

School of Economics and Management
University of Minho

Backtesting Value-at-Risk Models

Dissertation
Master in Finance
2016/2017

Supervisor:
Professor Nelson Areal

By:
Pedro Rodrigues PG29133

Acknowledges

I would like to thank Mr. Nelson Areal for his guidance, helpful comments, suggestions and patience through all the process, his help was fundamental. Also I would like to thank to my friends, girlfriend and family for their constant support and motivation.

Abstract

In the last decades, Value-at-Risk has become one of the most popular risk measurements techniques in the financial world. However, VaR models are only useful if they predict risk accurately. In order to evaluate the quality of the VaR estimates, it is necessary to perform appropriate and diverse backtesting methodologies.

In this study I test VaR estimates obtained from an unconditional parametric models (student-t generalized error, skewed student-t, pareto, and Weibull distributions) for four stock market indexes (DJIA, SP-500, Nikkei 225 and Dax 30) considering several different confidence levels. A rolling function procedure is applied to estimate the models parameters through maximum likelihood.

The performance of the VaR models is measured by applying several different tests of Unconditional Coverage, Independence and Conditional Coverage.

The results of the backtests provide some indication of the possible problems of the models, being the main one the independence property, leading us to conclude that they do not react well under high turbulent times, and consequently exceptions are auto correlated and come in clusters.

Title: Backtesting Value-at-Risk models

Author: Pedro Rodrigues

Department: Management

Major subject: Financial Risk

Keywords: Value-at-Risk, risk management, backtesting

Resumo

Durante as últimas décadas, Value-at-Risk tornou-se uma das medidas de risco mais populares na indústria financeira. Todavia, os modelos VaR só são úteis se conseguirem fazer uma previsão acertada do risco. De forma a avaliar a qualidade e precisão das estimativas de um modelo VaR, é necessário utilizar uma metodologia apropriada de avaliação.

A principal contribuição desta dissertação consiste em estudos empíricos, onde diversos modelos VaR paramétricos não condicionais são estimados para os quatro índices selecionados assumindo, para cada um, utilizando um leque de cinco distribuições: Student t, Generalized Error, Skewed Student t, Pareto e Weibull. Os parâmetros dos modelos são estimados por máxima verossimilhança através de uma janela rolante.

A performance das estimativas VaR é medida aplicando testes de cobertura incondicional, independência e cobertura condicional.

Os resultados da avaliação aos modelos mostrou alguns problemas, sendo o mais grave a falta de independência entre as exceções, levando-nos a concluir que os modelos não reagem bem durante períodos turbulentos, e conseqüentemente as exceções surge em grupos e estão bastante correlacionadas.

Palavras-chave: Value-at-Risk, risco financeiro, backtesting.

Index

Acknowledges.....	iii
Abstract	v
Resumo	vii
Index	ix
List of Figures	xi
1. Introduction	1
1.1 Background	1
1.2 Objective	2
1.3 Structure	3
2. Value-at-Risk Literature Review	4
2.1 History of VaR	4
2.2 Defining VaR	4
2.3 Different Approaches to VaR	7
2.4. Backtesting Procedures	11
2.4.1 Unconditional Coverage	11
2.4.2 Independence Test	12
2.4.3 Conditional Coverage	13
2.4.4 Multiple VaR levels	13
2.4.5 Loss Function	14
3. Methodology.....	15
3.1 Daily returns	15
3.2 VaR – Unconditional Parametric Estimation	15
3.3 Returns Distributions	16
3.3.1 Student t	16
3.3.2 Generalized Error	16
3.3.3 Skewed Student t	17
3.3.4 Pareto	18
3.3.5 Weibull	18

3.4 Backtests	18
3.4.1 Unconditional Coverage	19
3.4.2 Independence	20
3.4.3 Conditional Coverage	22
4. Data	24
5. Results	28
5.1 VaR Calculation	28
5.1.1 Unconditional Parametric Estimation	28
5.2 Backtests	39
5.2.1 Unconditional Coverage	39
5.2.2 Independence	43
5.2.3 Conditional Coverage	46
6. Conclusions	50
Bibliography	52

List of Figures

Figure 1: VaR for Normal distribution	6
Table 1: Basic stats of the four indexes	24
Figure 2: Dow Jones Industrial index returns	25
Figure 3: S&P 500 index returns	26
Figure 4: Nikkei 225 index returns	26
Figure 5: Dax 30 index returns	27
Figure 6: Dow Jones Industrials VaR Unconditional Parametric approach under Student t.....	29
Figure 7: S&P 500 VaR Unconditional Parametric approach under Student t.....	29
Figure 8: Nikkei 225 VaR Unconditional Parametric approach under Student t.....	30
Figure 9: Dax 30 VaR Unconditional Parametric approach under Student t.....	30
Figure 10: Dow Jones Industrials VaR Unconditional Parametric approach under GED.....	31
Figure 11: S&P 500 VaR Unconditional Parametric approach under GED.....	31
Figure 12: Nikkei 225 VaR Unconditional Parametric approach under GED.....	32
Figure 13: Dax 30 VaR Unconditional Parametric approach under GED.....	32
Figure 14: Dow Jones Industrials VaR Unconditional Parametric approach under SST.....	33
Figure 15: S&P 500 VaR Unconditional Parametric approach under SST.....	33
Figure 16: Nikkei 225 VaR Unconditional Parametric approach under SST.....	34
Figure 17: Dax 30 VaR Unconditional Parametric approach under SST.....	34
Figure 18: Dow Jones Industrials VaR Unconditional Parametric approach under Pareto.....	35
Figure 19: S&P 500 VaR Unconditional Parametric approach under Pareto	35
Figure 20: Nikkei 225 VaR Unconditional Parametric approach under Pareto	36
Figure 21: Dax 30 VaR Unconditional Parametric approach under Pareto	36
Figure 22: Dow Jones Industrials VaR Unconditional Parametric approach under Weibull.....	37
Figure 23: S&P 500 VaR Unconditional Parametric approach under Weibull.....	37
Figure 24: Nikkei 225 VaR Unconditional Parametric approach under Weibull.....	38
Figure 25: Dax 30 VaR Unconditional Parametric approach under Weibull.....	38
Table 2: Dow Jones Unconditional Coverage Test	40

Table 3: S&P 500 Unconditional Coverage Test	40
Table 4: Nikkei 225 Unconditional Coverage Test	41
Table 5: Dax 30 Unconditional Coverage Test	41
Table 6: Dow Jones Independence Test	44
Table 7: S&P 500 Independence Test	44
Table 8: Nikkei 225 Independence Test	45
Table 9: Dax 30 Independence Test	45
Table 10: Dow Jones Conditional Coverage Test	47
Table 11: S&P 500 Conditional Coverage Test	47
Table 12: Nikkei 225 Conditional Coverage Test	48
Table 13: Dax 30 Conditional Coverage Test	48

1. Introduction

1.1 Background

During the last two decades, Value-at-Risk (also known as VaR) became one of the most popular risk measurement techniques in finance. VaR is a method that aims to capture the downside risk of a single asset or a portfolio of assets. By definition, VaR measures the maximum loss in value of an asset/portfolio over a predetermined time for a given confidence interval.

Despite the common acceptance and use of VaR as a risk management tool, it has frequently been criticized for being incapable to produce reliable risk estimates. When implementing VaR systems, there will always be numerous simplifications and assumptions involved. Moreover, every VaR model, attempts to forecast future asset prices using historical market data, which won't necessarily reflect the future market environment.

Thus VaR models are useful only if they accurately predict future risk. In order to verify that the results acquired from VaR calculations are a consistent and reliable forecast, the models should always be submitted to backtest procedures with appropriate statistical methods. Backtesting is a procedure where actual profits and losses are compared to projected VaR estimates. Jorion (2005) refers to these tests as "reality checks". If the estimates of the VaR model are not accurate, the model should be re-examined for incorrect assumptions, inaccurate modelling or wrong parameters.

A series of different testing methods have been proposed for backtesting purposes. The first, and most basic test, such as Kupiec's (1995) point-of-proof (POF-test), focuses in examining the frequency of losses in excess of VaR. This called *failure rate* should be in line with the selected confidence level. For instance, if daily VaR estimates are computed at 95% confidence level for one year (with generally has 250 trading days), we would expect, on average, 12.5 VaR violations, or exceptions, to occur during this period, not more, not less. In the POF-test we would then examine whether the actual *failure rate* is reasonable when compared to the predicted amount.

In addition to the acceptable proportion of exceptions, another equally important aspect is to make sure that the observations exceeding VaR levels are independent and evenly spread over time. A good VaR model is one that is capable of avoiding exceptions clustering by reacting quickly to changes in financial assets volatilities and correlations. These types of tests, that take into account

the independence of exceptions were first suggested by Christoffersen (1998) and later by Hass (2001). These tests come as a complementary measure to the regulatory backtesting framework proposed by the Basel Committee (1996).

Backtesting is, or at least should be an integral part of VaR reporting in current risk management practices. Without a proper model validation process, one can never be sure that the VaR system yields accurate risk estimates. This subject is especially important in the current market environment where volatile market prices tend to make investors more interested in portfolio risk figures as losses accumulate. Although the common widespread and acceptance of VaR, it is known to have severe problems in estimating losses at times of turbulent markets. As a matter of fact, by definition, VaR measures the expected loss only under normal market conditions as said by Jorion (2005). This limitation is one of the major disadvantages of VaR and it makes model forecast and backtesting procedures very challenging and interesting as will be shown latter in this dissertation.

1.2 Objective

The main objective of this dissertation consists in performing an empirical study. However, in order to provide an exhaustive description about the modelling and the backtesting processes in the empirical part, I will first discuss VaR in general and the theory of the main models and backtesting methods currently available. In the literature review we aim to provide the reader with some of the most common procedures used to calculate VaR and to backtest it.

The objective of the empirical study here performed, is to uncurtain which of the tested distributions yield the best VaR estimates under an unconditional parametric approach, for each of the indexes here considered.

Methodological issues will not be covered in great detail, meaning that the reader is assumed to be familiar with statistical decision theory and related mathematics to some extent.

1.3 Structure

The thesis consists of seven chapters, of which the first one is the introduction. The second chapter starts with describing the basic idea behind VaR and gives some background and history on the

subject, also including some of the general advantages and shortcomings reported in the literature. The second chapter, also presents the forecasting procedures. Several methods are presented, and the advantages and disadvantages throughout the literature are shown. The aim of the discussion is to focus on the most common VaR estimation methods, and especially on those that will be applied later on. The last subsection of the second chapter focus on the backtesting methods. Some of the backtests available are presented in detail, but the discussion is by no means exhaustive, since it is impossible in this context to go through the panoply of different methods and their applications. I will discuss more in concert the methods applied later on, which are the ones mostly used in the literature.

The fourth chapter presents the detailed methodology behind the methods used in this dissertation, providing the formulas and procedures in detail. In fifth chapter data is presented, along with some descriptive statistics.

The sixth chapter forms the empirical part of the dissertation, and as such, can be considered the core of the study. Some of the methods presented in the preceding chapters, are applied to actual VaR calculations, and thereafter to their backtesting. The results are discussed in detail and the factors affecting the outcome are analysed thoroughly.

The seventh and last chapter is the conclusion. The most significant results are explained, discussed and compared to the existing literature.

2. Literature Review

2.1 History of VaR

In the past decades, risk management has evolved to a point where it is considered to be a distinct sub-field in the theory of finance. The growth of the risk management industry traces its roots back to the increased volatility of financial markets in 1970's. The breakdown of Bretton Woods system of fixed exchange rates and the fast pace on the development of new theories, such as adoption of Black-Scholes model, were among the important events that contributed to the revolution in risk management as stated by Dowd (1998). Another factor that is pointed by Dowd (1998) is simply the increase of the trading activity. For example, the average number of shares traded per day grew from 3.5 million in 1970 to 40 million in 1990. Equally impressive is the growth of the dollar value of outstanding derivatives positions, that went from \$1.1 trillion in 1986 to \$72 trillion in 1999 as presented by Jorion (2005). These factors combined with the unpredictable events of the 1990s, such as the financial disasters in Metallgesellschaft, Barings Bank, Orange County, Daiwa, "dot com bubble" and the "currency crisis" highlighted the danger allocated to financial positions and stimulated the need of improved internal risk management tools (Jorion 2005).

The mathematical roots of VaR calculation were developed already in the context of portfolio theory by Harry Markowitz and others in the 1950s. Financial institutions began to construct their own risk management models in the 1970s and 1980s, but it was not until the ground-breaking work from J. P. Morgan through the RiskMetrics model in 1994 that made VaR the industry-wide standard. (Dowd 1998). During this process, also regulators became interested in VaR. The Basel Accord of 1996 came as an important mark as it allowed banks to use their own internal VaR models to compute their regulatory capital requirement (Linsmeier & Pearson, 1996). Since then, VaR has grown to become the dominant measure of market risk and it is likely to gain even more acceptance in the near future as the methods are improved further, with new approaches being able to better estimate VaR.

2.2 Defining VaR

Companies face many different kinds of risks, such as market risk, liquidity risk, credit risk, counterparty risk, model risk, and estimation risk. VaR was original developed to measure market risk, which is caused by movements in the level or volatility of asset prices, but it was soon realized

that VaR methodology could also be applied to measure other types of risks, like liquidity risk and credit risk as stated by Jorion (2005). According to Dowd (1998) the market risk can be subdivided into four classes: interest rate risk, equity price risk, exchange risk and commodity price risk.

Philippe Jorion (2005) states that “VaR measures the worst expected loss over a given horizon under normal market conditions at a given level of confidence. For example, a bank might say that the daily VaR of its trading conditions, only one percent of the time, the daily loss will exceed \$1 million.”

The basic idea behind VaR is very straightforward since it gives a simple quantitative measure of portfolio’s downside risk. VaR has two important and appealing characteristics. The first one is that it provides a common consistent measure of risk for different positions and instrument types. Second, it takes into account the correlation between different risk factors. This property is absolutely essential whenever computing risk figures for a portfolio of more than one instrument Dowd (1998).

Assuming that asset returns are normally distributed, VaR may be illustrated graphically as in Figure 1. In mathematical terms, VaR is calculated as follows:

$$VaR_{\alpha} = \alpha * \sigma * W \quad (1)$$

The α indicates the selected confidence level, σ the standard deviation of the portfolio returns and W the initial portfolio value. An example seen in Jorion (2005), considers the situation where initial portfolio value is €100 million and the portfolio returns have an annual volatility of 20%. Calculating the 10 day VaR at a confidence level of 99% gives us the following result:

$$VaR_{99\%} = -2.33 * \sqrt{\left(\frac{10}{250}\right)} * €100M \approx -€9.3M \quad (2)$$

Where the square root in this function represents the 10-day time horizon assuming there are 250 trading days. As can be seen above, VaR computation can be very straightforward if normality is assumed to prevail. The simplicity of this assumption comes with some severe drawbacks which will be discussed shortly.

