as a risk measure. Section 2 gives an overview of the existing papers that are published in this area in order to differentiate this thesis from the rest of the studies. Section 3, 4 and 5 introduce the theoretical framework, methodology and empirical tests for the models used in this thesis. Section 6 describes the selection of the data sample used in this thesis. Section 7 presents the empirical findings and their interpretation. Section 8 concludes.

2. Literature Review

There has been an extensive research in the field of VaR. We chose to review the papers that are focused in the implementation of VaR methods to the specific stock indices that we included in this thesis.

Danielsson (2002) used 4 estimation methods: GARCH with normal distribution, GARCH with Student's t-distribution, Historical simulation and Extreme value theory. The data used was a representative foreign exchange, commodity, and equity datasets containing daily observations from the first recorded observation until the end of 1999 (S&P 500 Index, Hang Seng Index, Microsoft Stock Prices, Amazon Stock Prices, Ringgit pound exchange rates, Pound dollar exchange rates, Clean US government bond price index, Gold Prices and Oil prices). He estimated each model and dataset with a moving 300, 1.000, and 2.000 day estimation window, and forecasted risk one day ahead.

For the stock index that Danielsson used (S&P 500 Index), and for 300 and 1.000 days estimation window, GARCH with normal distribution showed the biggest underestimation of risk, whereas for the 2.000 day estimation window, the Historical Simulation and the Extreme Value Theory showed the highest underestimation. On the contrary, for 300 days estimation window, Historical Simulation and the Extreme Value Theory showed the highest overestimation of risk, whereas for the 1.000 and 2.000 day estimation window, the GARCH with Student's t-distribution showed the highest overestimation. The closest to the expected value was the Extreme Value Theory for 1.000 day estimation window (having a value of 0.99 when the expected value is 1). All the other models either underestimated the risk (causing exposure to the risk), or overestimated the risk (causing the allocation of capital where it is not necessary). All of these reported results correspond to the 99% confidence level.

He also showed that longer horizons provide us with more accurate forecasts, as well as lower risk volatility and therefore should be preferred when computing VaR. The reason for this could be the long cycles in volatility.

Using the same index (S&P 500 Index), 20 Year U.S. Treasury Bond Yield, 3 Month U.S. Treasury Bill Yield and Deutschemark Exchange Rate (DM per $) Figlewski (1997) compared the historical volatility (with moving average) and GARCH (1, 1) for forecasting volatility. He used monthly and daily prices for the stock index S&P500 for the period July 02, 1962 to December 29, 1995. The mean returns had an assumed value of zero. For the daily prices, the estimation horizon used was 1, 3, 12, 24 and 60 months (where one month was assumed to be 20 days), and the forecasting horizon 1, 3, 6, 12, and 24 months. The results showed that the size of the estimation horizon should be adequate to the forecasting horizon. The best volatility forecast for 1 month forecasting horizon is achieved with 1 month estimation horizon, for a 3 month horizon 12 months of historical sample give the best results and for 12 and 24 forecast horizon, the last 60 months of data have to be used for the best result.

The GARCH (1, 1) on the other hand is estimated using 5 years of historical data. As a result, it performs much better than the historic forecasts for every forecast period. The RMSE is smaller than the historic forecasts for every horizon. This result is exclusively for the equity market data, since GARCH (1, 1) did not have the best forecasts for the other type of markets (3 Month Treasury Bill Yield, 10 Year Treasury Bond Yield and Deutschemark Exchange Rate). This model also does not perform well with the monthly prices since it is very hard for the parameters to be estimated using data with such small frequency.

Alexander and Leigh (1997) used five major equity indices (DEM, FRF, GBP, JPY, USD) with data from 01/01/1993 to 06/10/1996. They analyzed the accuracy of different models such as moving average, exponentially weighted moving average (EWMA), and GARCH for forecasting volatility. Statistical and operational evaluation of these models showed that the moving average and the GARCH gave the best results for all the equity indices except for the US index. For this index the results were inversed, the best predictive performance was the one of the EWMA.

Angelidis et al. (2004) used five index portfolios: S&P 500, Nikkei 225, FTSE 100, CAC 40 and DAX 30 and implemented a number of volatility models with Normal, Student's t and Generalized Error Distribution. They estimated the models with four sample sizes: 500, 1000, 1500 and 2000 observations. The general conclusion is that the Student's t-distribution is more accurate than the other two distributions and that for every index there is a different type of mean model, volatility model and sample size that works best. There is not a specific rule that could be applied to all the indices and that would result in the best forecast.

Using the NASDAQ Composite Index, Kuester, Mittnik and Paolella (2006) examine different models, including existing models as well as some extensions. The data that they used comprised of daily closing prices of the index from 08/02/1971 (the date that the index was launched) to 22/07/2001. Their in-sample data was 1000 observations (approximately 4 years) and 6681 out of sample VaR forecasts. Serving as a benchmark was the three-zone framework suggested by the Basle Committee (1996) in which a VaR model is considered to be acceptable (green zone) if the number of violations of 1% VaR remains below the binomial (0.01) 95% quantile. A model is disputable (yellow zone) up to the 99.99% quantile and is flawed (red zone) whenever more violations occur. The historical simulation was rated as disputable (yellow category). However, using the GARCH volatility estimation, the results were improved, especially under the assumption of a Student-t distribution.

Angelidis and Benos (2004) implemented several volatility models (parametric and non-parametric) in order to estimate the 97.5% and 99% one-day VaR for long and short trading positions in the Greek Stock Market. They used daily prices from four equities (Alpha Bank, Commercial Bank, National Bank, Titan) and the ATHEX Composite Index for the period of January 2, 1991 to December 18, 2003. The models that they used were: parametric (Variance Covariance, RiskMetrics, GARCH, EGARCH, TARCH under Normal, Student-t and Skewed Student-t distribution), nonparametric (Historical), semi-parametric (Filtered Historical Simulation) and

Extreme Value Theory. The results show that for the 97.5% confidence level most of the models indicate good outcomes, with GARCH (1, 1), EGARCH (1, 1) and TARCH (1, 1) performing better for the long positions. For the 99% confidence level, the Filtered Historical Simulation method showed the best results and Extreme Value Theory acceptable results.

Another study from the Greek stock market is the one by Lambadiaris, Papadopoulou, Skiadopoulos and Zoulis (2003). They assessed the performance of historical and Monte Carlo simulation in calculating VAR, by using data from the Greek stock and bond market for the period July 17, 2000 to July 18, 2002. For both approaches, the one-day VAR was calculated for two separate portfolios: one consisting of stocks and one consisting of bonds. VAR was calculated separately for a 99% and 95% confidence level. The historical simulation VAR was calculated by using one-day rolling window of 100 and 252 observations (HS-100 and HS-252). In the stock portfolio case, the stock is assumed to follow a geometric Brownian motion with drift equal to zero. The volatility was estimated by using the moving average (MA-MC), the exponentially weighted moving average (EWMA-MC) and diagonal BEKK volatility estimators (BEKK-MC).

The results from the stock portfolio analysis for the HS and MC simulations show that for the 99% confidence level, all of the methods are acceptable and can be included in the ‘green’ BIS zone. For the 95% confidence level, the results are slightly different from the 99%. The expected exceptions (number of violations) are again not exceeded, so the methods are once again acceptable. However, the HS-252 was the most conservative and as a result the use of MC simulation as opposed to the HS is recommended for the computing of VAR for a stock portfolio.

Danielsson et al. (2001) argue that the regulations that were proposed by the Basel Committee in the Basel II “has failed to address many of the key deficiencies of the global financial regulatory system and even created the potential for new sources of instability”. Specifically, they claim that having a unique measure of risk such as VaR can disrupt the economic and financial system and make it less rather than more stable.

Similar to the opinion of Danielsson are the findings by Zikovic and Filer (2009). They investigate the relative performance of VaR and ES models using daily returns for sixteen stock market indices (eight from developed and eight from emerging markets) prior to and during the 2008 financial crisis. They introduced a new hybrid approach to joint estimation of Value at Risk (VaR) and Expected Shortfall (ES) for high quantiles of return distributions. The results from this research showed that only their proposed new hybrid and Extreme Value (EV) based VaR models provide adequate protection in both developed and emerging markets. The other models did not provide the anticipated results by way of underestimating the risk. The parametric models: normal simple moving average, RiskMetrics and GARCH provided the worst results for both developed and emerging markets. The authors believe that the wide use of these very popular models was the reason for the underestimation of risk in the financial institutions in the recent crisis.

From this overview of the existing literature, we can conclude that there is not a unique VaR model that would be ideal for different kind of data. The results rely on a very subjective choice of the parameters such as confidence level, forecast horizon, data, method, assumptions about the distribution etc. (Beder, 1995).

The more advanced models of the VaR are performing better than the original models, however the best model has not yet been found. A general conclusion is that the parametric methods work better than the non-parametric ones. This is especially true for the GARCH (1,1) models that address the heteroscedasticity problem, even more when a Student's t-distribution is assumed, given the fat-tail property of stock indices.

The purpose of this thesis is twofold. First, we implement several VaR models in order to estimate a one-day VaR forecast for 95% and 99% confidence level. Then, we evaluate these model as to their forecast accuracy using the three backtesting criteria (unconditional, independent and conditional coverage tests). The difference of this theses from the rest of the existing literature is that we use data from stock indices in US as well as Europe and compare the two categories.

The empirical research that we conduct has shown that the normal distribution tends to underestimate the risk, while the Student's t-distribution captures more accurately the extreme returns, resulting in more accurate results. This is emphasized even more when the leverage effect is taken into account. Because of this the EGARCH (1,1) yields the best results for most of the indices.

3. Theoretical Framework

Market risk is defined as the risk to a financial portfolio due to changes in market prices such as equity prices, foreign exchange rates, interest rates etc.

Value-at-Risk (VaR) is a simple risk measure that answers the following question: What loss is such that it will only be exceeded $p \times 100\%$ of the time in the next K trading days? (Christoffersen, 2012)

Mathematically:

$$Pr(\$Loss > -\$VaR) = p, \quad (1)$$

where p is the probability that the loss incurred will be larger than the VaR. p is usually $p = 5\%$ or $p = 1\%$ and K is one or ten trading days. These are the basic parameters of VaR.

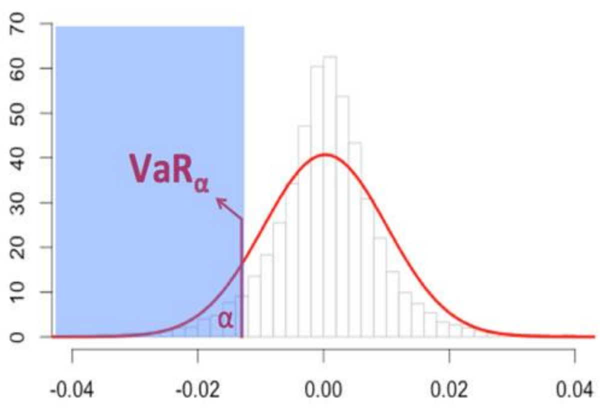
The BIS has selected p to be 1% and K to be 10 days for the purpose of measuring the adequacy of capital in banks, but many companies choose a one day VaR with probability of 5% or 1%, for internal purposes (Duffie and Pan, 1997). The choice of the confidence level depends on the risk attitude of the user, the more risk averse he is, the lower the probability of loss p (Alexander, 2008).

Since we are going to be using logarithmic daily returns,

$$Pr\ (R < -VaR) = p \quad (2)$$

In this case p is the probability that we will have a lower return than –VaR.

Figure 1 VaR for a confidence level a and normal distribution



The Basel's Committee recommends this measure for determining the capital adequacy of the banks. More specifically: "banks must hold capital equal to the potential loss on the institution's equity holdings as derived using internal value-at-risk models subject to the 99th percentile, one-tailed confidence interval of the difference between quarterly returns and an appropriate risk-free rate computed over a long-term sample period". As far as the method that should be used there is no specific recommendation, every financial institution can use its internal method, but backtesting is required, as well as regulatory audit.

The fact that Basil's Committee recommends the use of the 99^{th} percentiles means that the confidence level should be set at 99% or that the loss would exceed the VaR in only 1% of the cases.

Value-at-Risk is a number that represents the potential change in a portfolio's future value. How this change is defined depends on the horizon over which the portfolio's change in value is measured and the "degree of confidence" chosen by the risk manager (RiskMetrics, 1996).

3.1. Advantages and disadvantages of VaR

There are many advantages to VaR, making it one of the most popular risk measure used worldwide. One main advantage is that it can aggregate different types of risk in one single measure. As we have stated earlier, this is especially convenient for investors, as it makes it easier to understand and interpret. Another important advantage of VaR is that it is comparable. We can use it to compare the market risk for different types of securities or portfolios.

However, there are a few shortcomings to the VaR as a risk measure. Some of these are examined in detail by the existing literature. Such disadvantages are:

VaR only gives a forecast to a specific point in the distribution. If a loss is realized and it is beyond what the VaR forecasted, we have no information about how big this loss can be. Hence, VaR does not capture the extra risk in the tail of the distribution and does not provide us with any clue as to what to expect in this case (Taylor, 2008).

Another weakness of the VaR is that it is not coherent. This is shown by the non-additivity property of VaR, i.e. the VaR of a portfolio is greater than the sum of VaR of its components. This is a very important property for a risk measure because if VaR is non-additive, it could mean that a well-diversified portfolio shows signs of higher risk, which we know is not the case.

These disadvantages are presented by Artzner et al. (1999). They showed that VaR failed to satisfy the subadditivity and convexity property and is therefore a non-coherent risk measure.

According to them, a risk measure $\rho(\cdot)$ is coherent if it satisfies the four axioms:

1. Translation invariance: $\rho(X + a * r) = \rho(X) - a$, for all real numbers a.
2. Subadditivity: $\rho(X_1 + X_2) \leq \rho(X_1) + \rho(X_2)$,
3. Homogeneity: for all $\lambda \geq 0$, $\rho(\lambda X) = \lambda\rho(X)$,
4. Monotonicity: $X \leq Y$, we have $\rho(Y) \leq \rho(X)$,

Another concern with the VaR is that: if a loss is bigger than the VaR forecast, there is no way to compute the expected magnitude of the loss; VaR does not give give us any information about losses outside of the bound, which is an important information to have in a moment of crisis. They proposed a new alternative measure of risk that resolves these problems, called Conditional Value-at-Risk (CVar), also called Expected Shortfall (ES). It quantifies the losses and at the same time it is a coherent measure of risk.

3.2. Stylized facts

The daily logarithmic return of the stock indices is defined as:

$$R_{t+1} = ln(S_{t+1}) - ln(St), \tag{3}$$

where St is the price of the index at time t and S_{t+1} the price at time $t + 1$.

All financial returns have three stylized facts (Danielsson, 2011):

1. Volatility clusters

Volatility of the returns usually cluster together, periods of high volatility are followed by periods of high volatility and periods of low volatility are followed by periods of low volatility.

2. Fat tails

The returns are usually not normally distributed because of the outliers, either positive or negative that require fatter tails of the distribution in order to be correctly observed.

3. Nonlinear dependence

It explains the correlation of the returns.

The daily returns should have very little autocorrelation

$$Corr(R_{t+1}, R_{t+1-\tau}) \approx 0 \tag{4}$$

If the variance is measured by squared returns, it should present positive correlation with its own past. This should be even more obvious because of the short horizon that we will use (daily).

$$Corr(R^2_{t+1}, R^2_{t+1-\tau}) > 0, \tag{5}$$

for small τ.

3.3. Normal Distribution

The normal distribution is a continuous probability distribution. It is usually used as a first approach for describing random variables that tend to cluster around an average. The shape of the normal distribution is a bell curve and it is one of the most significant distributions in statistics. This distribution often applies in cases where it is less likely to occur extreme values than it is to occur average values.

A normal distribution can be determined by two moments: the mean and the variance (standard deviation). The average lies in the distribution peak. The standard deviation determines the shape of the distribution, if it is spread out, or if the majority of the surface will be concentrated near the top.

If x is a normal random variable, the probability density function (pdf) is

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(\frac{1}{2})[(x-\mu)/\sigma]^2}, \tag{6}$$

where:

μ – mean or expectation of the distribution,

σ – standard deviation.

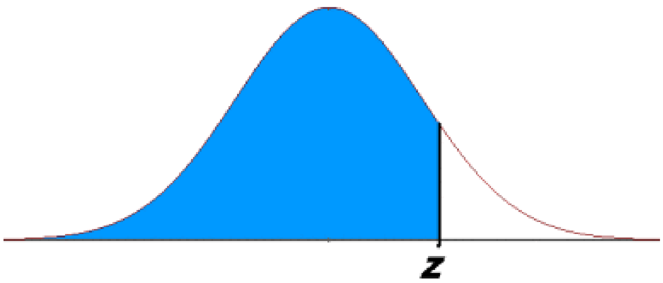
In most cases when we want to find the surface area under a normal distribution we will use the standard normal distribution. The standard normal distribution is a normal distribution with a mean of 0 and variance of 1.

To find the surface we need the cumulative distribution function ($\Phi(z)$). This distribution gives the probability that a standard normal random variable Z has a lower value than a specified value z:

$$\Phi(z) = Pr(Z < z) \tag{7}$$

It also provides the surface area under the density function of the standard normal random variable, which is to the left of the specified value z.

Figure 2 Cumulative Density Function – probability of being below the number z



Two additional moments of the normal distributions are known as skewness and kurtosis. Skewness characterizes the asymmetry of a distribution around its mean. The expression for skewness is given by:

$$s^3 = E[(r_t - \mu)^3] \quad (8)$$

For the normal distribution, skewness is zero. In practice, it is more convenient to work with the skewness coefficient, which is defined as:

$$\gamma = \frac{E[(r_t - \mu)^3]}{\sigma^3} \quad (9)$$

Kurtosis measures the relative peakedness or flatness of a given distribution. The expression for kurtosis is given by:

$$s^4 = E[(r_t - \mu)^4] \quad (10)$$

For the normal distribution, kurtosis is 3. As in the case of skewness, in practice, the kurtosis coefficient is more frequently used:

$$k = \frac{E[(r_t - \mu)^4]}{\sigma^4} \quad (11)$$

3.4. Student-t Distribution

The Student-t is a continuous probability distribution which is very similar to the Normal Distribution. The shape of this distribution is also a bell curve which depends on the degrees of freedom. As they increase, the distribution takes the shape of the Normal Distribution. As they decrease, it diverges from the shape of the Normal Distribution by way of demonstrating lower peak and fatter tails.

4. Methodology for calculating VaR

For calculating VaR we will use 2 methods: the first is a non-parametric method – The Historical Simulation, the second is a parametric method – the Variance / Covariance Matrix.

4.1. Historical Simulation

The Historical Simulation method uses historical data in order to construct the distribution for future Profit/Loss. It does not assume normal distribution but allows the historic data to shape its distribution. In addition, it assumes that the changes in the past will continue to be the same in the future as well. The securities are valued under different window lengths that are defined by the person doing the calculation. Typically, the lengths vary from three months to two years.

Once the distribution is defined, we rank the results from the highest profit to the lowest loss. If we use a confidence level of 99%, then we select the loss which is exceeded 1% of the time. This loss is the VaR.

Christoffersen (2012) explains the Historical Simulation in terms of portfolios where:

We consider a hypothetical portfolio consistent of n assets. If today, at time t, we own $N_{i,t}$ shares of asset i then the portfolio would be worth:

$$V_{PF,t} = \sum_{i=1}^{n} N_{i,t} S_{i,t} \tag{12}$$

To find the pseudo value of the portfolio, we use historical asset prices of the existing shares but we do not change the number of each share that we own. This is exactly why the "pseudo" part occurs, since normally we would change the number of each shares owned as time passes.

For instance, yesterday's pseudo portfolio value would be:

$$V_{PF,t-1} = \sum_{i=1}^{n} N_{i,t} S_{i,t-1} \tag{13}$$

The pseudo logarithmic daily return is:

$$R_{PF,t} = \ln(\frac{S_{PF,t}}{S_{PF,t-1}}) \tag{14}$$

Using these statements, we consider a sequence of pseudo logarithmic daily returns, using historical asset prices but keeping the number of shares unchanged:

$$\{R_{PF,t+1-\tau}\}_{\tau=1}^{m}, \tag{15}$$

where m – past sequence of the returns. The VaR with probability p ($p = 1 - confidence\ interval$) is then calculated as 100th percentile of the sequence of past portfolio returns:

$$VaR_{t+1}^{p} = -Percentile\left(\{R_{PF,t+1-\tau}\}_{\tau=1}^{m}\right) \tag{16}$$

We apply the same method with the difference being that we do not deal with portfolios but a single asset each time.

Historical Simulation uses the concept of rolling windows. First, we need to choose a window size that generally ranges from 3 months to two years. Then, the returns within this window are sorted in ascending order and the value that leaves $p\%$ of the observations on its left side and $(1-p)\%$ on its right side is the VaR. To compute the VaR the following day, the whole window is moved forward by one day and the entire procedure is repeated.

The Historical Simulation is easy to implement but has been criticized mostly because of its use of past data when predicting the future. The best predictor of future behavior is not the past behavior when it comes to financial instruments.

4.2. Parametric method – Variance / Covariance Matrix

Contrary to the Historic Simulation, the Variance/Covariance Matrix method assumes that the market factors have a certain type of distribution. (Jackson et al. 1997)

It uses historical time series analysis to estimate the variance, covariance and correlation coefficients for each item separately. Afterward, the risk of the portfolio is determined by the linear combination of all the factors.

Our portfolio consists of one index each time we compute VaR so we will not be using the covariance and correlation coefficients, but we will focus on the estimation of the variance.

The most commonly type of distribution assumed is the normal distribution. A standard property of the normal distribution is that outcomes less than or equal to 2.33 standard deviations below the mean occur only 1% of the time. The number 2.33 is the cumulative density function (CDF) of the standard normal distribution $\Phi(\cdot)$. $\Phi^{-1}(p)$ calculates the number such that p% of the probability mass is below $\Phi^{-1}(p)$.

For a 95% confidence level, the CDF is -1.65:

$$VaR = -1.65 * (standard\ deviation\ of\ change\ in\ asset\ value) \quad (17)$$

Respectively, for a 99% confidence level, the VaR is equal to -2.33 times the standard deviation of changes in asset value.

$$VaR = -2.33 * (standard\ deviation\ of\ change\ in\ asset\ value) \quad (18)$$

One of the biggest flaws of this method is that it assumes that all risk factors are normally distributed. This causes the fat-tails problem, i.e. the risk is underestimated under the assumption of normality, and does not take into consideration the fat-tails distribution.

If the Student's t-distribution is assumed, we use the CDF of this distribution and we multiply that number by the standard deviation that is estimated.

4.2.1. Markov Process

A Markov process is a specific type of stochastic process where only the present value of a variable is significant for predicting the future. The historical data of the variable is irrelevant. (Hull, 2006)

A particular type of Markov stochastic process is the Wiener process. This process has a mean change of 0 and a variance rate of 1 per year.

