LUND UNIVERSITY

School of Economics and Management

NEKN01 Master's Thesis I
2014-08-18

Measuring Risk for WTI Crude Oil

An application of Value-at-Risk

Authors:
Alexander Eriksson
Jonathan Ljungqvist

Supervisor:
Karl Larsson

Abstract

Crude oil is the most traded energy commodity in the world, and its price has a large impact on the everyday life of billions. Given the volatility of crude oil prices and its enormous effects on economies worldwide, there has been a growing demand for risk quantification and risk management for the market participants. The measurement known as Value-at-Risk (VaR) has become the industry standard for internal risk control among firms, financial institutions and regulators. This study will assess which VaR method is most effective to quantify the risk of price changes embedded in the West Texas Intermediate (WTI); used as a benchmark for oil prices in the USA. VaR will be estimated by using both parametric and non-parametric methods that will be backtested with the Christoffersen test. The parametric methods considered in this paper are the GARCH, EGARCH, and TGARCH models, estimated by considering the effects of using normal distribution versus student's t-distribution or the Generalized Error Distribution (GED). The non-parametric methods used in this paper are Basic Historical Simulation (BHS), Volatility-Weighted Historical Simulation (VWHS), and Age Weighted Historical Simulation (AWHS). This study also tries to answer if the optimal choice of VaR estimation method differs when evaluating WTI Crude Oil prices as opposed to the S&P 500 index. The parametric models had an in-sample of 1000 observations and estimated a one day-ahead VaR estimate over the period 2007-01-01 to 2013-12-31.The model was re-estimated every day in the period to a total of 1826 estimations. The non-parametric models had an in-sample of 1000, and the volatility calculated for VWHS used the RiskMetric approach. For both S&P 500 and WTI the non-parametric methods provided poor VaR estimates. The parametric models provided better results, the GARCH models with leptokurtic distribution was the most effective in capturing price volatility. GARCH(1,1) with GED provided the best result for WTI, while GARCH(2,1) with t-distribution was the more optimal model to capture volatility in the S&P 500 index. Thus, we conclude that different models are needed to accurately capture the risk depending on which benchmark is used.

Keywords: Market Risk, VaR, Crude oil, Forecasting

Table of Contents

1. Introduction

1.1 Background

Oil has been the main source of energy since the end of the 19th century, making crude oil the single most traded commodity in the world today, in terms of volume and level. The majority of global oil trade is made with crude oil rather than refined petroleum products such as gasoline or heating oil,[1] which are typically sold at prices set in consideration of local demand and supply.[2] Crude oil is also often regarded as a benchmark for the energy sector as a whole given that it can have a disproportionate impact on electricity and heating costs, as well as the supply and demand of all other energy commodities.[3] In fact, the spot price for crude oil is not only used for simple spot market transactions, but also used to settle future contracts, derivatives, and taxation by governments.[4] Today, the two most traded Crude oil benchmarks are West Texas Intermediate (WTI), which serves as a benchmark for oil prices in the USA, and Brent Crude which serves as the benchmark in Europe.[5]

Since the OPEC crises of the 1970s, both competition and deregulation has risen and consequently made oil markets increasingly free. As a result of this, the energy markets of today, in particular the crude oil market can be characterized as highly volatile. This volatility is largely driven by interactions between trading of the product and the supply and demand imbalances that can result from the state of the economy; this was not the case when prices were regulated.[6] Global events can also impact the price, for example, there was a rapid increase in price to $35 a barrel in response to the Iraqi invasion of Kuwait, only to decrease to $20 a barrel a couple of months later when Iraq was defeated.[7]

The high fluctuations resulting from deregulation can also have deep effects on the economy as a whole, especially for resource-based economies that are highly dependent

---

[1] Edwards, Davis W. *"Energy Trading and Investing."* (2010). p 126
2 Refined petroleum products such as heating oil and gasoline may however be bought on the Chicago Mercantile Exchange (CME) futures market.
[3] Edwards (2010). p. 127
4 Fattouh, Bassam. *"An anatomy of the crude oil pricing system"*. Oxford, England: Oxford Institute for Energy Studies, 2011. p 7
[5] Davis (2010) p.13
[6] Giot Pierre, and Sébastien Laurent. "*Market risk in commodity markets: a VaR approach.*" Energy Economics 25.5 (2003): p. 435-457. P.437
[7] Ibid (2003) p. 437

on oil export and government revenues. In 2012, when the Brent Crude Oil price was at $100 a barrel, Russia's Higher School of Economics warned that if the oil price was reduced to $80 a barrel, the government would quickly burn through its $60 billion rainy-day reserve fund to meet its budget obligations.[8] This is emphasized by Sadorsky who states that oil price fluctuations in these economies not only affect government budgets, but can also have lasting effects on macroeconomic variables and stock prices. Furthermore, he states that though changes in oil prices can have an impact on economic activity, changes in economic activity seldom have a large impact on oil prices.[9] Although, the latter point is disputable, especially considering the result of the great recession in 2008, when oil prices plummeted in unison with the rest of the greater economy; see Figure 1.

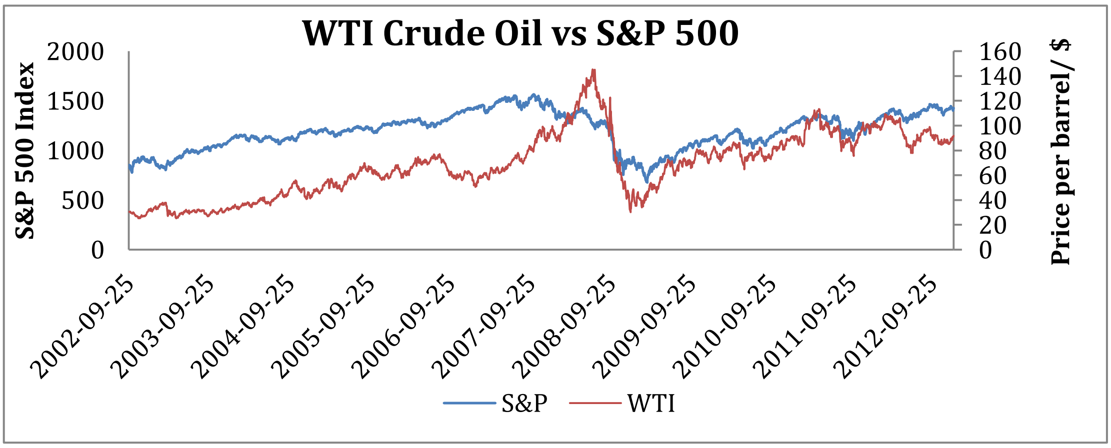


Figure 1- *Graph depicting WTI Crude Oil prices in $/per barrel and the S&P 500 composite index between 25-09-2002 and 31-12-2013. . Data Retrieved from Thomson Reuters DataStream - 2014-04-30*

Given the characteristics of crude oil prices and its potential effects, there has been a growing demand and need for risk quantification and risk management for the market participants. Especially, since it would allow both countries and firms to apply proper hedges to potentially absorb market shocks and reduce market risk by reducing volatility in earnings while maximizing return on investment. Hedging would allow firms and governments to manage the energy exposure of their energy supplies and forward contracts.[10] In addition, risk managers would be able to meet regulatory requirements that limit risk; all of which is outlined in the Basel agreements.[11]

[8] Buckley, Neil (2012-06-20)"*Economy: Oil dependency remains a fundamental weakness".*
<http://www.ft.com/cms/s/0/438712b2-b497-11e1-bb2e-00144feabdc0.html#axzz381n6dNp6>
Retrieved (2014-08-17)
[9] Sadorsky, Perry. "*Oil price shocks and stock market activity.*" Energy Economics 21.5 (1999): 449-469. P. 468
[10] Sadeghi, Mehdi, and Saeed Shavvalpour. *"Energy risk management and value at risk modeling."* Energy policy 34.18 (2006): 3367-3373. p. 3368.
[11] Ibid p. 3368

For the reasons stated above, the risk quantification of the crude oil market is essential for its market participants. This is why we will study the application of risk management on crude oil prices. The theoretical framework behind risk management is presented in 1.2, followed by a literature review on the subject and a outline of the objective of this thesis.

1.2 Risk Management

As a concept, Harry Markowitz introduced modern risk management in his paper "Portfolio Theory" in 1952. Now, over half a century later, risk management has become one of the most important areas within financial management. Recent financial downturns and the expansion of the derivatives market along with other financial markets have led to an increased focus on supervision and regulation.[12] As a result, the measurement known as Value-at-Risk (VaR) has been cultivated to become the industry standard for internal risk control among firms, financial institutions and regulators.[13] Cabedo and Moya define Value-at-Risk, - as a measure that *"determines the maximum loss a portfolio can generate over a certain holding period with a predetermined likelihood level"*.[14] In terms of crude oil, VaR measures the oil price change associated with a certain likelihood level, and it has become increasingly important when firms design their risk strategies.[15] VaR can also be seen as a way to measure market value exposure of assets.[16]

Though work on internal models to measure and aggregate risk across a whole institution was started in the 1960s and 1970s, it was in the 1990s that JP Morgan developed the concept of the VaR as a single measurement of the probability of losses at the firmwide level.[17] A development, which was driven by the regulators need for better control, the fact that there were many sources of risks and that technological advances made it possible to calculate these risks.[18] Since then, the measure has been consolidated further, as Basel regulators allowed banks to adopt internal VaR models,

[12] Dowd, Kevin. Measuring market risk. John Wiley & Sons, 2005. p. 1-4
[13] Ibid p. 9-10
[14] Cabedo, David J., Moya, Ismael. "*Estimating oil price- Value at Risk using the historical simulation approach".* Energy Economics 25.3 (2003): 239-253 p. 240.
[15] Sadeghi,& Shavvalpour. (2006). p. 3368
[16] Saunders, Anthony, and Linda Allen. "Credit risk measurement." 2nd John Wiley & Sons Inc. New York (2002). p. 4
[17] Ibid p. 9
[18] Jorion, Philippe. "*Value at risk: the new benchmark for managing financial risk*". Vol. 2. New York: McGraw-Hill, 2007 p. 25

after the original standardized method was criticized as being too conservative.[19] In conjunction with bank's VaR measure there would then be a market risk capital requirement, based on the number of times the actual loss exceeded or violated the VaR estimate. This meant that a required amount of capital was needed in order to maintain a certain level of market risk. [20]

1.2.1 Value-at-Risk

To illustrate the concept of VaR, we may define it as "the smallest loss *l* such that the probability of a future portfolio loss *L* that is larger than *l*, is less than or equal to $1\text{-}\alpha$."[21] This means that we expect to experience a loss greater than VaR with the probability $1\text{-}\alpha$ over a specified time horizon or holding period. In this thesis, a 1-day-ahead VaR forecast will be estimated, but another common length for the horizon is 20 days while the Basel regulations set a time horizon of 10-days.[22] In mathematical terms, the above VaR definition may be written according to EQ. 1.

$$VaR_\alpha = \min\{l\colon \Pr(L > l) \leq 1 - \alpha\} \quad \text{[23]} \qquad \textbf{EQ. 1}$$

Common choices for α, are 0.95 and 0.99, in which case we expect to experience a loss greater than the VaR estimate with a probability of 5% and 1% over the given time horizon.[24]