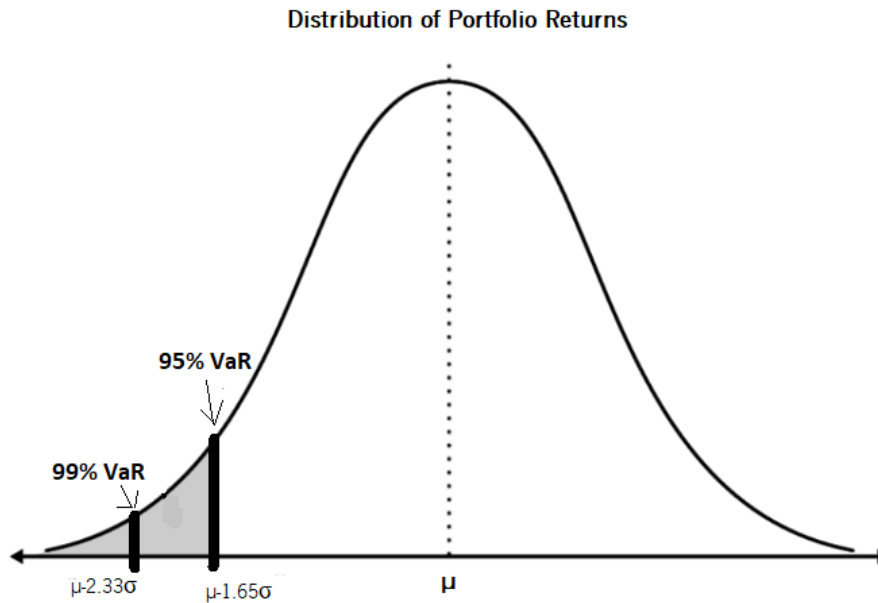


Figure 1: VaR for Normal distribution. The graph illustrates the VaR for two confidence levels when portfolio returns are normally distributed. The values of μ can be read from the standard normal distribution tables.

When interpreting VaR figures, it is essential to keep in mind the time horizon and the confidence level, since without them, it is not possible to interpret VaR numbers. The investors that have actively traded portfolios such as financial firms, typically use 1-day time horizon, whereas institutional investors and non-financial firms, typically use longer horizons. Linsmeier & Pearson (1996) suggest that firms should select the holding period according to the length of time it takes to liquidate the portfolio. On the other hand, they must also take into account the properties of the calculation method. If methods based on the approximations are used, then a relatively short time horizon should be applied.

As for the choice of confidence level depends on the purpose of the model. If the objective is as in this case, a dissertation, to test different approaches to VaR, one should choose a high confidence level, but not a very high one, in order to be able to capture enough VaR violations. When assessing capital requirements, the confidence level depends on the risk aversion of senior management, were risk averse managers usually choose higher confidence levels. One additional aspect is to consider the possibility of comparing different VaR confidence levels with estimates from other sources, as Dowd (1998) suggests.

2.3 Different Approaches to estimate VaR

VaR calculation methods are usually divided into parametric and non-parametric models. Parametric models assume that returns follow a known probability distribution, whereas non-parametric models do not (Ammann & Reich 2001).

In this section I briefly present the basics of the most common VaR calculation methods. The following discussion is meant to be mainly descriptive as the focus is on the strengths and weaknesses of each method. Thorough mathematical presentations are beyond the scope of this short review and will only be presented for the methods that I will use.

One of the simplest forms of VaR is the Unconditional Parametric estimation. It specifies a parametric model for the unconditional distribution of returns and derives VaR from the quantile of this distribution. Its parameters can be estimated using Maximum likelihood and different parametric distributions can be used.

Initially most of the classical financial models, namely the Markowitz's Portfolio Theory, the Capital Asset Pricing Model (CAPM) of Sharpe, Lintner and Mossin and the Black-Scholes' formula required the assumption of normality and finite variance of financial asset returns. The Gaussian hypothesis was not seriously questioned until the seminal papers of Mandelbrot (1963) and Fama (1965) which ignited numerous studies that latter found the empirical unconditional distribution of returns on financial data assets exhibit fatter tails and are more peaked around the center than would be predicted by a Gaussian distribution, i.e., the presence of leptokurtosis in the empirical distributions seems indisputable (Fama, 1965), and the variability of returns seems to be nonstationary. As a result, alternative distributions having such characteristics have been proposed as models for asset returns. Blattberg and Gonedes (1974) proposed Student t distribution, Box and Tiao (1962) and latter Nelson (1991) found evidence supporting the Generalized Error distribution, Wand et al. (2001) the exponential generalized beta distribution, Mitnik and Rachev (1993) the Weibull and double Weibull distribution, Fama (1965) Fama and Roll (1968) and So (1987) the Paretian distribution, Venkataraman (1997) the mixture of normal distributions as well as others.

Unconditional parametric is the natural analogue to historical simulation, which is using an empirical CDF to compute VaR instead of a model-estimated one. In particular, if the parametric model was unconditionally true, then it would be a better estimator than HS since it would have less uncertainty about the quantile than the non-parametric HS estimator.

The advantages and criticisms are identical to the parametric conditional VaR. The models are parsimonious and the parameters estimates are precise, yet finding a specification which necessarily includes the true distribution is a hard job as mentioned by Sheppard (2013) advises. Also, a distribution that may fit returns in one period, may be unfit in a different period. Finally returns may not even follow any known distribution as Bao et al. (2006) indicates.

J. P. Morgan back in 1992 released one of the VaR models that was most important to the spread of VaR usage. Its Riskmetrics released in 1992 became the most common and less restrictive conditional VaR, relying in a restricted generalized autoregressive conditional heteroscedasticity (GARCH) (1,1). This model includes no explicit mean model for returns and is only applicable to assets with returns that are close to zero or when the time horizon is short. This group has produced a surprisingly simple yet robust method for producing conditional VaR measure, but with some disadvantages being that, as Pafka & Kondor (2001) refer, it is based on the unrealistic assumption of normally distributed returns and completely ignores the presence of fat tails in the probability distribution, a most important feature of financial data. Also, several studies such as Danielsson & Vries (1997), Christoffersen (1998), and Engle & Manganelli (2004) have discovered relevant possible improvements when deviations from the relatively rigid RiskMetrics frameworks are explored, such as other models of the GARCH family.

Different types of autoregressive conditional heteroscedasticity models (ARCH), first introduced by Engle (1982), are commonly used to estimate risk. The ARCH model expanded into the GARCH model by Bollerslev (1986). These models capture the fluctuations in variance over time, which are present in most financial instruments. Another empirical observation is that the variance is usually higher during times of turmoil. Nelson (1991) created the exponential GARCH (EGARCH) model to capture this tendency. Nelson (1991) explains that an EGARCH model allows positive and negative shocks to have different effects on the estimated variance, he concludes, from empirical data, that the market volatility seem to react differently depending on the sign of the shocks, negative shocks usually result in periods of higher volatility compared to positive ones, and by his introduction of a third parameter the EGARCH allows the model to react differently depending on the different type of news.

The research in this area has been extensive. There is no definitive conclusion on which model in the GARCH family is best suited to forecast the volatility for specific types of securities. Most research has shown that a leptokurtic conditional distribution is better at describing the VaR

(Angelidis et al., 2004; Orchan and Köksal, 2011; Kösal, 2009; Aloui and Mabrouk, 2010). Usually the normal distribution is compared against Student t or the generalized error distribution. A majority of research have concluded that models, that take into account information asymmetry, outperform symmetric models (Hansen and Lunde, 2005; Aloui and Mabrouk, 2010). When compared to other families of volatility models, the GARCH models outperforms others such as Black and Sholes models, stochastic volatility models, regime switching models and grey theorem (Lehar et al., 2002; Kung and Yu, 2008; Lou et al., 2010; Pederzoli, 2006).

The limitations that comes with this procedure are that implementations require knowledge of a density family, and that all the dynamics of returns can be summarized by a time-varying mean and variance, and so higher order moments must be time invariant as referred by Bekiros & Georgoutsos (2005) and Sheppard (2013).

Back in 1937, Cornish and Sir Ronald Fisher introduced the Edgeworth expansion, that in its most-used inverse form, relates the cumulative distribution function of a normal distribution to some distribution of interest. It later served as support to the Cornish-Fisher Approximation, that splits the difference between a fully parametric model and a semi-parametric model. It shares the strengths of the semi-parametric distribution in that it can be accurate without a parametric assumption. Unlike the semi-parametric estimator, the Cornish-Fisher estimators are not necessarily consistent which may be a drawback, and also that estimates of higher order moments of standardized residuals may be problematic or, in extreme cases the moments may not even exist as suggested by Christoffersen (1998).

As the purest form of a nonparametric estimate of the unconditional VaR comes Historical Simulation also known as HS, it is widely used in practice because of its very easy to implement. No numerical optimization must be performed because parameters don't have to be estimated by maximum likelihood or any other method. Another advantage is its model free nature, because it does not rely on any particular parametric model such as the RiskMetrics model for variance and a normal distribution for the standardized returns. Historical Simulation lets the past data points speak fully about the distribution of next day's return without imposing any further assumptions. It's model free approach has the obvious advantage compared with model-based approaches that relying on a model can be misleading if the model is poor. Model-free nature also has several drawbacks of the method. Christoffersen (2012) describes mainly the that the choice of the data sample length is very ad hoc, and this influences the magnitude and dynamics of VaR. Another

drawback that comes from its model-free nature is that as this method does not rely on well - specified dynamic model, there is no theoretical correct way of extrapolation from the 1-day distribution to get the 10-day distribution other than finding more past data as Christoffersen (2012) points out. Sheppard (2013) and Wiener (1999) consider Historical Simulation of being an unreliable method for the same reasons. More recently other variants of the HS have been preferred as the Weighted Historical Simulation and the Filtered Historical Simulation.

In order to overcome some of the disadvantages associated with the Historical Simulation, Barone-Adesi, F, & Giannopoulos (1998) and Barone-Adesi Giannopoulos, & Vosper (1999) introduce Filtered Historical Simulation. The procedure is considered non-parametric apart from assumptions used in the estimation of residuals in the GARCH process. It is one of the most highly reputed methods nowadays and one of the most used ones that suffered different innovations by several authors like Christoffersen, Barone-Adesi and Giannopoulos that with their approaches were able to correct several bias and often outperform the first generation VaR models. The clear advantage over their fully parametric ARCH cousins, is that the quantile, and hence the VaR, will be consistent under weaker conditions since the density of the standardized residuals does not have to be assumed. According to Hartz, Mittnik, & Paoella (2006) this method is also numerically extremely reliable. Pritsker (2001) and Kuester, Mittnik, & Paoella (2005) demonstrate de superiority of the FHS method in VaR estimation. Also Adcock, Areal, & Oliveira, (2011) concluded that the FHS is an accurate method to in order to forecast VaR in the presence of non-normal returns. However, the handicap is that the quantiles may be poorly estimated especially if the confidence interval is very small. Another limitation is that their use is only justified if returns are generated by some location-scale distribution Sheppard (2013).

Another conditional VaR measure is the Conditional Autoregressive Value-at-Risk, also known as CaViaR, introduced by Engle & Manganelli (2004). They developed an ARCH-like model to directly estimate the conditional VaR using quantile regression as in the GARCH model but with a so called shock, usually with the HIT form. Engle & Manganelli (2004) point the usefulness of this on the fact that it does not specify a distribution of returns at any moment. Its use is justified under much weaker assumptions than other VaR estimators, also providing reasonable convergence of the unknown parameters, but also having the drawbacks that it may produce out of the order quantiles, and the parameters estimation is quite challenging as Sheppard (2013) suggest. Bao et al. (2006) showed that this models react well under all market circumstances, and have a good success as a VaR measure since this approach only focusses on the left tail.

2.4 Backtesting Procedures

In previous sub-chapter, different VaR methods were presented. The numerous shortcomings of these methods and the VaR in general are the most important reasons why the accuracy of the risk estimates should be tested. Brown (2008) states that a model is only as good as its backtest, so in order to evaluate the quality of the model estimates, it should always be intensively backtested.

Backtesting is a statistical procedure where the actual profits and losses are systematically compared to the correspondent VaR estimate given by the model. For example, if we use a confidence level of 99% to calculate the daily VaR, we expect an exception to occur once in every 100 days, on average. In the backtesting process we could statistically examine whether the frequency of exceptions over some specified time interval is according to the selected confidence level, these types of tests are known as tests of unconditional coverage. These tests are simple to implement since they do not take into account when the exceptions occurred.

It is not only important to know if a certain VaR model produces the correct amount of exceptions, but also if they are evenly spread over time i.e. are independent of each other as mentioned by Christoffersen (1998). The clustering of VaR exceptions indicates that the model is not good when capturing changes in the market volatility and correlations. Therefore, it is also important to examine the conditioning, or time variation of the data. For that we have the conditional coverage tests.

In the following sub-chapters I aim to present a short description of the most used VaR methods, and later, a more detailed description of the methods that I will be using.

2.4.1 Unconditional Coverage

Some of the first VaR backtests proposed, focused exclusively on the unconditional coverage property. The widely known test for unconditional coverage of Kupiec (1995), also known as the POF-tests (proportion of failures), measures whether or not, the number of exceptions is consistent with the selected confidence level. Under the null Hypothesis of the model being correctly accurate, the number of exceptions follows the binomial distribution. The POF-test of Kupiec has some shortcomings. Campbell (2005) points out that the test is statistically weak with sample sizes consistent with the current regulatory framework of the Basel Accord (1 year). Kupiec himself

recognized this lack of power. Another shortcoming is that it considers only the frequency of losses and not the time when they occur, and therefore, as a result it may fail to reject a model that produces clustered exceptions, and that's why backtesting should not rely only on the tests for unconditional coverage as stated by Campbell (2005) states.

2.4.2 Independence Test

As important it may be to have the correct number of exceptions, it is also important to have them spread evenly over time. Finger (2005) states that a good VaR model should be capable of reacting to changing volatility and correlations in a way that the exceptions occur independently of each other, whereas bad models tend to produce a sequence of consecutive exceptions.

One of the first tests of the independence property of VaR hit series, was the Engle (1982) Autoregressive Conditional Heteroskedasticity (ARCH) concerning if one of more data points in a series of which the variance of the current error term or innovation is a function of the actual sizes of the previous time periods error terms, if the null hypothesis is rejected, then it indicates that there are ARCH effect in sample data.

Christoffersen's (1998) Markov test, came as a reference for testing the independence property of VaR models. It examines whether the likelihood of a VaR violation depended on whether or not a VaR violation occurred in the previous day. A more recent approach to the independence testing was proposed by Christoffersen and Pelletier (2004). It uses the insight that if a VaR violations are completely independent from each other, then the amount of time that elapses between VaR violations should be independent of the amount of time that has elapsed since the last violation, therefore, VaR violations should not exhibit any kind of duration dependence. Christoffersen and Pelletier (2004) provide evidence that this independence test is more powerful than the previous Markov test since it is capable of capturing more forms of dependence without requiring additional information.

Campbell (2005) points out, that the independence test must describe the types of anomalies that it is going to look for when examining whether or not the independence property is satisfied. This is in what the Christoffersen's Markov test may fail at, because if, for example, a violation may be dependent not on yesterday's violation, but rather on last week's.

2.4.3 Conditional Coverage

For a VaR model to be considered accurate, according to Campbell (2005) and Jorion (2005), it must exhibit both the independence and the unconditional coverage property. Tests that jointly examine both, provide an opportunity to detect VaR models which are deficient in one of them. By this one could assume that these tests are substitutes of either the unconditional coverage property or the independence property alone. While joint tests have the property that they will eventually detect a VaR measure which violates either of these properties, this comes with the expense of a decreased ability to capture a VaR model which only violates one of the two properties. The joint test is hampered by the fact that one of the two violations it is designed to detect, namely violations of unconditional coverage, is actually satisfied by the VaR measure. The fact that one of the two properties is already satisfied makes it more difficult for the joint test to detect the inadequacy of the VaR model Christoffersen (2012).

2.4.4 Multiple VaR levels

The backtests that were discussed earlier focussed only on determining the adequacy of a VaR model at a single confidence level. In general, however, there is no need to restrict attention to a single VaR level. The unconditional coverage and independence property of a correct VaR model should hold for any confidence level. Several backtests have been proposed to test the entire forecast distribution, authors like Crnkovic and Drachman (1997), Diebold, Gunther and Tay (1998), as well as Berkowitz (2001).

The Crnkovic and Drachman (1997) has the insight that if the model is accurate, then the 1% VaR should be violated 1% of the time, the 5% VaR should be violated 5% of the time and so on. A VaR violation at any confidence level should be independent on violations at any other level.