A variable z follows the Wiener process if it has the following two properties:

1. The change Δz during a small period of time Δt is:

$$\Delta z = \epsilon\sqrt{\Delta t}, \quad (19)$$

where ϵ has a standard normal distribution $\phi(0,1)$.

2. The values of Δz for any two different short intervals of time, $\Delta t,$ are independent.

The mean change per unit time for a stochastic process is known as the drift rate and the variance per unit time is known as the variance rate. The basic Wiener process, dz, has a drift rate of 0 and a variance rate of 1. The drift rate of 0 means that the expected value of z at any future time is equal to its current value. The variance rate of 1 means that the variance of the change in z in a time interval of length T equals T. A generalized Wiener process for a variable x can be defined in terms of dz as:

$$dx = adt + bdz, \tag{20}$$

where a and b are constant.

Using this we can derive the stochastic process of an individual asset. The most widely used model of stock price behavior is:

$$dS = \mu Sdt + \sigma Sdz \tag{21}$$

The variable μ is the stock's expected rate of return. The variable σ is the volatility of the stock price. This model is known as geometric Brownian motion.

Since we use the daily returns, the model will be:

$$R_{t+1} = \mu_{t+1} + \sigma_{t+1} z_{t+1} \tag{22}$$

In addition, we assume that the mean of return μ_{t+1} is zero. Removing this term, we have:

$$R_{t+1} = \sigma_{t+1} z_{t+1} \tag{23}$$

What this tells us is that the distribution of the asset's returns will depend entirely on the volatility of the asset. We need to determine a model in order to estimate this parameter. Hence the name of this method, parametric method.

One property of volatility is that it is not directly observable, therefore it has to be estimated. (Tsay, 2005)

4.2.2. Ways for estimating volatility

There are a number of ways to estimate volatility:

- Moving Average

Moving average is the simplest approach. It simply calculates volatility as the unweighted standard deviation of a window of X trading days. Subsequently, this becomes the volatility forecast for all future periods.

$$\sigma_{t+1}^2 = \frac{1}{m}\sum_{i=1}^{m}(R_{t+1} - \bar{R})^2 \tag{24}$$

We assume that the mean of return is zero and that the volaitlity of the returns should present positive autocorrelation. Practically, this means that the future period will have the same type of volatility as did the recent period (high/low). We represent this relationship by computing the next day's volatility as the average of the most recent m observations, as in:

$$\sigma_{t+1}^2 = \frac{1}{m}\sum_{\tau=1}^{m}(R_{t+1-\tau})^2 \tag{25}$$

This equation gives equal weight to R_{t+1-1}^2, R_{t+1-2}^2,....., R_{t+1-m}^2.

According to Figlewski (1997), the estimation for volatility is more accurate when longer samples of past returns are used. For example, if we are using daily returns from the past 5 years, the estimated volatility will be more accurate that if we use the returns from the last year only. However, there is a trade off when it comes to the estimation window, since a very large window size gives equal weight to the most recent returns, as well as to the returns that are at the beginning of the window. This causes the method to be less accurate in the event of extreme losses or profits.

Additionally, the longer time horizon that we want to apply the volatility to, the more accurate it is. If the time horizon that we choose is 5 years from now, the volatility calculated with this method will be more accurate than if we choose a time horizon of 1 year.

- Exponentially Weighted Moving Average (EWMA)

Another way of estimating the volatility is by using an exponential weighted moving average (EWMA) of historic data. This model was created by JP Morgan's RiskMetrics system for market risk management.

The difference between this approach and the simple moving average approach is that this time the observations don't have the same weight. The most recent observations bring the highest weight and as we move back in time, the weight of every observation is reduced. This is an important change in the way that volatility is estimated because of two reasons. First, the events in the recent past can be translated to the estimated volatility much faster if they carry higher weights. Second, in the case of a momentary shock in the market, the influence it has on the volatility will degrade slowly over time as the weight is reduced. In the case where all observations carry the same weight, it can lead to relatively sudden changes in the volatility once the shock falls out of the measurement sample, which sometimes can be several months after it occurs.

The EWMA, sometimes called exponential smoother model, is computed as:

$$\frac{R_t^2 + \lambda R_{t-1}^2 + \lambda^2 R_{t-2}^2 + \cdots + \lambda^n R_{t-n}^2}{1 + \lambda + \lambda^2 + \cdots + \lambda^n} \tag{26}$$

The denominator converges to $\frac{1}{1-\lambda}$ as n $\rightarrow \infty$, so an infinite EWMA may be written as:

$$\sigma_{t+1}^2 = (1-\lambda)\sum_{\tau=1}^{\infty}(\lambda^{\tau-1}R_{t+1-\tau}^2) \tag{27}$$

The future volatility can also be written as:

$$\sigma_{t+1}^2 = \lambda\sigma_t^2 + (1-\lambda)R_t^2 \tag{28}$$

The model depends on the “smoothing” parameter λ $(0 < \lambda < 1)$, which is often referred to as the decay factor. This parameter determines the relative weights that are applied to the observations.

When estimating λ on a 480 time series, RiskMetrics found that the estimates were quite similar across the time series, using one optimal decay factor. Consequently, this optimal decay factor turned out to be $\lambda = 0.94$ for daily returns and $\lambda = 0.97$ for monthly returns. In this case, there is no need for an estimation of the parameter when computing volatility. We simply apply the EWMA equation (JP Morgan, 1996).

For our daily return, the volatility would be:

$$\sigma_{t+1}^2 = 0.94\sigma_t^2 + 0.06R_t^2 \tag{29}$$

- ARCH

Engle (1982) was the first to propose a new class of stochastic processes called autoregressive conditional heteroscedasticity (ARCH) processes. These are mean zero, serially uncorrelated processes with non-constant volatility conditional on the past and constant unconditional volatility. It means that the conditional volatility of the financial series evolves according to an autoregressive type process. Specifically, the ARCH models assume the volatility of the current error term to be a function of the previous time periods' error terms. If the volatility of the error terms is constant, then we have homoscedasticity. If the error terms do not have constant volatility, they are said to be heteroskedastic. Since the discovery that the volatility can be time-varying there have been conducted many studies about its properties. (Glosten et al. 1993)

According to Engle (1982), if a random variable y_t is drawn from the conditional density function $f(y_t|y_{t-1})$, the forecast of today's value based upon past information, under standard assumptions, is simply $E(y_t|y_{t-1})$, which depends upon the value of the conditioning variable y_{t-1}. The volatility of this one period forecast is given by $V\ (y_t|y_{t-1})$. Such an expression recognizes that the conditional forecast volatility depends upon past information and may therefore be a random variable.

We initially consider the first-order autoregression:

$$y_{t+1} = \gamma y_t + \epsilon_{t+1}, \qquad (30)$$

where ϵ is white noise with $V(\epsilon) = \sigma^2$. The conditional mean of y_t is γy_t while the unconditional mean is zero. The conditional volatility of y_{t+1} is σ^2 while the unconditional volatility of y_{t+1} is $\sigma^2/(1-\gamma^2)$.

The distinction between the conditional and unconditional volatility of a random variable is exactly the same as that of the conditional and unconditional mean. The conditional volatility of y_{t+1} may be denoted σ^2, which is written as:

$$\sigma_{t+1}^2 = V(y_{t+1}|y_t, y_{t-1}, \dots) = E[(y_{t+1} - E(y_{t+1}))^2|y_t, y_{t-1}, \dots] \qquad (31)$$

It is usually assumed that $E(y_{t+1}) = 0$, so

$$\sigma_{t+1}^2 = V(y_{t+1}|y_t, y_{t-1}, \dots) = E[{y_{t+1}}^2|y_t, y_{t-1}, \dots] \qquad (32)$$

This equation states that the conditional volatility of a zero mean normally distributed random variable y_{t+1} is equal to the conditional expected value of the square of y_{t+1}.

A model that allows the conditional volatility to depend on the past realizations of the series is the bilinear model:

$$y_{t+1} = \epsilon_{t+1} y_t \qquad (33)$$

Under the ARCH model, the 'autocorrelation in volatility' is modelled by allowing the conditional volatility of the error term, σ_t^2 to depend on the immediately previous value of the squared error (Brooks, 2008)

$$\sigma_{t+1}^2 = a_0 + a_1 y_t^2 \qquad (34)$$

This is an ARCH (1) model as it contains only a single lag on the squared error term, however it is possible to extend this to any number of lags, if there are q lags it is termed an ARCH (q) model:

$$\sigma_{t+1}^2 = a_0 + a_1 y_t^2 + a_2 y_{t-1}^2 + \cdots + a_q y_{t-q}^2 \qquad (35)$$

- GARCH

Bollerslev (1986) and Taylor (2008) independently generalized Engle's model to make it more realistic; the generalization was called GARCH (General Autoregressive Conditional Heteroscedasticity).

According to Bollerslev, let ε_{t+1} denote a real-valued discrete-time stochastic process, and ψ_t, the information set (σ-field) of all information through time t. The GARCH(p, q) process is then given by:

$$\varepsilon_{t+1}|\psi_t \sim N(0, h_{t+1}), \qquad (36)$$

$$h_{t+1}^2 = a_0 + \sum_{i=1}^{q} a_1 \varepsilon_t^2 + \sum_{i=1}^{p} \beta_i h_t^2 \qquad (37)$$

$$= a_0 + A(L)\varepsilon_t^2 + B(L)h_t,$$

where,

$p \geq 0,$ $\quad q > 0$

$a_0 > 0$ $\quad a_i \geq 0,$ $\quad i = 1, \ldots, q,$

$\beta_i \geq 0$ $\quad i = 1, \ldots, p$

p is the number of estimates of volatility rates (h_t) and q is the number of observations of the error terms (ε_t^2). From the previous relation (37), we can see that for $p = 0$, GARCH is equal to ARCH. Moreover, for $p = q = 0$, there is no correlation between the values of ε_t at different times. In this case, we would characterize this process as white noise.

The main difference from the ARCH process is that GARCH allows the use of not only the sample volatility, but the conditional volatility as well.

- GARCH (1,1)

One of the most used GARCH process is GARCH (1, 1), where:

$$h_{t+1}^2 = a_0 + a_1\varepsilon_t^2 + \beta_1 h_t^2, \tag{38}$$

$a_0 > 0, a_1 \geq 0, \beta_1 \geq 0$

In this case h_{t+1} is calculated from the first recent estimate of the volatility rate and the first recent observation on ε^2.

We can see that the EWMA discussed earlier is a particular case of GARCH (1,1) where $a_0 = 1, a_1 = 1 - \lambda,$ and $\beta_1 = \lambda.$

Since we deal with one index at a time we use the univariate GARCH model. In the case of a portfolio consisting of more than one asset, the multivariate GARCH model could be used for estimating the correlation between assets (Bollerslev, 1990).

We applied this method to our data in the form of:

$$\sigma_{t+1}^2 = a_0 + a_1 R_t^2 + \beta_1\sigma_t^2, \tag{39}$$

where σ_{t+1} is the volatility forecast for the period $t+1$, σ_t is the volatility forecast for the period t, R_t^2 is the squared return for the period t and a_0, a_1, and β_1 are the estimated parameters using the information up to period t.

- EGARCH

Nelson (1991) published the Exponential GARCH model (non-linear). It enables to capture the asymmetric effects (leverage effects), the different effects of positive and negative shocks on conditional volatility, i.e. an unexpected price drop increases volatility more than an equivalent unexpected price increase. The EGARCH model takes into consideration these effects by defining the conditional volatility not only by its size but also by its sign (positive/negative).

The EGARCH (p, q) model is described by:

$$\log(h_{t+1}^2) = a_0 + \sum_{i=1}^{q}\left(a_1\left|\frac{\varepsilon_t}{h_t}\right| + \gamma_1\frac{\varepsilon_t}{h_t}\right) + \sum_{i=1}^{p}(\beta_1\log(h_t^2)) \tag{40}$$

The most common type of this model is the EGARCH (1,1):

$$\log(h_{t+1}^2) = a_0 + a_1 \left|\frac{\varepsilon_t}{h_t}\right| + \gamma_1 \frac{\varepsilon_t}{h_t} + \beta_1 \log(h_t^2) \tag{41}$$

Again, when we applied this model to our data it took the form of:

$$\log(\sigma_{t+1}^2) = a_0 + a_1 \left|\frac{R_t}{\sigma_t}\right| + \gamma_1 \frac{R_t}{\sigma_t} + \beta_1 \log(\sigma_t^2) \tag{42}$$

where σ_{t+1} is the volatility forecast for the period $t+1$, σ_t is the volatility forecast for the period t, R_t is the return for the period t and a_0, a_1, β_1 and γ_1 are the estimated parameters using the information up to period t. γ_1 is the leverage parameter.

5. Backtesting

Regardless of the methods used for calculating VaR, one of the most important procedure is the backtesting. The objective of this procedure is to consider the ex-ante risk measure forecasts from the model and compare them with the ex post realized returns. In other words, examining to what extent was the data given by the VaR process, in fact consistent with the reality.

We use the three criteria suggested by Christoffersen (2012), in order to backtest VaR statistically: unconditional coverage test, independence test and conditional coverage test.

If we compare the VaR forecasts and the actual returns, we can describe the results as:

$$I_{t+1} = \begin{cases} 1, if\ R_{t+1} < -VaR_{t+1}^{p} \\ 0, if\ R_{t+1} \geq -VaR_{t+1}^{p} \end{cases} \tag{43}$$

where VaR_{t+1}^{p} denotes the loss that it will only be exceeded $p \times 100\%$ of the time.

If the loss on day $t+1$ exceeds the VaR forecast for that day, I_{t+1} returns 1, otherwise it returns a value of 0. When backtesting the VaR model, we construct a sequence of $\{I_{t+1}\}_{t=1}^{T}$ for T number of days when the loss exceeded the Var forecast.

Since we cannot predict the probability of VaR violations, the hit sequence I_{t+1} should be unpredictable and distributed independently over time, taking the form of a Bernoulli variable that takes the value 1 with probability p and the value 0 with probability $1 - p$:

$$H_0: I_{t+1} \sim \text{i.i.d. Bernoulli } (p) \tag{44}$$

This is the null hypothesis.

When conducting the backtesting, p will be 0.01 or 0.05, depending on the confidence level chosen for the VaR.

The probability mass function (pmf) f of this distribution, over the possible outcome p, is:

$$f(I_{t+1}; p) = (1-p)^{1-I_{t+1}} p^{I_{t+1}} \tag{45}$$

5.1. Unconditional coverage testing

We first want to test if the violations obtained from a particular risk model is considerably different from the coverage rate p. Christoffersen called this the unconditional coverage hypothesis. This test is concerned with whether or not the forecasted VaR is violated more (or less) than $p * 100\%$ of the time.

To test it, we start with the likelihood of an i.i.d. Bernoulli hit sequence (π). Since the Bernoulli distribution defines a discrete probability function instead of continuous, the likelihood equals the probability mass function (for continuous functions, we used the probability density function):

$$L(\pi) = \prod_{t=1}^{T} (1-\pi)^{1-I_{t+1}} \pi^{I_{t+1}} = (1-\pi)^{T_0} \pi^{T_1} \tag{46}$$

where T_0 and T_1 are the values of 0 and 1 that the hit sequence I_{t+1} take in the sample.

We can estimate π as $\hat{\pi} = T_1/T$. This is the maximum likelihood estimate of π. Substituting this in the likelihood function we have:

$$L(\pi) = (1 - \frac{T_1}{T})^{T_0} (\frac{T_1}{T})^{T_1} \tag{47}$$

Under the unconditional coverage null hypothesis, $\pi = p$ where p is the known VaR coverage rate, the likelihood function of which is:

$$L(p) = \prod_{t=1}^{T} (1-p)^{1-I_{t+1}} p^{I_{t+1}} = (1-p)^{T_0} p^{T_1} \tag{48}$$

Lastly, we can check the unconditional coverage hypothesis using a likelihood ratio test:

$$LR_{uc} = -2ln\left[\frac{L(p)}{L(\hat{\pi})}\right] \tag{49}$$

As the number of observations T goes to infinity, the test is distributed as a χ^2 with one degree of freedom. This likelihood ratio, or equivalently its logarithm, can then be used to compute a p-value, or using a significance level compare it to the critical value to decide whether to accept or reject the null hypothesis.

When running this statistical test, we are faced with two types of errors. Type 1 error refers to the possibility of rejecting a correct model and type 2 error to the possibility of not rejecting (accepting) an incorrect model. Best case scenario would be to eliminate the two types of errors.

If we reject the null hypothesis $\pi = p$, then in this case $\pi \neq p$. This can be interpreted in one of two ways: If $\pi > p$, losses in excess of the forecasted VaR occur more frequently than $p * 100\%$ of the time. This would suggest that the forecasted VaR measure understates the portfolio's actual level of risk. The opposite, $\pi < p$, would mean that the VaR measure is conservative and overestimates the risk.

5.2. Independence testing

Another problem occurs when the VaR violations are happening around the same time. Even if the models "pass" the unconditional coverage test, the violations also need to be tested for their clustering in time.

Cristoffersen (2012) uses the Markov test in order to reject a VaR with clustered violations.

The Markov test examines the likelihood of a VaR violation depending on a VaR violation that occurred on the previous day. If the VaR measure accurately reflects the underlying risk, then the probability of a violation today should be independent of whether or not there was a violation yesterday. Hence the name of this test, independence test.

We assume the hit sequence is dependent over time and that it can be described as a so-called first-order Markov sequence with transition probability matrix:

$$\Pi_1 = \begin{bmatrix} 1-\pi_{01} & \pi_{01} \\ 1-\pi_{11} & \pi_{11} \end{bmatrix} \quad (50)$$

where,

π_{01}, the probability of tomorrow being a violation given that today is not a violation:

$$\pi_{01} = \Pr(I_{t+1} = 1 | I_t = 0) \quad (51)$$

π_{11}, the probability of tomorrow being a violation given that today is also a violation:

$$\pi_{11} = \Pr(I_{t+1} = 1 | I_t = 1) \quad (52)$$

$1 - \pi_{01}$, the probability of tomorrow not being a violation, given that today is also not a violation

$$1 - \pi_{01} = \Pr(I_{t+1} = 0 | I_t = 0) \quad (53)$$

$1 - \pi_{11}$, the probability of tomorrow not being a violation given that today is a violation

$$1 - \pi_{11} = \Pr(I_{t+1} = 0 | I_t = 1) \quad (54)$$

The likelihood function in this case would be:

$$L(\Pi_1) = (1 - \pi_{01})^{T_{00}} \pi_{01}^{T_{01}} (1 - \pi_{11})^{T_{10}} \pi_{11}^{T11}, \quad (55)$$

where T_{ij}, $i, j = 0{,}1$ is the number of observations with a j following an i.

The maximum likelihood estimates for $\widehat{\pi_{01}}$ and $\widehat{\pi_{11}}$ are:

$$\widehat{\pi_{01}} = \frac{T_{01}}{T_{00} + T_{01}} \quad (56)$$

$$\widehat{\pi_{11}} = \frac{T_{11}}{T_{10} + T_{11}}$$

Since the probabilities have to sum to one, we have

$$\widehat{\pi_{00}} = 1 - \widehat{\pi_{01}} \quad (57)$$

$$\widehat{\pi_{10}} = 1 - \widehat{\pi_{11}}$$

The estimated transition probability matrix is:

$$\widehat{\Pi_1} \equiv \begin{bmatrix} \widehat{\pi_{00}} & \widehat{\pi_{01}} \\ \widehat{\pi_{10}} & \widehat{\pi_{11}} \end{bmatrix} = \begin{bmatrix} 1 - \widehat{\pi_{01}} & \widehat{\pi_{01}} \\ 1 - \widehat{\pi_{11}} & \widehat{\pi_{11}} \end{bmatrix} = \begin{bmatrix} \frac{T_{00}}{T_{00} + T_{01}} & \frac{T_{01}}{T_{00} + T_{01}} \\ \frac{T_{10}}{T_{10} + T_{11}} & \frac{T_{11}}{T_{10} + T_{11}} \end{bmatrix} \quad (58)$$

If the probabilities are independent over time, then the probability of a violation tomorrow does not depend on today being a violation or not, and so $\pi_{01} = \pi_{11} = \pi$. In this case the estimated transition probability matrix is:

$$\widehat{\Pi_1} = \begin{bmatrix} 1-\hat{\pi} & \hat{\pi} \\ 1-\hat{\pi} & \hat{\pi} \end{bmatrix} \quad (59)$$

Again, we can check the independence hypothesis using a likelihood ratio test:

$$LR_{ind} = -2ln\left[\frac{L(\widehat{\Pi})}{L(\widehat{\Pi_1})}\right] \quad (60)$$

In large samples, the LR_{ind} test is also distributed as a χ^2 with one degree of freedom. We use the p-value or the critical value to determine if we will accept or decline the hypothesis.

5.3. Conditional Coverage Testing

In order for the VaR to be measured accurately, it needs to satisfy both the unconditional coverage and the independence property. Therefore, we need to consider a test that examines these properties at the same time.

This kind of test is the conditional coverage test by Christoffersen (2012).

The likelihood ratio test that we are going to use for this test is:

$$LR_{cc} = -2ln\left[\frac{L(p)}{L(\widehat{\Pi_1})}\right], \quad (61)$$

which also follows the χ^2 distribution with one degrees of freedom.

As we can see, the LR_{cc} test takes the likelihood from the null hypothesis in the LR_{uc} test and combines it with the likelihood from the hypothesis in the LR_{ind} test. Hence:

$$LR_{cc} = LR_{uc} + LR_{ind} \quad (62)$$

In deciding on the significance level with which we are going to reject or accept each null hypothesis, there is a trade-off between rejecting a correct model, also known as Type I error and accepting an incorrect model, or Type II error. We decided to use a significance level of 10% for the p-value in order to avoid these errors.

6. Data

For the purpose of this thesis, we used Bloomberg for the gathering of the time series data for equity indices. Daily prices are collected for the 15 year period from September 26, 2001 to September 27, 2016. Non-trading days were excluded from the data set.