So why is VaR so popular? One of the main reasons is that it provides a common measure of risk across different portfolio types and risk factors, making it easy to compare the risks, while at the same time letting us aggregate the risks of different sub-positions into one measure of portfolio risk. Another positive attribute is that it gives a probabilistic measure by providing the probability of losses larger than VaR. Lastly, VaR is expressed in an easily understood unit of measure, namely 'lost money', which can easily be presented throughout the hierarchy of a firm, financial institution or the regulator. [25]

---

[19] *Fallon, William. Calculating value-at-risk. Wharton School, University of Pennsylvania, 1996.* p.1
[20] Basle Committee on Banking Supervision. "Supervisory Framework for the use of "Backtesting" in Conjunction with the Internal Models Approach to Market Risk Capital Requirements." 1996. p.2.
[21] Nilsson, Birger(2014) "*Value-at risk"*lecture notes in. *NEKN83/TEK180 spring 2014. Lund University* p.2
[22] Dowd (2005) p. 30
[23] Nilsson, (2014) "*Value-at- risk"* p. 2
[24] Dowd (2005) p. 29
[25] Ibid p. 12

The main methods or approaches in quantifying VaR can be put into three categories: Non-Parametric, Parametric and Extreme Value Theory (EVT). The essence of the non-parametric approach is that VaR estimates are simulated based on historical observed data without any distributional assumptions.[26] The parametric approach however, estimates risk by fitting probability curves on the data, then calculating the VaR measure from the fitted curve given by the chosen underlying distribution and standard deviation. [27] Examples of parametric models include fitting an underlying distribution that is conditional on an ARCH (Autoregressive Conditional Heteroskedasticity) or GARCH (Generalized Autoregressive Conditional Heteroskedasticity) volatility process, to the data. Lastly, there is the method known as Extreme Value Theory, which draws from both of the previous methods, but instead focuses on the extreme outcomes i.e. the largest losses.[28]

1.2.2 Potential Drawbacks in estimating VaR

Though there are many advantages to VaR as a risk measurement, it is not without its drawbacks and limitations. One limitation is that VaR estimates are very sensitive to model and assumption selection. It is very easy to incorrectly specify a model so it does not accurately capture the risk. This is referred to as 'model risk', meaning there is a risk that the model is not capturing the risk it is designed to capture. This can be the result of bad assumptions, model limitations, poorly estimated parameters or inadequate understanding by the people using the model. Thus potentially rendering the model useless and propagating a financial disaster for the firm or entity in question. [29]

This may especially pertain to the commodity market as a whole and the crude oil market in particular, as modeling risk is a complex task given that the markets are characterized as having highly fluctuating prices. It is thus imperative to choose a model and assumptions that are best able to account for such attributes. Also, given that the regulators punish financial institutions for poorly estimating VaR models (e.g. underestimation or overestimation) by inflicting higher capital charges, it has become increasingly important to estimate VaR accurately.[30]

---

[26] Dowd (2005) p. 83
[27] Ibid p. 151
[28] Nilsson, Birger (2014) "*Extreme value theory for VaR estimation*" lecture notes in. *NEKN83/TEK180 spring 2014. Lund University.* p. 1
[29] Dowd (2005) p. 31
[30] Ibid p. 328

Another drawback is the possibility of implementation risk, where theoretically similar models give different VaR estimates because of the way they are implemented. Such risk has the potential of leaving people exposed to a greater risk than anticipated, should they take the model too seriously. This is a problem, which can be inherently common when using VaR as a risk estimate, given that it does not indicate the size of the loss, other than the fact that it is larger than VaR.[31]

It is important to keep these drawbacks and limitations of VaR in mind when performing VaR analysis, since the repercussions of choosing an incorrect estimate, as the result of selecting an inappropriate model can be very large. Therefore, in the following two sections previous research will be reviewed, followed by the objective of this thesis, along with its delimitations.

1.3 Literature review

There has been a variety of research on the risk quantification of crude oil prices, and on commodities in general, but nothing to date has been entirely conclusive in procuring a standard method for the quantification of risk. The reasons being that oil price volatility is a complex function of a range of factors such as expansions and downturns,[32] [33] energy demand and supply chocks,[34] inventory holding[35] and, movement in exchange rates[36] and interest rates[37]; all of which affect oil price movements, thus making risk hard to forecast.

Cabedo and Moya did one of the earlier studies into the risk quantification of oil, using VaR on Brent crude oil prices for the period from January 1992 to December 1999 with their out-of-sample forecast between 1998 and 1999.[38] In their paper, they find that

[31] Dowd (2005) p. 31
[32] Kilian, Lutz, and Cheolbeom Park. "*The impact of oil price shocks on the us stock market*.*" *International Economic Review* 50.4 (2009): 1267-1287. p. 1267
[33] Balke, Nathan S., Stephen PA Brown, and Mine K. Yücel. "*Oil Price Shocks and US Economic Activity.*" *Resources of The Future* (2010): 10-37.
[34] Kilian, Lutz. *"Not all oil price shocks are alike: Disentangling demand and supply shocks in the crude oil market."* The American Economic Review (2009): 1053-1069. p. 1053
[35] Hamilton, James D. "*Understanding crude oil prices*." No. w14492. National Bureau of Economic Research, 2008. P.15-16
[36] Alquist, Ron, and Lutz Kilian. *"What do we learn from the price of crude oil futures?."* Journal of Applied Econometrics 25.4 (2010): 539-573.
[37] Killian & Cheolbeom (2009) p.28
[38] Cabedo and Moya (2003) p. 1

Historical simulation ARMA Forecasting (HSAF) provides the best model to estimate VaR, in comparison to the Basic Historical Simulation (BHS). The reason being, that HSAF gives a more flexible VaR quantification, which better fits continuous price movements. They also find that the parametric GARCH(1,1)-forecasting method underperforms, as it overestimates the maximum price change. Sadeghi and Shavvalpour[39] arrive at similar conclusions in their paper which compares the HSAF method with the variance-covariance method that was proposed by Hull and White in 1998.[40] The variance-covariance method is based on ARCH and GARCH modelling where potential losses are assumed to be proportional to the return standard deviation. Sadeghi and Shavvalpour uses a GARCH(1,1) model with weekly OPEC prices from 1997 to 2003, and assume that values of the standard deviation have a normal distribution. Though they assess that VaR estimated through the variance-covariance methodology is above actual price changes for the whole out-of-sample forecast, they conclude that HSAF proves to be more efficient in comparison to the variance-covariance method, due to the high variation above actual changes. They also conclude that VaR is a reliable measure of oil price risk for anyone who is concerned with oil price volatility; whether it is a firm, a financial institution or a policy maker.[41]

Costello found that the semi-parametric GARCH model with historical simulation is superior to the HSAF in estimating VaR forecasts for Brent Crude Oil, over the period spanning from 20th of May, 1987 to 18th of January, 2005.[42] They used the first five years as in-sample period to estimate the data and the rest as the out-of-sample investigative period. The reason being that, unlike Cabedo and Moya who assume normality and that oil prices are independently and identically distributed (i.i.d.).[43] Costello makes oil prices conditional on GARCH, which allows the forecasting to capture time-varying volatility. The use of this method is further supported by Giot and Laurent's[44] findings of volatility clustering in oil prices.[45] Costello further notes that the variance-covariance method failed because of the assumption of normal distribution,

[39] Sadeghi, Mehdi, and Saeed Shavvalpour. "*Energy risk management and value at risk modeling.*" Energy policy 34.18 (2006): 3367-3373.
[40] Hull, John, and Alan White. *"Incorporating volatility updating into the historical simulation method for value-at-risk." Journal of Risk* 1.1 (1998): 5-19.
[41] Sadeghi & Shavvalpour (2006) p.3373
[42] Costello, Alexandra, Ebenezer Asem, and Eldon Gardner. "*Comparison of historically simulated VaR: Evidence from oil prices.*" Energy economics 30.5 (2008): 2154-2166.
[43] Cabedo and Moya 2003 p. 242
[44] Giot & Laurent (2003) p. 437
[45] Costello (2008) p. 2154-2157

which according to Barone-Adesi produces poor VaR estimates in a GARCH setup.[46] When considering risk management measurements, extreme events occur more often and are larger than what is often forecasted when using normal distribution.[47] Subsequently, much of the research today prefers the use of conditional models which apply Exponentially Weighted Moving Average (EWMA) and GARCH, instead of unconditional methods. One of the reasons for this is that EWMA and GARCH characterize asset returns with conditional heteroskedasticity, which is based on the assumption that estimates are more efficient when more weight is put on the most recent observations in the data set.[48]

Aghayev and Rizvanoghlu, tested the performance of GARCH(1,1) with normal distribution and Generalized Error Distribution (GED), Threshold GARCH(1,1) with GED and different EWMA models as a predictor for a 20-day VaR forecast of Azeri light crude oil, produced in Azerbaijan, starting from 17th June, 2002 to 18th June, 2013, with the last 1000 observations as the out-of-sample period. They found that the GARCH(1,1) with GED outperformed GARCH with normal distribution(GARCH-N) in the out-of-sample forecast. The reason was that the GARCH-N model underestimated the market risk of the commodity. They found no difference in the out-of-sample forecast between GARCH(1,1)-GED, and EWMA, but the GARCH model performed a better in-sample forecast. They also found some evidence of asymmetric leverage effect and that TGARCH(1,1) provided more parsimonious VaR estimates.[49]

Fan, who calculated VaR for daily spot WTI prices found that GARCH (1,1)-GED outperformed GARCH(1,1)-N and HSAF over the period 1986-2006, with the last year as the out-of-sample period.[50] This is similar to Xiliang and Xi, who conclude that the GARCH-GED is the best model for WTI Crude Oil at a low confidence level (95%) while GARCH-N is better at high confidence levels (99%); they used WTI prices from 21st of May, 1987 to 18th of November, 2008; with the out-of-sample period from 19th of

---

[46] Barone-Adesi, Giovanni, Kostas Giannopoulos, and Les Vosper. "*VaR without correlations for portfolios of derivative securities.*" *Journal of Futures Markets* 19.5 (1999): 583-602. P 586

[47] Hendricks, Darryll. "Evaluation of value-at-risk models using historical data." Federal Reserve Bank of New York Economic Policy Review 2.1 (1996): 39-69. p. 50

[48] Dowd (2005) p. 83

[49] Aghayeva, Huseyn, and Islam Rizvanoghlub. "*Understanding the crude oil price Value at Risk: the Case of Azeri Light.*" Available at SSRN 2402622 (2014).

[50] Fan, Ying, et al. "Estimating 'Value at Risk' of crude oil price and its spillover effect using the GED-GARCH approach." Energy Economics 30.6 (2008): 3156-3171.

October, 2004 to 18th of November, 2008.[51] Hung on the other hand, estimated VaR for WTI Crude Oil by using GARCH with the heavy tailed distribution, and compared it with the GARCH-N model and GARCH with student's t-distribution model.[52] In his findings, he concluded that the GARCH-t model was the least accurate, while GARCH-N proved more efficient at low confidence intervals. The forecast from the GARCH-HT model was the most accurate and most efficient risk measure.