The advantage of this test is that it provides additional power in identifying inaccurate models as Campbell (2005) refers. However, this approach needs a wide data series. According to Crnkovic and Drachman (1997), their test requires observations of at least four years in order to obtain reliable estimates, which can be an obstacle for short sets of data.

2.4.5 Loss Function

Until now, all the backtests presented focused only in whether a hit occurred or not. Despite the fact that it plays a prominent role, its information is limited, and one may be more interested, for example in the severity of VaR exceptions as referred by Colletaz, Hurlin and Perignon (2013).

Lopez (1998 and 1999) introduced a method to examine this aspect of VaR estimates. The idea behind the test, is to gauge the performance of VaR models by how well they minimize a loss function that represents the evaluator's concerns. Unlike most other backtesting methods, loss function approach is not based on the hypothesis-testing framework. Dowd (2006) argues that this makes loss functions attractive for backtesting with relatively small amount of observations.

The advantage of this method is its flexible approach. It can be tailored to address specific concerns of the evaluator since the loss function may take different forms. The main downfall is that it relies on the correct assumption about returns distribution. In case the distribution is incorrectly defined, the backtest results become distorted. Observing a score that exceeds the benchmark value could imply either an inaccurate VaR model or wrong assumptions about the stochastic behaviour of profits and losses as Campbell (2005) points out. Lopez (1999) recognizes some of the problems associated with this backtest based on the loss function and points that this test cannot be used to statistically classify a model as accurate or inaccurate, but rather to monitor the relative accuracy of the model and to compare different models with each other. It should be used as complement of the hit based methods.

A more recent approach to this method is made by Colletaz, Hurlin and Perignon (2013). Their approach is based on their concept of "super exception" which is defined by a much smaller coverage probability, e.g., 0.2%, similarly to Lopez (1998 and 1999) tests, this one should also be accompanied with hypothesis-testing methods.

3. Methodology

This section starts with the specification of the data used and gives an in-depth description of the applied methods for VaR estimation and for the backtesting procedures. The evaluation criteria used is comprehensively stated.

3.1 Daily Returns

The daily returns can be calculated as,

$$r = \frac{P_t - P_{t-1}}{P_{t-1}} \quad (3)$$

were,

r = Stock return

P_t = Price in period t

P_{t-1} = Price in period $t - 1$

3.2 VaR – Unconditional Parametric Estimation

This form of VaR specifies a parametric model for the unconditional distribution of returns and derives the VaR from the α -quantile of the designated distribution. When $r_t \sim N(\mu, \sigma^2)$, the α -quantile VaR is,

$$VaR = -\mu - \sigma \Phi^{-1}(\alpha) \quad (4)$$

Where $\Phi(\cdot)$ is the Cumulative Density Function (CDF) of a standard normal (and so $\Phi^{-1}(\cdot)$ is the inverse CDF)

and the parameters are directly estimated using maximum likelihood with the estimators being,

$$\hat{\mu} = T^{-1} \sum_{t=1}^T r_t \quad (5)$$

$$\hat{\sigma}^2 = T^{-1} \sum_{t=1}^T (r_t - \hat{\mu})^2 \quad (6)$$

3.3 Returns Distributions

3.3.1 Student-t

The probability density function for student's t-distribution is defined as,

$$f(t) = \frac{\Gamma\left(\frac{v+1}{2}\right)}{\sqrt{v\pi}\Gamma\left(\frac{v}{2}\right)} \left(1 + \frac{t^2}{v}\right)^{-\frac{1}{2}(v+1)} \quad (7)$$

With $\nu = n - 1$ degrees of freedom and Γ as the gamma function defined by,

$$\Gamma(z) = \int_0^{\infty} t^{z-1} e^{-t} dt \quad (8)$$

3.3.2 Generalized Error

The Generalized Error Distribution is a symmetrical unimodal member of the exponential family.

The domain of the p.d.f. is $x \in [-\infty, \infty]$ and the distribution is defined by,

$$f_v(x) = \frac{v \exp\left\{-\frac{1}{2}\left|\frac{x}{\lambda}\right|^v\right\}}{\lambda 2^{1+\frac{1}{v}} \Gamma\left(\frac{1}{v}\right)}, v > 0, x \in \mathbb{R}, \quad (9)$$

Where,

$$\lambda = \left[2^{-\left(\frac{2}{v}\right)} \frac{\Gamma\left(\frac{1}{v}\right)}{\Gamma\left(\frac{3}{v}\right)} \right]^{\frac{1}{2}} \quad (10)$$

and $\Gamma(\cdot)$ denotes the Gamma function.

3.3.3 Skewed Student-t

The probability density function of skewed student distribution is defined as,

$$f(x|\eta, \lambda) = \begin{cases} bc \left(1 + \frac{1}{\eta - 2} \left(\frac{a + bx}{1 - \lambda}\right)^2\right)^{-\left(\frac{\eta+1}{2}\right)}, & x < -\frac{a}{b}, \\ bc \left(1 + \frac{1}{\eta - 2} \left(\frac{a + bx}{(1 + \lambda)^2}\right)\right)^{-\left(\frac{\eta+1}{2}\right)}, & x \geq -\frac{a}{b}, \end{cases} \quad (11)$$

Where $2 < \eta < \infty$ is the degrees of freedom, and $-1 < \lambda < 1$ is the skewness parameter. The constants a , b , and c are given by,

$$a = 4\lambda c \frac{\eta - 2}{\eta - 1} \quad (12)$$

$$b^2 = 1 + 3\lambda^2 - a^2 \quad (13)$$

$$c = \frac{\Gamma\left(\frac{\eta + 1}{2}\right)}{\sqrt{\pi(\eta - 2)\Gamma\left(\frac{\eta}{2}\right)}} \quad (14)$$

3.3.4 Pareto

The probability density function of the pareto distribution is defined as,

$$f(x) = \begin{cases} \frac{\alpha x_m^\alpha}{x^{\alpha+1}}, & x \geq x_m \\ 0, & x < x_m \end{cases} \quad (15)$$

Where x_m is the (necessarily positive) minimum possible value of x .

3.3.5 Weibull

The probability density function of skewed student distribution is defined as,

$$f(x) = \frac{\beta}{\eta} \left(\frac{x - \gamma}{\eta} \right)^{\beta-1} e^{-\left(\frac{x-\gamma}{\eta} \right)^\beta} \quad (16)$$

where,

$$f(x) \geq 0, x \geq 0 \text{ or } \gamma, \beta > 0, \eta > 0, -\infty < \gamma < \infty \quad (17)$$

and, β is the shape parameter / Weibull slope, η is the scale parameter and γ is the location parameter.

3.4 Backtests

Defining the hit sequence:

A VaR_{t+1}^α measure promises that the actual return will only be worse than the VaR_{t+1}^α forecast $\alpha \times 100\%$ of the time. When compared a time series of past ex ante VaR forecasts and past ex post returns, according to Christoffersen it is possible to define a hit sequence of VaR violations as,

$$I_{t+1} = \begin{cases} 1, & \text{if } r_{t+1} < -VaR_{t+1}^\alpha \\ 0, & \text{if } r_{t+1} \geq -VaR_{t+1}^\alpha \end{cases} \quad (18)$$

The hit sequence returns a 1 on day t+1 if the loss on that day was larger than the VaR number predicted in advance for that day. If the VaR was not violated, then HIT sequence returns a 0. When performing a backtest on the risk model, it is necessary to construct a sequence $\{I_{t+1}\}_{t=1}^T$ across T days indicating when the past violations occurred.

Defining the Null Hypothesis:

The null hypothesis states that if we are using a perfect VaR model, then given all the information available to us at the time the VaR forecast is made, we should not be able to predict whether the

VaR will be violated. The forecast of the probability VaR violation should simply be α every day. If we can predict VaR violations, then that information should be used to construct a better model, meaning the hit sequence should be completely unpredictable and therefore distributed independently over time as a Bernoulli variable that takes the value of 1 with probability α and the value 0 with probability $(1-\alpha)$, being,

$$H_0: \mathbb{Y}_{t+1} \sim i. i. d. \text{ Bernoulli } (\alpha)$$

The Bernoulli distribution function is written as,

$$f(I_{t+1}; \alpha) = (1 - \alpha)^{1-I_{t+1}} \alpha^{I_{t+1}} \quad (19)$$

3.4.1 Unconditional Coverage

As explained above, the unconditional coverage tests account to check if the fraction of violations obtained for the estimated model \mathbb{Y} , is significantly different from the promised fraction, α . This is the unconditional coverage hypothesis. To perform this test it is necessary to use the likelihood of a i.i.d. Bernoulli (\mathbb{Y}) hit sequence,

$$L(\mathbb{Y}) = \prod_{t=1}^T (1 - \mathbb{Y})^{1-I_{t+1}} \mathbb{Y}^{I_{t+1}} = (1 - \mathbb{Y})^{T_0} \mathbb{Y}^{T_1} \quad (20)$$

Where T_0 and T_1 are the number of zeros and ones in the sample. It is possible to then estimate \mathbb{Y} from $\hat{\mathbb{Y}} = \frac{T_1}{T}$ that is the observed fraction of violations in the sequence. Plugging the Maximum Likelihood (ML) estimates back into the likelihood function it is possible to obtain the optimized likelihood as,

$$L(\hat{\mathbb{Y}}) = \left(1 - \frac{T_1}{T}\right)^{T_0} \left(\frac{T_1}{T}\right)^{T_1} \quad (21)$$

It is then possible to check the unconditional coverage of the null hypothesis that $\mathbb{Y}=\alpha$, where, as above, α is the known VaR coverage rate, we get the likelihood as,

$$L(\alpha) = \prod_{t=1}^T (1 - \alpha)^{1-I_{t+1}} \alpha^{I_{t+1}} = (1 - \alpha)^{T_0} \alpha^{T_1} \quad (22)$$

Then it is possible to check the unconditional coverage hypothesis using the likelihood ratio test,

$$LR_{uc} = -2 \ln \left[\frac{L(\alpha)}{L(\hat{\Psi})} \right] \quad (23)$$

Alternately, it is possible to calculate the p-value associated with the test statistic above. The p-value is defined as the probability of obtaining a sample that converges even less to the null hypothesis than the sample we actually got given that the null hypothesis is true. In that case, the P-value is calculated as,

$$p - value = 1 - F_{x_1^2}(LR_{uc}) \quad (24)$$

Where $F_{x_1^2}(\bullet)$ denotes the cumulative density function of a x^2 variable with one degree of freedom. If the p-value is below the desired significance level, then we reject the null hypothesis.

3.4.2 Independence test

As explained above, if VaR violations are clustered, then the investor can essentially predict that if today is a violation, then tomorrow is more than $\alpha \cdot 100\%$ likely to be a violation as well. An investor needs to establish a test that will be able to reject the VaR with clustered violations. To this end is necessary to 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 of,

$$\Pi_1 = \begin{bmatrix} 1 - \Psi_{01} & \Psi_{01} \\ 1 - \Psi_{11} & \Psi_{11} \end{bmatrix} \quad (25)$$

The transition probabilities mean that conditional on today being a nonviolation, meaning $I_t = 0$, then the probability of tomorrow being a violation given today is also a violation defined by,

$$\Psi_{11} = \Pr(I_{t+1} = 1 | I_t = 1) \quad (26)$$

Similarly, the probability of tomorrow being a violation given today is not a violation can be defined by,

$$\Psi_{01} = \Pr(I_{t+1} = 1 | I_t = 0) \quad (27)$$

The first order Markov property refers to the assumption that only today's outcome matters for tomorrow's outcome being that the exact sequence of past hits does not matter, only the value of I_t matters.

As only two outcomes are possible, the two probabilities Ψ_{01} and Ψ_{11} describe the entire process. The probability of a nonviolation following a nonviolation is $1 - \Psi_{01}$, and the probability of a nonviolation following a violation is $1 - \Psi_{11}$.

Given a sample of T observations, then the likelihood function of the first-order Markov process as,

$$L(\Pi_1) = (1 - \Psi_{01})^{T_{00}} \Psi_{01}^{T_{01}} (1 - \Psi_{11})^{T_{10}} \Psi_{11}^{T_{11}} \quad (28)$$

Where T_{ij} , $i, j = 0, 1$ is the number of observations with a j following an i . Taking first derivatives with respect to Ψ_{01} and Ψ_{11} and setting these derivatives to zero, it is possible to solve for the maximum likelihood estimates,

$$\hat{\Psi}_{01} = \frac{T_{01}}{T_{00} + T_{01}} \quad (29)$$

$$\hat{\Psi}_{11} = \frac{T_{11}}{T_{10} + T_{11}} \quad (30)$$

Using then the fact that the probabilities have to sum to the one, we have,

$$\hat{\Psi}_{00} = 1 - \hat{\Psi}_{01} \quad (31)$$

$$\hat{\Psi}_{10} = 1 - \hat{\Psi}_{11}$$

Which gives the matrix of estimated transition probabilities,

$$\hat{\Pi}_1 = \begin{bmatrix} \hat{\Psi}_{00} & \hat{\Psi}_{01} \\ \hat{\Psi}_{10} & \hat{\Psi}_{11} \end{bmatrix} = \begin{bmatrix} 1 - \hat{\Psi}_{01} & \hat{\Psi}_{01} \\ 1 - \hat{\Psi}_{11} & \hat{\Psi}_{11} \end{bmatrix} = \begin{bmatrix} \frac{T_{00}}{T_{00} + T_{01}} & \frac{T_{01}}{T_{00} + T_{01}} \\ \frac{T_{10}}{T_{10} + T_{11}} & \frac{T_{11}}{T_{10} + T_{11}} \end{bmatrix} \quad (32)$$

Allowing for dependence in the hit sequence corresponds to allowing Ψ_{01} to be different from Ψ_{11} .

If, the HIT tests are independent over time, then the probability of a violation tomorrow does not depend on today being a violation or not, and is possible to write $\Psi_{01} = \Psi_{11} = \Psi$. Under independence, the transition matrix is,

$$\hat{\Pi} = \begin{bmatrix} 1 - \hat{\Pi} & \hat{\Pi} \\ 1 - \hat{\Pi} & \hat{\Pi} \end{bmatrix} \quad (33)$$

And is possible to test the independence hypothesis that $\Psi_{01} = \Psi_{11}$ using a likelihood ratio test

$$LR_{ind} = 2 \ln \left[\frac{L(\hat{\Pi})}{L(\hat{\Pi}_1)} \right] \sim \chi_1^2 \quad (34)$$

Where $L(\hat{\Pi})$ is the likelihood under the alternative hypothesis from the LR_{uc} test.

The test used is the one proposed by Christoffersen and Pelletier (2004).

3.4.3 Conditional coverage

After testing if the VaR violations are independent and the average number of violations is correct, I will perform the Christoffersen et al. (2001) that jointly tests for independence and correct coverage using the conditional coverage test as stated,

$$LR_{cc} = -2 \ln \left[\frac{L(p)}{L(\hat{\Pi}_1)} \right] \sim \chi_2^2 \quad (35)$$

Which corresponds to testing that $\Psi_{01} = \Psi_{11} = p$.

The LR_{cc} tests takes the likelihood from the null hypothesis in the LR_{uc} test and combines it with the likelihood from the alternative in the LR_{ind} test. Therefore,

$$LR_{ind} = 2 \ln \left[\frac{L(\hat{\Pi})}{L(\hat{\Pi}_1)} \right] \sim \chi_1^2 \quad (36)$$

$$\begin{aligned} LR_{cc} &= -2 \ln \left[\frac{L(p)}{L(\hat{\Pi}_1)} \right] \quad (37) \\ &= -2 \ln \left[\{L(p)L(\hat{\Pi})\} \left\{ \frac{L(\hat{\Psi})}{L(\hat{\Pi}_1)} \right\} \right] \\ &= -2 \ln \left[\frac{L(p)}{L(\hat{\Pi})} \right] - 2 \ln \left[\frac{L(\hat{\Pi})}{L(\hat{\Pi}_1)} \right] \\ LR_{cc} &= LR_{uc} + LR_{ind} \end{aligned}$$

So that the joint test of conditional coverage can be calculated by simply summing the two individual tests for unconditional coverage and independence.