The data used in this thesis consists of the major US and European stock indices. More specifically,

1. For U.S. market:

- **S&P 500 Composite**: The S&P 500 was the first U.S. market-cap-weighted stock market index. Today, it's the basis of many listed and over-the-counter investment instruments. This world-renowned index includes 500 of the top companies in leading industries of the U.S. economy and captures approximately 80% coverage of available market capitalization. (Standard and Poor's, 2004)

- **Nasdaq Composite**: Launched in 1971, the NASDAQ Composite Index is a broad based Index. Today, the Index includes over 3,000 securities, more than most other stock market indices. The NASDAQ Composite is calculated under a market capitalization weighted methodology index. (Nasdaq, 2016)

2. For European market:

- **Euro Stoxx**: The EURO STOXX 50 Index, Europe's leading blue-chip index for the Eurozone, provides a blue-chip representation of supersector leaders in the Eurozone. The index covers 50 stocks from 12 Eurozone countries. The Index is licensed to financial institutions to serve as underlying for a wide range of investment products such as Exchange Traded Funds (ETF), Futures and Options and structured products. (Bloomberg, 2016)

- **FTSE 100**: The FTSE 100 is a market-capitalization weighted index of UK-listed blue chip companies. The index is part of the FTSE UK Series and is designed to measure the performance of the 100 largest companies traded on the London Stock Exchange that pass screening for size and liquidity. (FTSE Group, 2012)

- **DAX 30 Performance**: The DAX Index tracks the segment of the largest and most important companies – known as blue chips – on the German equities market. It contains the shares of the 30 largest and most liquid companies admitted to the FWB Frankfurt Stock Exchange in the Prime Standard segment. The DAX represents about 80% of the aggregated prime standard's market cap. (Deutsche Börse Group, 2012)

- **France CAC 40**: The CAC 40 is a free float market capitalization weighted index that reflects the performance of the 40 largest and most actively traded shares listed on Euronext Paris, and is the most widely used indicator of the Paris stock market. The index serves as an underlying for structured products, funds, exchange traded funds, options and futures. Stocks are screened to ensure liquidity and selected and free float weighted to ensure that the index is investable. (Euronext, 2016)

- **ATHEX Composite**: The Athens Stock Exchange (ATHEX) General Index is a capitalization-weighted index of Greek stocks listed on the Athens Stock Exchange. It comprises a reliable measure, on a real time basis, of the performance of the shares of the companies that trade in the Big Cap segment of the Athens Exchange. (Athens Exchange Group, 2016)

The analysis is limited to a simple portfolio that consists of one index each time we compute VaR.

Our data set is split into two sections:

1. In-sample period, which we will use in order to compute the VaR forecasts for the out of sample period. This period includes the data from September 26, 2001 to December 29, 2006.
2. Out of sample period, which we will use in the backtesting to compare the forecasts to the realized returns, and determine the accuracy of the methods used. This period includes the data from January 03, 2007 to September 27, 2016.

The reason that we chose the out of sample period to be 2007-2016 is because we wanted to test the various VaR methods in much more volatile times than usual. Having chosen the 2007 as a starting point in our study, it coincides with the start of the most recent financial crisis that also started in the summer of 2007.

The in-sample period was chosen proportionally to the out of sample period, having enough historic data for the estimation of the methods, but not too much so as to contain redundant information.

Descriptive statistics for the time series are provided in the table below:

Table 1 Descriptive Statistics for logarithmic returns: Sample 26/09/2001 – 27/09/2016
The probability in the Jarque-Bera Test indicates the p-value for the null hypothesis Ho: the data comes from normal distribution.

Parameters	S&P500	NASDAQ	EURO STOXX	FTSE 100	DAX	CAC	ATHEX
N of observations	3777	3777	3847	3791	3813	3840	3710
Mean	0.02%	0.03%	-0.00%	0.00%	0.02%	0.00%	-0.03%
Median	0.06%	0.09%	0.01%	0.04%	0.08%	0.03%	0.02%
Maximum	10.95%	11.15%	10.43%	9.38%	10.79%	10.59%	13.43%
Minimum	-9.46%	-9.58%	-9.01%	-9.26%	-7.43%	-9.47%	-17.71%
Std. Dev.	1.23%	1.40%	1.51%	1.21%	1.53%	1.49%	1.92%
Skewness	-0.22	-0.09	-0.02	-0.12	0.00	0.00	-0.35
Kurtosis	12.27	8.07	7.57	9.55	7.36	8.01	9.74
Jarque-Bera	13567.77	4056.98	3354.31	6801.20	3024.79	4029.71	7110.24
Probability	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)

From these statistics we can draw a few conclusions:

1. With a few exceptions, the indices are quite similar.
2. The mean in most cases is 0 and in the case where it is not, we can see that it is much smaller than the volatility, so we can assume that it is 0 as we did in most of the models.
3. The most volatile index is the ATHEX (std. dev. = 1.92%), and the least volatile is the FTSE 100 (std. dev. = 1.21%), making it the safest market in our sample.
4. All of the indices with the exception of DAX and CAC have negative skewness, which means that the left tail of the distribution is longer than the right one. This could mean that in these markets there is a greater chance of negative outcomes. In the case of DAX and CAC, the skewness is very close to 0, meaning that their distribution is symmetric.
5. The kurtosis is more than 3 for all of the indices, indicating a leptokurtic distribution, which is characterized by fatter tails. This could affect the estimation of VaR when normal distribution is assumed and instead a distribution with fatter tails should be preferred (as the Student's t-distribution).
6. The above observation is confirmed by the Jarque-Bera test for normality. The probability measure is 0.00 for all the indices, causing the rejection of the hypothesis for normality.
7. The lowest return for all indices was recorded in the fall of 2008, as a result of the financial crisis, with the exception of ATHEX which had its lowest return in the summer of 2015 with the imposed capital controls.

We conducted the ADF-GLS unit root test for determining the stationarity of the time series. As we can see from the tables 11-24 in Appendix A most of the time series show the presence of unit root in the level. On the other hand, when using the 1st difference of the time series in the test, the unit root does not exist, and instead the time series show stationarity. This fact validates the use of the logarithmic differences instead of the price levels in all of the models.

We also conducted a Ljung-Box Q-test in order to determine whether the returns are autocorrelated. The null hypothesis is that the data are independently distributed, i.e. the returns are not autocorrelated. A logical value (h) of 1 indicates the rejection of the null hypothesis for no autocorrelation, $h = 0$ indicates the non-rejection of the null hypothesis. The p-value indicates the strength at which the test rejects or does not reject the null hypothesis for 5, 10 and 20 lags.

In the appendix section B1, the sample ACF and PACF are presented (Figures 3-9).

Table 2 Ljung-Box Q-Test for logarithmic returns: Sample 26/09/2001 – 27/09/2016

Parameters		S&P500	NASDAQ	EURO STOXX	FTSE 100	DAX	CAC	ATHEX
h		1	1	1	1	1	1	1
	5 lags	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
p-value	10 lags	0.00%	0.46%	0.00%	0.00%	0.17%	0.00%	0.00%
	20 lags	0.00%	0.00%	0.00%	0.00%	0.18%	0.00%	0.00%

For all of the indices the test indicates the rejection of the null hypothesis H0: no autocorrelation. All of the logical values indicate the rejection decision and all of the p-values are less than 10% for all the lags reported.

We used the Engle's ARCH test which assesses the null hypothesis that the returns exhibit no conditional heteroscedasticity (ARCH effects), against the alternative that an ARCH model describes the series. Again we use the logical value and the p -value to determine the presence of heteroscedasticity.

We de-meaned the returns and squared them. We used two lags for the ARCH test. We did this because an ARCH (2) is locally equivalent to a GARCH (1,1) model.

In the appendix section B2, the sample ACF and PACF of the squared returns are presented (Figures 10-16).

Table 3 Engle's ARCH test for logarithmic returns: Sample 26/09/2001 – 27/09/2016

Parameters	S&P500	NASDAQ	EURO STOXX	FTSE 100	DAX	CAC	ATHEX
h	1	1	1	1	1	1	1
p-value	0	0	0	0	0	0	0

As we can see from the logical value h and the p -value, in all of the cases there is heteroscedasticity. This indicates that a GARCH model would be a fitting model for the variance.

7. Empirical Findings

We apply the aforementioned methods to all the stock indices and produce one-day ahead VaR forecasts for 95% and 99% confidence level for the period 03/01/2007 to 27/09/2016. The in-sample period 26/09/2001 – 29/12/2006 is used to estimate the parameters and then with the use of rolling window the out of sample VaR is forecasted. The concept of the rolling analysis is used because the environment surrounding the time series changes constantly and so the parameters of the models should change accordingly (Zivot and Wang, 2003). We estimate the parameters by using all of the information available up until the moment of estimation.

For the Historic Simulation and Moving Average we use a rolling window of 100 and 250 trading days.

For the GARCH (1, 1) and EGARCH (1, 1) we use the AR (1) as our conditional mean model. We also use the normal and Student's t-distribution.

The forecasting horizon is chosen to be one day, which is also the most widely used horizon in the literature. In some circumstances, a horizon of 10 days is also used, which is derived from the one-day horizon with a simple calculation, multiplying the daily VaR by the square root of 10. Nevertheless, this is not always the right way for deriving the 10 days VaR, according to Danielsson (2002), since it violates some of the distribution assumptions such as: returns are normally distributed, volatility is independent over time and the volatility is identical across all time periods.

Subsequently, the models are tested using Christoffersen's criteria: the unconditional coverage test, the independence test and the conditional coverage test. Using the p-value we decide to accept or reject the hypothesis of each test. The significance level is 10%.

The main empirical results are presented in Tables 4-10.

In the appendix C, figures 17-51 show the plot of the returns of every index, along with the VaR methods applied.

7.1. S&P500

Table 4 contains the VaR results and the backtesting results for S&P 500. The estimation period is 26/09/2001 to 29/12/2006 and the forecast period 03/01/2007 to 27/09/2016. The number of observations for the backtesting is therefore 2452. The

expected number of violations for 95% confidence level is 123 and for 99% the expected violations are 25.

As we can see from table 4:

At 95% confidence level the highest number of violations is made using EGARCH (1,1) with normal distribution (170) and the lowest number with GARCH (1,1) with Student's t-distribution (103). Although low, this method is not the closest to the expected number of violations, which is provided by EGARCH (1,1) with Student's t-distribution (130 when expecting 123). When using different estimation horizons, Historical Simulation performs better with longer horizon (250 observations). However, this is not the case with Moving Average since it is more accurate with shorter horizons (100 horizons). When compared, Moving Average and Exponentially Weighted Moving Average perform quite worse than GARCH (1,1) and EGARCH (1,1). There is substantial difference between normal and Student's t-distribution. In all of the cases, the normal distribution underestimates the risk, providing more violations than expected, and in some of the cases the Student t distribution overestimates risk, providing less violations than expected. The independence criterion is passed by almost all methods. The only methods that pass the conditional coverage criterion are Historical Simulation with 100 observations, GARCH (1,1) with Student's t-distribution and EGARCH (1,1) with Student's t-distribution. The Historical Simulation however does not pass the unconditional coverage criterion at 10%, but it does at 5%. Out of these three methods, EGARCH (1,1) scores the highest p-value (36.16%).

At 99% confidence level the highest number of violations is made using Moving Average with 250 observations (76) and the lowest with GARCH (1,1) with Student's t-distribution (17). The closest to the expected number of violations is EGARCH (1,1) with Student's t-distribution (30 when expecting 25). The estimation horizon provides the same results as in 95%, longer horizons are more accurate for Historical Simulation and shorter for Moving Average. GARCH (1,1) and EGARCH (1,1) outperform the Moving Average and the Exponentially Weighted Moving Average again. The normal distribution again has very poor results, passing only the independence criterion in both cases and having very low p-value for the unconditional criterion (0.00%). The independence criterion is passed by all the models, except for Moving Average with 250 observations. The GARCH (1,1) with Student's t-distribution overestimates risk with lower than expected violations. Regardless of this fact, it passes the unconditional criterion along with EGARCH (1,1) with Student's t-distribution. EGARCH (1,1) with Student's t-distribution had the highest p-value (38.71%).

7.2. NASDAQ

Table 5 contains the VaR results and the backtesting results for NASDAQ. The estimation period is 26/09/2001 to 29/12/2006 and the forecast period 03/01/2007 to 27/09/2016. The number of observations for the backtesting is therefore 2452. The

expected number of violations for 95% confidence level is 123 and for 99% the expected violations are 25.

As we can see from table 5:

At 95% confidence level the highest number of violations is made using EGARCH (1,1) with normal distribution (159) and the lowest number was with GARCH (1,1) with Student's t-distribution (112). The closest to the expected number of violations is provided by EGARCH (1,1) with Student's t-distribution (131 when expecting 123), although slightly underestimated. The better estimation horizon depends on the method used, with Historical Simulation performing better with shorter horizon (100 observations) and Moving Average with longer horizons (250 observations). The Moving Average and Exponentially Weighted Moving Average perform worse than GARCH (1,1) with both distributions, but not than EGARCH (1,1) with normal distribution. There is a significant difference between normal and Student's t-distribution. The normal distribution has shown that it does not perform well, underestimating the risk with GARCH (1,1) and EGARCH (1,1), whereas the Student's t-distribution overestimates risk in the case of GARCH (1,1). The independence criterion is passed by some of the methods. The methods that pass the conditional coverage criterion are Historical Simulation with 100 observations and GARCH (1,1) with Student's t-distribution, with the Historical Simulation having the highest p-value (25.11%).

At 99% confidence level the highest number of violations is made using Moving Average with 250 observations (70) and the lowest with GARCH (1,1) with Student's t-distribution, which is also the closest to the expected number of violations (22 when expecting 25). Interestingly, the better estimation horizon is the opposite of the one using the 95% confidence level. In this case, the Historical Simulation using 250 observations is slightly more accurate than the Historical Simulation using 100 observations. The Moving Average has a somewhat better performance when estimated using 100 observations. As far as the violations frequency in the sample, the Moving Average and the Exponentially Weighted Moving Average are less accurate than the GARCH (1,1) and EGARCH (1,1) methods, but in terms of statistical backtesting, they don't differ much from these methods when the normal distribution is assumed. The results are different when the Student's t-distribution is applied. In this case the GARCH (1,1) and EGARCH (1,1) are the only methods that pass the unconditional and conditional criteria. Although the GARCH (1,1) with Student's t-distribution overestimates risk, it does pass all of the criteria and has the highest p-value for the conditional coverage criterion (71.55%).

7.3. EURO STOXX

Table 6 contains the VaR results and the backtesting results for EURO STOXX. The estimation period is 26/09/2001 to 29/12/2006 and the forecast period 02/01/2007 to 27/09/2016. The number of observations for the backtesting is therefore 2501. The

expected number of violations for 95% confidence level is 125 and for 99% the expected violations are 25.

As we can see from table 6:

At 95% confidence level the highest number of violations is made using EWMA (157) and the lowest number with GARCH (1,1) with Student's t-distribution (107). Again, closest to the expected number of violations is EGARCH (1,1) with Student's t-distribution (128 when expecting 123). When deriving VaR from Historical Simulation and Moving Average the longer estimation horizon gives better results than the shorter, both in terms of violations and in statistical backtesting, with Historical Simulation using 250 observations even passing the unconditional coverage test. The GARCH (1,1) and EGARCH (1,1) with normal distribution don't perform much better than Moving Average and Exponentially Weighted Moving Average. The Student's t-distribution fixes this problem in the case of EGARCH (1,1), where it estimates the frequency of violations very close to the expected one. This is not the case of GARCH (1,1) where it causes overestimation of the risk, so that it provides less violations than expected, and in that way failing the unconditional coverage test at 10%. This only leaves the EGARCH (1,1) passing all the tests with highest p-value for the conditional coverage test (51.18%).

At 99% confidence level the highest number of violations is made using Moving Average with 250 observations (63). The lowest and also closest to the expected number is EGARCH (1,1) with Student's t-distribution (31 when expecting 25). The estimation horizon gives different results according to the method used, in Historical Simulation we would prefer the longer horizon and in Moving Average, the shorter, although neither of them passes the unconditional and conditional coverage criteria. The same goes for the GARCH (1,1) and EGARCH (1,1) using the normal distribution. The p-value for these methods, as well as Moving Average and Exponentially Weighted Moving Average is 0.00%. However, using the Student's t-distribution, the accuracy of the forecasts is very improved. Both GARCH (1,1) and EGARCH (1,1) pass all the tests, with EGARCH (1,1) having a slightly higher p-value (34.56%).

7.4. FTSE100

Table 7 contains the VaR results and the backtesting results for FTSE 100. The estimation period is 26/09/2001 to 29/12/2006 and the forecast period 02/01/2007 to 27/09/2016. The number of observations for the backtesting is therefore 2462. The expected number of violations for 95% confidence level is 123 and for 99% the expected violations are 25.

As we can see from table 7:

At 95% confidence level the highest number of violations is made using EWMA (166) and the lowest number with GARCH (1,1) with Student's t-distribution (114). Once again the closest to the expected number of violations is EGARCH (1,1) with Student's t-distribution (131 when expecting 123). Longer forecasting horizons seem to be yielding better results, with Historical Simulation using 250 observations again passing the unconditional coverage criterion. The normal distribution has disappointing results, diverging largely from the expected number of violations. Although GARCH (1,1) with Student's t-distribution overestimates the risk, it still manages to pass all of the three tests along with the EGARCH (1,1) with Student's t-distribution, which has the highest p-value for the conditional coverage test (71.00%).

At 99% confidence level the highest number of violations is made using both Moving Average with 250 observations and EWMA (63). The lowest number and closest to the expected number is GARCH (1,1) with Student's t-distribution (25, the same number as expected). The best estimation horizon is not the same for the two methods that differentiate the estimation horizon. Specifically, the Historical Simulation gives us better results when using longer horizon, and Moving Average when using shorter. The difference is very slight since none of these models along with the EWMA and GARCH (1,1) and EGARCH (1,1) with normal distribution pass any of the tests. Once again, the Student's t-distribution proves to be the best both for the number of violations recorded, as well as backtesting statistically. Although GARCH (1,1) predicts the same number of violations as expected and has a p-value of almost 100% for the unconditional coverage test, EGARCH (1,1) has the highest p-value (72.91%) for the total conditional coverage criterion.

7.5. DAX

Table 8 contains the VaR results and the backtesting results for DAX. The estimation period is 26/09/2001 to 29/12/2006 and the forecast period 02/01/2007 to 27/09/2016. The number of observations for the backtesting is therefore 2474. The expected number of violations for 95% confidence level is 124 and for 99% the expected violations are 25.

As we can see from table 7:

At 95% confidence level the highest number of violations is made with Moving Average using 250 observations (156) and the lowest number with GARCH (1,1) with Student's t-distribution (108). EGARCH (1,1) with Student's t-distribution again comes very close to predicting the number of violations (124, the same as expected). The estimation horizon this time is universally better when it is smaller for both methods. However, this is not enough for them to pass any of the tests. The normal distribution has a very low score for the unconditional coverage, once again proving that it underestimates risk. On the other hand, the Student's t-distribution when used with GARCH (1,1) it slightly overestimates risk, but comes close enough to pass the tests, and when used with EGARCH (1,1) it perfectly estimates the risk, locating the correct number of violations as expected and a high p-value for the conditional coverage test (60.88%).

At 99% confidence level the highest number of violations is made with the Moving Average with 100 observations (59) and the lowest number with GARCH (1,1) with

Student’s t-distribution (15). Again, very close to the expected number of violations is EGARCH (1,1) with Student’s t-distribution (26 when expecting 25). The better estimation horizon in both Historical Simulation and Moving Average is the opposite than what it is for 95%. This time the longer horizon gives us better results by a very small difference. In all of the models there is a big underestimation of risk, including the models that assume normal distribution. In the case of GARCH (1,1) using a Student’s t-distribution there is a significant overestimation of risk. Again, the optimal combination is that of EGARCH (1,1) with Student’s t-distribution, since it is very close to finding the right amount of violations according to the sample, but it also passes all of the three tests (p-value for the conditional coverage test 73.48%).

7.6. CAC

Table 9 contains the VaR results and the backtesting results for CAC. The estimation period is 26/09/2001 to 29/12/2006 and the forecast period 02/01/2007 to 27/09/2016. The number of observations for the backtesting is therefore 2494. The expected number of violations for 95% confidence level is 125 and for 99% the expected violations are 25.

As we can see from table 8:

At 95% confidence level the highest number of violations is made using Exponentially Weighted Moving Average (160) and the lowest using GARCH (1,1) with Student’s t-distribution (112). The closest to the expected number of violations is the EGARCH (1,1) with Student’s t-distribution (134 when expecting 125). The shorter estimation window is slightly better when using Historical Simulation and Moving Average. In the case of the Historical Simulation it is good enough to pass the unconditional coverage criterion at 10%. It does not however pass the other 2 tests. The methods that do pass all of the tests are again GARCH (1,1) and EGARCH (1,1) both when Student’s t-distribution is assumed. In the case of GARCH (1,1) we detect a small overestimation, which causes it to have a smaller p-value for the conditional coverage test than the EGARCH (1,1) (55.39%).

At 99% confidence level the highest number of violations is with the Moving Average with 250 observations (60) and the lowest number with GARCH (1,1) with Student’s t-distribution (16). EGARCH (1,1) with Student’s t-distribution is fairly close to the expected number (32 when expecting 25). The estimation horizon is different for the Historical Simulation and Moving Average. In the case of the Historical Simulation the number of violations is slightly more accurate when using longer horizons, and in the case of Moving Average when using shorter. Still none of these passes the unconditional and conditional criteria. GARCH (1,1) passes the independence and conditional coverage test, but because of the overestimation that it presents it does not pass the conditional coverage at 10%. This is succeeded only by the EGARCH (1,1) ((p-value for the conditional coverage test 26.12%).

7.7. ATHEX

Table 9 contains the VaR results and the backtesting results for ATHEX. The estimation period is 26/09/2001 to 29/12/2006 and the forecast period 02/01/2007 to 27/09/2016. The number of observations for the backtesting is therefore 2399. The expected number of violations for 95% confidence level is 120 and for 99% the expected violations are 24.

As we can see from table 9:

At 95% confidence level the highest number of violations is made using Historical Simulation with 100 observations (153) and the lowest using GARCH (1,1) with Student's t-distribution (103). The closest to the expected number of violations is for one more time the EGARCH (1,1) with Student's t-distribution (110 when expecting 120). The better length of the estimation period is different for the two methods that we are testing, for Historical Simulation the more accurate is the estimation window of 250 days and for Moving Average the one with 100 days. In this index the normality assumption with the GARCH (1,1) is one that offered us good results, along with the Student's t-distribution, which caused overestimation of the risk. These methods pass all of the tests, with EGARCH (1,1) with Student's t-distribution having once again the highest p-value for the conditional coverage (63.99%), since it comes very close to predicting the violations, regardless that it overestimates the risk.

At 99% confidence level the highest number of violations is made using the Moving Average with 250 observations (57) and the lowest using GARCH (1,1) with Student's t-distribution (20). EGARCH (1,1) with Student's t-distribution comes very close to the expected number, having missed only by one (25 when expecting 24). For this confidence level the length of the estimation window is consistent for both methods, showing us more accurate results for the shorter window, i.e. 100 observations. The normality assumption causes an underestimation of the risk, passing only the independence criterion. The Student's t-distribution overestimates risk when used with GARCH (1,1), although it still passes the three criteria. The EGARCH (1,1) with the Student's t-distribution estimates risk just as expected, and passes all three of the criteria with flying colors (p-value for the conditional coverage test 75.23%).