Unconditional models such as the BHS inherently dwells on the theory that asset returns come from an i.i.d. distribution. It is this notion that Longin believes to be the true drawback of unconditional models.[53] In comparison, Pritsker underlines the unconditional models' inability to incorporate heteroskedastic behavior, market dynamics and the risk factor distribution.[54] In addition, unconditional VaR models suffer from the incapability of identifying risk factors that thoroughly underestimates risk, which can be of substantial size as it is slow to react to extreme changes.[55] Nonetheless, even though conditional models are more popular in recent research, unconditional models such as BHS still remain the most used method among financial institutions. One reason being, that banks are exposed to numerous risks and thus want to avoid too volatile day-to-day risks, which parametric methods tend to produce.[56]

When examining existing research, it is also important to observe that even though models such as GARCH(1,1) with GED are considered superior, due to the fat tails seen in many of the cases analyzed, there is the added risk that results are an outcome of the data chosen and more specifically the period considered. Given that WTI, Brent and OPEC crude oil prices move symbiotically, there can be marginal differences, given that Brent is representative of European oil prices, WTI of the US, and OPEC prices are based on a Basket of oil prices. In addition, a model that is found to be superior in a period where oil prices are relatively stable does not have to be in a period of high volatility.

---

[51] Xiliang, Zhao, and Zhu Xi. *"Estimation of Value-at-Risk for Energy Commodities via CAViaR Model."* Cutting-Edge Research Topics on Multiple Criteria Decision Making. Springer Berlin Heidelberg, 2009. 429-437.

[52] Hung, Jui-Cheng, Ming-Chih Lee, and Hung-Chun Liu. *"Estimation of value-at-risk for energy commodities via fat-tailed GARCH models." Energy Economics* 30.3 (2008): 1173-1191.

[53] Longin, Francois M. *"From value at risk to stress testing: The extreme value approach."* Journal of Banking & Finance 24.7 (2000): 1097-1130.

[54] Pritsker, M. "Evaluating Value-at-Risk Methodologies: Accuracy versus Computational Time in 'Model Risk: concepts, calibration and pricing'." (2000).

[55] Dowd (2005) p. 100

[56] Pérignon, Christophe, and Daniel R. Smith. *"Diversification and value-at-risk."* Journal of Banking & Finance 34.1 (2010): 55-66. p. 55

Both of the early studies by Cabedo & Moya and Sadeghi & Shavvalpour that found the non-parametric HSAF method to be superior involved out-of-sample forecasting between 1998-1999 and 1997-2003 respectively, unlike later studies that considered periods closer to pre- or post-financial crisis which found that conditional GARCH models were superior.

In order to examine if there is a difference in quantification of risk between different benchmarks, the Standard & Poor's 500 Composite Index will be used. The reason being that S&P 500 is the most widely used benchmark for the US equity market and has proven to reflect the fundamentals in the US large cap equity markets.[57] Considering earlier research for S&P 500, Awartani and Valentina tested several different GARCH models for S&P 500, and their predictive power. They concluded that asymmetric GARCH models outperformed the symmetric GARCH models.[58] Angelidis on the other hand investigated, using several GARCH models, which model produced the best VaR estimates for several stock indices including S&P 500.[59] Considering the period between 9th of July, 1987 to 18th October, 2002, they found that the mean equation in the GARCH estimation did not play an important role when forecasting VaR. EGARCH(1,1) with a student's t-distribution produced the best results, but the authors also found that the GED distribution produced acceptable results when having a 99% confidence interval. However, they rejected the use of a normal distribution, as it produced inaccurate results for all models.

The introduction of conditional techniques as indicated by Stefsos and Kalyvas, is a good step towards producing accurate VaR estimates.[60] Still, the question remains -is it truly better than the unconditional techniques? If so, what is the proper distribution that should be used and how do we best backtest the result in order to know which one is best? In addition, how does risk quantification differ between a highly volatile data set such as crude oil to a less volatile data set such as S&P 500 and is it important to use

[57]Investopia. "*Standard & Poor's 500 index– S&P 500*" retrieved from http://www.investopedia.com/terms/s/sp500.asp Accessed: 2014-08-13

[58] Awartani, Basel, and Valentina Corradi. "*Predicting the volatility of the S&P-500 stock index via GARCH models: the role of asymmetries.*" *International Journal of Forecasting* 21.1 (2005): 167-183

[59] Angelidis, Timotheos, Alexandros Benos, and Stavros Degiannakis. *"The use of GARCH models in VaR estimation." Statistical Methodology* 1.1 (2004): 105-128.

[60] Sfetsos, A., and L. Kalyvas. "*Are conditional Value-at-Risk models justifiable?.*" Applied Financial Economics Letters 3.2 (2007): 129-132.

different models of quantification when calculating the risk of each one? These are the questions that this paper aims to answer.

1.4 Objective

The specific aim of this thesis is to evaluate the use of non-parametric and parametric Value-at-Risk methods for WTI Crude Oil in order to answer the below specific questions

(i) In calculation of a 99% 1-day-ahead VaR, what is the best method according to the statistical backtest?
   a. Are parametric methods better than non-parametric methods?
   b. If so, what underlying distributional assumptions are the best?

(ii) Do choices in optimal VaR estimation models differ between WTI Crude Oil and S&P 500?

Given the objective, and specifically based on existing literature on VaR; the hypotheses are that:

H1: Parametric methods using a fatter tailed distribution are generally more effective in providing a realistic VaR estimate than Non-parametric methods because they are more accommodating to changes in the market volatility.

H2: The choice of VaR estimation model will differ depending on the benchmark considered because statistical properties stemming from the variation of externally affecting factors or variables.

Ultimately, by evaluating H1 and H2 this study hopes to contribute greater knowledge pertaining to the risk quantification of WTI Crude Oil spot prices during and after the financial crises.

1.4.1 Delimitation

The focus in this thesis is on the non-parametric and parametric approaches. For the non-parametric methods, Basic-, Age Weighted-, and Volatility Weighted Historical Simulation will be used. Together these constitute the most commonly used non-parametric methods and should be able to showcase non-parametric models' ability to quantify risk for WTI Crude Oil. For the parametric methods, VaR will be estimated conditional on a GARCH, EGARCH and TGARCH volatility process with underlying normal distribution, student's t-distribution and Generalized Error Distribution (GED), all of which will be explained in further detail in 2.2. Methods that depend on Extreme Value Theory (EVT), such as the Peaks over Threshold (PoT) and Generalized Extreme

Value (GEV) methods, are excluded and the reason for this is twofold. First, based on existing literature, EVT is uncommon when calculating VaR for crude oil. Second, our data sample is not large enough for the EVT models to generate an accurate estimate.[61]

The period of interest for forecasting the 1-day-ahead VaR for WTI Crude Oil is 1st of January, 2007 to 31st of December, 2013, which makes up the out-of-sample period. The in-sample-period for the estimation of the parameters is the 1000 observations prior to 1st of January, 2007. We have chosen this period because we are interested in the oil price risk during the turbulent time leading up to and around the Lehman Brothers collapse on the 15th of September, 2008, as well as the years following the 2008 financial crisis. Consequently, the paper hopes to investigate how well the methods are able to account for the extreme fall in oil prices seen immediately after the crisis and the highly volatile prices seen in the market shortly thereafter. For this reason, the standard confidence interval of α=0.99 is used given that our interest lies within the extreme events over that period, while α=0.95 was excluded. Additionally, the forecast horizon or holding period is one day, because it is the most common period generally used and banks use this horizon to approximate the 10-day-ahead VaR for regulatory purposes by multiplying the 1-day-ahead VaR by the square root of 10.[62]

In oil markets, it can also be of interest to estimate VaR for both the left and the right tail of the distribution. Which tail is of interest, depends on ones location in the production pipeline. A logistics company is not interested in the same tail as an oil drilling company when it comes to risk quantification, since an increase in the oil price will depress the margins for the logistics company, but increase it for the oil producer. Therefore, the tail of interest for the VaR analysis depends on whether the institution considered has a short or long position. Our analysis will be that of an oil producer which means that a sudden sharp decrease in the oil price can produce a VaR violation, but a sudden increase cannot.

Nilsson, Birger(2014) "*Extreme value theory for VaR estimation*" lecture notes in. *NEKN83/TEK180 spring 2014. Lund University*, p. 4

[62] Dowd (2005) p. 30 & 52

2. Methods to Estimate VaR

This section, describes each of the chosen methods used in this thesis. First, the non-parametric methods (BHS, AWHS, VWHS) are presented, followed by the parametric methods (GARCH, TGARCH and EGARCH) and the different distributions (Normal Dist., Student-t Dist. and GED), along with a comparative discussion of the models. After the models are presented, the backtesting methods (Christoffersen and Basel test) used to examine which of the models are superior, are explained.

2.1 Non-parametric Methods

The non-parametric methods for estimating Value-at-Risk builds on the assumption that recent past values can be used to forecast risks over the near future. The non-parametric methods used in this paper are Basic Historical Simulation, Volatility-Weighted Historical Simulation and Age Weighted Historical Simulation as mentioned earlier.

2.1.1 Basic Historical Simulation (BHS)[63]

Basic historical simulation also known as the standard approach is the simplest way of calculating VaR. Given a rolling in-sample window of 250, and a 99% confidence interval, the value at risk is the value of the 2.5 largest loss. It is however impossible to take a fraction of a loss, which means that VaR is the value of the third largest loss in the estimation window. Therefore, a violation will occur if we observe a loss larger than the third largest in-sample loss in the first out of sample observation.

2.1.2 Age Weighted Historical Simulation (AWHS)[64]

While BHS gives the same probability weights to all observations, i.e. 1/N, the AWHS, which was suggested by Boudoukh, Richardson and Whitelaw in 1998, instead assigns different weights to observations depending on how recent the observation is. The BHS can therefore be seen as a special case of AWHS where all weights are the same. In the AWHS method, older observations are given a lower weight and the reason is intuitive, newer observations are more relevant for forecasting than older observations. In equation 2, which is used to calculate the weights, λ is the decaying factor and decides how fast older observations become irrelevant.

$$\omega_1 = \frac{(1-\lambda)}{(1-\lambda^N)} \qquad \textit{EQ. 2}$$

$$\omega_2 = \lambda\omega_1$$

[63] Dowd (2005) p. 84-85
[64] Ibid p. 93-94

$$\omega_3 = \lambda^2\omega_1 = \lambda\omega_2$$

$$\omega_N = \lambda^{N-1}\omega_1$$

$$\sum_{i=1}^{N}\omega_i = \sum_{i=1}^{N}\lambda^{i-1}\omega_1 = \omega_1(1+\lambda+\cdots+\lambda^{N-1}) = \omega_1\frac{1-\lambda^N}{1-\lambda}=1$$

In the illustration above, ω_1 is the probability weight given to the newest observation and ω_n is given to the oldest in the in-sample. After the weights are calculated for all observations in the in-sample period, the losses are then ranked from largest to smallest loss in the sample. The cumulative probability is then calculated and the VaR estimate is the smallest loss, where the probability of observing a lager loss is smaller or equal to *(1-α).* Despite similarities with the BHS method, when using the AWHS method, recent large losses will impact the VaR estimate more than large losses further back in time.

2.1.3 Volatility Weighted Historical Simulation (VWHS)[65]

The idea of Volatility-weighted Historical Simulation was first suggested by Hull and White, and is built on the premise of updating return information to take into account recent changes in volatility, in order to account for the common problem of volatility clustering.[66] When using the BHS model and last month's market volatility was 2%, and this month's market volatility is 3%, then last month's data will help understate the changes expected to be seen this month[67]. This will lead to an underestimation of tomorrow's risk, and to solve this, we update historical returns to reflect changes in volatility.