4. Data

The data used is composed of daily prices of four international stock market indices, from 3 different continents downloaded from Thomson Reuters Eikon's Datastream. I needed a time period as long as possible, in order to capture a large number of volatile periods. The indices selected were the Dow Jones Industrials, S&P 500, Nikkei 225 and Dax 30. The Dax 30 is the index with shorter history of values available, starting in 1/1/1964 till 25/11/2016 corresponding to almost 52 years. We used the same time period for all other indices. For the purpose of this dissertation, I will only work with the daily returns. After the calculation of the returns for each index, I ended up with 12581 returns observations for the Dow Jones Industrials, 12599 for S&P 500, 12254 for Nikkei 225 and 12712 for Dax 30. The difference in the number of observations is due to the different number of trading days per market.

Table 1 provide basic descriptive statistics for the chosen indexes.

Index	Dow Jones Ind.	S&P 500	Nikkei 225	Dax 30
Number of observations	12581	12599	12254	12712
Minimum	-0.226105	-0.204669	-0.149009	-0.128116
Maximum	0.110803	0.1158	0.132359	0.11402
1 st Quantile	-0.004732	-0.004522	-0.00546	-0.005968
3 rd Quantile	0.005376	0.005239	0.006294	0.006722
Mean	0.000282	0.000305	0.000241	0.000256
Median	0.000361	0.000404	0.00049	0.000452
Variance	0.000104	0.000105	0.000156	0.000152
Standard deviation	0.01022	0.010269	0.012489	0.012348
Skewness	-0.757673	-0.641088	-0.352827	-0.123724
Kurtosis	26.26849	20.89739	8.632922	6.924438
P-value normality test	$2.2 * 10^{-16}$	$2.2 * 10^{-16}$	$2.2 * 10^{-16}$	$2.2 * 10^{-16}$

Table 1: Basic stats of the four indexes.

Kurtosis provide a measure of the thickness of the tails of the distribution, the higher the value, the higher the probability of extreme events. The positive kurtosis is according to Lee & Lee & Lee

(2000) called leptokurtic and lower kurtosis meaning a negative excess is called platykurtic. The skewness in all indices is negative, which indicates that the returns are skewed to the left. This means another departure from the assumption of normality.

A formal normality test for each index is presented using the Jarque-Bera test. As we can observe the test rejects the hypothesis of normality for all indices.

Next we can see the daily returns for each index, and in all indexes it is visible some volatility clusters, and some of the large absolute returns are visible in the graph.

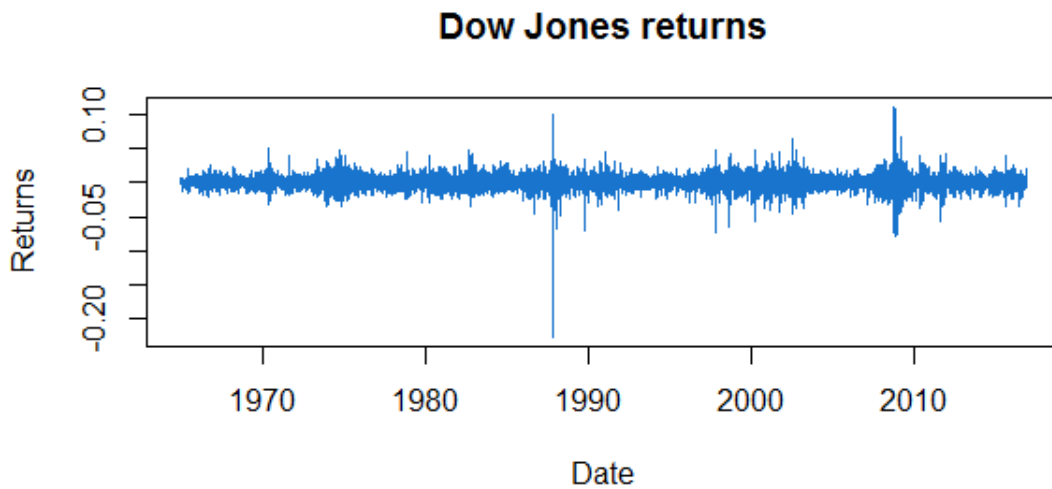


Figure 2: Dow Jones Industrials daily returns from 1/1/1964 till 25/11/2016.

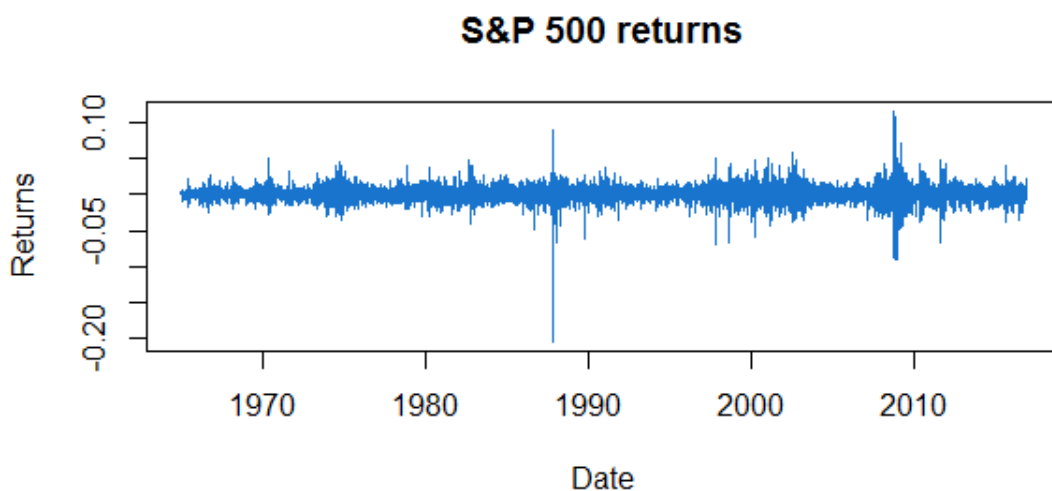


Figure 3: S&P 500 daily returns from 1/1/1964 till 25/11/2016.

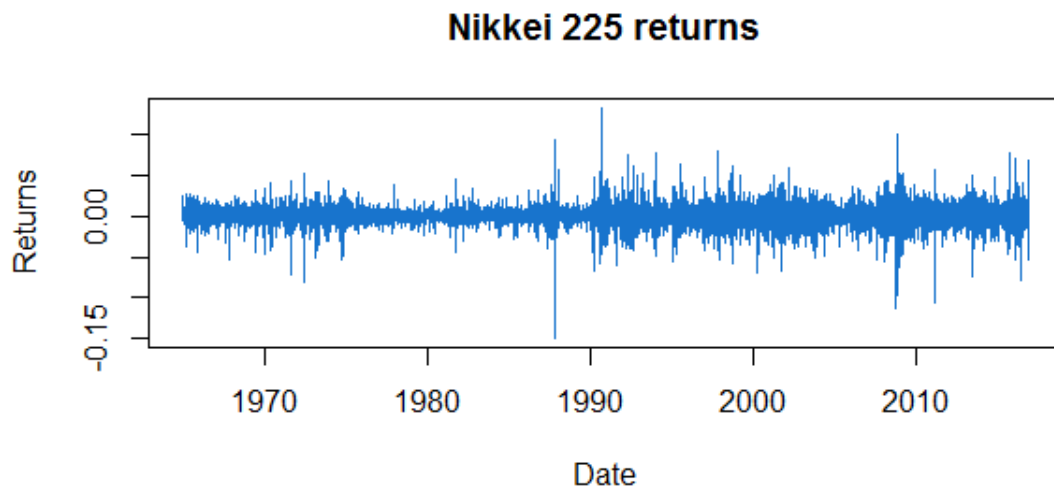


Figure 4: Nikkei 225 daily returns from 1/1/1964 till 25/11/2016.

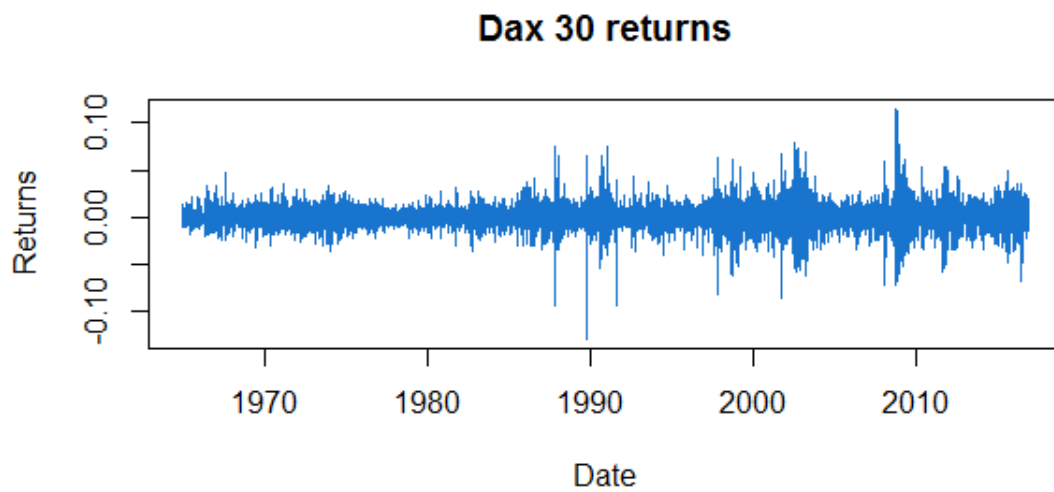


Figure 5: Dax 30 daily returns from 1/1/1964 till 25/11/2016.

5. Results

5.1 VaR calculation

With many different models and approaches possible, the choice that VaR users face is the choice of the right one that matches their purpose best. The approaches should make estimates that fit best the future distribution of returns. If an overestimation of the VaR measure is made, then investors end up with an overestimate of risk and therefore restraining their investments opportunities. On the other side, if an underestimation is made, investors may fail to cover incurred losses.

In this dissertation the VaR for the four selected indexes, being them Dow Jones Industrials, S&P 500, Nikkei 225 and Dax 30 is estimated using the Unconditional Parametric approach. The performance of the parametric approaches depends a lot on the distribution assumed for returns, so I will test different distributions as Student t, Generalized Error, Skewed Student t, Pareto and Weibull and with confidence levels of 95%, 97.5% and 99%.

The amount of data used and complexity of the calculations involved goes beyond the capabilities of more common software like Excel, so the tool that will be used to perform the calculations is the R software, through Rstudio interface.

5.1.1 Unconditional Parametric Estimation

The Unconditional Parametric approach using the maximum likelihood estimation was performed through a rolling function procedure also known as rolling window where for the Student t, Generalized Error, and Skewed Student t had a rolling window with a width of 2500 observations consisting of roughly of 10 years, while the Pareto and Weibull have a smaller one, of 1500 because this distribution can only be calculated using the days where the returns were negative. The confidence levels used were some of the most commonly used in literature and some of the recommended by Christoffersen (2012), that are 95%, 97.5% and 99%.

The figures bellow illustrate the different Unconditional Parametric VaR estimates of the different distributions assumed in comparison with the returns of each index for the several confidence levels. It is already possible to see that the Skewed Student t may be doing an underestimation of risk and the Pareto and Weibull probably may be doing an overestimation. Just by the graphs it is not possible to make further analysis neither be certain of the stated before, so in the next sections

empirical backtest studies are presented and will examine the models in detail and show if they can be called adequate or not.

In the next pages, the graphs with the Unconditional Parametric estimation of VaR for each index and the tree levels of confidence.

Dow Jones Industrials - Unconditional Parametric - Student t

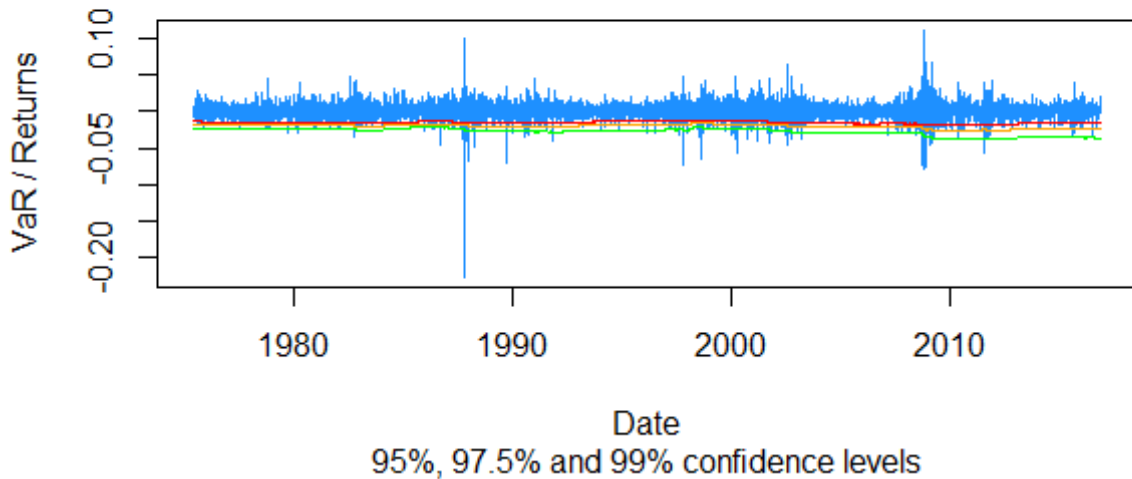


Figure 6: Dow Jones Industrials returns compared to estimated VaR for the Unconditional Parametric approach under Student t distribution from 26/06/1975 till 25/11/2016.

S&P 500 - Unconditional Parametric - Student t

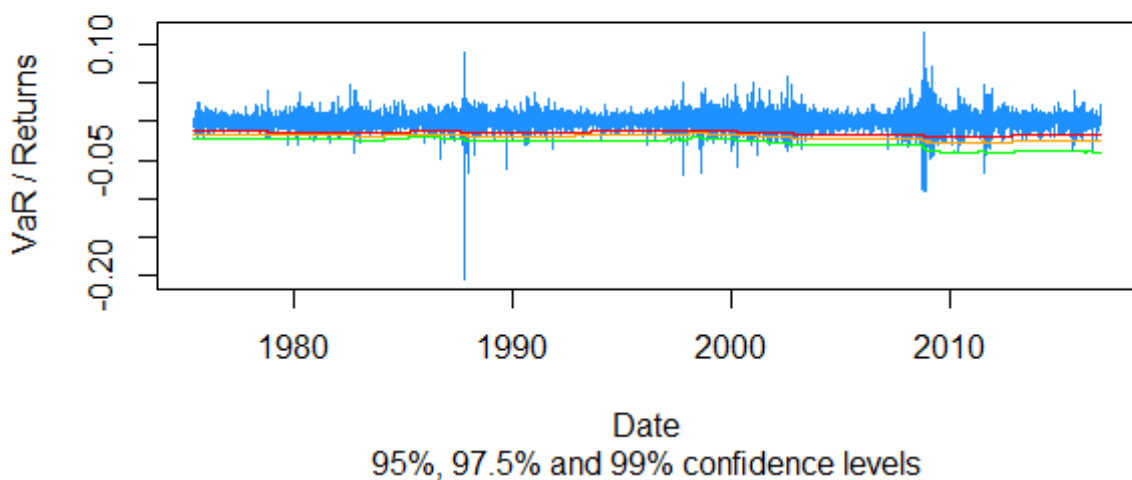


Figure 7: S&P 500 returns compared to estimated VaR for the Unconditional Parametric approach under Student t distribution from 26/06/1975 till 23/11/2016.