Table 4 VaR results for S&P500 Index. *VaR is computed by: Historical Simulation based on the 100 and 250 past observations (HS-100, HS-250), Moving Average based on the 100 and 250 past observations (MA-100, MA-250), Exponentially Weighted Moving Average (EWMA), GARCH (1,1) with normal and Student-t distribution (GARCH – normal, GARCH – t) and EGARCH (1,1) with normal and Student-t distribution (EGARCH – normal, EGARCH – t). The table presents the average VaR, the number of violations (number of times that the actual realized losses exceeded the VaR), along with the frequency of the violations in the out of sample data, the p-value of each backtest criterion: unconditional coverage LR_{UC}, independence LR_{IND} and conditional coverage LR_{CC}. The p-values in bold indicate that the method satisfies the criterion with 10% confidence level.*

Model	Average VaR	No of violations (Frequency)	LR_{UC}	LR_{IND}	LR_{CC}
		VaR Results for 95%			
HS-100	-1.96%	144 (5.87%)	5.33%	**58.33%**	**13.31%**
HS-250	-2.02%	141 (5.81%)	9.55%	1.99%	1.66%
MA-100	-1.89%	146 (5.95%)	3.50%	**42.28%**	7.87%
MA-250	-1.95%	152 (6.19%)	0.85%	7.16%	0.62%
EWMA	-1.84%	155 (6.32%)	0.39%	**16.60%**	0.59%
GARCH - normal	-1.85%	144 (5.87%)	5.34%	7.24%	3.08%
GARCH - t	-2.20%	103 (4.20%)	6.21%	**48.23%**	**13.71%**
EGARCH - normal	-1.75%	170 (6.93%)	0.00%	**21.13%**	0.00%
EGARCH - t	-1.99%	130 (5.30%)	**49.69%**	**20.98%**	**36.16%**
		VaR Results for 99%			
HS-100	-3.05%	40 (1.63%)	0.39%	**17.05%**	0.62%
HS-250	-3.45%	39 (1.60%)	0.68%	**15.48%**	0.93%
MA-100	-2.68%	68 (2.77%)	0.00%	**93.29%**	0.00%
MA-250	-2.75%	76 (3.09%)	0.00%	3.81%	0.00%
EWMA	-2.61%	66 (2.69%)	0.00%	**38.86%**	0.00%
GARCH - normal	-2.62%	62 (2.52%)	0.00%	**73.40%**	0.00%
GARCH - t	-3.53%	17 (0.69%)	**10.62%**	**62.20%**	**24.08%**
EGARCH - normal	-2.47%	65 (2.65%)	0.00%	**54.02%**	0.00%
EGARCH - t	-3.16%	30 (1.22%)	**28.25%**	**38.85%**	**38.71%**

Table 5 VaR results for NASDAQ Index. *VaR is computed by: Historical Simulation based on the 100 and 250 past observations (HS-100, HS-250), Moving Average based on the 100 and 250 past observations (MA-100, MA-250), Exponentially Weighted Moving Average (EWMA), GARCH (1,1) with normal and Student-t distribution (GARCH – normal, GARCH – t) and EGARCH (1,1) with normal and Student-t distribution (EGARCH – normal, EGARCH – t). The table presents the average VaR, number of violations (number of times that the actual realized losses exceeded the VaR), along with the frequency of the violations in the out of sample data., and the p-value of each backtest criterion: unconditional coverage* LR_{UC}*, independence* LR_{IND} *and conditional coverage* LR_{CC}*. The p-values in bold indicate that the method satisfies the criterion with 10% confidence level.*

Model	Average VaR	No of violations (Frequency)	LR_{UC}	LR_{IND}	LR_{CC}
		VaR Results for 95%			
HS-100	-2.13%	139 (5.66%)	**13.64%**	**45.99%**	**25.11%**
HS-250	-2.19%	143 (5.83%)	6.52%	1.11%	0.73%
MA-100	-2.08%	154 (6.28%)	0.50%	**65.58%**	1.78%
MA-250	-2.13%	152 (6.19%)	0.85%	0.71%	0.00%
EWMA	-2.03%	151 (6.15%)	1.09%	3.91%	0.47%
GARCH - normal	-2.05%	149 (6.07%)	1.78%	4.69%	0.84%
GARCH - t	-2.29%	112 (4.56%)	**31.92%**	**10.28%**	**16.09%**
EGARCH - normal	-1.98%	159 (6.48%)	0.12%	**12.26%**	0.16%
EGARCH - t	-2.14%	131 (5.34%)	**44.12%**	0.32%	0.96%
		VaR Results for 99%			
HS-100	-3.38%	44 (1.79%)	0.00%	5.06%	0.00%
HS-250	-3.70%	43 (1.75%)	0.00%	**22.32%**	0.15%
MA-100	-2.94%	68 (2.77%)	0.00%	**16.54%**	0.00%
MA-250	-3.01%	70 (2.85%)	0.00%	**19.65%**	0.00%
EWMA	-2.87%	60 (2.44%)	0.00%	**66.95%**	0.00%
GARCH - normal	-2.89%	56 (2.28%)	0.00%	**54.57%**	0.00%
GARCH - t	-3.52%	22 (0.89%)	**60.27%**	**52.79%**	**71.55%**
EGARCH - normal	-2.80%	57 (2.32%)	0.00%	**57.59%**	0.00%
EGARCH - t	-3.27%	33 (1.34%)	**10.21%**	**34.26%**	**16.75%**

Table 6 VaR results for EUROSTOXX Index*. VaR is computed by: Historical Simulation based on the 100 and 250 past observations (HS-100, HS-250), Moving Average based on the 100 and 250 past observations (MA-100, MA-250), Exponentially Weighted Moving Average (EWMA), GARCH (1,1) with normal and Student-t distribution (GARCH – normal, GARCH – t) and EGARCH (1,1) with normal and Student-t distribution (EGARCH – normal, EGARCH – t). The table presents the average VaR, number of violations (number of times that the actual realized losses exceeded the VaR), along with the frequency of the violations in the out of sample data., and the p-value of each backtest criterion: unconditional coverage* LR_{UC}*, independence* LR_{IND} *and conditional coverage* LR_{CC}*. The p-values in bold indicate that the method satisfies the criterion with 10% confidence level.*

Model	Average VaR	No of violations (Frequency)	LR_{UC}	LR_{IND}	LR_{CC}
		VaR Results for 95%			
HS-100	-2.43%	144 (5.75%)	8.92%	5.28%	3.62%
HS-250	-2.45%	141 (5.63%)	**15.11%**	0.68%	0.93%
MA-100	-2.36%	149 (5.95%)	3.26%	0.36%	0.15%
MA-250	-2.39%	148 (5.91%)	4.04%	0.00%	0.00%
EWMA	-2.32%	157 (6.27%)	0.47%	**48.00%**	1.44%
GARCH - normal	-2.33%	154 (6.15%)	1.02%	**60.84%**	3.25%
GARCH - t	-2.66%	107 (4.27%)	8.98%	**41.28%**	**16.96%**
EGARCH - normal	-2.26%	148 (5.91%)	4.05%	**93.20%**	**12.21%**
EGARCH - t	-2.45%	128 (5.11%)	**78.74%**	**26.03%**	**51.18%**
		VaR Results for 99%			
HS-100	-3.80%	42 (1.67%)	0.18%	**23.08%**	0.39%
HS-250	-3.95%	37 (1.47%)	2.44%	**29.17%**	4.57%
MA-100	-3.34%	56 (2.23%)	0.00%	**52.94%**	0.00%
MA-250	-3.38%	63 (2.51%)	0.00%	0.52%	0.00%
EWMA	-3.28%	51 (2.03%)	0.00%	**14.50%**	0.00%
GARCH - normal	-3.29%	48 (1.91%)	0.00%	**17.04%**	0.00%
GARCH - t	-3.72%	33 (1.31%)	**12.58%**	**34.74%**	**19.93%**
EGARCH - normal	-3.20%	52 (2.07%)	0.00%	**13.72%**	0.00%
EGARCH - t	-3.73%	31 (1.23%)	**24.59%**	**37.76%**	**34.56%**

Table 7 VaR results for FTSE100 Index. *VaR is computed by: Historical Simulation based on the 100 and 250 past observations (HS-100, HS-250), Moving Average based on the 100 and 250 past observations (MA-100, MA-250), Exponentially Weighted Moving Average (EWMA), GARCH (1,1) with normal and Student-t distribution (GARCH – normal, GARCH – t) and EGARCH (1,1) with normal and Student-t distribution (EGARCH – normal, EGARCH – t). The table presents the average VaR, number of violations (number of times that the actual realized losses exceeded the VaR), along with the frequency of the violations in the out of sample data, and the p-value of each backtest criterion: unconditional coverage* LR_{UC}, *independence* LR_{IND} *and conditional coverage* LR_{CC}. *The p-values in bold indicate that the method satisfies the criterion with 10% confidence level.*

Model	Average VaR	No of violations (Frequency)	LR_{UC}	LR_{IND}	LR_{CC}
		VaR Results for 95%			
HS-100	-1.99%	142 (5.76%)	8.76%	0.00%	0.00%
HS-250	-2.01%	136 (5.52%)	**24.04%**	0.97%	1.78%
MA-100	-1.90%	157 (6.37%)	0.25%	0.00%	0.00%
MA-250	-1.93%	147 (5.97%)	3.17%	0.79%	0.29%
EWMA	-1.86%	166 (6.74%)	0.00%	0.99%	0.00%
GARCH - normal	-1.88%	148 (6.01%)	2.54%	**69.99%**	7.64%
GARCH - t	-2.08%	114 (4.63%)	**39.44%**	**45.30%**	**52.20%**
EGARCH - normal	-1.80%	155 (6.29%)	0.45%	**67.79%**	1.62%
EGARCH - t	-1.93%	131 (5.32%)	**46.95%**	**68.75%**	**71.00%**
		VaR Results for 99%			
HS-100	-3.08%	44 (1.78%)	0.00%	0.00%	0.00%
HS-250	-3.28%	40 (1.62%)	0.42%	0.36%	0.00%
MA-100	-2.69%	61 (2.47%)	0.00%	0.00%	0.00%
MA-250	-2.74%	63 (2.55%)	0.00%	0.00%	0.00%
EWMA	-2.63%	63 (2.55%)	0.00%	2.65%	0.00%
GARCH - normal	-2.66%	49 (1.99%)	0.00%	**35.18%**	0.00%
GARCH - t	-3.18%	25 (1.01%)	**93.88%**	**25.52%**	**52.19%**
EGARCH - normal	-2.55%	52 (2.11%)	0.00%	**92.22%**	0.00%
EGARCH - t	-2.89%	26 (1.05%)	**78.18%**	**45.62%**	**72.91%**

Table 8 VaR results for DAX Index*. VaR is computed by: Historical Simulation based on the 100 and 250 past observations (HS-100, HS-250), Moving Average based on the 100 and 250 past observations (MA-100, MA-250), Exponentially Weighted Moving Average (EWMA), GARCH (1,1) with normal and Student-t distribution (GARCH – normal, GARCH – t) and EGARCH (1,1) with normal and Student-t distribution (EGARCH – normal, EGARCH – t). The table presents the average VaR, number of violations (number of times that the actual realized losses exceeded the VaR), along with the frequency of the violations in the out of sample data., and the p-value of each backtest criterion: unconditional coverage LR_{UC}, independence LR_{IND} and conditional coverage LR_{CC}. The p-values in bold indicate that the method satisfies the criterion with 10% confidence level.*

Model	Average VaR	No of violations (Frequency)	LR_{UC}	LR_{IND}	LR_{CC}
		VaR Results for 95%			
HS-100	-2.34%	144 (5.82%)	6.76%	0.00%	0.00%
HS-250	-2.39%	148 (5.98%)	2.94%	0.00%	0.00%
MA-100	-2.26%	153 (6.18%)	0.90%	0.75%	0.00%
MA-250	-2.29%	156 (6.30%)	0.41%	0.00%	0.00%
EWMA	-2.22%	154 (6.22%)	0.70%	8.20%	0.58%
GARCH - normal	-2.23%	148 (5.98%)	2.95%	**75.59%**	8.90%
GARCH - t	-2.56%	108 (4.36%)	**13.91%**	**72.37%**	**31.47%**
EGARCH - normal	-2.30%	153 (6.18%)	0.90%	**37.30%**	2.23%
EGARCH - t	-2.57%	124 (5.01%)	**97.79%**	**31.93%**	**60.88%**
		VaR Results for 99%			
HS-100	-3.58%	43 (1.73%)	0.00%	**77.73%**	0.36%
HS-250	-3.79%	37 (1.49%)	2.10%	**58.33%**	6.00%
MA-100	-3.19%	59 (2.38%)	0.00%	**22.92%**	0.00%
MA-250	-3.23%	55 (2.22%)	0.00%	0.78%	0.00%
EWMA	-3.14%	55 (2.22%)	0.00%	**11.37%**	0.00%
GARCH - normal	-3.15%	47 (1.89%)	0.00%	**91.00%**	0.00%
GARCH - t	-4.01%	15 (0.60%)	3.37%	**66.87%**	9.58%
EGARCH - normal	-3.20%	50 (2.02%)	0.00%	**99.11%**	0.00%
EGARCH - t	-3.83%	26 (1.05%)	**80.07%**	**45.73%**	**73.48%**

***Table 9 VaR results for CAC Index**. VaR is computed by: Historical Simulation based on the 100 and 250 past observations (HS-100, HS-250), Moving Average based on the 100 and 250 past observations (MA-100, MA-250), Exponentially Weighted Moving Average (EWMA), GARCH (1,1) with normal and Student-t distribution (GARCH – normal, GARCH – t) and EGARCH (1,1) with normal and Student-t distribution (EGARCH – normal, EGARCH – t). The table presents the average VaR, number of violations (number of times that the actual realized losses exceeded the VaR) along with the frequency of the violations in the out of sample data, and the p-value of each backtest criterion: unconditional coverage LR_{UC}, independence LR_{IND} and conditional coverage LR_{CC}. The p-values in bold indicate that the method satisfies the criterion with 10% confidence level.*

Model	Average VaR	No of violations (Frequency)	LR_{UC}	LR_{IND}	LR_{CC}
		VaR Results for 95%			
HS-100	-2.40%	143 (5.73%)	**10.00%**	0.94%	0.89%
HS-250	-2.42%	146 (5.85%)	5.64%	0.00%	0.00%
MA-100	-2.35%	150 (6.01%)	2.40%	2.39%	0.61%
MA-250	-2.38%	153 (6.13%)	1.19%	0.00%	0.00%
EWMA	-2.30%	160 (6.41%)	0.19%	**37.93%**	0.53%
GARCH - normal	-2.34%	155 (6.21%)	0.72%	**90.13%**	2.69%
GARCH - t	-2.60%	112 (4.49%)	**23.54%**	**30.80%**	**29.42%**
EGARCH - normal	-2.25%	152 (6.09%)	1.52%	**80.00%**	5.07%
EGARCH - t	-2.39%	134 (5.37%)	**39.83%**	**49.39%**	**55.39%**
		VaR Results for 99%			
HS-100	-3.77%	41 (1.64%)	0.31%	**24.16%**	0.64%
HS-250	-3.93%	39 (1.56%)	0.89%	**26.55%**	1.77%
MA-100	-3.32%	53 (2.12%)	0.00%	**90.10%**	0.00%
MA-250	-3.36%	60 (2.40%)	0.00%	0.31%	0.00%
EWMA	-3.26%	49 (1.96%)	0.00%	**96.96%**	0.00%
GARCH - normal	-3.30%	48 (1.92%)	0.00%	**93.67%**	0.00%
GARCH - t	-4.02%	16 (0.64%)	5.42%	**64.94%**	**14.12%**
EGARCH - normal	-3.18%	43 (1.72%)	0.00%	**21.92%**	0.21%
EGARCH - t	-3.61%	32 (1.28%)	**17.34%**	**36.16%**	**26.12%**

Table 10 VaR results for ATHEX Index. *VaR is computed by: Historical Simulation based on the 100 and 250 past observations (HS-100, HS-250), Moving Average based on the 100 and 250 past observations (MA-100, MA-250), Exponentially Weighted Moving Average (EWMA), GARCH (1,1) with normal and Student-t distribution (GARCH – normal, GARCH – t) and EGARCH (1,1) with normal and Student-t distribution (EGARCH – normal, EGARCH – t). The table presents the average VaR, number of violations (number of times that the actual realized losses exceeded the VaR), along with the frequency of the violations in the out of sample data, and the p-value of each backtest criterion: unconditional coverage* LR_{UC}*, independence* LR_{IND} *and conditional coverage* LR_{CC}*. The p-values in bold indicate that the method satisfies the criterion with 10% confidence level.*

Model	Average VaR	No of violations (Frequency)	LR_{UC}	LR_{IND}	LR_{CC}
		VaR Results for 95%			
HS-100	-3.51%	153 (6.37%)	0.29%	0.00%	0.00%
HS-250	-3.60%	143 (5.96%)	3.58%	0.00%	0.00%
MA-100	-3.52%	138 (5.75%)	9.83%	0.00%	0.00%
MA-250	-3.53%	144 (6.00%)	2.87%	0.00%	0.00%
EWMA	-3.48%	143 (5.96%)	3.59%	0.57%	0.24%
GARCH - normal	-3.47%	131 (5.46%)	**30.74%**	**28.57%**	**33.60%**
GARCH - t	-3.86%	103 (4.29%)	**10.40%**	**77.89%**	**25.65%**
EGARCH - normal	-3.35%	143 (5.96%)	3.59%	**60.06%**	9.64%
EGARCH - t	-3.70%	110 (4.58%)	**34.48%**	**98.29%**	**63.99%**
		VaR Results for 99%			
HS-100	-6.06%	36 (1.50%)	2.17%	**56.94%**	6.12%
HS-250	-5.95%	39 (1.62%)	0.47%	**16.13%**	0.69%
MA-100	-4.99%	52 (2.16%)	0.00%	**44.71%**	0.00%
MA-250	-4.99%	57 (2.37%)	0.00%	0.22%	0.00%
EWMA	-4.92%	48 (2.00%)	0.00%	**34.25%**	0.00%
GARCH - normal	-4.91%	46 (1.91%)	0.00%	**90.05%**	0.00%
GARCH - t	-5.96%	20 (0.83%)	**39.93%**	**56.19%**	**59.25%**
EGARCH - normal	-4.74%	46 (1.91%)	0.00%	**90.05%**	0.00%
EGARCH - t	-5.68%	25 (1.04%)	**83.69%**	**46.80%**	**75.23%**

8. Conclusion

We estimated the VaR for seven stock indices from U.S. and European markets: S&P 500, Nasdaq, Euro Stoxx, FTSE 100, DAX, CAC and ATHEX. The methods that we chose were Historical Simulation and Variance-Covariance with volatility estimated by: Moving Average, Exponentially Weighted Moving Average, GARCH (1,1) and EGARCH (1,1). We computed one day forecasts for VaR for 95% and 99% confidence level. For Historical Simulation and Moving Average we used two different estimation windows of 100 and 250 observations. GARCH (1,1) and EGARCH (1,1) were estimated using the normal and the Student's t-distribution.

The Historical Simulation was not very precise in the accuracy of the forecasts. There was only one case where it passed all of the three criteria. This was in the case of the NASDAQ index, as it is also suggested by Kuester et al (2006). The estimation sample size has proved quite inconsistent across indices as well as confidence levels, many times giving different preference for different confidence level in the same index. There is not one particular pattern for us to reach a conclusion as to what sample size is better, although the empirical results seem to give more preference to the longer sample size (250 observations).

The Variance-Covariance method with volatility estimated by Moving Average did not pass the conditional or unconditional criterion for any index. In facts, in most of the indices it provided the least accurate forecasts, largely underestimating the risk. This is especially true for the 99% confidence level and Moving Average using 250 observations, which provided the highest number of violations across all indices with the exception of DAX. In that case, the Moving Average using 100 observations performed the worse. This is also an indication about the sample size of the estimation window, concluding that perhaps a smaller size would not yield as inaccurate results as the bigger sample size. The bad forecasting ability of this method may be the reason why it is not that often tested in the existing literature.

The Variance-Covariance method with volatility estimated by Exponentially Weighted Moving Average was also one of the worst performing methods. It only passed the independence test in some of the cases, and it did not pass the conditional or unconditional criterion. In the case of FTSE 100 it resulted with the highest number of violations for both confidence levels. For EUROSTOXX and CAC it was the worst performing model only for the 95% confidence level, for 99% confidence level it was the Moving Average with 250 observations as mentioned before. For the U.S. stock indices, it did not have the worst performance as it did in the aforementioned European ones. This is consistent with the findings of Alexander & Leigh (1997).

When comparing the normal and the Student's t-distribution we can clearly see that for all indices and all confidence levels the normal distribution performed better than the previous methods analyzed but it still underestimated risk, causing it not to be the best choice for a VaR model. The Student's t-distribution on the other hand performed much better than the normal, although in some cases it overestimated the risk (as confirmed by Billio and Pelizzon (2000) and Guermat and Harris (2002)), causing the unnecessary allocation of capital for the purpose of risk management. This is consistent with the descriptive statistics of the data which rejected the hypothesis of normality, and the kurtosis which also pointed towards the need for a leptokurtic distribution.

The Variance-Covariance method with volatility estimated by GARCH (1,1) with normal distribution passed all the tests only in the case of ATHEX, giving a good estimate of the number of violations. The Variance-Covariance method with volatility estimated by EGARCH (1,1) with normal distribution did not pass the unconditional or conditional coverage criterion in any of the indices for any of the confidence level.

The Variance-Covariance method with volatility estimated by GARCH (1,1) with Student's t-distribution painted a totally different picture. It passed all of the three criteria in almost all of the indices. In the cases where it did not pass the unconditional coverage test is because we chose the 10% level of p-value as a point for accepting or rejecting the null hypothesis. Had we been less conservative and had chosen a 5% level for the p-value it would have passed all the tests in all of the indices. These problems are resolved when using an EGARCH (1,1) with Student's t-distribution for the estimation of volatility. In almost all of the cases it forecasted the exact number of violations as expected. This fact, along with the statistical backtesting results made it the most accurate method in all of the indices and for all of the confidence levels. Clearly this points out to the significance of the leverage effect, as most of the leverage parameters that were estimated had a negative sign. This points out to the fact that negative returns have bigger impact to the volatility than positive impact. These results are not consistent with the study by Angelidis and Benos (2015).

The general conclusion is that the biggest difference comes from the distribution assumed. For data that stray from the normal distribution different distributions have to be assumed. The sample size does not play a part in the accuracy of the models, since the inaccurate ones continued to be inaccurate even when the sample size was changed. The leverage effect exists and it is very important in estimating the volatility.

These conclusions are shared by all of the indices, regardless of the financial market in which they are traded, whether it's in the U.S. or Europe. The U.S. indices share a few characteristics, as well as Euro Stoxx and FTSE 100. Aside from this, the results in general are the same for every index.

Bibliography

Alexander, C.O. and Leigh, C.T., 1997. On the covariance matrices used in value at risk models. *The Journal of Derivatives*, *4*(3), pp.50-62.

Alexander, C., 2008. *Value-at-risk models*. John Wiley & Sons.

Angelidis, T. and Benos, A., 2015. Value-at-risk for Greek stocks. *Multinational Finance Journal, pp. 67-104.*

Angelidis, T., Benos, A. and Degiannakis, S., 2004. The use of GARCH models in VaR estimation. *Statistical methodology*, *1*(1), pp.105-128.

Artzner, P., Delbaen, F., Eber, J.M. and Heath, D., 1999. Coherent measures of risk. *Mathematical finance*, *9*(3), pp.203-228.