Assuming a historical sample of T losses, the rescaled losses are denoted as l_t^*, and are calculated as stated below:

$$l_1^* = \left(\frac{\sigma_{T+1}}{\sigma_1}\right) l_1$$

$$\vdots$$

$$l_{t-1}^* = \left(\frac{\sigma_{T+1}}{\sigma_{t-1}}\right) l_{t-1}$$

$$l_t^* = \left(\frac{\sigma_{T+1}}{\sigma_t}\right) l_t \qquad \textbf{\textit{EQ. 3}}$$

[65] Dowd (2005) p. 94-95
[66] Hull & White (1998) p. 5
[67] Ibid p. 5

Where l_t is the historical loss at t, σ_{t+1} is the GARCH or Exponentially Weighted Moving Average (EWMA) forecasted volatility for the asset at $t+1$, made at t, and σ_T is the volatility at time T. In this paper, to forecast the volatility we use the RiskMetric approach, introduced and developed by JP Morgan,[68] in order to sidestep the parameter estimation, which is needed in the GARCH approach. To do this, we start by calculating the EWMA conditional variance:

$$\sigma_t^2 = (1-\lambda)e_{t-1}^2 + \lambda\sigma_{t-1}^2 \quad For\ t=1, 2,..., n \qquad \textbf{EQ. 4}$$ [69]

Where $\lambda=0.94$ is a fixed constant set as the standard RiskMetric value for daily data, e_{t-1}^2 is the observed error variance , and σ_{t-1}^2 is the conditional variance for the previous period, *t-1.*

After obtaining the EWMA conditional variance, the square root is taken to get the EWMA Conditional Standard deviation. This is then used to construct the volatility scaled losses by using Equation 3. The actual returns are then replaced with the volatility-adjusted returns and VaR is estimated using the standard approach.

2.2 Parametric Methods

Parametric methods estimate risk by fitting a probability distribution function over the data and then inferring the risk measure from the fitted curve. As these models use additional information derived from the distribution function, they are in many ways more powerful than non-parametric methods. It is however crucial to use the right distribution function in-order to accurately mimic the behavior of the data. [70]

Simply fitting a distribution unconditionally to the data ignores the fact that returns exhibit volatility clustering, which can lead to excess kurtosis. That is, an underestimation of the risk during a volatile period, and an overestimation during a calm period.[71] Taking volatility clustering into account, we fit a distribution of returns that is conditional on an assumed volatility process, which itself is consistent with volatility clustering. This could be done by for example fitting a distribution conditional

[69] Riskmetrics, T. M. "JP Morgan Technical Document." (1996).p. 82
[70] Dowd (2005) p. 151
[71] Ibid p 152-153

on a EWMA or GARCH process, which both exhibit tail heaviness and volatility clustering.[72]

This paper tests the GARCH, EGARCH, and TGARCH models using normal distribution, student's t-distribution and Generalized Error Distribution (GED). Each method and distribution will be presented below starting with the general GARCH method.

2.2.1 GARCH (1, 1)[73]

The most used model for estimating conditional volatility is the GARCH (Generalized Autoregressive Conditional Heteroskedasticity) model. This model was the work of Bollerslev[74], who built its premise on the work of Engle[75]. The model expresses the conditional variance as a function of previous error terms and variances, consequently accounting for volatility clustering. That is, if the current period exhibits high variance, then the next period will also be expected to have high variance, given that we use the information in the current period. By using the conditional variance, the one-step-ahead VaR estimate will account for volatility clustering. Below is the formula for the conditional variance:

$$\sigma_t^2 = \omega + \alpha_1\epsilon_{t-1}^2 + \beta_1\sigma_{t-1}^2 \qquad \textbf{\textit{EQ. 5}}$$

$$\sigma_t^2 = \alpha_0 + \alpha_1\varepsilon_{t-1}^2 + \alpha_2\varepsilon_{t-2}^2 + \cdots + \alpha_q\varepsilon_{t-q}^2 + \lambda_1\sigma_{t-1}^2 + \lambda_2\sigma_{t-2}^2 + \cdots + \lambda_q\sigma_{t-q}^2$$
$$= \alpha_0 + \sum_{i=1}^{p}\alpha_i\,\varepsilon_{t-i}^2 + \sum_{j=1}^{q}\lambda_j\sigma_{t-j}^2$$

This conditional variance is then used in order to forecast the one-step-ahead volatility that is used in VaR.

$$\mathrm{VaR}_\alpha(\mathrm{L})_{t+1} = \mu + \sigma_{T+1}Z_{\alpha,} \qquad \textbf{EQ. 6}$$

When estimating a univariate time series like this, there are two main components, the variance equation- explained above-, and the mean equation. We have so far omitted the mean equation, but when estimating a GARCH model the mean equation should be

[72] Dowd (2005) p 153
[73] Enders, Walter. *"Applied econometric time series."* John Wiley & Sons, 2010. P 126-131
[74] Bollerslev, Tim. "*Generalized autoregressive conditional heteroskedasticity.*" Journal of econometrics 31.3 (1986): 307-327.
[75] Engle, Robert F. *"Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation."* Econometrica: Journal of the Econometric Society (1982): 987-1007.

specified. A common choice for the mean equation is to have an AR(1) process; see EQ. 7.

$$y = \delta + \theta y_{t-1} + \varepsilon_t \qquad \textit{EQ. 7}$$

The value of y is equal to some constant δ, the error term ε_t, and θ times the previous value of y. The purpose of this study is to find the model that produces the best estimates of VaR, and AR(1) is a standard choice for financial time series. Because of this, we will test all the GARCH models with first just a constant, and then a constant with an AR(1) term. After this we evaluate if there is autocorrelation in the standardized residuals, which aims to test the validity of the mean equation. If there is autocorrelation then the mean equation needs to be re-specified[76].

The standardized residuals, $\hat{v}_t$, see *EQ. 8*, are obtained to check the validity of the GARCH model.

$$\hat{v}_t = \frac{\hat{\varepsilon}_t}{\sqrt{\hat{h}_t}} \qquad \textbf{EQ. 8}$$

$\hat{v}_t$ is tested for serial correlation using the Ljung-Box test. If H_o is rejected, meaning there is serial correlation, the mean equation needs to be re-specified. After an acceptable mean equation is established, the validity of the variance equation should be checked.[77] This is done by applying the same procedure to *EQ. 9.*

$$\hat{v}_t = \frac{\hat{\varepsilon}_t}{\hat{h}_t} \qquad \textit{EQ. 9}$$

If H_0 is rejected *in EQ. 9*, the variance equation is not valid, and needs to be re-specified.

The GARCH model implies that negative and positive shocks have the same effect on volatility. Yet, often in financial data, negative shocks of the same magnitude as positive shocks will cause higher volatility. The inclination for volatility to decline when return increase and to increase when returns decline can be referred to as 'leveraged effects'.[78] It is therefore reasonable to estimate models that are not symmetric in the way they react to negative and positive shocks. We will test two such models in this thesis, namely the EGARCH and TGARCH models.

[76] Enders(2010) p.138
[77] Ibid p. 131-132
[78] Ibid p. 155

2.2.2 Threshold- GARCH (1, 1)[79]

Developed by Glosten, Jaganathan and Runkle in 1993, the TGARCH model tries to capture the phenomenon explained above, by creating a threshold where shocks above and below the threshold have different effects on volatility. By adding a dummy variable when we have negative shocks, the model can capture if there is any asymmetry in the shocks effect on volatility. Consider the TGARCH process depicted in *EQ.10*.

$$\sigma_t^2 = \omega + \beta_1 \sigma_{t-1}^2 + \alpha_1 \epsilon_{t-1}^2 + d_1 \gamma_1 \epsilon_{t-1}^2 \text{ , If } \varepsilon_t < 0 \text{ then d=1, otherwise d=0} \qquad \textbf{EQ. 10}$$

If γ_1 is equal to zero, it would imply that we have symmetry in the effect that shocks have on the conditional variance. If instead $\gamma_1 > 0$, a negative shock will have a larger effect on the conditional variance then a positive shock.

2.2.3 Exponential-GARCH (1, 1)[80]

Introduced by Nelson in 1991, the second model that allows for the asymmetric effects is the Exponential-GARCH. When considering the EGARCH process depicted in *EQ. 11*, there are three things worth noting.[81]

$$\log(\sigma_t^2) = \omega + \beta_1 \log(\sigma_{t-1}^2) + \alpha_1 \left| \frac{\epsilon_{t-1}}{\sigma_{t-1}} \right| + \gamma_1 \frac{\epsilon_{t-1}}{\sigma_{t-1}} \qquad \textbf{EQ. 11}$$

First, the conditional variance is in logarithmic form, meaning that the estimated coefficients are positive, as stated above. Second, by not using ϵ_{t-1}^2, as is done in the TGARCH model and instead using the standardizedϵ_{t-1}, Nelson argues that it gives a better interpretation of the size and persistence of the shocks. Lastly, the EGARCH allows for leveraged effects. If $\frac{\epsilon_{t-1}}{\sigma_{t-1}}$>0, then the effect of the shock on the log of the conditional variance is $\alpha_1 + \gamma_1$. If $\frac{\epsilon_{t-1}}{\sigma_{t-1}} < 0$, then the effect is $-\alpha_1 + \gamma_1$.[82]

2.2.4 Distributions

The GARCH models are estimated by Maximum Likelihood (ML). In order for ML to work, distributional assumptions about the conditional error terms have to be established.[83] This thesis will use three such distributions; the normal distribution,

[79] Enders(2010) p. 155
[80] Ibid p. 156
[81] Ibid p. 156
[82] Ibid p. 156
[83] Ibid p.211

Student's t-distribution and the General Error Distribution (GED). The discussion about distributions can easily become very technical, and therefore only brief explanation of the differences will be made here. For a more mathematical explanation the reader can refer to the source material.[84]

Normal distribution

The normal distribution is the most commonly used distribution when doing statistical tests and it exhibits many simple properties. The main advantage is that the whole distribution can be explained with only two parameters, namely mean and variance.[85] The VaR estimate under normal distribution is given by *EQ. 12* and the probability density function of the normal distribution can be seen in figure 2.

$$\mathrm{Var}_{\alpha}(\mathrm{L})_{t+1} = \mu + \sigma_{T+1} Z_{\alpha,} \qquad \textbf{\textit{EQ. 12}}$$

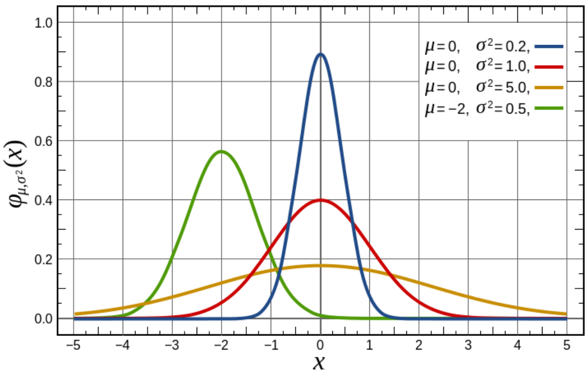


Figure 2: *Normal distribution*[86]

In this study, a confidence level of 99% is used, which means that the critical value, Z_{α} is 2.326. The forecasted standard deviation, σ_{T+1}, will depend on the chosen GARCH model. Consequently, there is a possibility that this distribution will not provide accurate estimates of VaR since financial instruments usually exhibits fat tail distribution characteristics.