Nikkei 225 - Unconditional Parametric - Student t

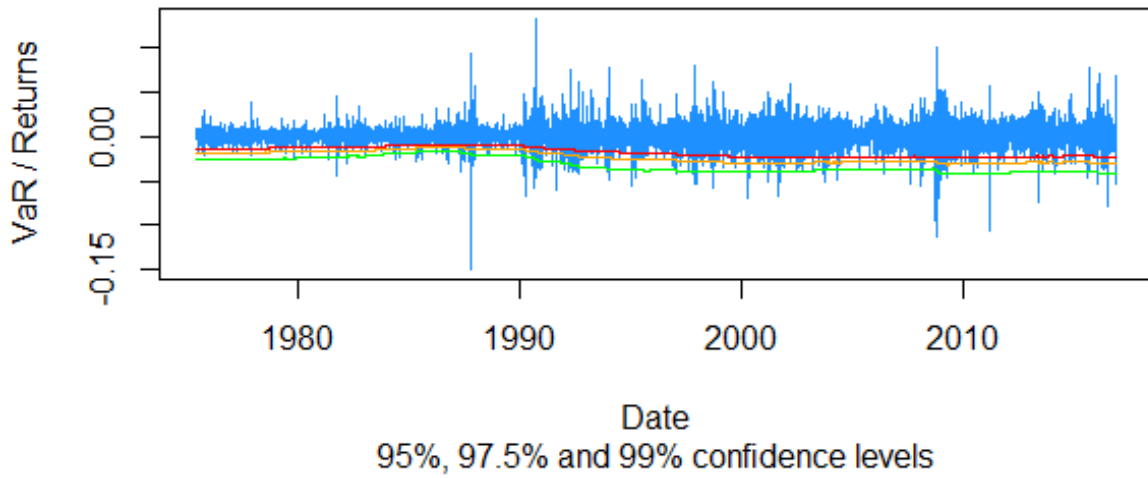


Figure 8: Nikkei 225 returns compared to estimated VaR for the Unconditional Parametric approach under Student t distribution from 28/08/1975 till 23/11/2016.

Dax 30 - Unconditional Parametric - Student t

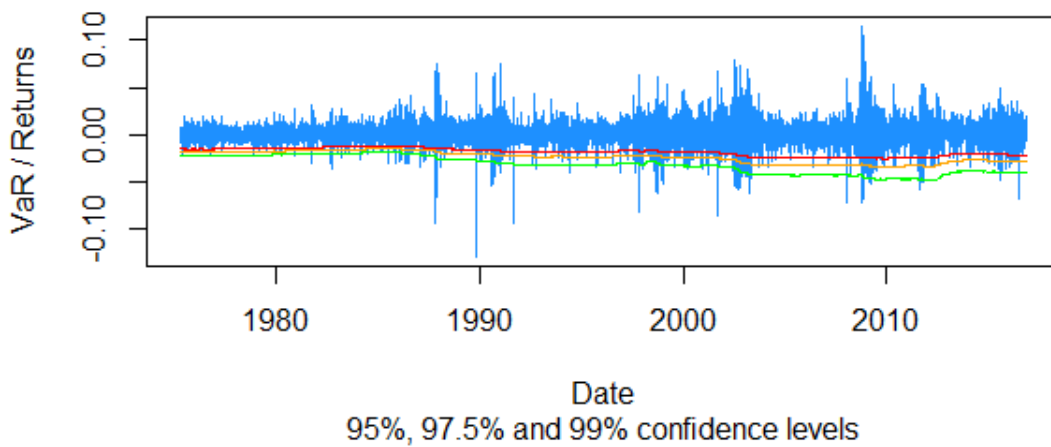


Figure 9: Dax 30 returns compared to estimated VaR for the Unconditional Parametric approach under Student t distribution from 18/04/1975 till 23/11/2016.

Dow Jones Industrials – Unconditional Parametric Generalized

Error

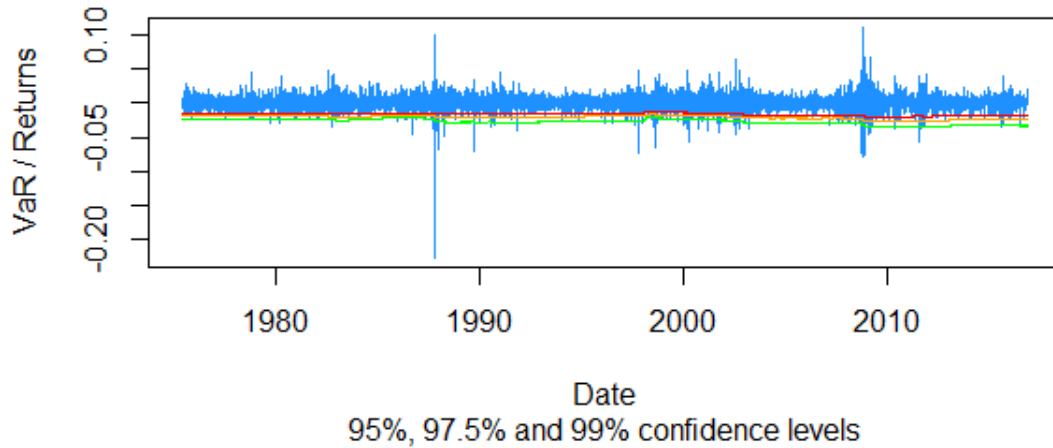


Figure 10: Dow Jones Industrials returns compared to estimated VaR for the Unconditional Parametric approach under Generalized Error distribution from 26/06/1975 till 23/11/2016.

S&P 500 - Unconditional Parametric - Generalized Error

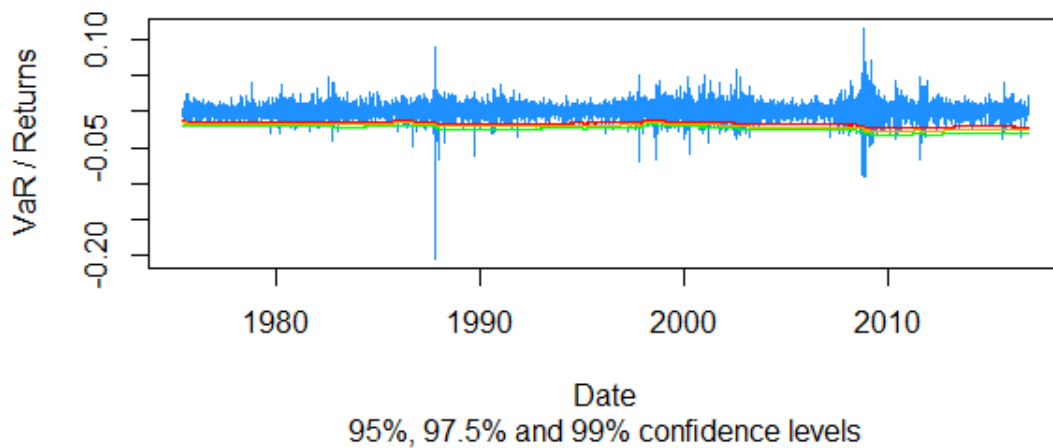


Figure 11: S&P 500 returns compared to estimated VaR for the Unconditional Parametric approach under Generalized Error distribution from 26/06/1975 till 23/11/2016.

Nikkei 225 - Unconditional Parametric - Generalized Error

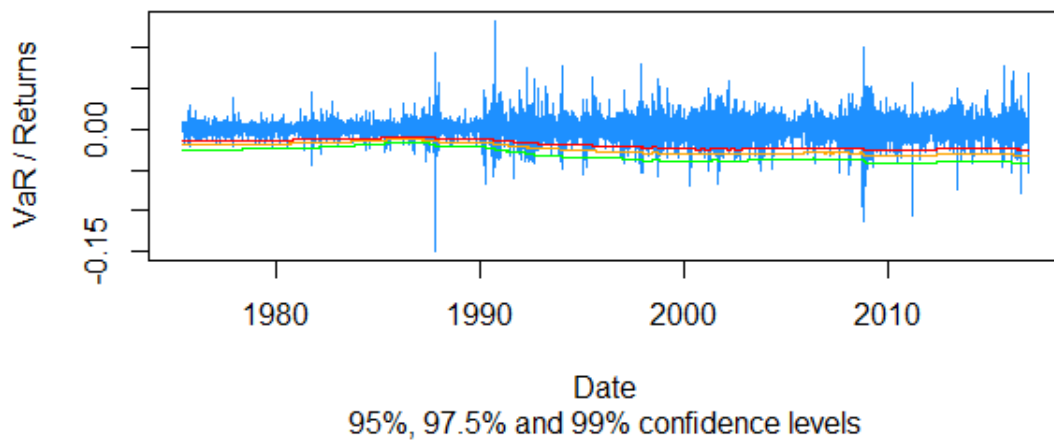


Figure 12: Nikkei 225 returns compared to estimated VaR for the Unconditional Parametric approach under Generalized Error distribution from 28/08/1975 till 23/11/2016.

Dax 30 - Unconditional Parametric - Generalized Error

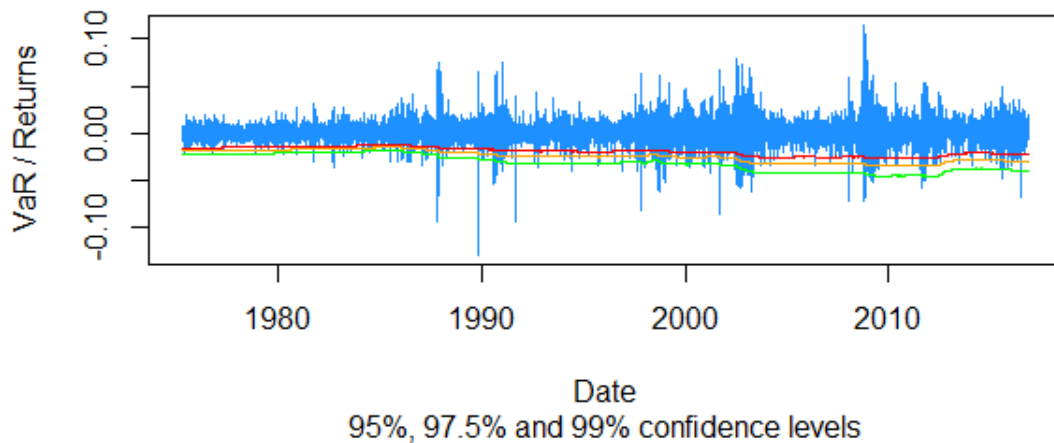


Figure 13: Dax 30 returns compared to estimated VaR for the Unconditional Parametric approach under Generalized Error distribution from 18/04/1975 till 23/11/2016.

Dow Jones Industrials – Unconditional Parametric – Skewed

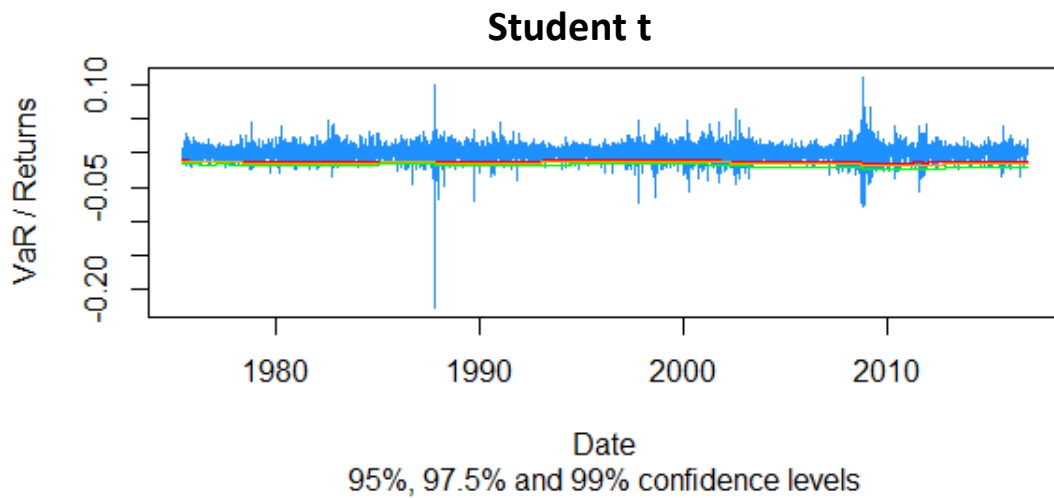


Figure 14: Dow Jones Industrials returns compared to estimated VaR for the Unconditional Parametric approach under Skewed Student t distribution from 26/06/1975 till 23/11/2016.

S&P 500 - Unconditional Parametric - Skewed Student t

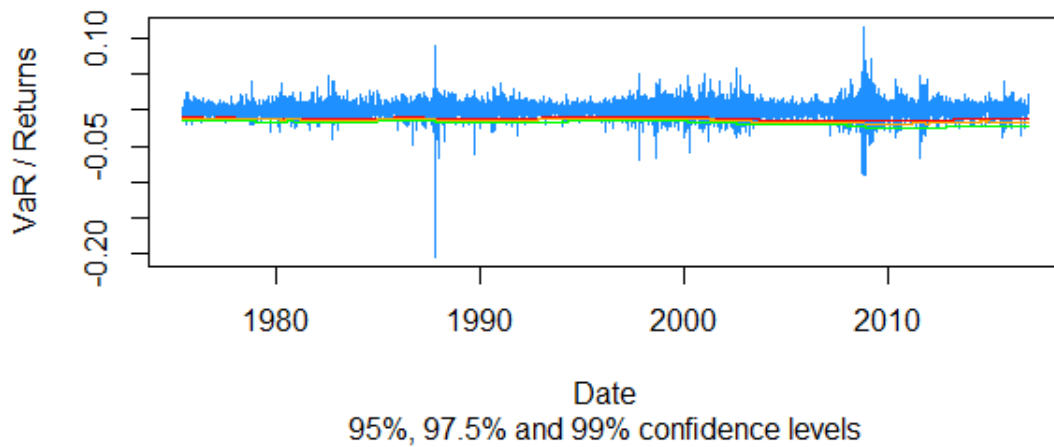


Figure 15: S&P 500 returns compared to estimated VaR for the Unconditional Parametric approach under Skewed Student t distribution from 26/06/1975 till 23/11/2016.

Nikkei 225 - Unconditional Parametric - Skewed Student t

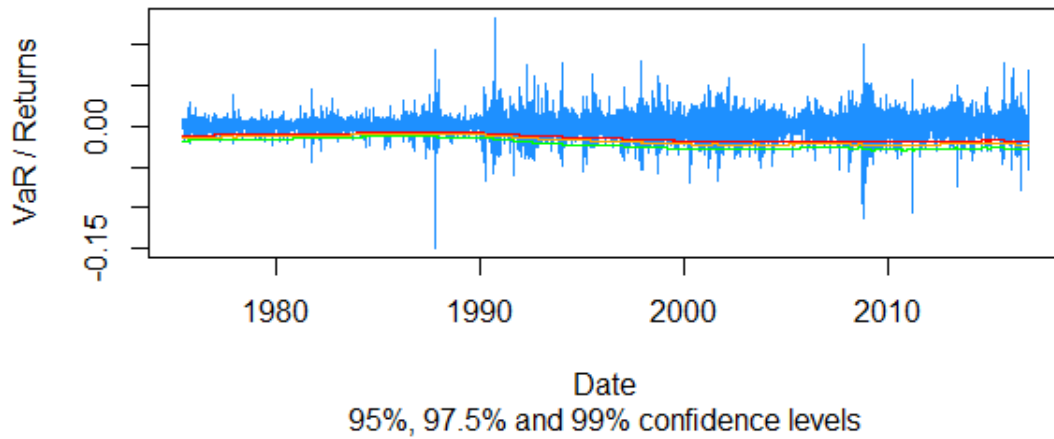


Figure 16: Nikkei 225 returns compared to estimated VaR for the Unconditional Parametric approach under Skewed Student t distribution from 28/08/1975 till 23/11/2016.