Basle Committee on Banking Supervision, 1995. *An Internal Model–Based Approach to Market Risk Capital Requirements. Bank for International Settlement*

Basle Committee on Banking Supervision, 2004. *International convergence of capital measurement and capital standards: a revised framework*. Bank for International Settlements.

Basle Committee on Banking Supervision, 1996. *Overview of the amendment to the capital accord to incorporate market risks*. Bank for International Settlements.

Beder, T.S., 1995. VAR: Seductive but dangerous. *Financial Analysts Journal*, pp.12-24.

Billio, M. and Pelizzon, L., 2000. Value-at-risk: a multivariate switching regime approach. *Journal of Empirical Finance*, *7*(5), pp.531-554.

Bollerslev, T., 1986. Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics*, *31*(3), pp.307-327.

Bollerslev, T., 1990. Modelling the coherence in short-run nominal exchange rates: a multivariate generalized ARCH model. *The review of economics and statistics*, pp.498-505.

Brooks, C., 2014. *Introductory econometrics for finance*. Cambridge university press.

Christoffersen, P.F., 2012. *Elements of financial risk management*. Academic Press.

Crouhy, M., Galai, D. and Mark, R., 2006. *The essentials of risk management* (Vol. 1). New York: McGraw-Hill.

Danielsson, J., Embrechts, P., Goodhart, C., Keating, C., Muennich, F., Renault, O. and Shin, H.S., 2001. An academic response to Basel II. *Special Paper-LSE Financial Markets Group*.

Daníelsson, J., 2011. *Financial risk forecasting: the theory and practice of forecasting market risk with implementation in R and Matlab* (Vol. 588). John Wiley & Sons.

Danıelsson, J., 2002. The emperor has no clothes: Limits to risk modelling. *Journal of Banking & Finance*, *26*(7), pp.1273-1296

Deutsche Börse Group, 2012. DAX INDEX Factsheet. *Deutsche Börse Group*

Duffie, D. and Pan, J., 1997. An overview of value at risk. *The Journal of derivatives*, *4*(3), pp.7-49.

Engle, R.F., 1982. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica: Journal of the Econometric Society*, pp.987-1007.

Euronext, 2016. CAC 40 Factsheet. *Euronext*

Fallon, W., 1996. *Calculating value-at-risk* (pp. 96-49). Wharton School, University of Pennsylvania.

Figlewski, S., 1997. Forecasting volatility. *Financial markets, institutions & instruments*, *6*(1), pp.1-88.

FTSE Group, 2012. FTSE All Share Index Factsheet. *London: FTSE Group. Retrieved*, *4*.

Gallati, R., 2003. *Risk management and capital adequacy*. McGraw Hill Professional.

Glosten, L.R., Jagannathan, R. and Runkle, D.E., 1993. On the relation between the expected value and the volatility of the nominal excess return on stocks. *The journal of finance*, *48*(5), pp.1779-1801.

Guermat, C. and Harris, R.D., 2002. Forecasting value at risk allowing for time variation in the variance and kurtosis of portfolio returns. *International Journal of Forecasting*, *18*(3), pp.409-419.

Hull, J.C., 2006. *Options, futures, and other derivatives*. Pearson Education India.

Jackson, P., Maude, D.J. and Perraudin, W., 1997. Bank capital and value at risk. *The Journal of Derivatives*, *4*(3), pp.73-89.

Kuester, K., Mittnik, S. and Paolella, M.S., 2006. Value-at-risk prediction: A comparison of alternative strategies. *Journal of Financial Econometrics*, *4*(1), pp.53-89.

Lambadiaris, G., Papadopoulou, L., Skiadopoulos, G. and Zoulis, Y., 2003. VAR: history or simulation. *Risk*, *16*(9), pp.122-127.

Morgan, J.P., 1996. Reuters. *RiskMetrics—Technical Document*.

Nelson, D.B., 1991. Conditional heteroskedasticity in asset returns: A new approach. *Econometrica: Journal of the Econometric Society*, pp.347-370.

Standard and Poor's, 2004. S&P 500 fact sheet.

Taylor, S.J., 2008. *Modelling financial time series*. world scientific.

Taylor, J.W., 2008. Estimating value at risk and expected shortfall using expectiles. *Journal of Financial Econometrics*, *6*(2), pp.231-252.

Tsay, R.S., 2005. *Analysis of financial time series* (Vol. 543). John Wiley & Sons.

www.nasdaq.com

www.bloomber.com

www.athexgroup.gr

Zikovic, S. and Filer, R.K., 2009. Hybrid historical simulation VaR and ES: performance in developed and emerging markets.

Zivot, E. and Wang, J., 2003. Rolling Analysis of Time Series. In *Modeling Financial Time Series with S-Plus®* (pp. 299-346). Springer New York.

Appendix A: ADF-GLS unit root tests

***Table 11 ADF-GLS unit root test for S&P 500 prices**. The null hypothesis is H0: there is unit root. We compare the absolute value of the t-Statistic with the Test critical values at 1%, 5% and 10% level. When the t-Statistic is higher than the critical values, we reject the hypothesis, otherwise we accept it.*

Null Hypothesis: S_PCOMP has a unit root
Exogenous: Constant
Lag Length: 1 (Automatic - based on SIC, maxlag=29)

		t-Statistic
Elliott-Rothenberg-Stock DF-GLS test statistic		0.700444
Test critical values:	1% level	-2.565575
	5% level	-1.940908
	10% level	-1.616643

***Table 12 ADF-GLS unit root test for S&P 500 returns**. The null hypothesis is H0: there is unit root. We compare the absolute value of the t-Statistic with the Test critical values at 1%, 5% and 10% level. When the t-Statistic is higher than the critical values, we reject the hypothesis, otherwise we accept it.*

Null Hypothesis: D(S_PCOMP) has a unit root
Exogenous: Constant
Lag Length: 23 (Automatic - based on SIC, maxlag=29)

		t-Statistic
Elliott-Rothenberg-Stock DF-GLS test statistic		-2.678501
Test critical values:	1% level	-2.565578
	5% level	-1.940908
	10% level	-1.616643

***Table 13 ADF-GLS unit root test for NASDAQ prices.** The null hypothesis is H0: there is unit root. We compare the absolute value of the t-Statistic with the Test critical values at 1%, 5% and 10% level. When the t-Statistic is higher than the critical values, we reject the hypothesis, otherwise we accept it.*

Null Hypothesis: NASCOMP has a unit root
Exogenous: Constant
Lag Length: 0 (Automatic - based on SIC, maxlag=29)

		t-Statistic
Elliott-Rothenberg-Stock DF-GLS test statistic		1.493368
Test critical values:	1% level	-2.565574
	5% level	-1.940908
	10% level	-1.616643

Table 14 ADF-GLS unit root test for NASDAQ returns*. The null hypothesis is H0: there is unit root. We compare the absolute value of the t-Statistic with the Test critical values at 1%, 5% and 10% level. When the t-Statistic is higher than the critical values, we reject the hypothesis, otherwise we accept it.*

Null Hypothesis: D(NASCOMP) has a unit root
Exogenous: Constant
Lag Length: 0 (Automatic - based on SIC, maxlag=29)

		t-Statistic
Elliott-Rothenberg-Stock DF-GLS test statistic		-61.89388
Test critical values:	1% level	-2.565575
	5% level	-1.940908
	10% level	-1.616643

Table 15 ADF-GLS unit root test for EUROSTOXX prices*. The null hypothesis is H0: there is unit root. We compare the absolute value of the t-Statistic with the Test critical values at 1%, 5% and 10% level. When the t-Statistic is higher than the critical values, we reject the hypothesis, otherwise we accept it.*

Null Hypothesis: DJEURST has a unit root
Exogenous: Constant
Lag Length: 0 (Automatic - based on SIC, maxlag=29)

		t-Statistic
Elliott-Rothenberg-Stock DF-GLS test statistic		-2.276021
Test critical values:	1% level	-2.565563
	5% level	-1.940906
	10% level	-1.616644

Table 16 ADF-GLS unit root test for EUROSTOXX returns*. The null hypothesis is H0: there is unit root. We compare the absolute value of the t-Statistic with the Test critical values at 1%, 5% and 10% level. When the t-Statistic is higher than the critical values, we reject the hypothesis, otherwise we accept it.*

Null Hypothesis: D(DJEURST) has a unit root
Exogenous: Constant
Lag Length: 29 (Automatic - based on SIC, maxlag=29)

		t-Statistic
Elliott-Rothenberg-Stock DF-GLS test statistic		-1.552763
Test critical values:	1% level	-2.565568
	5% level	-1.940907
	10% level	-1.616644

Table 17 ADF-GLS unit root test for FTSE100 prices*. The null hypothesis is H0: there is unit root. We compare the absolute value of the t-Statistic with the Test critical values at 1%, 5% and 10% level. When the t-Statistic is higher than the critical values, we reject the hypothesis, otherwise we accept it.*

Null Hypothesis: FTSE100 has a unit root
Exogenous: Constant
Lag Length: 0 (Automatic - based on SIC, maxlag=29)

		t-Statistic
Elliott-Rothenberg-Stock DF-GLS test statistic		-1.109256
Test critical values:	1% level	-2.565572
	5% level	-1.940908
	10% level	-1.616643

Table 18 ADF-GLS unit root test for FTSE100 returns*. The null hypothesis is H0: there is unit root. We compare the absolute value of the t-Statistic with the Test critical values at 1%, 5% and 10% level. When the t-Statistic is higher than the critical values, we reject the hypothesis, otherwise we accept it.*

Null Hypothesis: D(FTSE100) has a unit root
Exogenous: Constant
Lag Length: 25 (Automatic - based on SIC, maxlag=29)

		t-Statistic
Elliott-Rothenberg-Stock DF-GLS test statistic		-1.841163
Test critical values:	1% level	-2.565576
	5% level	-1.940908
	10% level	-1.616643

Table 19 ADF-GLS unit root test for DAX prices*. The null hypothesis is H0: there is unit root. We compare the absolute value of the t-Statistic with the Test critical values at 1%, 5% and 10% level. When the t-Statistic is higher than the critical values, we reject the hypothesis, otherwise we accept it.*

Null Hypothesis: DAXINDX has a unit root
Exogenous: Constant
Lag Length: 0 (Automatic - based on SIC, maxlag=29)

		t-Statistic
Elliott-Rothenberg-Stock DF-GLS test statistic		0.163522
Test critical values:	1% level	-2.565569
	5% level	-1.940907
	10% level	-1.616644

Table 20 ADF-GLS unit root test for DAX returns*. The null hypothesis is H0: there is unit root. We compare the absolute value of the t-Statistic with the Test critical values at 1%, 5% and 10% level. When the t-Statistic is higher than the critical values, we reject the hypothesis, otherwise we accept it.*

Null Hypothesis: D(DAXINDX) has a unit root
Exogenous: Constant
Lag Length: 21 (Automatic - based on SIC, maxlag=29)

		t-Statistic
Elliott-Rothenberg-Stock DF-GLS test statistic		-2.487480
Test critical values:	1% level	-2.565572
	5% level	-1.940908
	10% level	-1.616643

Table 21 ADF-GLS unit root test for CAC prices*. The null hypothesis is H0: there is unit root. We compare the absolute value of the t-Statistic with the Test critical values at 1%, 5% and 10% level. When the t-Statistic is higher than the critical values, we reject the hypothesis, otherwise we accept it.*

Null Hypothesis: FRCAC40 has a unit root
Exogenous: Constant
Lag Length: 0 (Automatic - based on SIC, maxlag=29)

		t-Statistic
Elliott-Rothenberg-Stock DF-GLS test statistic		-2.241200
Test critical values:	1% level	-2.565564
	5% level	-1.940906
	10% level	-1.616644

Table 22 ADF-GLS unit root test for CAC returns*. The null hypothesis is H0: there is unit root. We compare the absolute value of the t-Statistic with the Test critical values at 1%, 5% and 10% level. When the t-Statistic is higher than the critical values, we reject the hypothesis, otherwise we accept it.*

Null Hypothesis: D(FRCAC40) has a unit root
Exogenous: Constant
Lag Length: 25 (Automatic - based on SIC, maxlag=29)

		t-Statistic
Elliott-Rothenberg-Stock DF-GLS test statistic		-2.771353
Test critical values:	1% level	-2.565568
	5% level	-1.940907
	10% level	-1.616644

Table 23 ADF-GLS unit root test for ATHEX prices. *The null hypothesis is H0: there is unit root. We compare the absolute value of the t-Statistic with the Test critical values at 1%, 5% and 10% level. When the t-Statistic is higher than the critical values, we reject the hypothesis, otherwise we accept it.*

Null Hypothesis: GRAGENL has a unit root
Exogenous: Constant
Lag Length: 1 (Automatic - based on SIC, maxlag=29)

		t-Statistic
Elliott-Rothenberg-Stock DF-GLS test statistic		-0.480849
Test critical values:	1% level	-2.565586
	5% level	-1.940909
	10% level	-1.616642

Table 24 ADF-GLS unit root test for ATHEX returns*. The null hypothesis is H0: there is unit root. We compare the absolute value of the t-Statistic with the Test critical values at 1%, 5% and 10% level. When the t-Statistic is higher than the critical values, we reject the hypothesis, otherwise we accept it.*

Null Hypothesis: D(GRAGENL) has a unit root
Exogenous: Constant
Lag Length: 17 (Automatic - based on SIC, maxlag=29)

		t-Statistic
Elliott-Rothenberg-Stock DF-GLS test statistic		-6.616319
Test critical values:	1% level	-2.565589
	5% level	-1.940910
	10% level	-1.616642

Appendix B.1: Sample autocorrelation function (ACF) and partial autocorrelation function (PACF) of returns

Figure 3 ACF & PACF of S&P500 returns

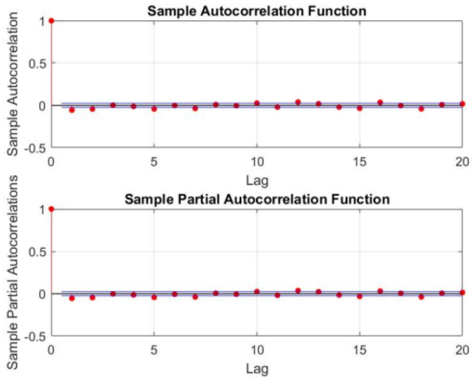


Figure 4 ACF & PACF of NASDAQ returns

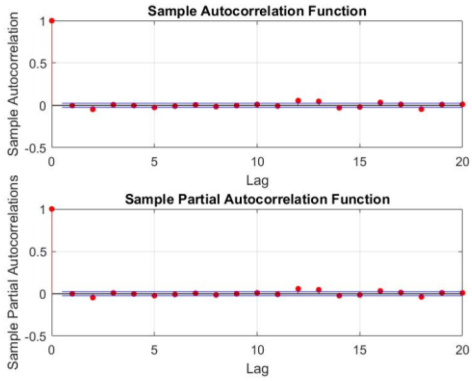


Figure 5 ACF & PACF of EUROSTOXX returns

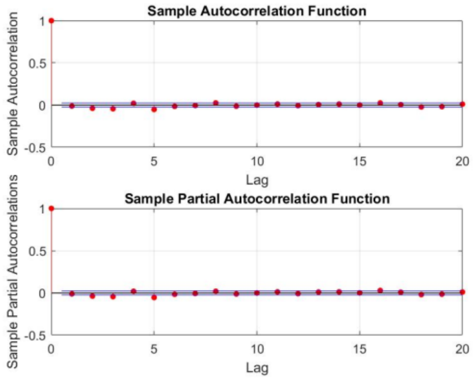


Figure 6 ACF & PACF of FTSE 100 returns

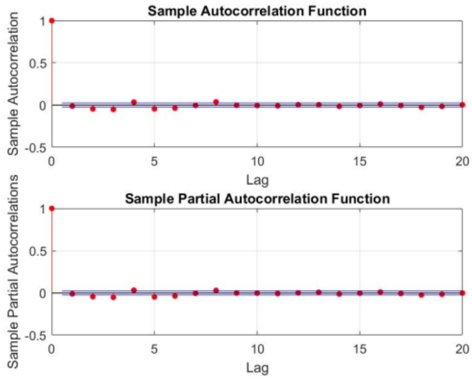


Figure 7 ACF & PACF of DAX returns

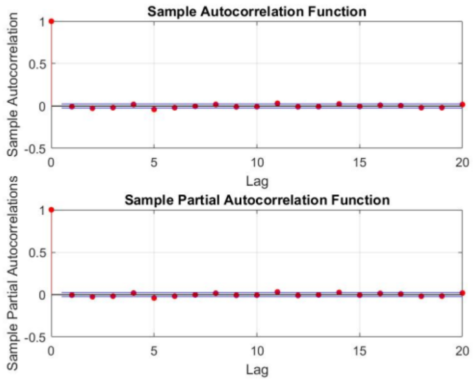


Figure 8 ACF & PACF of CAC returns

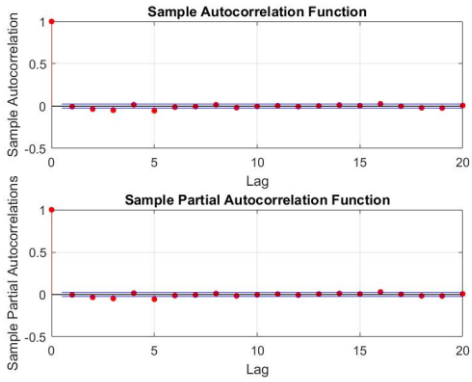
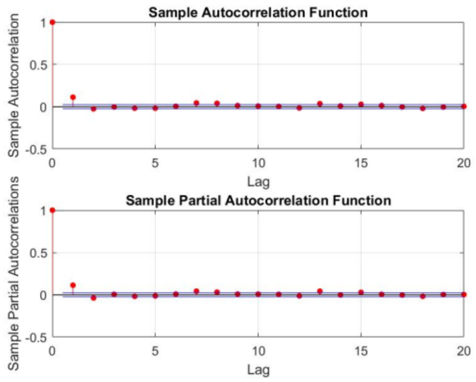


Figure 9 ACF & PACF of ATHEX returns



Appendix B.2: Sample autocorrelation function (ACF) and partial autocorrelation function (PACF) of squared returns

Figure 10 ACF & PACF of S&P500 squared returns

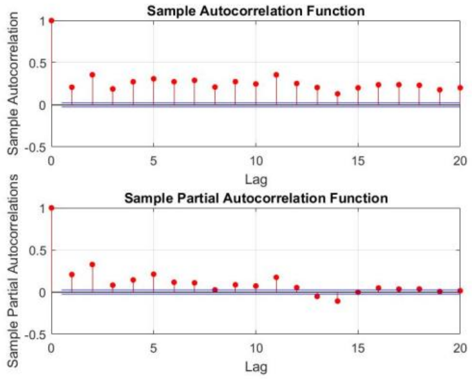


Figure 11 ACF & PACF of NASDAQ squared returns

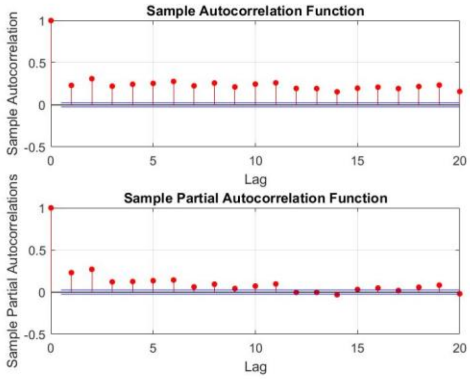


Figure 12 ACF & PACF of EUROSTOXX squared returns

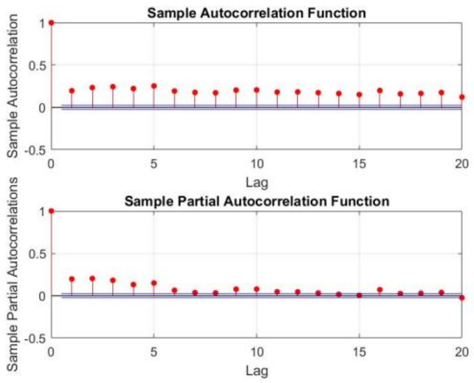


Figure 13 ACF & PACF of FTSE100 squared returns

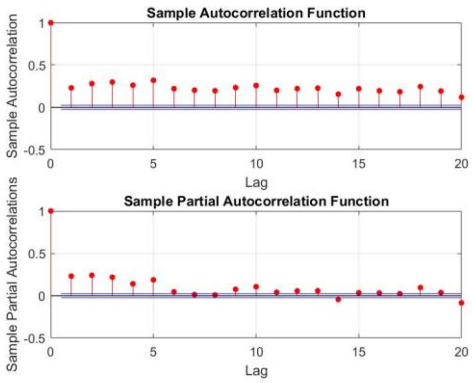


Figure 14 ACF & PACF of DAX squared returns

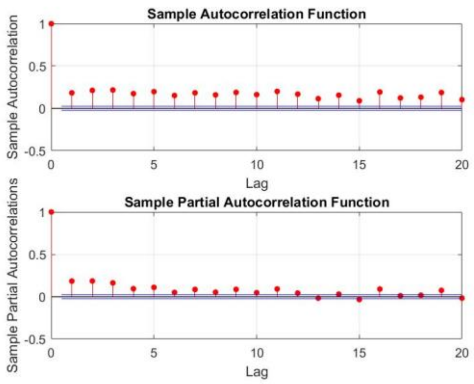


Figure 15 ACF & PACF of CAC squared returns

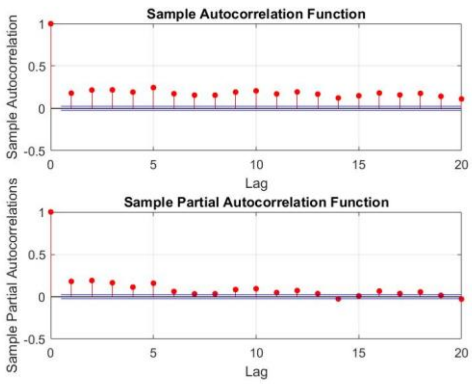
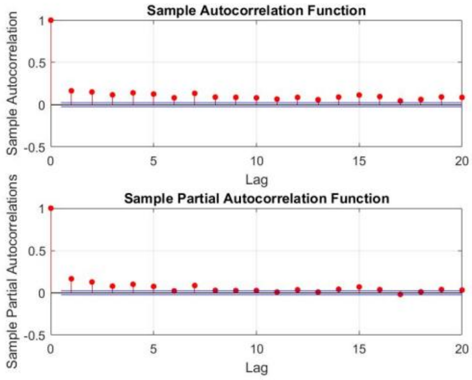