Student's t- distribution

The student's t-distribution is closely related to the normal distribution with the exception that it can account for kurtosis or fat tails. Financial data often exhibits excess kurtosis and this is why the t-distribution is frequently used when modelling financial instruments behavior. The parameter that determines the fatness of the tails is the

[84] Hamilton, James Douglas. "*Time series analysis*." Vol. 2. Princeton: Princeton university press, 1994
[85] Verbeek, Marno. "*A guide to modern econometrics*." John Wiley & Sons, 2012 4th ed. p. 454.
[86] Retrieved from <http://en.wikipedia.org/wiki/File:Normal_Distribution_PDF.svg> Accessed: 2014-08-01

degree of freedom (d.f.) parameter *V*. When d.f. = $V \rightarrow \infty$, the t-distribution will approach normal distribution.[87] The VaR estimation under student's t-distribution is given by *EQ. 13*, where *T* has the prefix V_t. This means that the critical value is not fixed at 2.326, as is the case with normal distribution. Instead, it takes into consideration the degree of freedom at *t*, as is illustrated by figure 3, which shows the differences between normal (blue) and t distribution(red) by graphing their probability density function where the student's t-distribution has *v=1*.[88]

$$\mathrm{Var}_{\alpha}(\mathrm{L})_{t+1} = \mu + \sigma_{T+1}\mathrm{T}_{\alpha,v_t} \qquad \textit{EQ. 13}$$

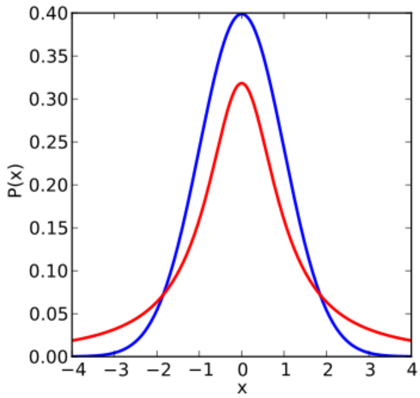


Figure 3: *Student's t-distribution*[89]

As can be seen, the t distribution has more weight in the tails than the normal distribution. Given that financial assets have a distribution function where the rate of return is fat-tailed; it would make sense to model VaR with a t-distribution rather than a normal distribution; especially if the asset in question has a higher probability of a large loss than is indicated by a normal distribution.[90]

Generalized error distribution

The third and final distribution used in this study is the Generalized Error Distribution (GED). The Generalized Error Distribution (GED) was first introduced by Subbotin in 1923 and depends on the so called 'shape parameter β'. This parameter is similar to the degrees of freedom of the t-distribution since it decides the fatness of the tails. The t-distribution can only produce fatter tails compared to the normal distribution unlike the GED, which can indicate either thinner or fatter tails depending on the shape parameter.

[87] Verbeek (2012) p. 457
[88] Please refer to appendix 8.2 for figure showing the differences between critical values for the different distributions.
[89] Retrieved from <http://en.wikipedia.org/wiki/File:T_distribution_1df_enhanced.svg> Accessed: 2014-08-01
[90] Enders (2010) p.157-158

For example, $\beta=2$ means that the function follows a normal distribution, $\beta<2$ indicate that the distribution has fatter tails than the normal distribution and $\beta>2$ means thinner tails.[91] This relationship is illustrated by figure 4, which shows how GED can produce a pointy or totally flat curve at the mean depending on β. The VaR estimate under GED is described in *EQ. 14*. Similarly to the t-distribution, the critical value G_{α,β_t} is not fixed.[92]

$$\mathrm{Var}_{\alpha}(L)_{t+1} = \mu + \sigma_{T+1} G_{\alpha,\beta_t} \qquad \textbf{\textit{EQ. 14}}$$

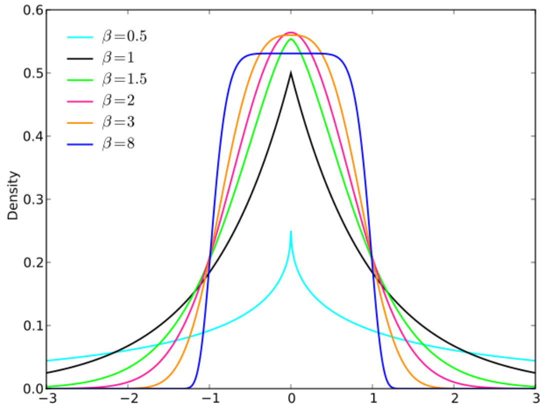


Figure 4: *GED distribution*[93]

Brief summary of the distributions

For this thesis, we have chosen the three most common distributions used in estimating VaR. In addition to the short descriptions above, we have placed a graph of the change in critical values for the GARCH(1,1) model for WTI in Appendix 8.2 to further stress their differences. The method used by e-views to calculate the maximum likelihood is not discussed further in this paper, but interested readers can consult the e-views user guide or Hamilton[94] for further details.

2.3 General Discussions of methodologies[95]

There are both pros and cons in using the non-parametric and the parametric method. A huge advantage of the non-parametric methods is that it is intuitive and simple since it does not depend on any parametric assumptions.[96] This means that it does not need to explicitly model fat tails, skewedness or any other feature that can cause problems for

[91] Fan (2008) p.3159
[92] Please refer to appendix 8.2 for figure showing how the critical value for the GARCH model changes with the different distributions over time.
[93] Retrieved from <http://en.wikipedia.org/wiki/File:Generalized_normal_densities.svg> Accessed: 2014-08-01
[94] Hamilton (1994) p.482
[95] Dowd (2005) p 99-100, p182
[96] ibid p 99

parametric methods as it is solely reliant on the empirical loss distribution. Furthermore, the non-parametric approach can accommodate any kind of instrument, and its result is easy to understand and communicate to senior managers, supervisors or rating agencies. The lack of needed assumptions and the ease of communicating its result is why the non-parametric methods are popular. [97]

The main problem with the non-parametric method however, is that it is too heavily dependent on historical data. This is also the root of many of its problem. Firstly, it is constrained by the largest loss in the data sample. This is especially true for BHS, since it is impossible for BHS to forecast a VaR larger than a loss in its in-sample. This problem is somewhat fixed in the VWHS method, but the problem still remains as the largest loss in the sample is more or less constrained by all non-parametric methods.[98] The second problem is the so called ghost effect, which entails that there is a change in the VaR estimates due to some significant observations falling out of the estimation window. This problem is significant in the BHS method, since all observations are given the same weight irrespective of where the observation is in the in-sample period. This problem is not as great for the AWHS method, since it gives lower weight to observations near the end of the observation window. So, when observation finally fall out of the in-sample period, its impact on VaR is not as great as it would be in the case of the BHS method. The third problem with the non-parametric methods is that they are slow to react when there is new market information. BHS for example is not well suited to handle large losses which are unlikely to recur. This is because the observation would dominate the VaR estimate until it falls out of the sample, only to create ghost effect. This problem is not as prevalent for AWHS and VWHS since the observations effect on the VaR estimate will decrease gradually.[99] Lastly, BHS and AWHS do to some extent underestimate the risk during calm periods and overestimate during turbulent times. This is an advantage for VWHS, as it lets us obtain VaR estimates that can exceed maximum loss in our data set. Thus, enabling the historical returns to be scaled upwards in periods of high volatility. This means that applying the VWHS method can produce VaR estimates that actually exceed the largest loss in previous historical losses.[100]

[97] Dowd (2005) p. 99-100
[98] Ibid p. 100
[99] Ibid p.99
[100] Ibid p.95

One of the important decisions when applying the non-parametric method is to choose the right sample length for the in-sample period. A common rule of thumb is that at least 500 observations are needed to get a fairly accurate risk measurement. However, it is important to understand that a sample window that is too long will cause the same type of problems as those with aged data which was explained above. In addition, with a long in-sample period, new information will not contribute as much to the estimate as it is slower to react. Despite these potential problems, the non-parametric methods are widely used and very attractive in-terms simplicity. They offer reasonable results under simplistic assumptions such as normality and under stable market conditions. The drawbacks explained earlier of the non-parametric method in combination with oil characteristics as a highly volatile commodity makes it important to complement the non-parametric method with other parametric methods. [101]

While the non-parametric has its strength in not having to make distributional assumptions, the parametric methods require these assumptions. Misspecifying the assumptions for the parametric method can be potentially disastrous since it can produce highly inaccurate results in times of distress. If the distributional assumptions are correctly specified however, it will provide better VaR estimates than a non-parametric method since it uses additional information inherent in the assumption. Therefore, the difficulty for parametric methods lies with the choice of distributional assumptions since different assets may have different needs in calculating the parameters. In order to make the right distributional assumption a number of factors need to be taken into consideration. Is the data skewed to some tail? And does it exhibit any kurtosis? If the data seem to have some kurtosis for example, it might be valuable to check several different fat tail distributions. Obtaining good results from one specified model does not mean that the model is perfect or its assumptions. It is therefore important to compliment any testing with additional models, but also to try different specifications in order check their sensitivity. [102]

Having examined and explained the different models and potential pros and cons, the next section will delve into the two backtesting methods used to test which model is the best.

[101] Dowd (2005) p.100
[102] Ibid p.182

2.4 Backtesting

After the VaR estimates have been obtained from the out-of-sample forecast, it is important to evaluate which model most accurately captures the volatility of the returns. In this study, two methods will be used. Firstly, the Christoffersen backtesting method and secondly, the regulatory method used under the Basel accord to test the accuracy of the internal based risk models.

2.4.1 Christoffersen[103]

Developed in 1998, Christoffersen extended Kupiec's pioneering unconditional VaR coverage test from 1995 to include a conditional VaR coverage test when backtesting.[104] Before the test itself can be explained however, a definition is needed for the hit sequence of VaR violations. To do this, we start by defining VaR_{t+1}^{p} as a number constructed at *t,* such that the probability of observing a portfolio loss at *t+1* that is larger than the VaR_{t+1}^{p} forecast is given by the probability *p*. Having made this definition, we can use observed ex-ante VaR forecasts and ex-post losses by defining the hit sequence of VaR violations as:[105]

$$V_{t+1} = \begin{Bmatrix} 1, if\ PL_{t+1} > VaR_{t+1}^{p} \\ 0, if\ PL_{t+1} \leq VaR_{t+1}^{p} \end{Bmatrix} \qquad \textit{\textbf{EQ. 15}}$$

The observation in the hit sequence is equal to 1 on *t+1* if the actual loss is greater than forecasted VaR at *t+1*, and 0 if the forecasted VaR is not violated. When backtesting the model, a hit sequence,$\{V_{t+1}\}_{t=1}^{T}$ is then created over a backtesting period of *T* observations.