Dax 30 - Unconditional Parametric - Skewed Student t

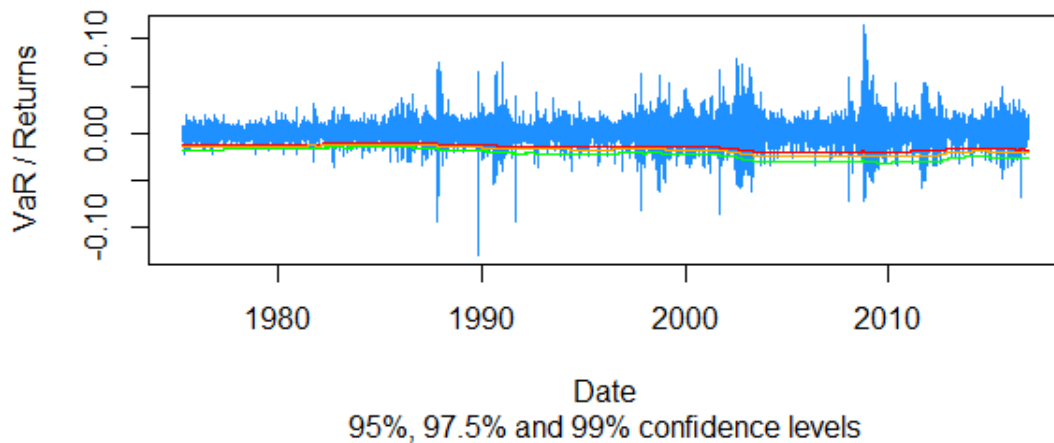


Figure 17: Dax 30 returns compared to estimated VaR for the Unconditional Parametric approach under Skewed Student t distribution from 18/04/1975 till 23/11/2016.

Dow Jones Industrials - Unconditional Parametric - Pareto

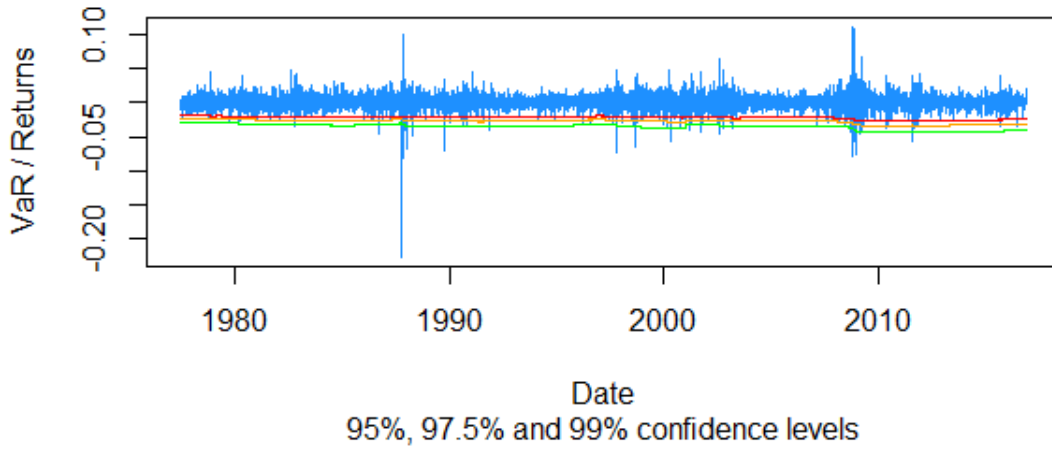


Figure 18: Dow Jones Industrials returns compared to estimated VaR for the Unconditional Parametric approach under Pareto distribution from 26/06/1975 till 23/11/2016.

S&P 500 - Unconditional Parametric - Pareto

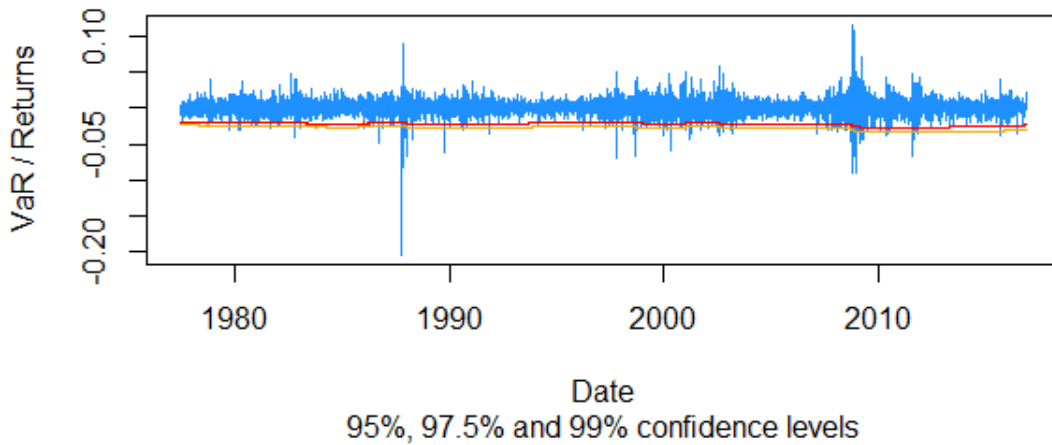


Figure 19: S&P 500 returns compared to estimated VaR for the Unconditional Parametric approach under Pareto distribution from 26/06/1975 till 23/11/2016.

Nikkei 225 - Unconditional Parametric - Pareto

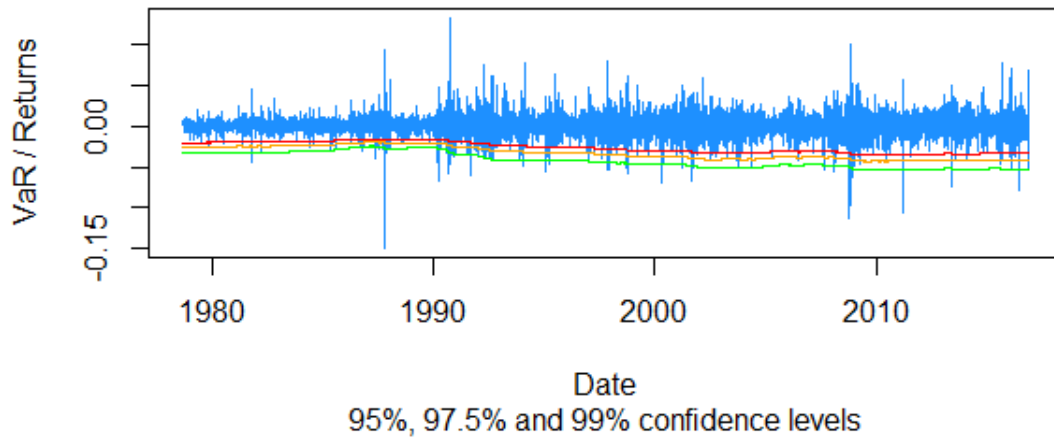


Figure 20: S&P 500 returns compared to estimated VaR for the Unconditional Parametric approach under Pareto distribution from 28/08/1975 till 23/11/2016.

Dax 30 - Unconditional Parametric - Pareto

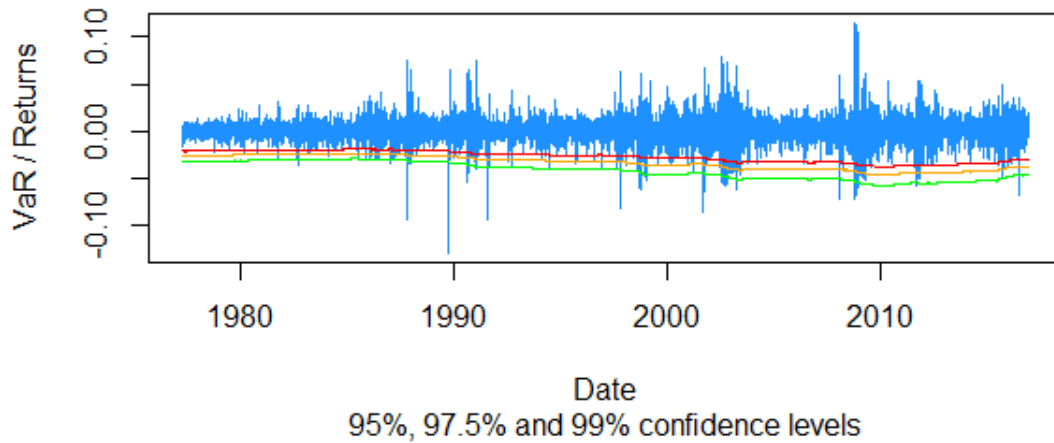


Figure 21: S&P 500 returns compared to estimated VaR for the Unconditional Parametric approach under Pareto distribution from 18/04/1975 till 23/11/2016.

Dow Jones Industrials - Unconditional Parametric - Weibull

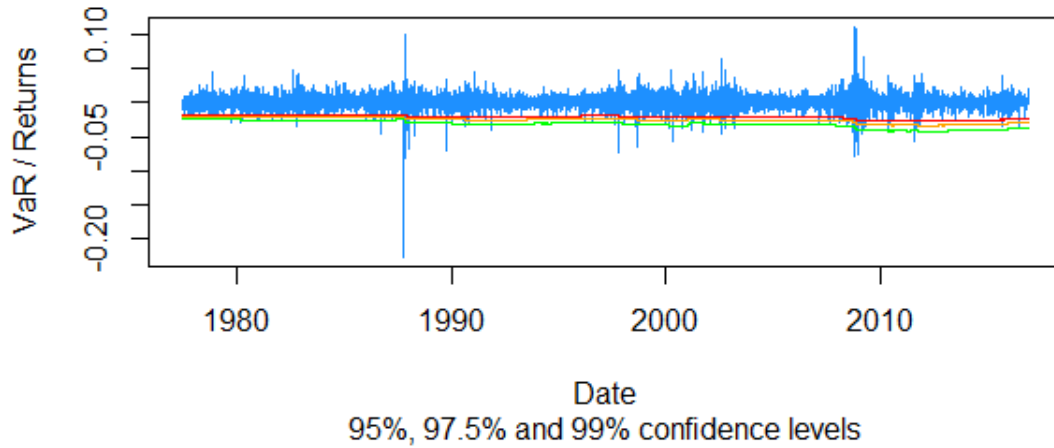


Figure 22: Dow Jones Industrials returns compared to estimated VaR for the Unconditional Parametric approach under Weibull distribution from 26/06/1975 till 23/11/2016.

S&P 500 - Unconditional Parametric - Weibull

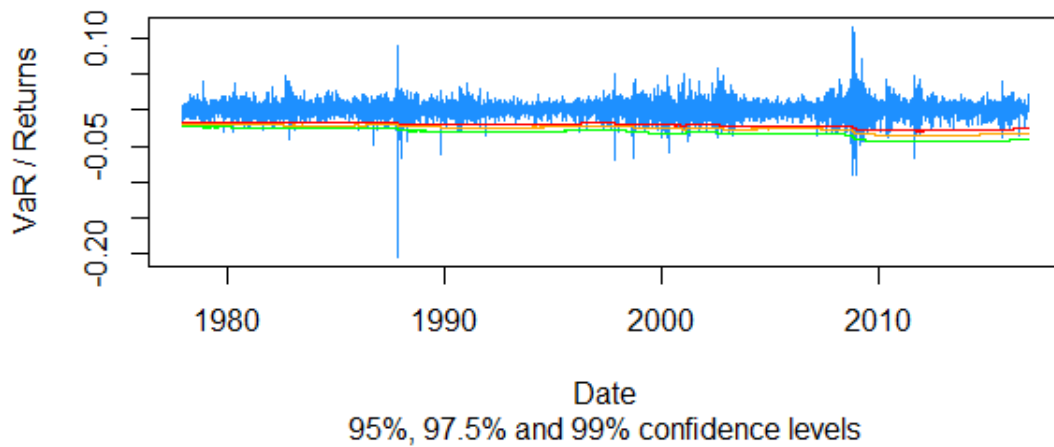


Figure 23: S&P 500 returns compared to estimated VaR for the Unconditional Parametric approach under Weibull distribution from 26/06/1975 till 23/11/2016.

Nikkei 225 - Unconditional Parametric - Weibull

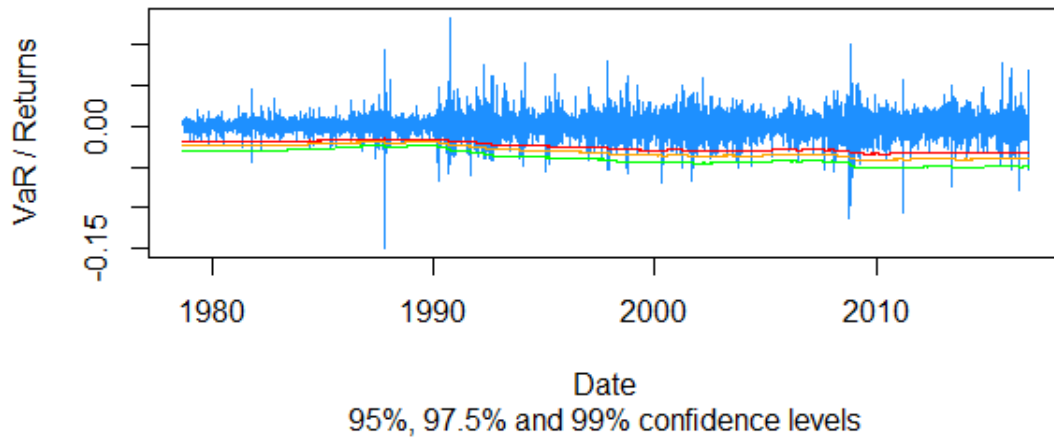


Figure 24: Nikkei 225 returns compared to estimated VaR for the Unconditional Parametric approach under Weibull distribution from 28/08/1975 till 23/11/2016.

Dax 30 - Unconditional Parametric - Weibull

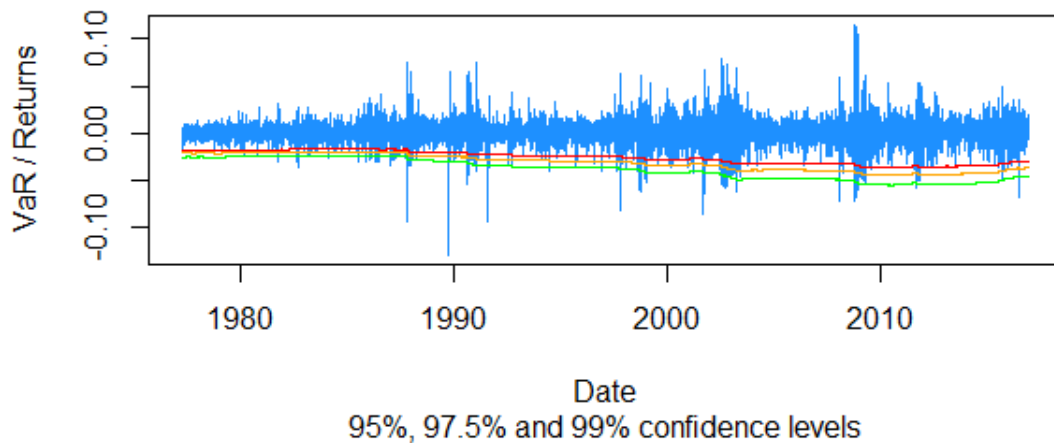


Figure 25: Dax 30 returns compared to estimated VaR for the Unconditional Parametric approach under Weibull distribution from 18/04/1975 till 23/11/2016.

5.2 Backtests

5.2.1 Unconditional Coverage Test

As explained before the unconditional coverage test of Christoffersen (1998) accounts for the number of exceptions is more or less than the selected confidence level would indicate. In the next two pages, table 3, 4, 5 and 6 present the results of Christoffersen tests of conditional coverage for $p=0,05$, $p=0,025$ and $p=0,01$. The tables show the number of expected exceptions and the actual one, then the Likelihood Ratio statistic (LR) for the unconditional coverage is presented and the decision for H_0 : Correct number of exceptions.