Figure 16 ACF & PACF of ATHEX squared returns



Appendix C: PLOTS

C.1. S&P 500

Figure 17 VaR Estimation for S&P using the Historical Simulation

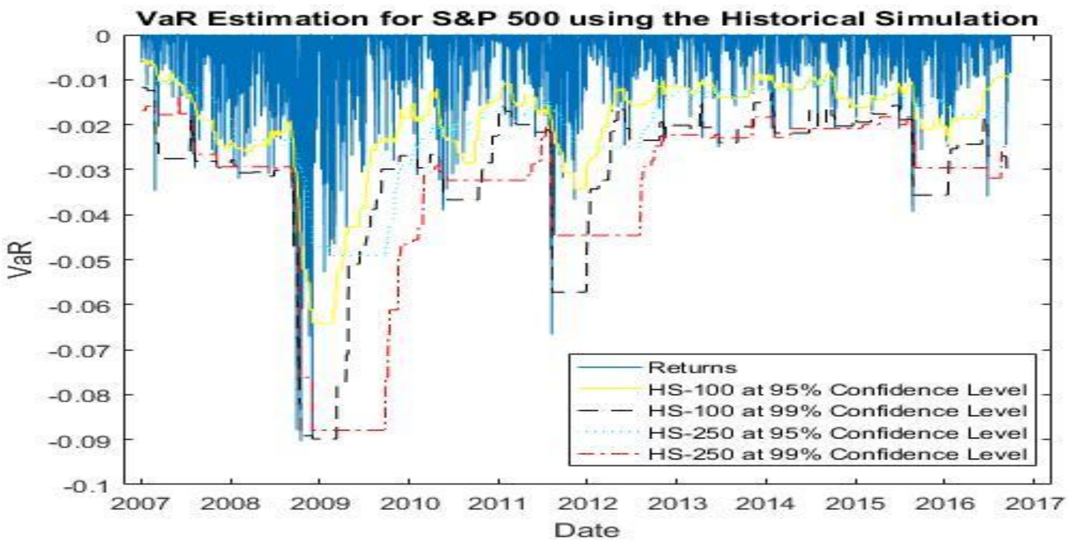


Figure 18 VaR Estimation for S&P using the Moving Average

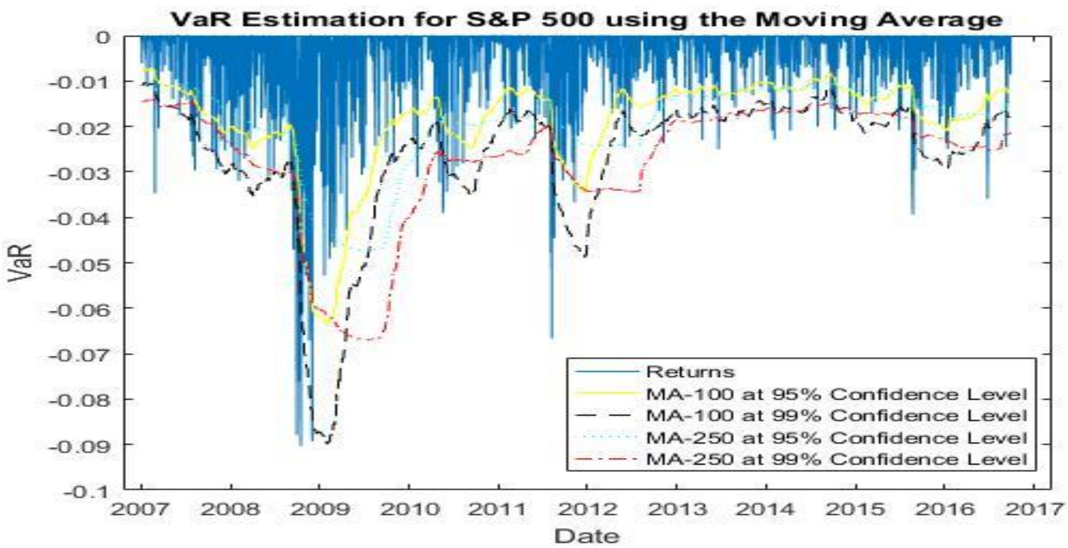


Figure 19 VaR Estimation for S&P using the Exponentially Weighted Moving Average

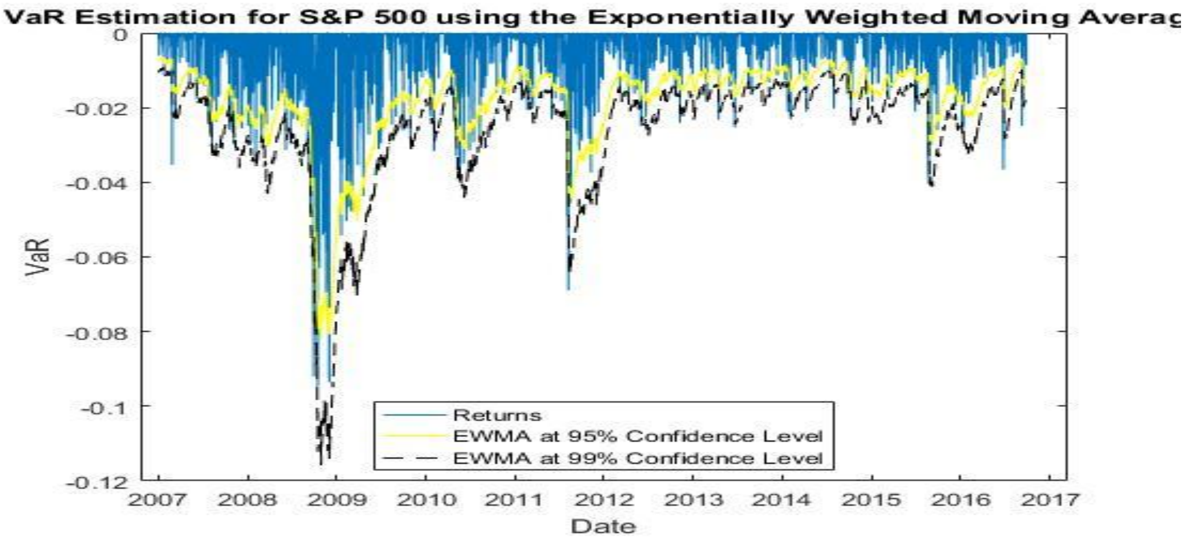


Figure 20 VaR Estimation for S&P using the GARCH (1,1)

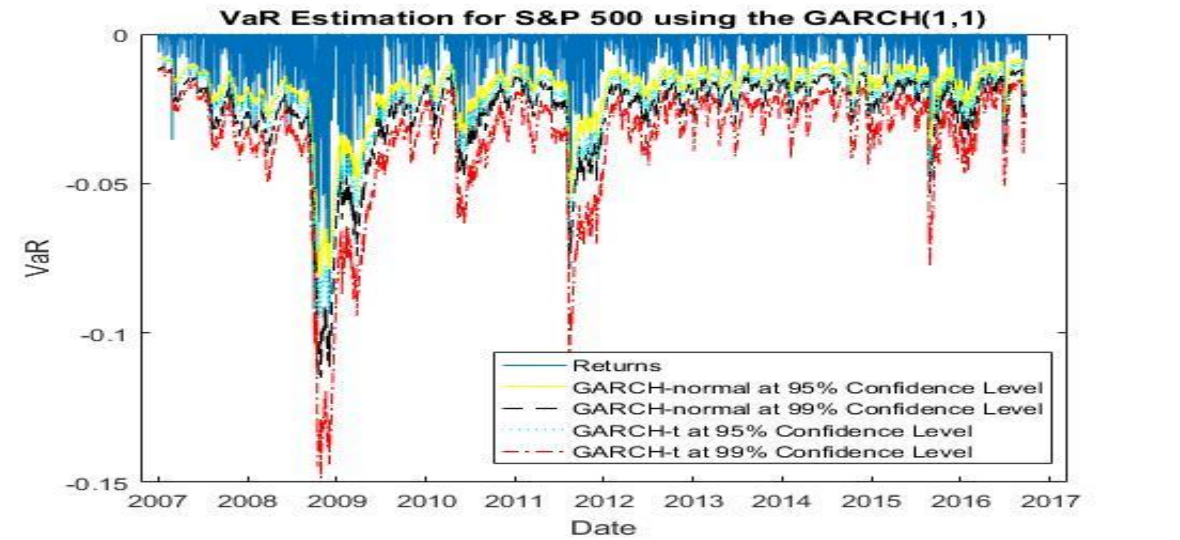
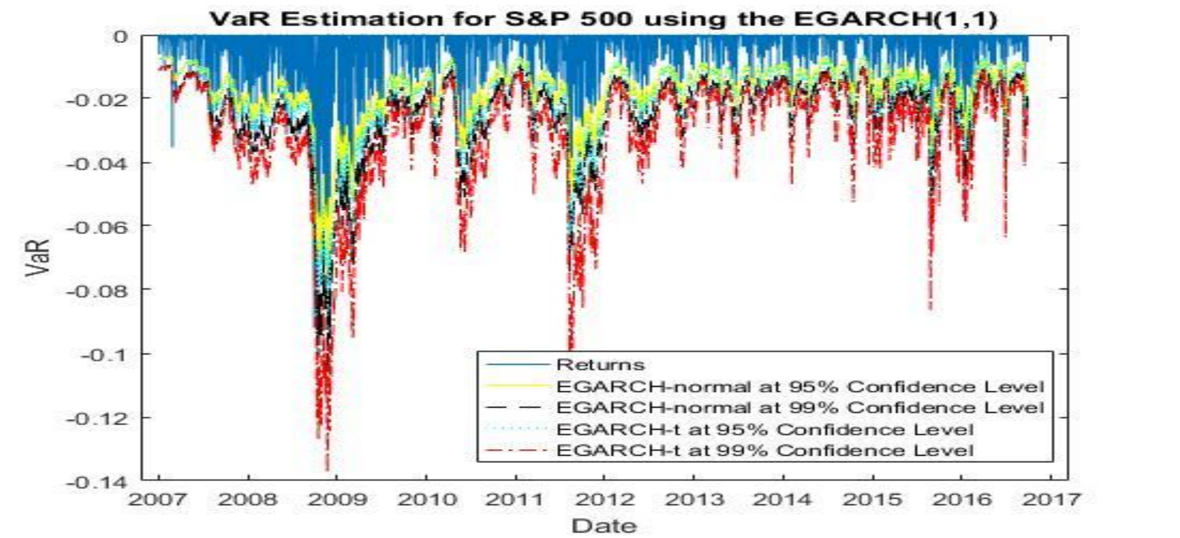


Figure 21 VaR Estimation for S&P using the EGARCH (1,1)



C.2. NASDAQ

Figure 22 VaR Estimation for NASDAQ using the Historical Simulation

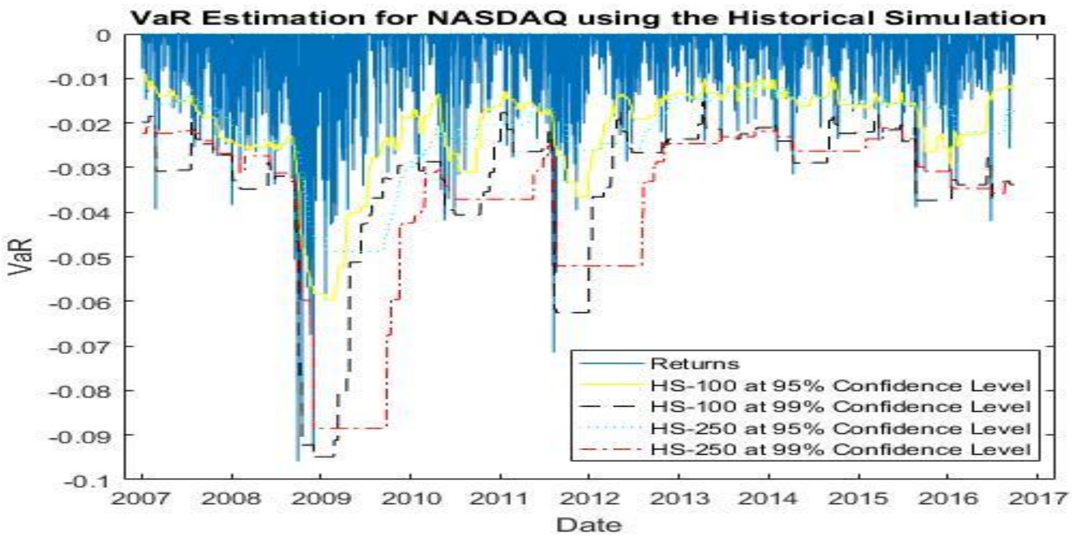


Figure 23 VaR Estimation for NASDAQ using the Moving Average

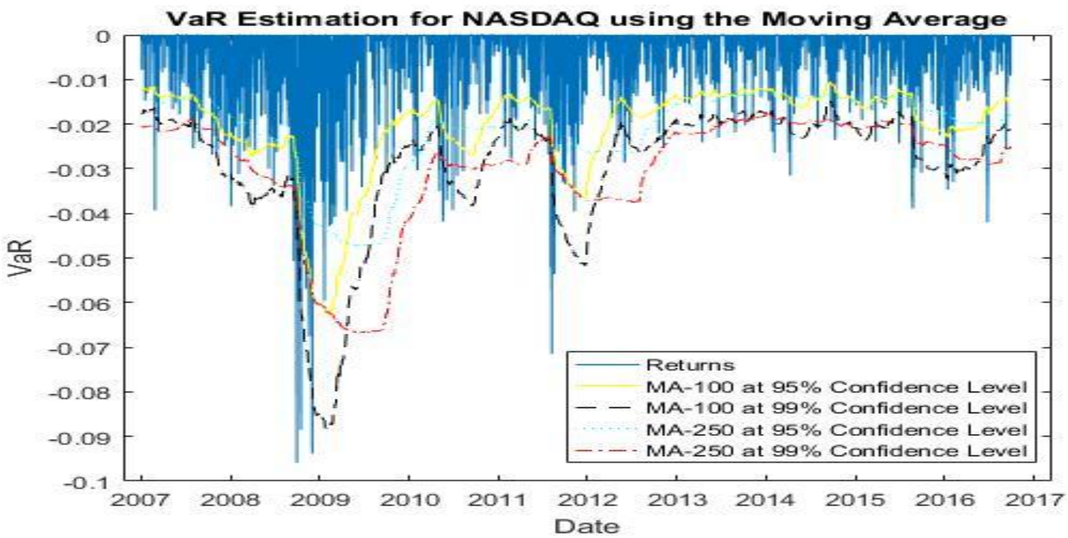


Figure 24 VaR Estimation for NASDAQ using the Exponentially Weighted Moving Average

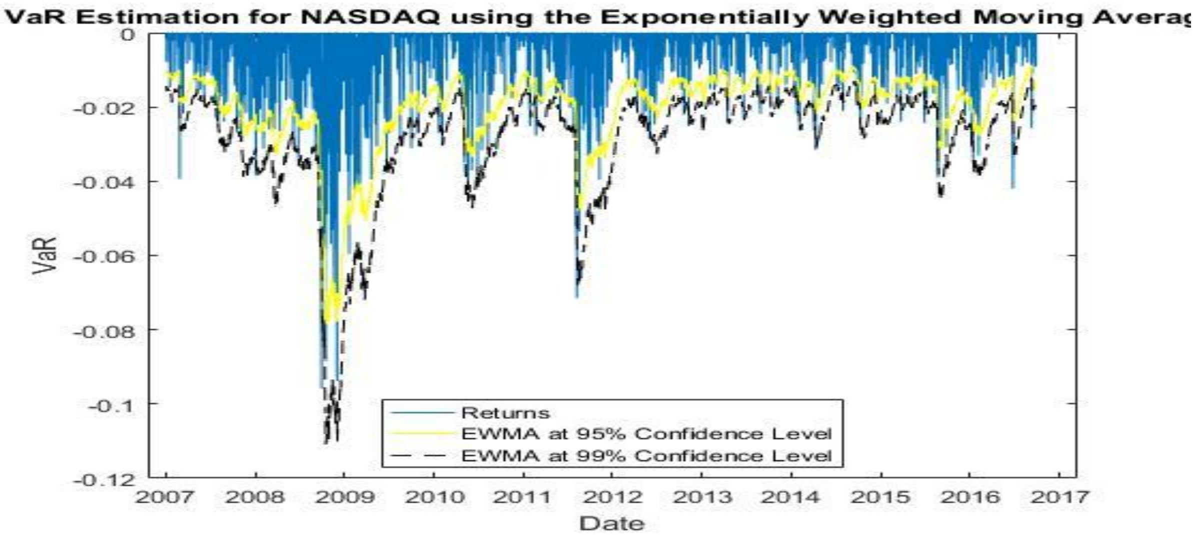


Figure 25 VaR Estimation for NASDAQ using the GARCH (1,1)

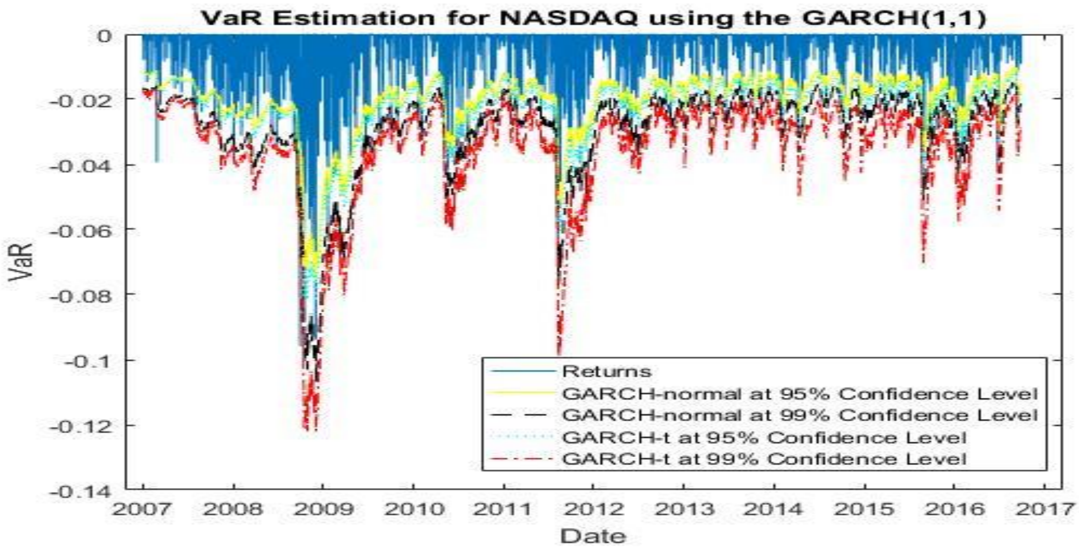
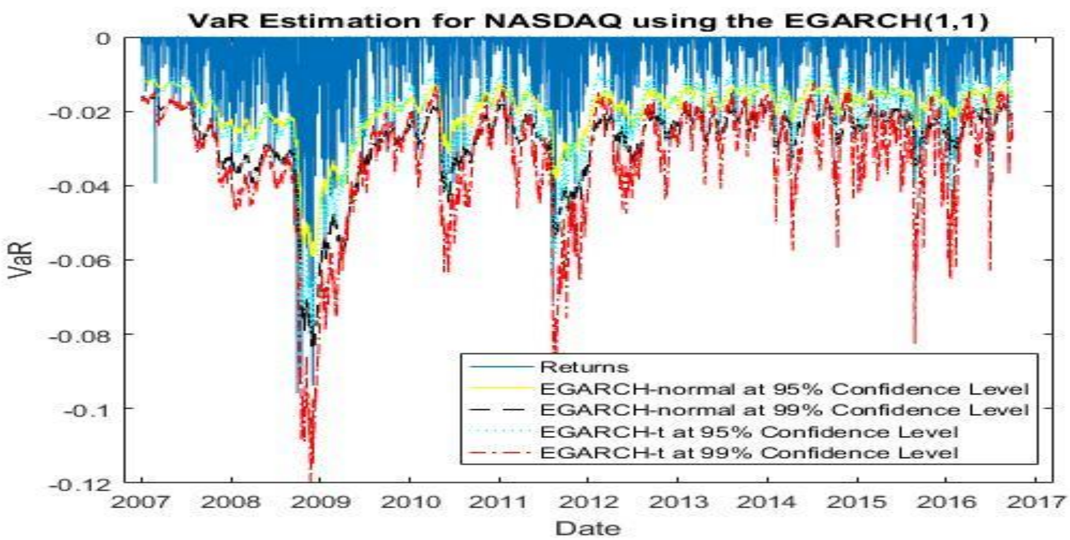


Figure 26 VaR Estimation for NASDAQ using the EGARCH (1,1)



C.3. EUROSTOXX

Figure 27 VaR Estimation for EUROSTOXX using the Historical Simulation

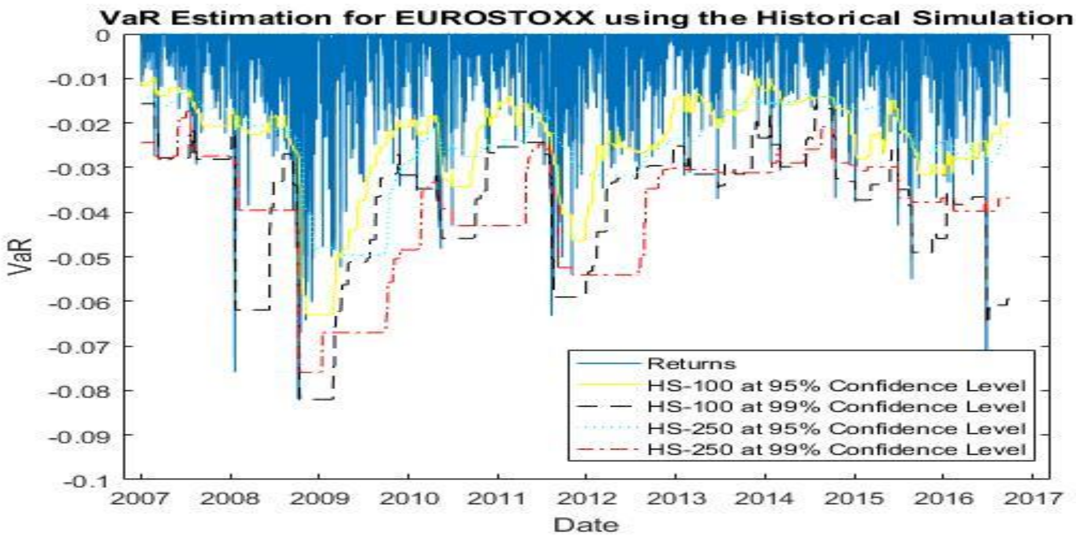


Figure 28 VaR Estimation for EUROSTOXX using the Moving Average

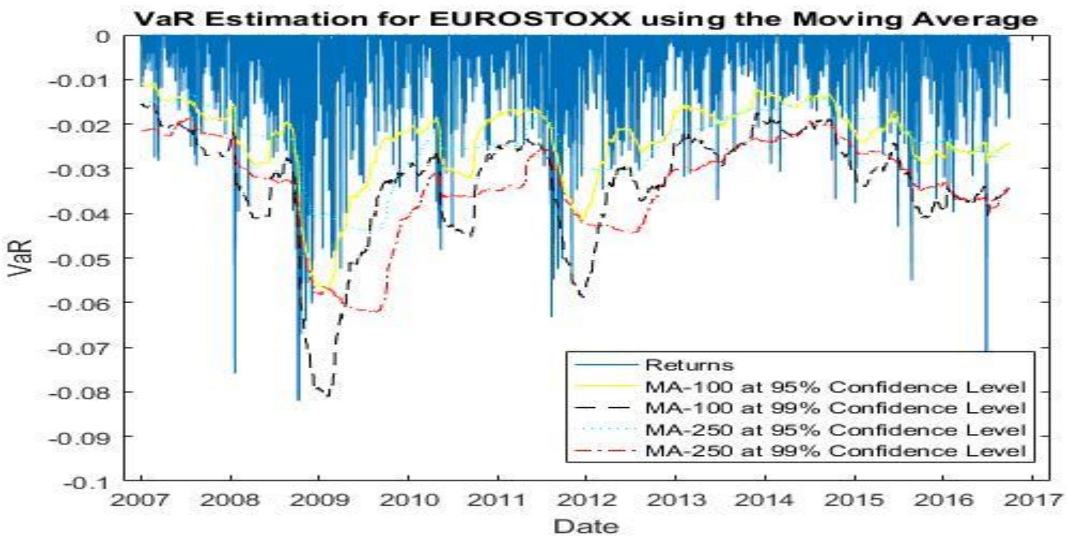


Figure 29 VaR Estimation for EUROSTOXX using the Exponentially Weighted Moving Average