As previously specified in the beginning, Christoffersen extended Kupiec's test to include two parts, the unconditional coverage and the conditional coverage. The unconditional coverage measures if the probability on average of observing a violation is p. Written in mathematical terms, a risk model has correct unconditional coverage if Pr $(V_{t+1} = 1)$=*p*. If the model is not however, it has either over-/underestimated the VaR estimate. The conditional coverage on the other hand, measures if the risk model gives a VaR hit with probability *p* irrespective of what information is available on the day before. In

[103] Christoffersen, P. F. *"Backtesting, Prepared for the Encyclopedia of Quantitative Finance, R. Cont."* (2008).
[104] Ibid p. 2
[105] Ibid p.3

mathematical terms, the risk model has correct conditional coverage if $Pr_t\,(V_{t+1} = 1) = p$.[106]

Performing a backtest of VaR is essentially the same as testing if the hit series follows a Bernoulli distribution, where the null-hypothesis is given in *EQ. 16:*

$$H_0: V_{t+1} \sim i.i.d\ Bernoulli(p) \qquad \textbf{EQ. 16}$$

Where, p will be 0.01 or 0.05 depending on the coverage rate. If the risk model is correctly specified the hit sequence will produce a 1 with probability 1% or 5% over the string of observations. In the following part of this section, the unconditional and conditional coverage tests will be explained in detail.[107]

Unconditional coverage[108]

In the unconditional coverage test, a likelihood ratio test is used to check if the expected number of violations, p, is the same as the actual number of violations. To do this, we first define the likelihood under the null-hypothesis as $\mathrm{L(p)} = (1-\mathrm{p})^{t_0}\mathrm{p}^{t_1}$, where t_0 is the number of non-violations in the hit series, t_1 is number of violation in the series, and P is the expected number of violations under the null-hypothesis (i.e. H_0: E [V_t] $=p$). The alternative hypothesis is defined as $\mathrm{L}(\widehat{\pi}) = (1-\widehat{\pi})^{t_0}\widehat{\pi}^{1}$, where $\widehat{\pi}$ is the actual probability of observing a violation in the hit series and is mathematically defined as $\widehat{\pi} = \mathrm{t}_1/\mathrm{T}$. Combining these we can estimate the log likelihood function in accordance with EQ. 17.

$$LR_{uc} = -2\,ln[L(p)/L(\widehat{\pi})] \sim \chi^2 \quad d.f.\ 1 \qquad \textbf{\textit{EQ. 17}}$$

$$H_0 : p = \widehat{\pi}$$
$$H_1 : p \neq \widehat{\pi}$$

The basic idea behind this test is to evaluate the distance between the unconstrained likelihood *L(π)* and the constrained likelihood *L(p).* If we fail to reject the null hypothesis then expected number of violations, p, is not statistically different from the actual number of violations.[109]

[106] Christoffersen(2008) p.3
[107] Ibid p.3
[108] Ibid p.4
[109] Ibid p.4

Conditional coverage[110]

The conditional coverage part of the Christoffersen test is used to check whether the statement $\Pr_t(V_{t+1} = 1) = p$ is true. If it is not true, then there are clustering effects in the series, and the violations in the hit series are not conditionally independent. Ideally, violations should be completely random, because if there is clustering effects then the risk manager knows that there is an increased probability of observing a violation in *t+1* given that there was a violation at t.[111] To analyze if there is clustering effects, we use the likelihood function. If we assume that the hit sequence is dependent over time, we can express the transition from one state to another using the probability matrix below.

$$\prod\nolimits_1 = \begin{bmatrix} 1-\pi_{01} & \pi_{01} \\ 1-\pi_{11} & \pi_{11} \end{bmatrix}$$

$$\begin{aligned} \pi_{01} &= Pr_t(V_t = 0,\ V_{t+1} = 1) \\ \pi_{11} &= Pr_t\ (V_t = 1,\ V_{t+1} = 1) \\ (1-\pi_{11}) = \pi_{10}\ \ &is\, Pr_t\ (V_t = 1,\ V_{t+1} = 0) \\ (1-\pi_{01}) = \pi_{00}\ \ &is\, Pr_t\ (V_t = 0,\ V_{t+1} = 0) \end{aligned}$$

Knowing that we have a non-violation in *t*, then π_{01} is the probability of observing a violation in *t+1*. [112]We define the likelihood function under the alternative hypothesis in *EQ. 18.*

$$L(\prod\nolimits_1) = (1-\pi_{01})^{T_{00}}\pi_{01}{}^{T_{01}}(1-\pi_{11})^{T_{10}}\ \pi_{11}{}^{T_{11}} \qquad \textbf{\textit{EQ. 18}}$$

T_{00} is number of observations that have a non-violation followed by a non-violation. Taking first derivatives w.r.t π_{01} and π_{11} we solve for Maximum Likelihood estimates:

$$\hat{\pi}_{01} = \frac{T_{01}}{T_{00}+T_{01}}, \quad \hat{\pi}_{11} = \frac{T_{11}}{T_{10}+T_{11}}$$

If the violations are independent, it means that a violation tomorrow does not depend on whether there is a violation today. In mathematical terms if $\pi_{11} = \pi_{01} = \pi$ and we use this restriction for the restricted likelihood ratio, then it the same as the unconditional test.

$$L(\hat{\pi}) = (1-\hat{\pi})^{t_0}\hat{\pi}^{t_1}$$

[110] Christoffersen(2008) p 5.
[111] Ibid p.5
[112] Ibid p.4

The Christoffersen test for independence is based on combining the restricted and unrestricted models in order to evaluate if there is independence in the violations. The likelihood ratio test can be seen in *EQ. 19*.[113]

$$LR_{indep} = -2\,ln\left[L(\hat{\pi})/L\left(\prod\ _1\right)\right] \sim \chi^2 \quad df.1 \qquad \textbf{\textit{EQ. 19}}$$

These models are then used to create the Christoffersen combined test in *EQ. 20*, which checks the validity of the model.

$$LRcc = LRuc + LR_{IND} \sim \chi^2 \quad df.2 \qquad \textbf{\textit{EQ. 20}}$$

2.4.2 Basel Backtest[114]

The second test used in this thesis was created by the Basel committee in order to validate banks internal models after the 1996 amendment of the Basel I accord. By using the 250 last $VaR_{0.99}$ estimates and actual losses, they assess the model by placing it either in the green, yellow or red zone depending on how many violations have occurred in the last 250 days (see table below). For example, a 1-day 99% VaR would be expected to have 2.5 violations in a period, but the Basel accord accepts a model with up to 4 violations. If the model surpasses four violations, then a penalty would be added giving the bank a higher market capital charge since their model underestimates the risk in the underlying asset.

Basel Accord Penalty ZONE	Number of Violations	Increase in Scaling Factor, k
Green Zone	0	0.00
	1	0.00
	2	0.00
	3	0.00
	4	0.00
Yellow Zone	5	0.40
	6	0.50
	7	0.65
	8	0.75
	9	0.85
Red Zone	10 or more	1.00

Table 1- *Basel Accord penalty zones from backtesting*[115]

113Christoffersen (2008) p5

[114] Supervisory Framework for the use of "*Backtesting" In Conjunction with the Internal Models Approach to Market Risk Capital Requirements"* Basle Committee on Banking Supervision 1996.

[115] Basel (1996) p.14

However, this study will not investigate which model is best in optimizing capital requirements. The Basel test is merely easy to incorporate and gives an insight to which model is satisfactory according to the Basel accord.

3. Implementation of the model

In this section, we will go through the implementation of the models along with the data handling for both WTI Crude Oil and S&P 500 along with their descriptive statistics. This will be followed by a description of how to interpret the results before the results themselves are presented.

For the non-parametric methods we use an in-sample size of 1000. As explained in 2.3, there is a tradeoff when choosing the right in-sample length; too small we do not get consistent estimates in our parameters; too long we include observations which are not relevant to the current market conditions. Cabedo and Moya used 1250 observations in their study.[116] Having this many observations in the in-sample will make the model slow to react to new information. However, when having a high confidence level, a small sample will produce inaccurate estimates and therefore a large sample is needed.[117] The 1000 observations we believe provides a good balance between these problems. The non-parametric methods where estimated with Excel and VBA. Excel was also used for creating the Christoffersen test as well as the Basel back test.

Our estimated GARCH models have an in-sample of 1000 observations, which were used to forecast the one day ahead conditional standard deviation. The GARCH model is re-estimated each trading day over the period 2006-12-29 to 2013-12-30 to obtain 1826 forecasts. The model was re-estimated since we suspected that the parameters significance and size will changes over this period. We believe this change to be especially frequent during the crisis and therefore we re-estimate the model to provide more accurate results. This was done by creating a loop in Eviews and resulted in 1826 different GARCH estimations and forecasts.

[116] Cabedo, J. Moya (2003) p 244-245
[117] Hendricks (1996) p 44

Data

In this paper, we use two time series data sets containing daily observations of WTI crude oil spot price and S&P 500 composite index retrieved from Thomas Reuters DataStream. This resulted in 2940 observations respectively covering the period from 9 September 2002 to the 31 December 2013. The daily observations are then converted by taking the log differences in order to obtain a Profit and Loss (P/L) series according to EQ 21.

$$R_t = log(P_t) - log(P_{t-1}) = \log(\frac{P_t}{P_{t-1}}) \qquad \textbf{EQ. 21}$$

By converting the raw data over the sample period, we obtain the stationary series showed in figure 5 and 7.[118]

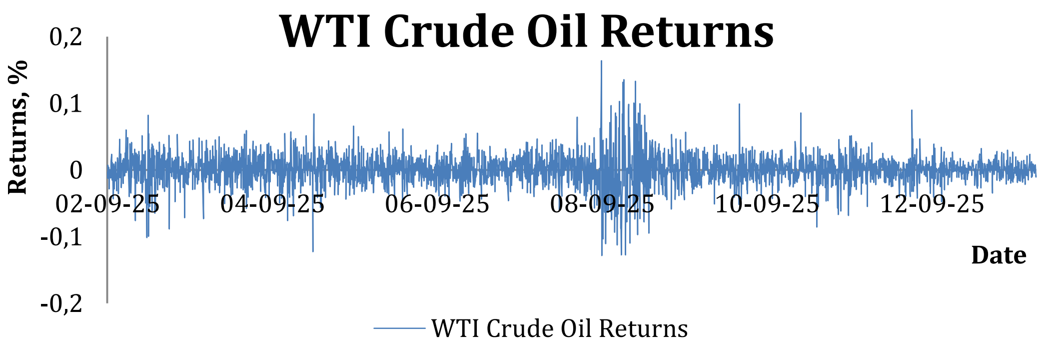


Figure 5- *WTI Crude Oil Returns* (Data Retrieved from Thomson Reuters DataStream - 2014-04-30)