Dow Jones Industrials															
95%						97.5%					99%				
	Expected	Actual	LR test	P-value	Decision	Expected	Actual	LR test	P-value	Decision	Expected	Actual	LR test	P-value	Decision
Student t	504	535	1.962796	0.1612146	Fail to Reject H0	252	268	1.017837	0.3130327	Fail to Reject H0	100	105	0.1735459	0.6769795	Fail to Reject H0
Generalized Error	504	475	1.795397	0.18027	Fail to Reject H0	252	233	1.510538	0.219057	Fail to Reject H0	100	113	1.432854	0.2312992	Fail to Reject H0
Skewed Student t	504	852	211.3461	0	Reject H0	252	585	330.6975	0	Reject H0	100	337	346.6681	0	Reject H0
Pareto	479	191	234.4836	0	Reject H0	239	90	125.5557	0	Reject H0	95	42	38.76329	4.784411e-10	Reject H0
Weibull	479	229	169.4076	0	Reject H0	239	121	73.54595	0	Reject H0	95	60	15.66776	7.550038e-05	Reject H0

Table 2: Unconditional Coverage Test for the Unconditional Parametric estimation of Dow Jones Industrials.

S&P 500															
95%						97.5%					99%				
	Expected	Actual	LR test	P-value	Decision	Expected	Actual	LR test	P-value	Decision	Expected	Actual	LR test	P-value	Decision
Student t	504	588	13.68762	0.0002158732	Reject H0	252	300	8.661897	0.003249328	Reject H0	100	109	0.6255699	0.6255699	Fail to Reject H0
Generalized Error	504	580	1.31845	0.2508704	Fail to Reject H0	252	302	9.384843	0.002187867	Reject H0	100	183	54.22861	1.785239e-13	Reject H0
Skewed Student t	504	905	272.9087	0	Reject H0	252	626	404.166	0	Reject H0	100	359	401.3363	0	Reject H0
Pareto	480	210	201.236	0	Reject H0	237	134	54.83506	0	Reject H0	96	50	27.07672	1.955385e-07	Reject H0
Weibull	475	252	132.1334	0	Reject H0	224	155	24.86887	1.311173e-13	Reject H0	96	67	9.310428	0.002278533	Reject H0

Table 3: Unconditional Coverage Test for the Unconditional Parametric estimation of S&P 500.

Nikkei 225															
95%						97.5%					99%				
	Expected	Actual	LR test	P-value	Decision	Expected	Actual	LR test	P-value	Decision	Expected	Actual	LR test	P-value	Decision
Student t	487	615	32.42073	1.241552e-08	Reject H0	243	319	21.68634	3.210692e-06	Reject H0	97	142	17.94534	2.273408e-05	Reject H0
Generalized Error	487	528	3.41751	0.06450824	Fail to Reject H0	243	284	6.444848	0.01112746	Reject H0	97	150	24.47734	7.518908e-07	Reject H0
Skewed Student t	487	930	337.4565	0	Reject H0	243	644	467.6133	1.735855e-10	Reject H0	97	388	499.3622	0	Reject H0
Pareto	449	261	97.3629	0	Reject H0	224	137	40.74344	0	Reject H0	89	69	5.335423	0.02089627	Fail to Reject H0
Weibull	449	283	74.34314	0	Reject H0	242	173	22.89656	6.136537e-07	Reject H0	89	90	0.0001123171	0.9915442	Fail to Reject H0

Table 4: Unconditional Coverage Test for the Unconditional Parametric estimation of Nikkei 225.

Dax 30															
95%						97.5%					99%				
	Expected	Actual	LR test	P-value	Decision	Expected	Actual	LR test	P-value	Decision	Expected	Actual	LR test	P-value	Decision
Student t	510	609	18.85174	1.412824e-05	Reject H0	255	338	24.97886	5.796251e-07	Reject H0	102	162	30.10444	2.273408e-05	Reject H0
Generalized Error	510	561	5.081087	0.02418821	Reject H0	255	308	10.47827	0.001207866	Reject H0	102	170	37.97773	5.995204e-15	Reject H0
Skewed Student t	510	909	268.3366	0	Reject H0	255	656	453.1092	0	Reject H0	102	408	527.8461	0	Reject H0
Pareto	485	230	174.6743	0	Reject H0	242	134	59.63802	1.14353e-14	Reject H0	97	56	20.78305	5.143627e-06	Reject H0
Weibull	285	267	123.1213	0	Reject H0	242	173	22.89656	1.709574e-06	Reject H0	97	86	1.346912	0	Fail to Reject H0

Table 5: Unconditional Coverage Test for the Unconditional Parametric estimation of Dax 30.

As presented in table 2, the Student t and Generalized Error distribution perform best for Dow Jones Industrials, making, in terms of unconditional coverage, a fairly good estimation for all levels of confidence having a good prediction of exceptions, suggesting that they fit better the Dow Jones returns probability distribution. On the other hand, Skewed Student t made the biggest error, with exceptions being very underestimated and is therefore considered inadequate. The Pareto and Weibull distributions make similar results both with some overestimation of risk and therefore are considered inadequate for all confidence levels.

In table 3 we can see the S&P 500 results for the unconditional coverage test, and here again Student t and Generalized Error distributions performs overall best but the results are not as good as for the Dow Jones Industrials, in fact the Student t Unconditional Parametric Model is only valid for 99% VaR and for the Generalized Error one, only the 95% VaR, making both this models always an underestimation of risk. Again, as in Dow Jones Industrials, it can be seen that Skewed t distribution makes a severe underestimation of risk. Pareto and Weibull again make an overestimation of risk, and although for the $p=0,01$ in the Weibull distribution the gap between the predicted number of exceptions and the actual is not very big, the model is still considered inadequate.

The table 4 show the results for Nikkei 225 and it is where the Unconditional Parametric model for the Weibull distribution gets the overall result for the 99% VaR, only missing by one the exception, which, considering that this test accounts only for the unconditional coverage which alone is not enough to validate a model, is an excellent result with a Likelihood Ratio test of 0.0001123171. As for Student t, Generalized Error and Skewed Student t distributions, they all underestimate the risk, and here again the last one having the worst results overall. In Generalized Error distribution for a 90% VaR a valid model is obtained as in the previews two indexes. The Pareto and Weibull here again overestimate risk, but obtain acceptable results in the higher confidence levels (99%).

For the Dax 30 index (table 5), the Unconditional Parametric approach performed worst with only one model not being rejected. The first tree distributions did as in the other indexes, an underestimation of risk with Generalized Error distribution, that although was rejected the null hypothesis still produced a Likelihood Ratio of 5.081087, in line with the previews results it is one of the best estimations in terms of conditional coverage. The Pareto and Weibull made an overestimation of risk as in all indexes, and as for Nikkei 225, the Unconditional Parametric

approach through the Weibull distribution provided the best result, with a valid model for the $p=0,01$.

5.2.2 Independence test

Failure test that examine if the number of exceptions is correct is not enough as explained before. It is also important to examine if the exceptions are spread evenly over time, or they occur in clusters. To examine this, I conduct the Christoffersen's (2004) interval forecast test. The Null Hypothesis states: duration between exceedances have no memory. The estimated Weibull parameter b is presented as well as the result of the Null Hypothesis in the following tables.

	Dow Jones Industrials								
	95%			97,50%			99%		
	b	P-value	Decision	b	P-value	Decision	b	P-value	Decision
Student t	0.74313	0	Reject H0	0.6309642	0	Reject H0	0.5296511	0	Reject H0
Generalized Error	0.7294691	0	Reject H0	0.603897	0	Reject H0	0.5447902	0	Reject H0
Skewed Student t	0.8192536	0	Reject H0	0.7424613	0	Reject H0	0.6696607	0	Reject H0
Pareto	0.5976097	0	Reject H0	0.5208795	0	Reject H0	0.4173752	0	Reject H0
Weibull	0.6269891	0	Reject H0	0.5564018	0	Reject H0	0.4620558	0	Reject H0

Table 6: Christoffersen's (2004) Independence Test for the Unconditional Parametric estimation of Dow Jones Industrials.

	S&P 500								
	95%			97,50%			99%		
	b	P-value	Decision	b	P-value	Decision	b	P-value	Decision
Student t	0.7332928	0	Reject H0	0.6203527	0	Reject H0	0.522745	0	Reject H0
Generalized Error	0.6901373	0	Reject H0	0.6147362	0	Reject H0	0.5732411	0	Reject H0
Skewed Student t	0.8204729	0	Reject H0	0.7424613	0	Reject H0	0.6439167	0	Reject H0
Pareto	0.4126516	0	Reject H0	0.5978832	0	Reject H0	0.4126516	0	Reject H0
Weibull	0.6153383	0	Reject H0	0.5486223	0	Reject H0	0.4752031	0	Reject H0

Table 7: Christoffersen's (2004) Independence Test for the Unconditional Parametric estimation of S&P 500.

	Nikkei 225								
	95%			97,50%			99%		
	b	P-value	Decision	b	P-value	Decision	b	P-value	Decision
Student t	0.7541309	0	Reject H0	0.6077712	0	Reject H0	2	1	Fail to Reject H0
Generalized Error	0.7185645	0	Reject H0	0.5967511	0	Reject H0	2	1	Fail to Reject H0
Skewed Student t	0.8339562	0	Reject H0	0.7647547	0	Reject H0	0.6530396	0	Reject H0
Pareto	0.6002065	0	Reject H0	0.5340701	0	Reject H0	0.4927001	0	Reject H0
Weibull	0.6065283	0	Reject H0	0.5595792	0	Reject H0	0.5078649	0	Reject H0

Table 8: Christoffersen's (2004) Independence Test for the Unconditional Parametric estimation of Nikkei 225.

	Dax 30								
	95%			97,50%			99%		
	b	P-value	Decision	b	P-value	Decision	b	P-value	Decision
Student t	0.7192889	0	Reject H0	0.639801	0	Reject H0	0.5296694	0	Reject H0
Generalized Error	0.7027968	0	Reject H0	0.6240941	0	Reject H0	0.5340516	0	Reject H0
Skewed Student t	0.7742832	0	Reject H0	0.7359569	0	Reject H0	0.6530396	0	Reject H0
Pareto	0.5816759	0	Reject H0	0.5157459	0	Reject H0	0.4592076	0	Reject H0
Weibull	0.6083599	0	Reject H0	0.6083599	0	Reject H0	0.5358759	0	Reject H0

Table 9: Christoffersen's (2004) Independence Test for the Unconditional Parametric estimation of Dax 30.

As can be seen, the Unconditional Parametric approach results of the independence test show poor results, meaning that the losses are clustered, and that today's exception will probably increase the next day's probability of exception, and therefore the model did not perform well under volatility clusters resulting in clusters of exceptions, which is generally a bad thing for investors. However, two models, for the 99% confidence level were not rejected, the Generalized Error and Student t distributions for the Nikkei 225, all others, for all confidence levels failed the independence test.

5.2.3 Conditional Coverage test

Now that I have conducted the tests for coverage and independence separately, the Christoffersen's independence and unconditional coverage can be combined into a joint test of conditional coverage. The following tables, 10, 11, 12 and 13, present the joint test results for each index.

	Dow Jones Industrials								
	95%			97,50%			99%		
	LR test	P-value	Decision	LR test	P-value	Decision	LR test	P-value	Decision
Student t	40.03999	2.02035e-09	Reject H0	27.82078	9.094813e-07	Reject H0	19.14225	6.971274e-05	Reject H0
Generalized Error	38.30657	4.806557e-09	Reject H0	29.65215	3.640132e-07	Reject H0	18.2869	0.0001069179	Reject H0
Skewed Student t	255.2239	0	Reject H0	370.2088	0	Reject H0	372.4357	0	Reject H0
Pareto	251.718	0	Reject H0	135.4233	0	Reject H0	50.27061	1.213041e-11	Reject H0
Weibull	186.1339	0	Reject H0	91.49454	0	Reject H0	27.79847	9.196838e-07	Reject H0

Table 10: Christoffersen's Conditional Coverage Test for the Unconditional Parametric estimation of Dow Jones Industrials.

	S&P 500								
	95%			97,50%			99%		
	LR test	P-value	Decision	LR test	P-value	Decision	LR test	P-value	Decision
Student t	59.95843	9.55902e-14	Reject H0	45.26248	1.483806e-10	Reject H0	22.77189	1.135395e-05	Reject H0
Generalized Error	56.38015	5.717649e-13	Reject H0	48.16175	3.481848e-11	Reject H0	84.58711	0	Reject H0
Skewed Student t	319.8513	0	Reject H0	452.7738	0	Reject H0	448.09	0	Reject H0
Pareto	235.0076	0	Reject H0	126.4315	0	Reject H0	42.03641	7.44577e-10	Reject H0
Weibull	161.4033	0	Reject H0	88.99553	0	Reject H0	24.4813	4.830066e-06	Reject H0

Table 11: Christoffersen's Conditional Coverage Test for the Unconditional Parametric estimation of S&P500.

	Nikkei 225								
	95%			97,50%			99%		
	LR test	P-value	Decision	LR test	P-value	Decision	LR test	P-value	Decision
Student t	97.55519	0	Reject H0	121.3522	0	Reject H0	67.7870	1.887379e-15	Reject H0
Generalized Error	79.04659	0	Reject H0	88.03504	0	Reject H0	75.16795	0	Reject H0
Skewed Student t	379.4292	0	Reject H0	523.1725	0	Reject H0	581.2755	0	Reject H0
Pareto	54712.99	0	Reject H0	76.64829	0	Reject H0	35.53361	1.922975e-08	Reject H0
Weibull	143.2781	0	Reject H0	57.90178	1.311173e-13	Reject H0	31.87125	1.20018e-07	Reject H0

Table 12: Christoffersen's Conditional Coverage Test for the Unconditional Parametric estimation of Nikkei 225.

	Dax 30								
	95%			97,50%			99%		
	LR test	P-value	Decision	LR test	P-value	Decision	LR test	P-value	Decision
Student t	70.74757	4.440892e-16	Reject H0	58.2526	2.24154e-13	Reject H0	56.39986	1.887379e-15	Reject H0
Generalized Error	57.90431	2.667866e-13	Reject H0	42.37714	6.279454e-10	Reject H0	65.50002	7.155668e-10	Reject H0
Skewed Student t	346.4604	0	Reject H0	507.5747	0	Reject H0	581.2482	0	Reject H0
Pareto	205.5352	0	Reject H0	85.84239	0	Reject H0	45.70105	1.922975e-08	Reject H0
Weibull	160.2144	0	Reject H0	48.27265	3.294032e-11	Reject H0	30.98239	0	Reject H0

Table 13: Christoffersen's Conditional Coverage Test for the Unconditional Parametric estimation of Dax 30.

The results of the joint test are not surprising given the overall poor performance of the Unconditional parametric approach in the Unconditional Coverage test, along with the bad results of the Independence test, it was expectable that the joint tests would also reveal a bad performance no matter the assumption of the distribution of returns or the confidence level, all models end up being considered inadequate. We should keep in mind that as referred previously, the fact that one of the two properties is already satisfied makes it more difficult for the joint test to detect the inadequacy of the VaR model as pointed by Campbell (2005).

6. Conclusions

“In short, we ought to be able to identify most bad models, but the worrying issue is whether we can find any good ones.” Dowd (2006)

VaR has become one of the most popular methods for measuring financial risk. Every VaR model uses historical market data to forecast future portfolio performance. In addition, the models rely on approximations and assumptions that do not necessarily hold in every situation. In fact, history may not repeat itself. Since the methods are far from perfect, there is a good reason to question the accuracy of the estimated VaR models.