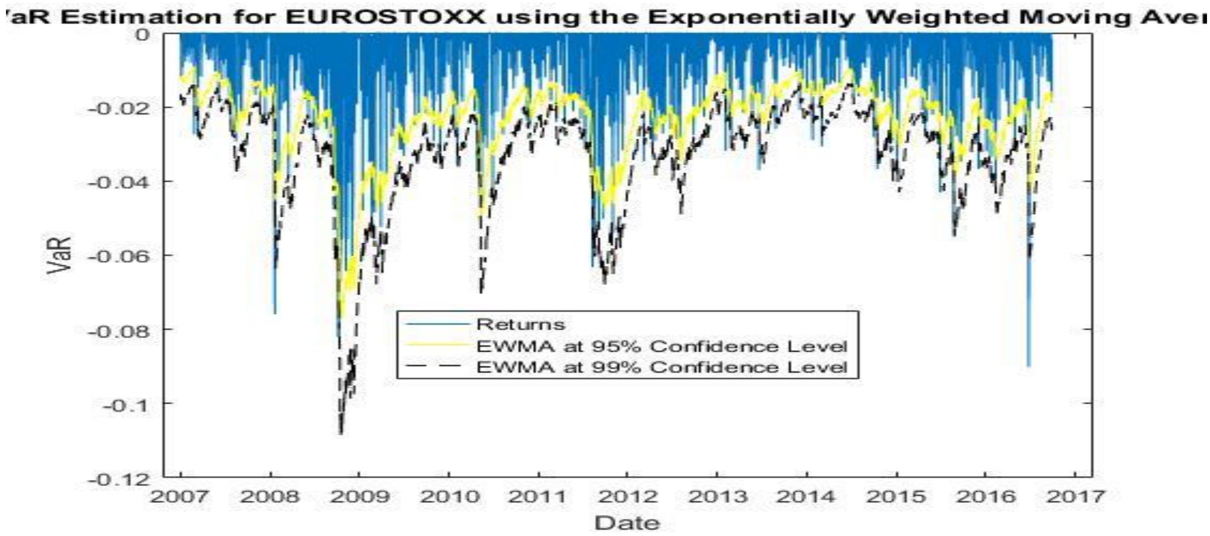


Figure 30 VaR Estimation for EUROSTOXX using the GARCH (1,1)

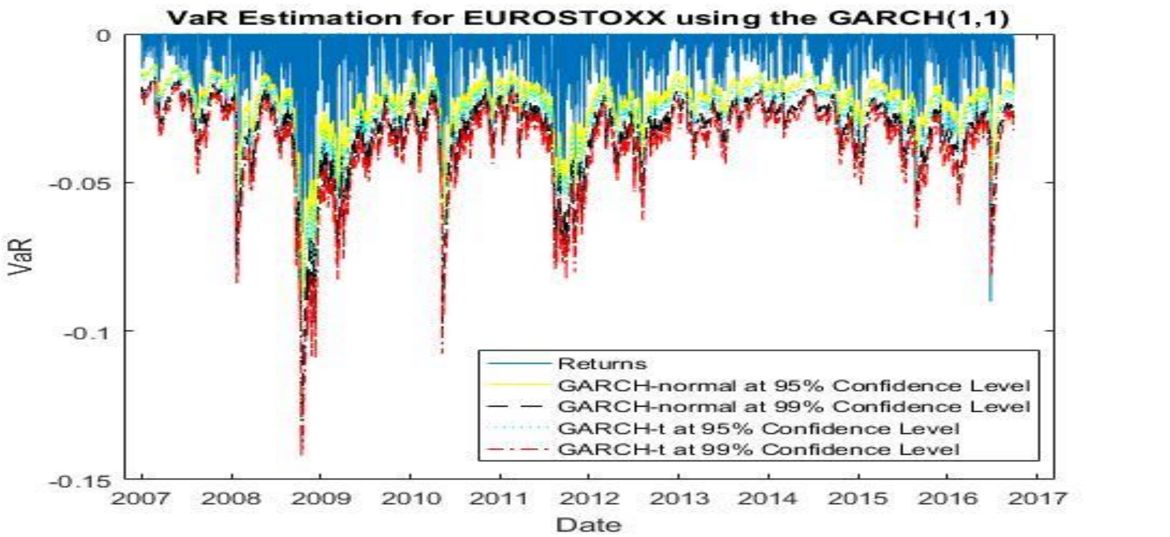
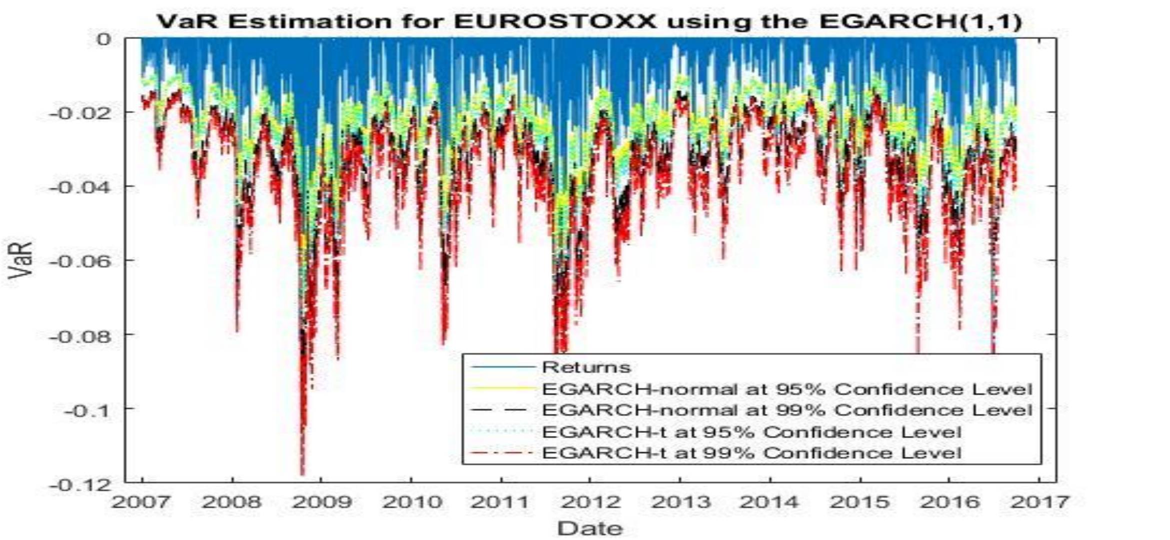


Figure 31 VaR Estimation for EUROSTOXX using the EGARCH (1,1)



C.4. FTSE

Figure 32 VaR Estimation for FTSE 100 using Historical Simulation

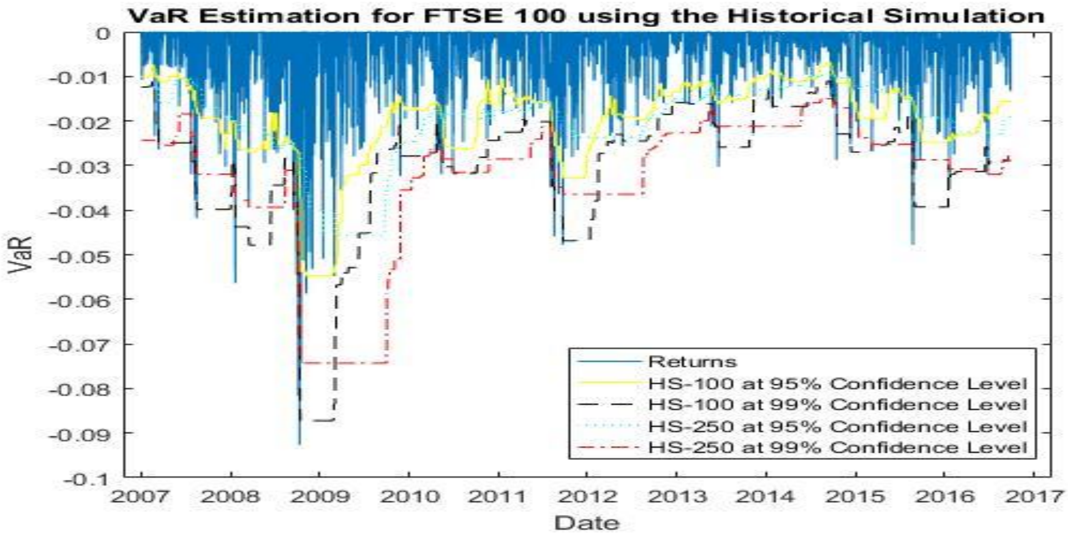


Figure 33 VaR Estimation for FTSE 100 using Moving Average

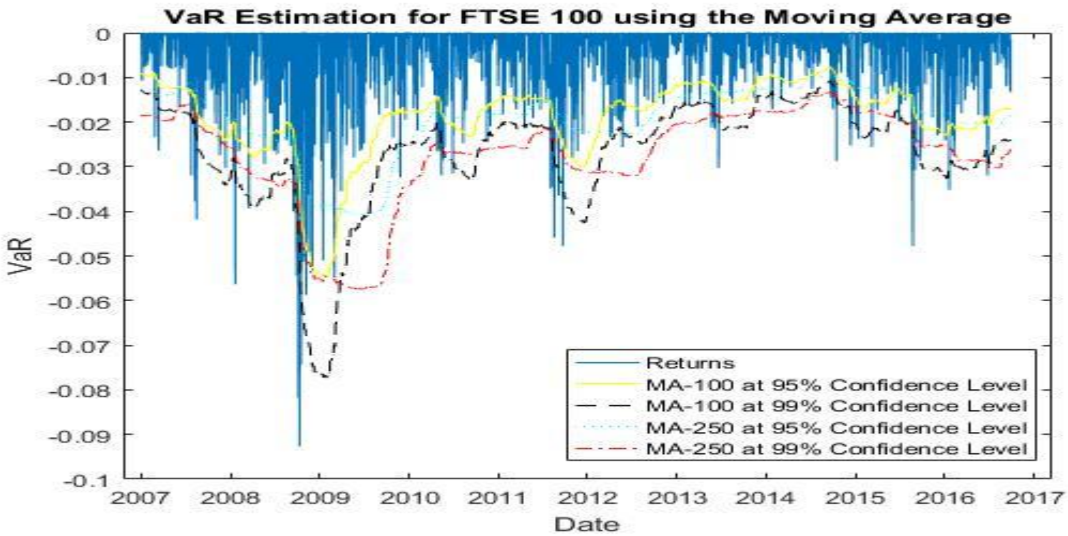


Figure 34 VaR Estimation for FTSE 100 using Exponentially Weighted Moving Average

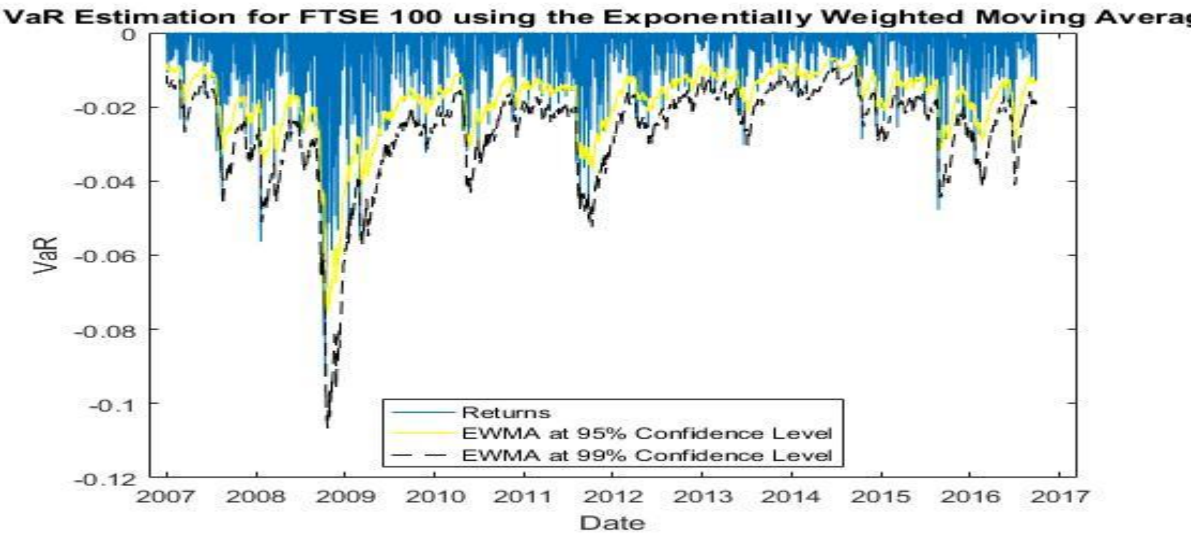


Figure 35 VaR Estimation for FTSE 100 using GARCH (1,1)

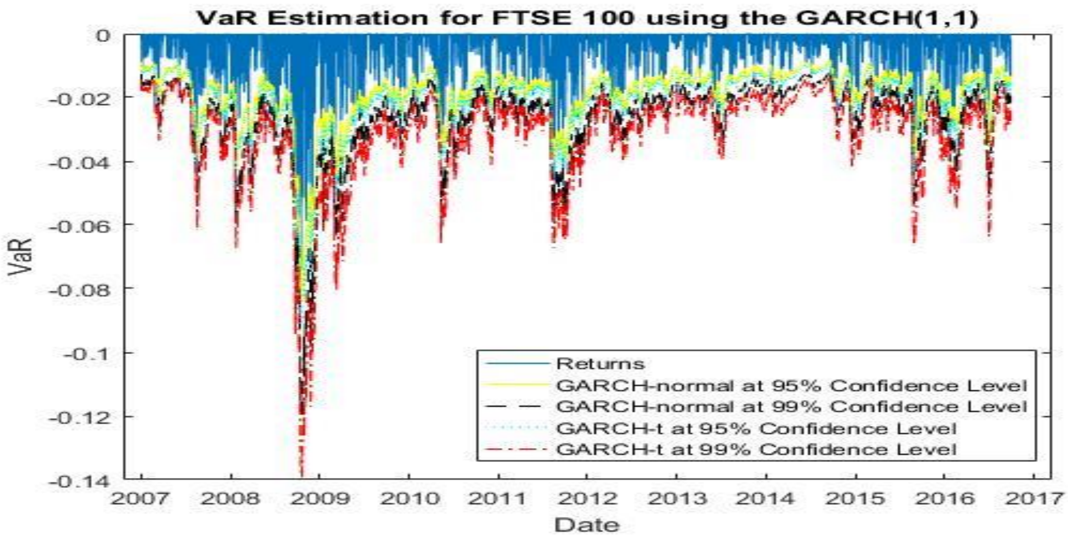
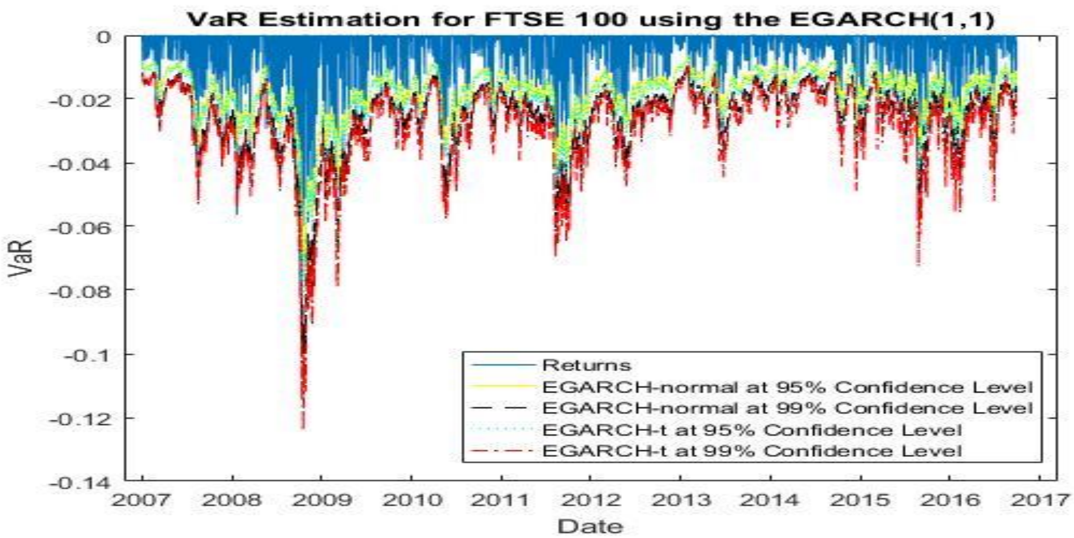


Figure 36 VaR Estimation for FTSE 100 using EGARCH (1,1)



C.5. DAX

Figure 37 VaR Estimation for DAX using Historical Simulation

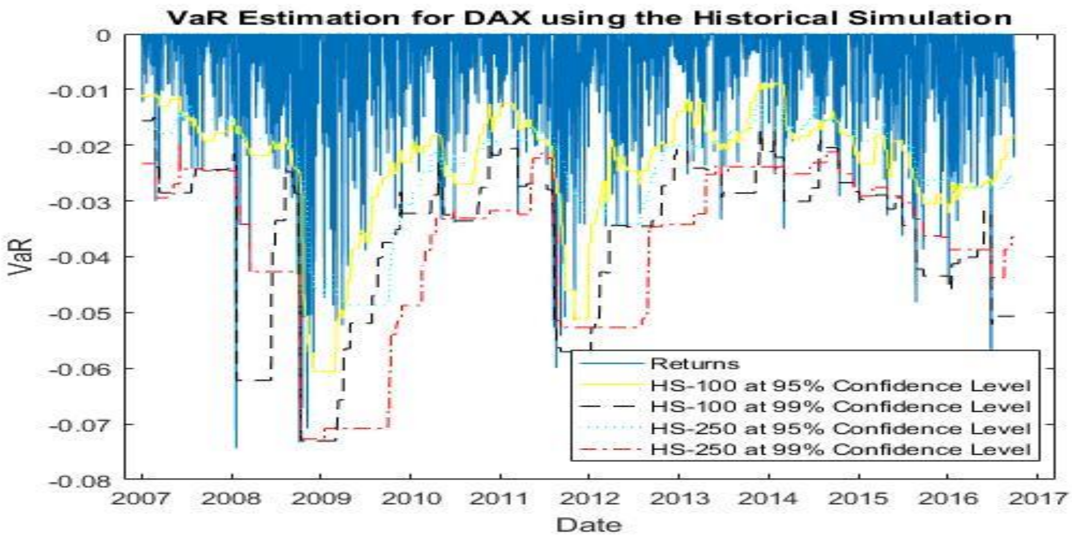


Figure 38 VaR Estimation for DAX using Moving Average

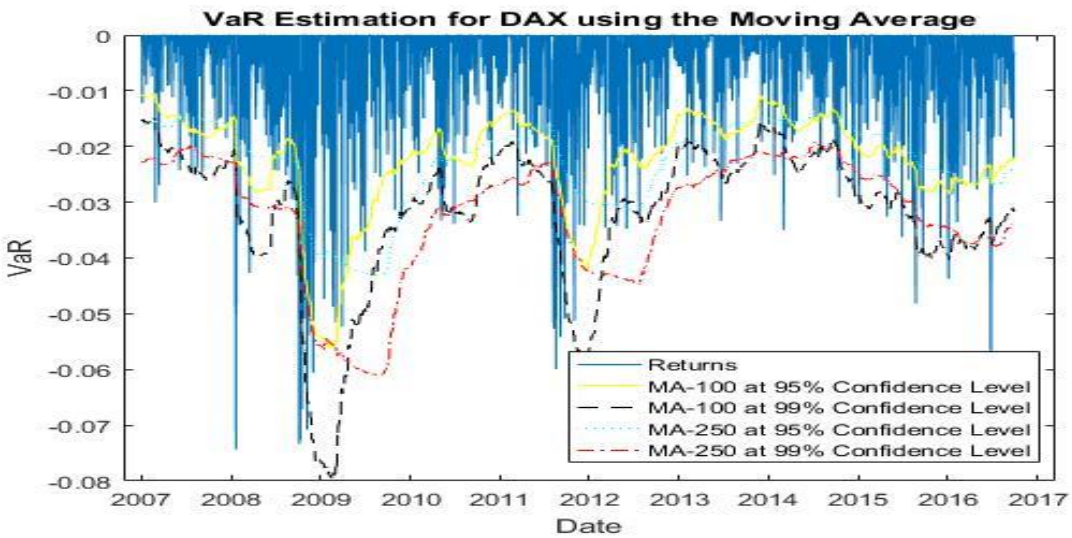


Figure 39 VaR Estimation for DAX using Exponentially Weighted Moving Average

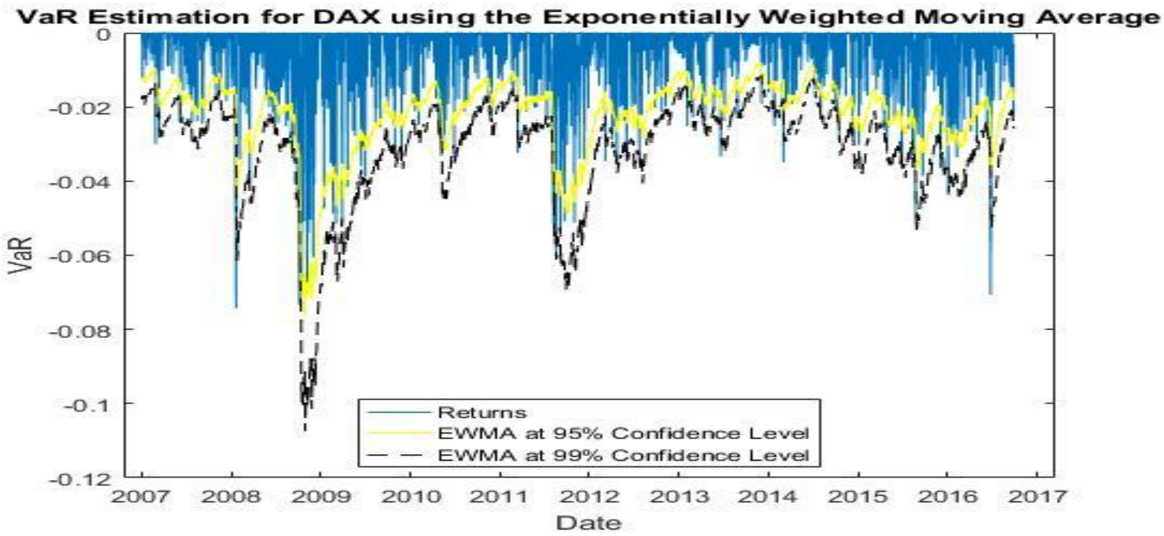


Figure 40 VaR Estimation for DAX using GARCH (1,1)