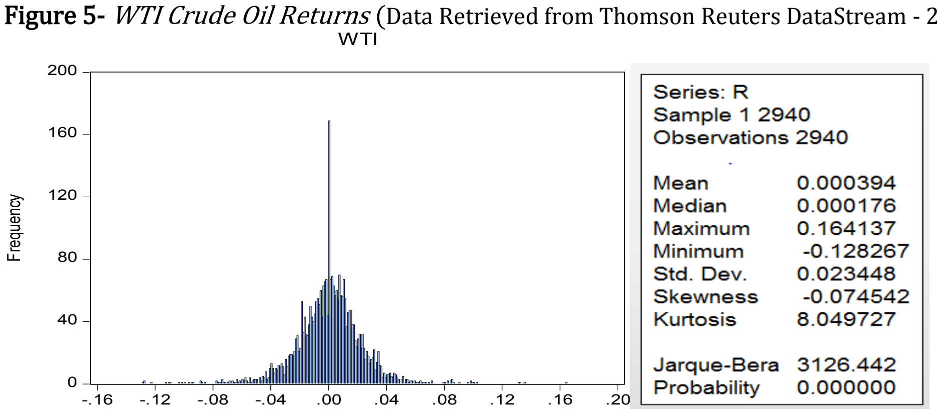


Figure 6- *Descriptive statistics for WTI Crude Oil*

[118] See appendix 8.3 for stationarity test

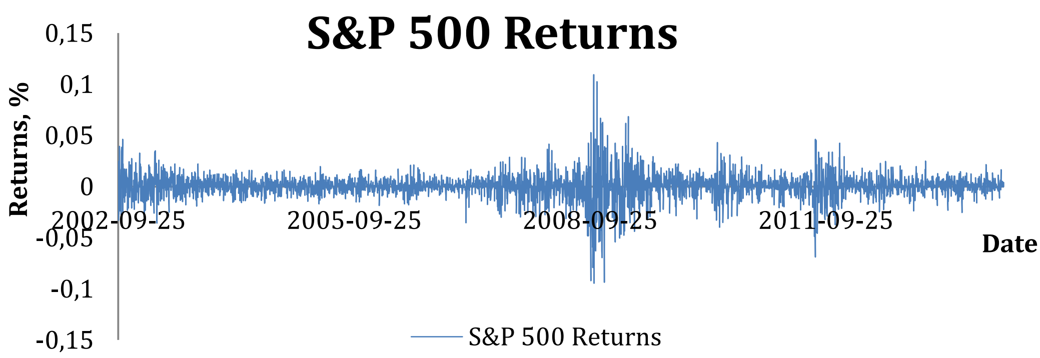


Figure 7- *S&P 500 Composite Index Returns* (Data Retrieved from Thomson Reuters DataStream - 2014-04-30)

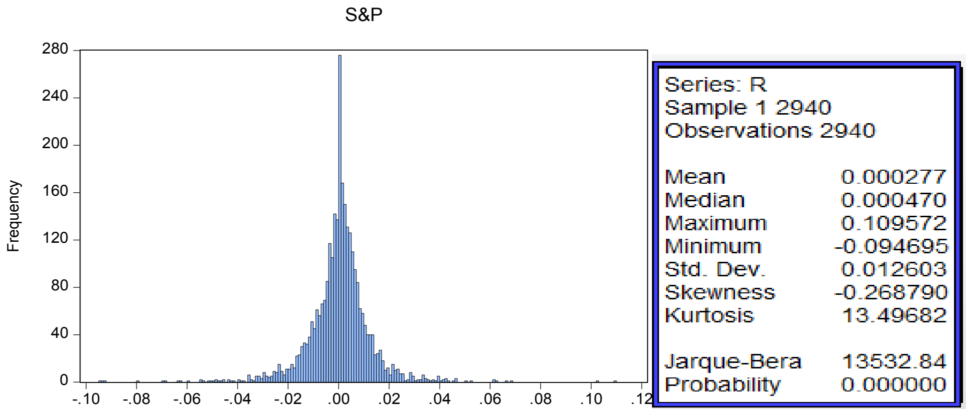


Figure 8- *Descriptive statistic for S&P 500 Composite Index*

As can be seen from figure 6, WTI has a kurtosis of 8.0497. The kurtosis value indicate the fatness of the tails, and under normal distribution the value is 3.The Jarque-Bera test is used to test for normality and the null hypothesis is that the series is normally distributed. Since we reject the null hypothesis we conclude that the series is not normally distributed. The skewedness factor is close to zero and therefore we conclude that the series does not seem to exhibit any notable asymmetries. In conclusion, the WTI P/L series seems to have fat tails, but no asymmetry.

S&P produces similar results since it has excess kurtosis and rejects the null hypothesis for the Jarque-Bera test. It appears to have somewhat more skewedness, however, a rule of thumb is that skewedness between +0.5 and -0.5 is classified as roughly symmetric. [119] We therefore classify both series as approximately symmetric with fat tails over the whole period.

[119] Brown, Stan (2012-12-27) *"Measures of Shape: Skewness and Kurtosis"* retrieved from http://www.tc3.edu/instruct/sbrown/stat/shape.htm Accessed: (2014-08-10)

How to interpret the results

Our total out-of-sample continuous forecast period is 1826 observations. With α=0.99 we expect that a correctly specified model will have a total of 18.26 violations in its hit series. To evaluate this, we apply the Christoffersen test to see if the value of the produced hit series is statistically different from the expected value of 18.26. The values of these tests are presented in table 2 and 3 for WTI, and table 4 and 5 for S&P 500.

It is important to choose a correct confidence level when performing the Christoffersen test, since there is a tradeoff between type I and type II errors. If the confidence level is set to too parsimonious, the test will reject a correct specified model too often, but if it is set too high we will accept an incorrectly specified model too frequently. In this paper, a 10% confidence interval is chosen because it gives a nice tradeoff between these types of errors.[120] This means that we will fail to reject the null hypothesis of a correctly specified model if the Christoffersen test is below 2.706 when we have one degree of freedom. This is the case when checking the components of the Christoffersen independently. When performing the combined test, the degree of freedom is two which means we reject the null hypothesis if the value of the test is above 4.605.

[120] Christoffersen (2008) p. 9

4. Results

The last section explained the different methods and their implementation. In the following section, the results from the backtests are presented. First, the results for WTI Crude Oil will be explored, starting with the non-parametric models, and then followed by the parametric models. Subsequently, the equivalent results will be presented for S&P 500 for a comparative study in the analysis part of the thesis.

4.1 WTI Crude Oil

4.1.1 Non-parametric Methods

Examining the non-parametric models in table 2 and figure 9, the VWHS model performs well for the unconditional part of the test, but fails the test for independence. This means that it does not accurately estimate VaR when there is a period of high volatility. This can be seen in the Basel test, as there are 20 violations over a period of 250 days. From the results for the BHS and AWHS models, we can reject both according to the Christoffersen test and the Basel test.

	BHS	VWHS	AWHS
Violations	29	21	32
Independence	16.307	10.372	9.627
Uncond. Coverage	5.426	0.399	8.545
Cond. Coverage	21.733	10.771	18.173
Basel	28	20	28

Table 2- *Non-Parametric backtest results for WTI Crude Oil*

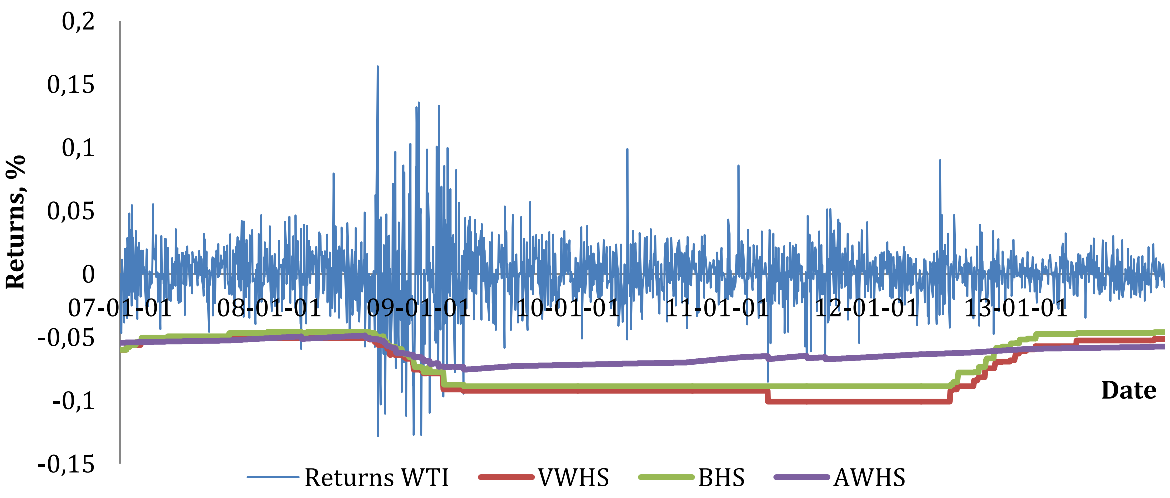


Figure 9- *Shows Non-parametric methods for WTI Crude Oil*

4.1.2 Parametric Methods

Estimating VaR with GARCH(1,1) and normal distribution produces fairly good results according to the Christoffersen test, since the null-hypothesis cannot be rejected, as seen in table 3. This is true, both when using GARCH(1,1) with and without an AR(1) term as both result in a unconditional and conditional coverage, which is below the critical values of 2.706 and 4.605 respectively. However, when using the Basel test, both models have a total of 8 violations in any 250 day period, and therefore lie exceptionally close to the red zone, which constitutes a poor VaR measurement. In fact, all of the models with an underlying normal distribution show poor results according to the Basel test. Comparing the Christoffersen result of the symmetric GARCH(1,1) model to the asymmetric TGARCH(1,1) and EGARCH(1,1) models, it becomes clear that adding asymmetric models with a normal distribution adds no or little value when estimating VaR for WTI crude oil, as all models reject the null-hypothesis of the combined Christoffersen test.

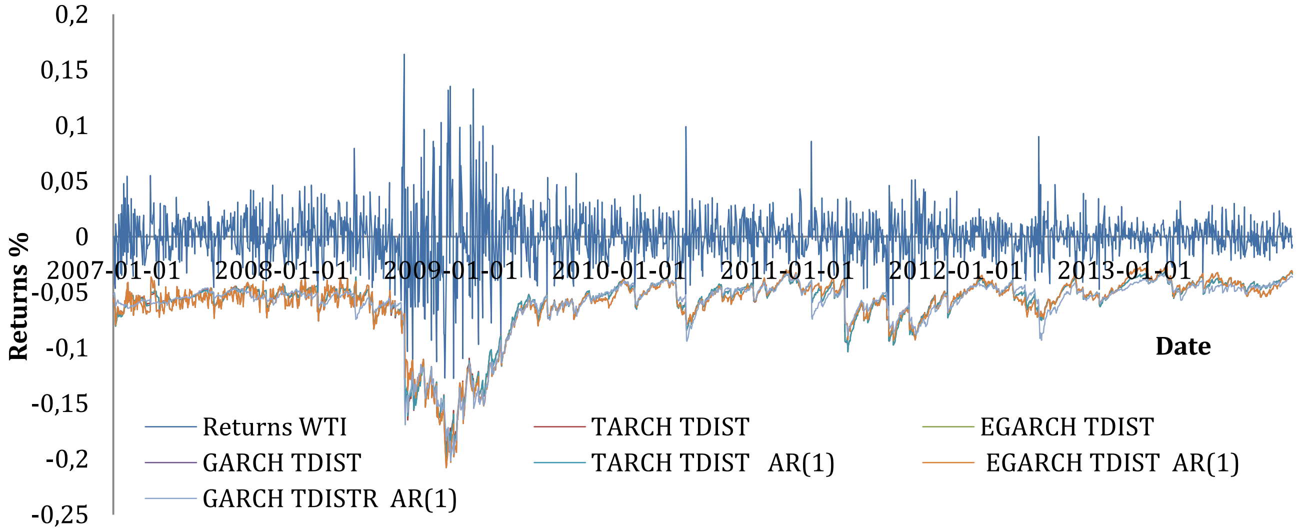


Figure 10- *Shows Parametric GARCH models with student's t-distribution for WTI Crude Oil*

The student's t-distribution on the other hand, estimates a relatively parsimonious VaR forecast. This can be seen in figure 10, where the VaR estimates consistently lie above actual losses for all models and therefore produce fewer violations. As a result, none of the models have more than 5 violations in any given 250 day period according to the Basel test. If we consider the Christoffersen test however, the standard GARCH(1,1) and EGARCH(1,1) including the AR(1) with t-distribution, produces poor results as the null-hypothesis is rejected. This is subsequently the case for the other models, which, even if they don't reject the null-hypothesis, lie extremely close. The reason for this is that the models overestimate the risk and therefore protract fewer violations as a result. Thus explaining why the number of violation ranges between 10 and 12 instead of being closer to the expected number of 18.26 violations.