This dissertation examined the one-day predictive power of the Unconditional Parametric VaR models assuming different returns distributions such as Student t, Generalized Error, Skewed Student t, Pareto and Weibull. The comparison is made on four markets through one of the main indexes in each, being: Dow Jones Industrial Average, S&P 500, Nikkei 225 and Dax 30.

The unconditional coverage approach of the backtest results gave some adequate results, with the distributions Student t and Generalized Error having always some underestimation of exceptions, but achieving, solely under the unconditional coverage, some models that were considered adequate, in fact for the Dow Jones Industrials the models were considered adequate for all confidence levels, which is a good indicator that meets what Chu-Hsiung & Shan-Shan (2006) stated: “In contrast, the Student-t distribution is a relatively simple distribution that is well suited to deal with the fat tailed and leptokurtic features.”

On the other hand, the models based on the Pareto and Weibull distribution that make only a new VaR estimation when the previews day had a negative return, tend to overestimate risk similarly to Mittnik and Rachev (1993). In general, these last two distributions had some fair estimations, mainly in the Dax 30 and Nikkei 225, with the last one worth mentioning that for the 99% confidence level, the model obtained an excellent estimation of VaR exceptions.

The Skewed Student t distribution obtained the worst results always making a severe underestimation of risk showing that this is nowhere near a good fit of returns distribution for any of the studied indexes.

The independence tests revealed the real flaw of the models, where all were rejected. This result revealed that the models were incapable of reacting to changing volatility and correlations resulting exceptions that do not occur independently of each other. The models produce a sequence of

consecutive exceptions, which is a problem for investors because all his losses may occur in subsequent periods. Inevitably, turbulent market times, where the volatility is high cause problems in estimating parameters that should describe future market movements. Bao et. al (2006) also found the unconditional parametric models to work well under tranquil times, but different risk models to work better under turbulent times. Since all VaR models rely on historical market data, this issue not only concerns the case at hand but VaR models in general. Abnormal market behaviour is always a thought task for VaR models in general.

The contrast from the results of the different backtests performed, comes as an example that a weak backtesting methodology can mislead in relation to the actual adequacy of a VaR model. Show again that as Brown (2008) refers, the VaR is only as good as its backtests.

The results for the conditional coverage test that reject all models came as no surprise. To obtain any good result here, it would be necessary that at least some models to would be considered adequate in the independence test, to be then valid here in the joint test. Other methodologies of estimating VaR, like the ones referred in the literature review, may be more suited specially predicting volatility clusters, therefore having a better performance in the independence property which is essential in any VaR model.

Whatever the empirical framework for future backtesting will be, the most important lesson to learn from this dissertation is to understand the weaknesses of VaR models calculation. As empirical research proves, VaR figures should never be considered to be 100% accurate, no matter how sophisticated the systems are. However, if the users of VaR know the flaws associated with VaR, the method can be a very useful tool in risk management especially because there are no serious contenders that could be used as alternatives for VaR.

Bibliography

Adcock, C., Areal, N., & Oliveira, B. (January of 2011). Value-at-Risk forecasting ability of filtered historical simulation for non-Normal GARCH returns. Working paper.

Aloui, C., Mabrouk, S., (2010). Value-at-risk estimations of energy commodities via longmemory, asymmetry and fat-tailed GARCH models. *Energy Policy* 38, 2326–2339.

Ammann, M. & Reich, C. (2001), Value-at-Risk for Nonlinear Financial Instruments – Linear Approximation or Full Monte-Carlo? University of Basel, WWZ/Department of Finance, Working Paper No 8/01.

Angelidis, T., Benos, A., Degiannakis, S., (2004). The use of GARCH models in VaR estimation. *STAT METHODOL* 1, 105–128.

Bao, Y., Lee, T.-H. and Saltoglu, B. (2006), Evaluating predictive performance of value-at-risk models in emerging markets: a reality check. *J. Forecast.*, 25: 101–128. doi:10.1002/for.977

Barone-Adesi, G., F, B., & Giannopoulos, K. (August of 1998). Don't look back. *Risk*, 11, 100-104.

Barone-Adesi, G., Giannopoulos, K., & Vosper, L. (August of 1999). A simplified approach to the conditional estimation of value at risk. *Journal of Futures Markets*, 19, 583-602.

Barone-Adesi, G., Giannopoulos, K., & Vosper, L. (March of 2002). Backtesting derivative portfolios with filtered historical simulation (FHS). *European Financial Management*, 8(1), 31–58.

Beder, T. (September/October of 1995). VaR: seductive but dangerous. *Financial Analysts Journal*, 51(5), 12-24.

Bekiros, S. D., & Georgoutsos, D. A. (July of 2005). Estimation of Value-at-Risk by extreme value and conventional methods: a comparative evaluation of their predictive performance. *Journal of International Financial Markets, Institutions and Money*, 15(3), 209-228.

Berkowitz, J. (February of 2001). Testing Density Forecasts With Applications to Risk Management. *Journal of Business and Economic Statistics*, 19(4), 465-474.

Berkowitz, J., Christoffersen, P., & Pelletier, D. (2008). Evaluating Value-at-Risk Models with desk level data. *Management Science*, 2213-2227.

- Berkowitz, J., Christoffersen, P., & Pelletier, D. (2008). Evaluating Value-at-Risk Models with Desk-Level Data. *Management Science*, 2213-2227.
- Berkowitz, J., Christoffersen, P., & Pelletier, D. (28 of August of 2011). Evaluating Value-at-Risk models with desk-level data. *Obtido de Management Science*:
<http://pubsonline.informs.org/doi/abs/10.1287/mnsc.1080.0964>
- Blattberg, R. C. e Gonedes, N. J. (1974), A comparison of the Stable and Student distributions as statistical methods for stock prices, *Journal of Business*, 47, pp. 244-280;
- Bollerslev, T. (February of 1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31, 307-327.
- Box, G. E. P. e Tiao, G. C. (1973), *Bayesian inference in statistical analysis*, Addison-Wesley, Reading, Massachusetts;
- Brown, A. (2008), *Private Profits and Socialized Risk – Counterpoint: Capital Inadequacy*, Global Association of Risk Professionals, June/July 08 issue.
- Campbell, S. (2005), *A Review of Backtesting and Backtesting Procedure*, Finance and Economics Discussion Series, Divisions of Research & Statistics and Monetary Affairs, Federal Reserve Board, Washington D.C.
- Candelon, B., Colletaz, G., & Tokpavi, S. (January of 2011). Backtesting Value-at-Risk: A GMM duration based test. *Journal of Financial Econometrics*, 9(2), 314-343.
- Christoffersen, P. F. (November of 1998). Evaluating Interval Forecasts. *International Economic Review*, 39(4), 841.
- Christoffersen, P. F. (2012). *Elements of Financial Risk Management (Vol. 2)*. Oxford: Elsilver.
- Christoffersen, P., & Pelletier, D. (February of 2004). Backtesting Value-at-Risk: A Duration-Based Approach. *Journal of Financial Econometrics*, 2(1), 84-108.
- Christoffersen, P., Hahn, J., & Inoue, A. (February of 2001). Testing and comparing value-at-risk measures. *Journal of empirical finance*, 8(3), 325–342.
- Chu-Hsiung, Lin, and Shen Shan-Shan. (2006). "Can the student-t distribution provide accurate value at risk?" *The Journal of Risk Finance*, 7(3): 292–300.

- Climent, F. J. (January of 2013). Extreme value theory versus traditional GARCH approaches applied to financial data: a comparative evaluation. *Quantitative Finance*, 13(1), 45-63.
- Colletaz G., Hurlin C. and Perignon C. (2013), The Risk Map: a new tool for Risk Management, forthcoming in *Journal of Banking and Finance*.
- Crnkovic, C., & Drachman, J. (1997). In *VaR: Understanding and Applying Value-at-Risk*. Quality Control.
- Danielsson, J., & Vries, C. G. (July of 1997). Extreme returns, tail estimation, and Value at Risk.
- Dendramis, Y., Spungin, G. E., & Tzavalis, E. (November of 2014). Forecasting VaR models under Different Volatility Processes and Distributions of Return Innovations: Forecasting VaR. *Journal of Forecasting*, 33(7), 515-531.
- Diebold, F. X., Gunther, T., & Tay, A. (February of 1998). Evaluating Density Forecasts with Applications to Financial Risk Management. *International Economic Review*, 39(4), 863-883.
- Dowd, K. (1998). *Beyond value at risk: the new science of risk management* (Vol. 1). California: Wiley.
- Dowd, K. (2006), *Retrospective Assessment of Value-at-Risk*. *Risk Management: A Modern Perspective*, pp. 183-202, San Diego, Elsevier.
- Engle, R. F. (November of 1982). Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. *Econometrica*, 50(5), 987-1008.
- Engle, R. F., & Manganelli, S. (November of 2004). CAViaR: Conditional Autoregressive Value at Risk by Regression Quantiles. *Journal of Business & Economic Statistics*, 22(4), 367-381.
- Escanciano, J. C., & Olmo, J. (April of 2010). Backtesting Parametric Value-at-Risk With Estimation Risk. *Journal of Business & Economic Statistics*, 28, 36-51.
- Escanciano, J. C., & Olmo, J. (November of 2011). Robust Backtesting Tests for Value-at-risk Models. *Journal of Financial Econometrics*, 9(1), 132-161.
- Escanciano, J. C., & Pei, P. (February of 2012). Pitfalls in backtesting Historical Simulation VaR models. *Journal of Banking & Finance*, 36(8), 2233-2244.

Fama, Eugene F. (1965), The behavior of stock prices, *Journal of Business*, XXXVIII (1), January, pp. 34-105;

Fama, Eugene F. e Roll, Richard (1968), Some properties of symmetric stable distributions, *Journal of The American Statistical Association*, vol. 63, No 323, September, pp. 817-37;

Fernandez C., Steel M.F.J. (2000); On Bayesian Modelling of Fat Tails and Skewness, Preprint

Finger, C. (2005), Back to Backtesting, *Research Monthly*, May 2005, RiskMetrics Group.

Franco, C., Horvath, L., & Zakoian, J.-M. (September of 2011). Merits and Drawbacks of Variance Targeting in GARCH Models. *Journal of Financial Econometrics*, 9(4), 619-656.

Gourieroux, C., & Jasiak, J. (2015). *The Econometrics of Individual Risk: Credit, Insurance, and Marketing (Vol. 1)*. Princeton: Princeton University Press.

Haas, M. (2001), *New Methods in Backtesting*, Financial Engineering, Research Center Caesar, Bonn.

Hansen, P.R., Lunde, A., (2005). A forecast comparison of volatility models: does anything beat a GARCH(1,1)? *Journal of Applied Econometrics* 20, 873–889.

Hartz, C., Mittnik, S., & Paoletta, M. (2006). Accurate Value-at-risk Forecasting Based on the Normal-GARCH Model. *Computational Statistics & Data Analysis*, 51, 2295-2312.

Jorion, P. (1997). *Value-at-Risk (Vol. 1)*. USA: McGraw-Hill.

Jorion, P. (2006). *Value at Risk: The New Benchmark for Managing Financial Risk*.

Köksal, B., (2009). A Comparison of Conditional Volatility Estimators for the ISE National 100 Index Returns. *Journal of Economic and social research* 11, 1.

Kuester, K., Mittnik, S., & Paoletta, M. S. (2005). Value-at-Risk Prediction: A Comparison of Alternative Strategies. *Journal of Financial Econometrics*, 4(1), 53-89.

Kung, L.-M., Yu, S.-W., (2008). Prediction of index futures returns and the analysis of financial spillovers—A comparison between GARCH and the grey theorem. *European Journal of Operational Research* 186, 1184–1200.

Kupiec, P. (1995), Techniques for Verifying the Accuracy of Risk Management Models, *Journal of Derivatives* 3:73-84.

- Lee, C. F., Lee, J. C., & Lee, A. A., (2000), *Statics for business and financial economics*, 2nd ed. World Scientific
- Linsmeier, J. & Pearson, N.D. (1996), *Risk Measurement: An Introduction to Value at Risk*, Working Paper 96-04, University of Illinois at Urbana-Champaign.
- Lan, H., Nelson, B. L., & Staum, J. (October of 2010). A Confidence Interval Procedure for Expected Shortfall Risk Measurement via Two-Level Simulation. *Operations Research*, 58(5), 1481-1490.
- Leavens, D. H. (1945). Divesification of investments. *Trusts and Estates*, 80(5), 469-473.
- Lehar, A., Scheicher, M., Schittenkopf, C., (2002). GARCH vs. stochastic volatility: Option pricing and risk management. *Journal of Banking and Finance* 26, 323–345.
- Lönnbark, C. (September of 2013). On the role of the estimation error in prediction of expected shortfall. *Journal of Banking & Finance*, 37(3), 847-853.
- Lopez, J. (1998), *Methods for Evaluating Value-at-Risk Estimates*, *Economic Policy Review*, October 1998, 119-64.
- Lopez, J. (1999), *Regulatory Evaluation of Value-at-Risk Models*, *Journal of Risk* 1, 37-64.
- Luo, C., Seco, L.A., Wang, H., Wu, D.D., (2010). Risk modeling in crude oil market: a comparison of Markov switching and GARCH models. *Kybernetes* 39, 750–769.
- Mandelbrodt, Benoit B. (1963), *The variation of certain speculative prices*, *The Journal of Business*, 36, pp. 394-419;
- Marshall, C., & Siegel, M. (July of 1996). *Value-at-Risk: Implementing a Risk Measurement Standard*. *SSRN Electronic Journal*, 4(3).
- Mittnik, Stefan e Rachev, Svetlozar T. (1993), *Modeling asset returns with alternative stable models*, *Econometric Reviews* 12, pp. 261-330;
- Mittnik, S., M. S. Paoletta, and S. T. Rachev. (2002). *Stationarity of stable power-GARCH processes*. *Journal of Econometrics* 106, 97-107.
- Nelson, D. B. (1991). *Conditional heteroscedasticity in asset returns: a new approach*. *Econometrica* 59, 347-370.

- Orhan, M., Köksal, B., (2011). A comparison of GARCH models for VaR estimation. *Eswa* 2012, 2582–3592.
- Pafka, S., & Kondor, I. (October of 2001). Evaluating the RiskMetrics methodology in measuring volatility and Value-at-Risk in financial markets. *Physica A: Statistical Mechanics and its Applications*, 209(1-2), 305-310.
- Pederzoli, C., (2006). Stochastic volatility and GARCH: a comparison based on UK stock data. *European Journal of Finance* 12, 41–59.
- Pritsker, M. (January of 2001). The Hidden Dangerous of Historical Simulation. (F. R. Board, Ed.) Working Paper.
- Rachev, Svetlozar e Mitnik, Stefan (2000), *Stable Paretian models in finance*, John Wiley & Sons, New York;
- Sheppard, K. (2013). *Financial Econometric Notes*. University of Oxford.
- So, J. C. (1987), The sub-Gaussian distribution of currency futures: stable Paretian or non-stationarity, *Review of Economics and Statistics*, pp. 100-107.
- Venkataraman, S. (1997). Value at risk for a mixture of normal distributions: the use of quasi-Bayesian estimation techniques. *Economic Perspectives*, March, 2-13, Federal Reserve Bank of Chicago.
- Wang, K., C. Fawson, C. Barrett, and J. B. McDonald. (2001). A flexible parametric GARCH model with an application to exchange rates. *Journal of Applied Econometrics* 16, 521-536.
- Wiener, Z. (1999), *Introduction to VaR (Value-at-Risk), Risk Management and Regulation in Banking*, Kluwer Academic Publishers, Boston.
- White, J. H. (January of 1998). Value at risk when daily changes in market variables are not normally distributed. *Journal of Derivatives*, 5(3), 9-19.
- Ziggel, D., Berens, T., Weiß, G. N., & Wied, D. (November of 2014). A new set of improved Value-at-Risk backtests. *Journal of Banking & Finance*, 48, 29-41.