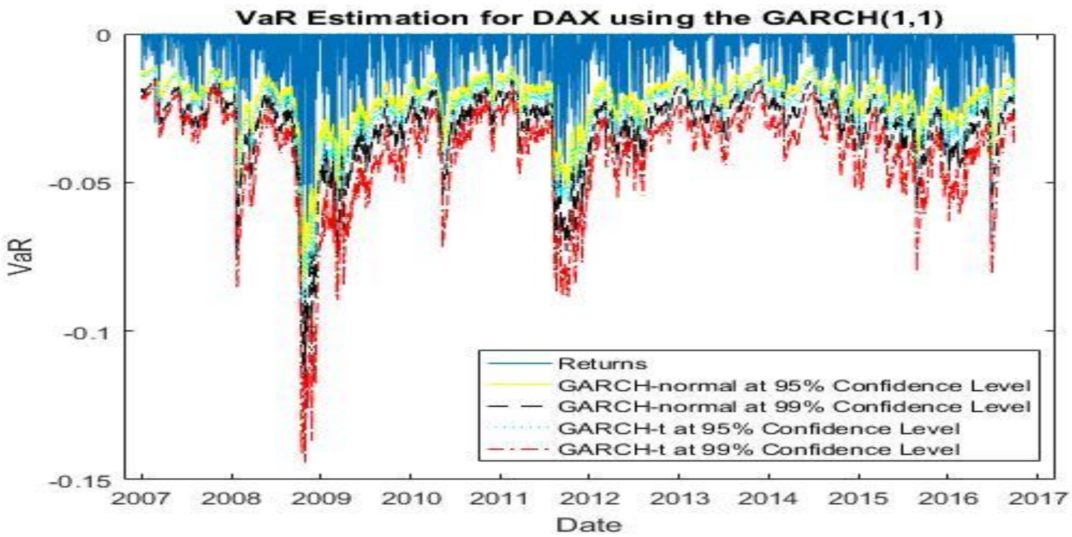
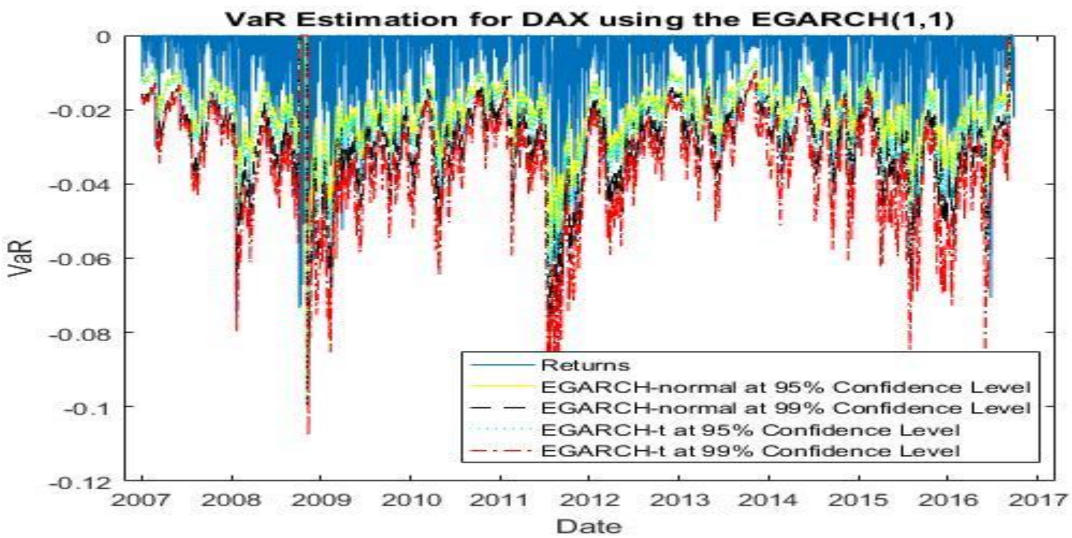


Figure 41 VaR Estimation for DAX using EGARCH (1,1)



C.6. CAC

Figure 42 VaR Estimation for CAC using Historical SImulation

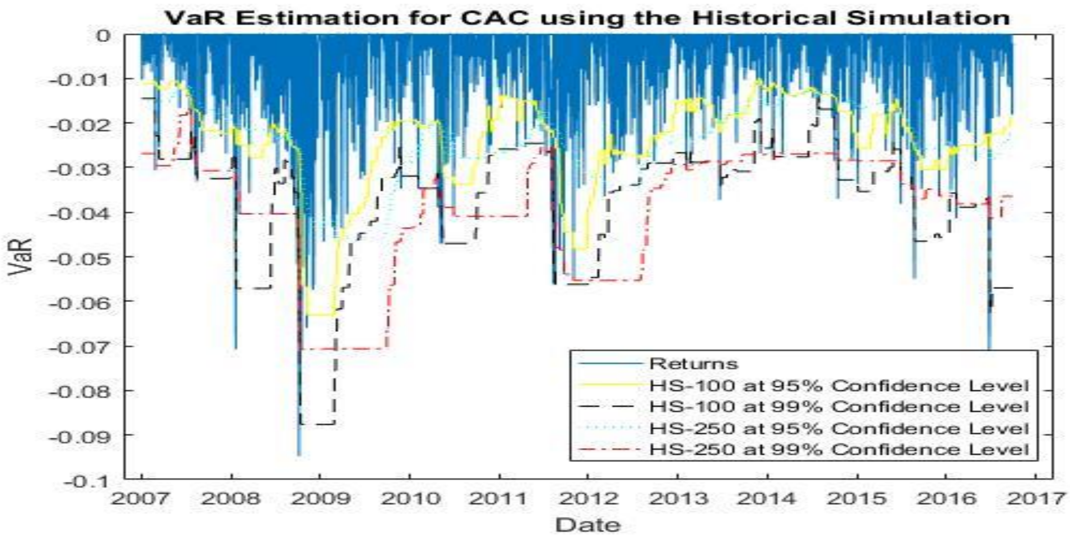


Figure 43 VaR Estimation for CAC using Moving Average

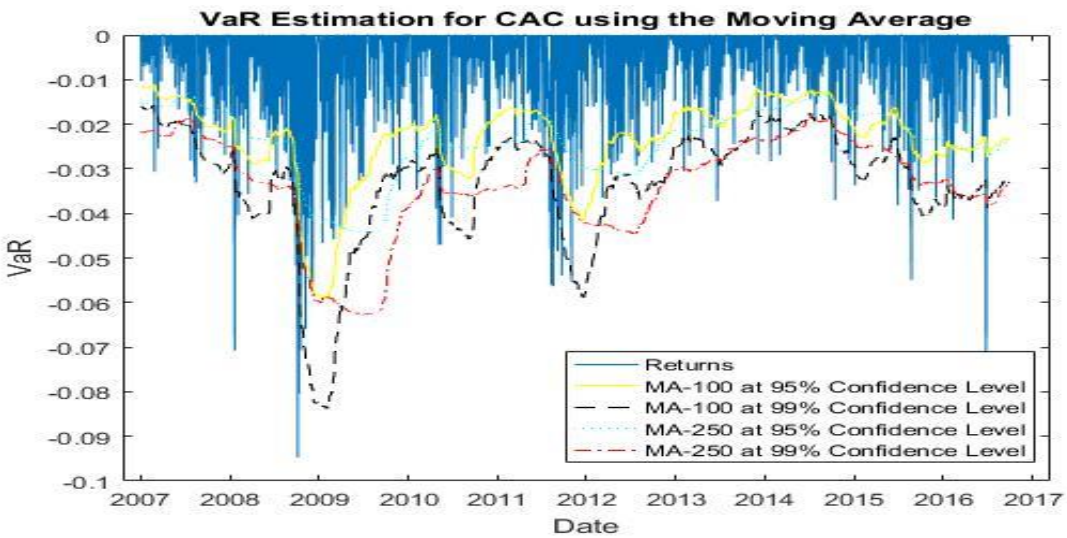


Figure 44 VaR Estimation for CAC using Exponentially Weighted Moving Average

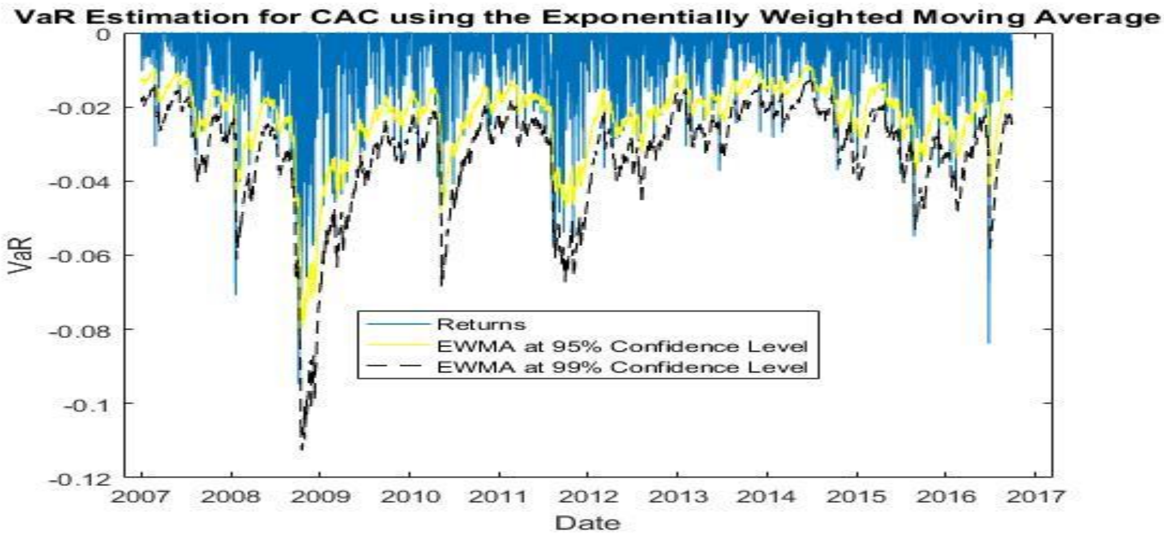


Figure 45 VaR Estimation for CAC using GARCH (1,1)

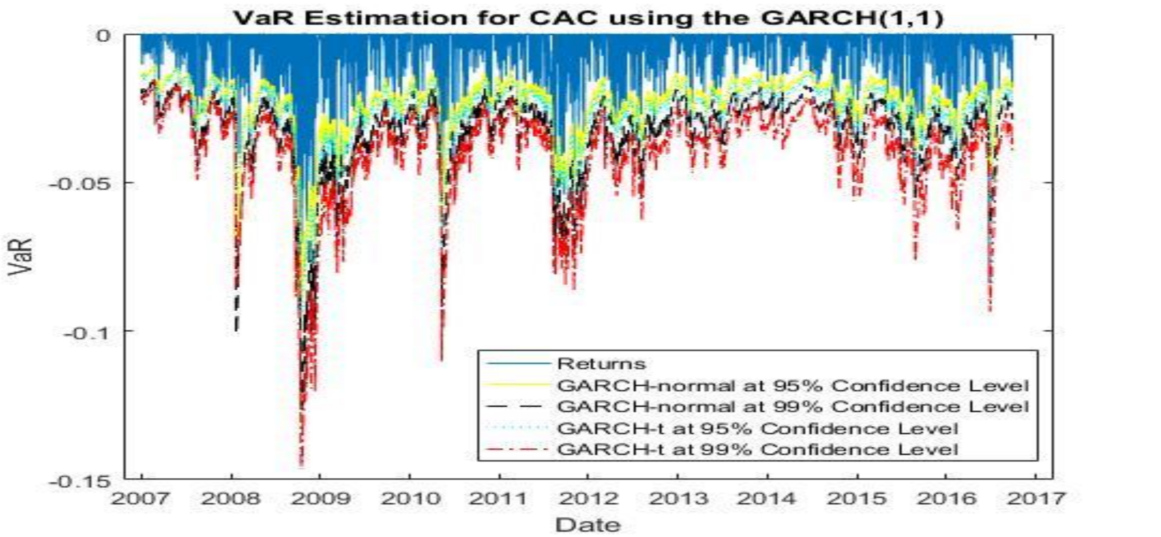
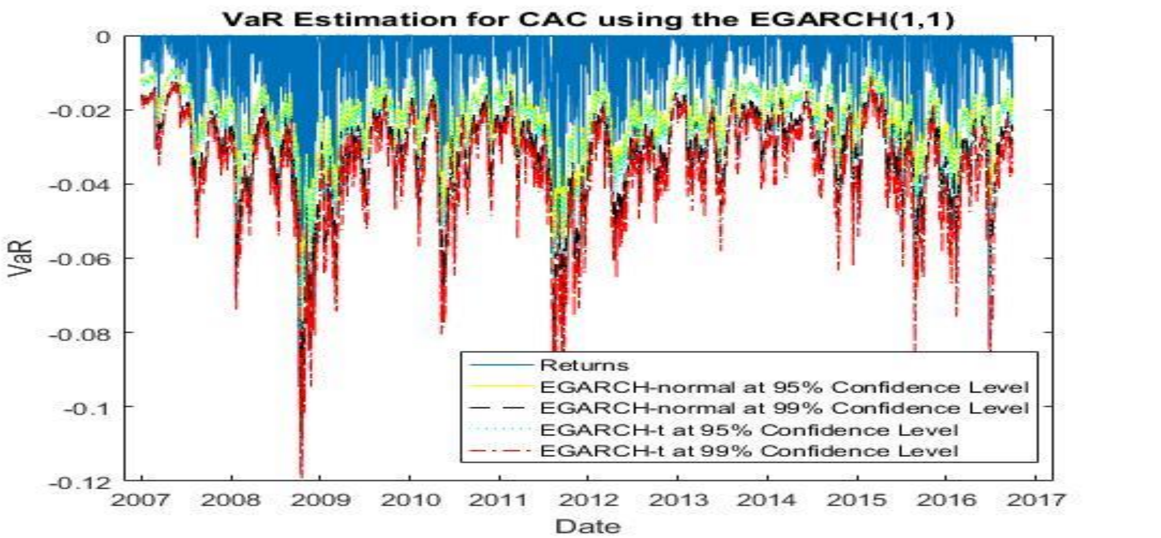


Figure 46 VaR Estimation for CAC using EGARCH (1,1)



C.7. ATHEX

Figure 47 VaR Estimation for ATHEX using Historical Simulation

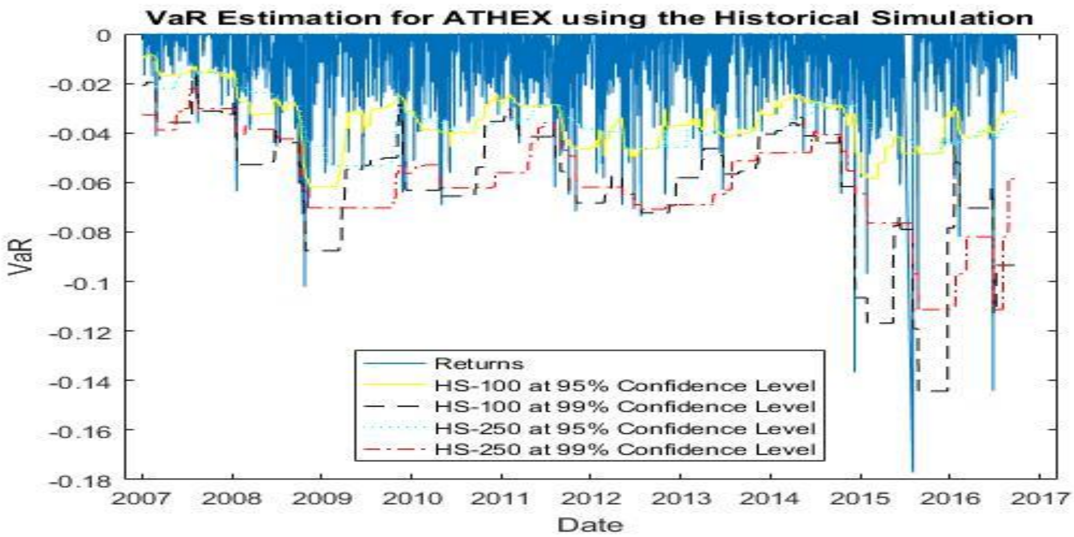


Figure 48 VaR Estimation for ATHEX using Moving Average

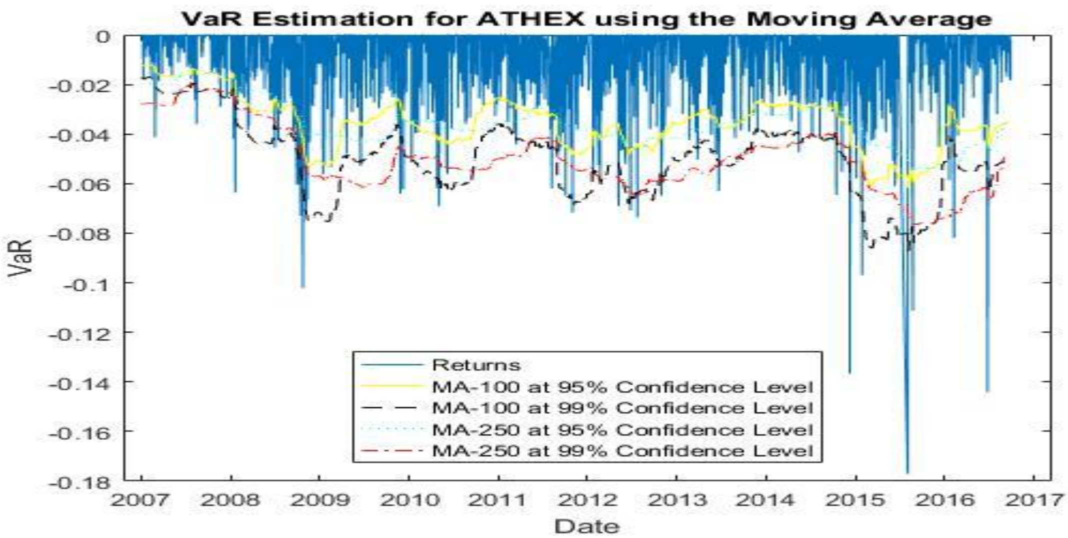


Figure 49 VaR Estimation for ATHEX using Exponentially Weighted Moving Average

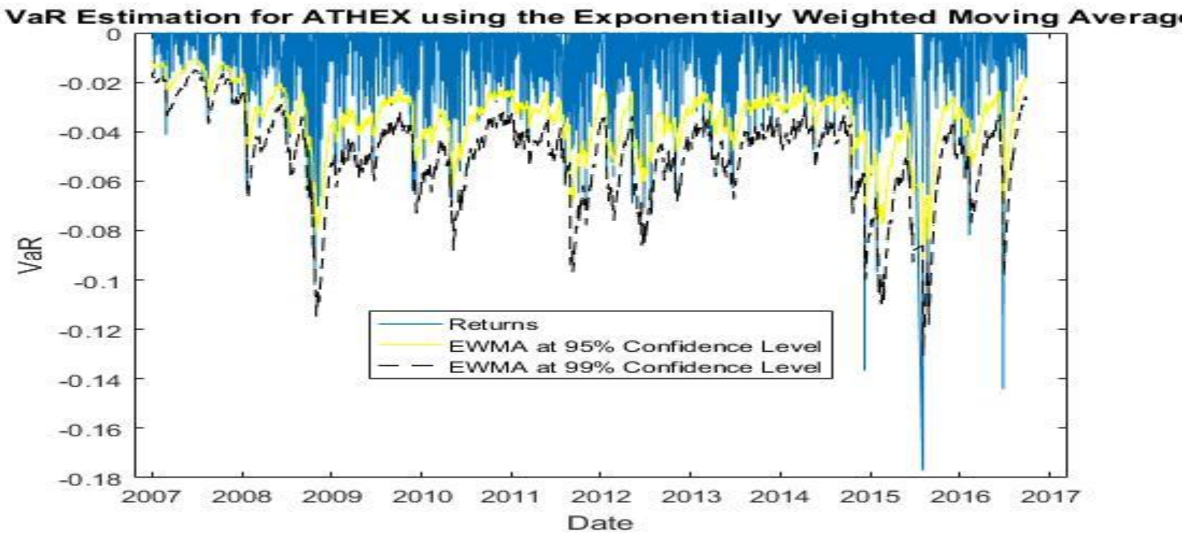


Figure 50 VaR Estimation for ATHEX using GARCH (1,1)

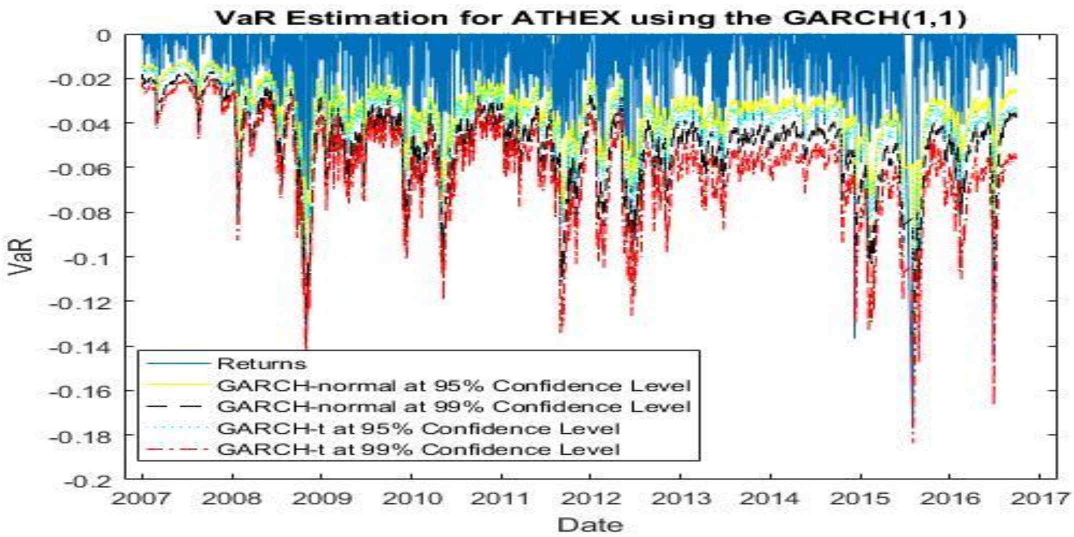


Figure 51 VaR Estimation for ATHEX using EGARCH (1,1)