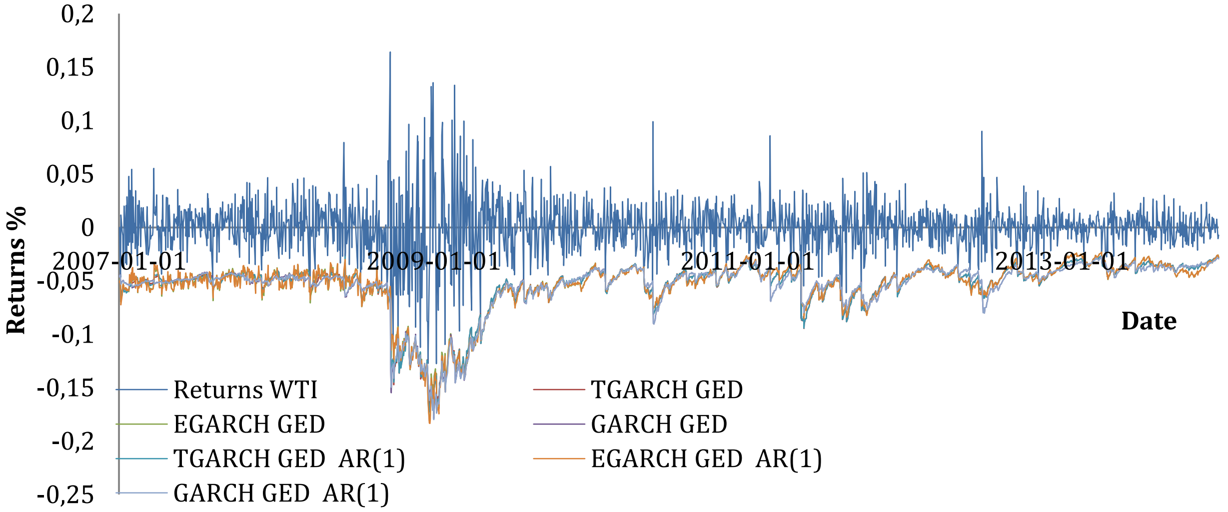


Figure 11- *Shows Parametric GARCH models with Generalized Error Distribution for WTI Crude Oil*

Lastly, considering the different models with an underlying Generalized Error Distribution, it becomes clear that this distribution seem to capture risk with better accuracy than both the normal distribution and student's t-distribution. Especially, given that the number of violations in the Christoffersen test is closer to the optimum of 18.26, for all models. The difficulty however, pertains to which of the models are the best at most accurately estimating VaR. The asymmetric GARCH models produce largely the same results as the symmetric GARCH. By considering the number of violations however, the EGARCH models might produce 2 or 3 too many, while the TGARCH models aren't far behind. Additionally, by examining the histogram, there are few signs of skewedness. Therefore, accounting for asymmetric effects in the volatility of WTI does not necessarily result in a better VaR estimates. It might thus, be most sensible to choose one of the simpler GARCH(1,1) models, and in this case GARCH(1,1) with an AR(1) process and the GED.

	GARCH n-dist.	GARCH n-dist. AR(1)	TGARCH n-dist.	TGARCH n-dist. AR(1)	EGARCH n-dist.	EGARCH n-dist. AR(1)
Violations	23	23	28	26	29	30
Independence	0.587	0.587	0.872	0.751	0.936	1.002
Uncond. Coverage	1.154	1.154	4.523	2.938	5.426	6.399
Cond. Coverage	1.740	1.740	5.395	3.689	6.362	7.401
Basel	8	8	9	12	11	12

	GARCH t-dist.	GARCH t-dist. AR(1)	TARCH t-dist.	TGARCH t-dist. AR(1)	EGARCH t-dist.	EGARCH t-dist. AR(1)
Violations	10	10	12	12	12	10
Independence	0.110	0.110	0.159	0.159	0.159	0.110
Uncond. Coverage	4.506	4.506	2.459	2.459	2.459	4.506
Cond. Coverage	4.616	4.616	2.618	2.618	2.618	4.616
Basel	3	5	4	4	4	3

	GARCH GED	GARCH GED AR(1)	TGARCH GED	TGARCH GED AR(1)	EGARCH GED	EGARCH GED AR(1)
Violations	19	18	18	19	20	21
Independence	0.400	0.358	0.358	0.400	0.443	0.489
Uncond. Coverage	0.031	0.003	0.003	0.031	0.164	0.399
Cond. Coverage	0.430	0.362	0.362	0.430	0.607	0.888
Basel	5	5	6	6	7	8

Table 3- *Parametric backtest results for various GARCH models and underlying distributions on WTI Crude Oil*

4.2 S&P 500

4.2.1 Non-parametric Methods

In consideration of another benchmark for comparative purposes, S&P's 500 Composite Index is added, as mentioned in the objective. By first examining the results for non-parametric methods in table 4, it is obvious that neither BHS, AWHS, nor VWHS is able to accurately capture the risk in S&P 500. The cause of this is that all models grossly fail both the Christoffersen combined test and Basel test.

	BHS	VWHS	AWHS
Violations	43	35	53
Independence	5.456096911	4.655063643	5.941586276
Uncond. Coverage	24.54559978	12.23856584	39.91862228
Cond. Coverage	30.00169669	16.89362948	45.86020855
Basel	26	20	30

Table 4- *Non-Parametric Backtest for S&P 500*

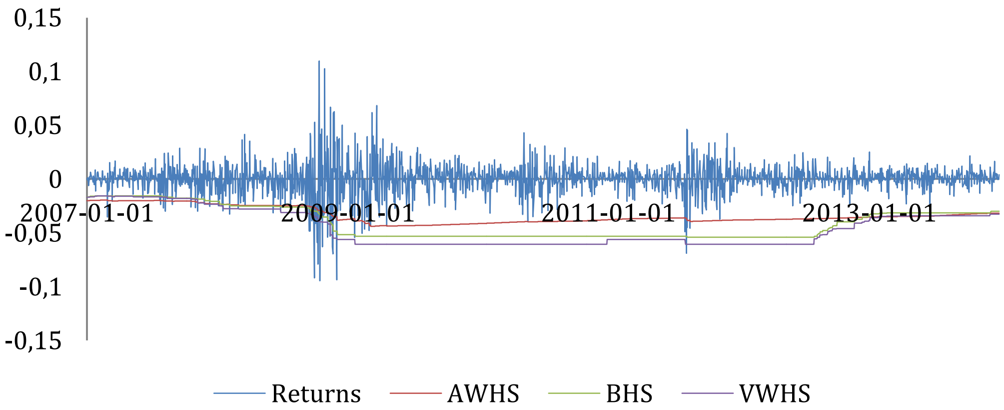


Figure 10- *Non-parametric methods for S&P 500*

4.2.2 Parametric Methods

When applying the GARCH(1,1) model for S&P 500, autocorrelation is obtained in the squared standardized residual. This is an indication of a misspecified variance equation. Therefore, a GARCH (2,1) model is applied instead, after results showed that it had no autocorrelation in the squared standardized residuals. By changing the GARCH specification, the variance equation is correctly specified and further testing is made possible. By applying the same parametric models on S&P 500, as for WTI, the following results in table 5 are obtained.

	GARCH n-dist.	GARCH n-dist. AR(1)	TGARCH n-dist.	TGARCH n-dist. AR(1)	EGARCH n-dist.	EGARCH n-dist. AR(1)
Violations	53	54	51	54	56	56
Independence	3.169	0.271	0.149	0.271	0.372	0.372
Uncond. Coverage	44.182	46.373	39.919	46.373	50.867	50.867
Cond. Coverage	47.351	46.644	40.067	46.644	51.240	51.240
Basel	16	16	13	15	17	18

	GARCH t-dist.	GARCH t-dist. AR(1)	TARCH t-dist.	TGARCH t-dist. AR(1)	EGARCH t-dist.	EGARCH t-dist. AR(1)
Violations	19	19	20	23	23	19
Independence	0.400	0.400	0.443	0.587	0.587	0.400
Uncond. Coverage	0.031	0.031	0.164	1.154	1.154	0.031
Cond. Coverage	0.430	0.430	0.607	1.740	1.740	0.430
Basel	12	9	8	9	9	12

	GARCH GED	GARCH GED AR(1)	TGARCH GED	TGARCH GED AR(1)	EGARCH GED	EGARCH GED AR(1)
Violations	28	29	33	32	40	41
Independence	0.872	0.936	1.215	1.142	1.792	0.007
Uncond. Coverage	4.523	5.426	9.715	8.545	19.540	21.159
Cond. Coverage	5.395	6.362	10.930	9.687	21.332	21.166
Basel	10	10	11	11	14	14

Table 5- *Parametric backtest results for the different GARCH (2, 1) models with S&P's 500 Composite Index*

By first examining the results for the normal distribution, it is evident that the normal distribution is poor in combination with standard GARCH(2,1), EGARCH(2,1) and TGARCH(2,1), as a risk measure for S&P 500. Especially, after all models failed both the Christoffersen test and the Basel test. This is however, not surprising since the descriptive statistics showed that the series exhibit excess kurtosis, and does not follow a normal distribution according to the Jarque-Bera test.

The poor results are also evident for the Generalized Error Distribution, as it too fails both the Christoffersen test and the Basel test for all parametric models; though, not as severely as the normal distribution. Last of all, by observing the results for the student's t-distribution; it appears that the log first differences of S&P 500 in conjunction with the t-distribution offer the best VaR estimation method. The problem however, lies with choosing the right GARCH model, as they all produce similar results. Although, in a similar fashion to WTI, it is preferable to choose the simpler GARCH model, which in this case would be the standard GARCH(2,1) with AR(1) as it produces less violations in the Basel test compared to GARCH(2,1) without the AR(1) term.

5. Analysis & Discussion

In this chapter, the results for WTI Crude Oil will be analyzed and discussed together with previous research. These results will then be compared to S&P 500, followed by potential extensions of this study.

West Texas Intermediate Crude Oil

Only the VWHS model of the non-parametric methods produced an acceptable result according to the Christoffersen backtest. However, when analyzing the VWHS model according to the Basel test, the model failed because of the 250 day period around the financial crisis, where there were 20 violations.[121] This shows that the RiskMetrics method may be a poor way of estimating the volatility in the VWHS model, and that other methods should be tested instead. One of which, is the semi-parametric GARCH(1,1) model with Historical Simulation that was proposed Costello. It also indicates, that the VWHS model is slow to react to sudden market conditions, even though it performed better than the AWHS model and the BHS model in that regard. Nevertheless, the slowness to react to new information is in line with previous discussions about the drawbacks of the non-parametric methods. Though it was unexpectedly slow for the AWHS model; especially, given that it performed worse than the BHS model, despite having a low decay factor. Consequently, a high decay factor in the AWHS model might have been more appropriate for the model to produce more accurate results, regardless of the non-parametric methods being rejected in favor of the parametric methods.

It was expected that the parametric methods would be superior to the non-parametric methods. Not only, because they allowed flexibility in assumptions, but because they had proven reliable in previous literature. Nonetheless, there were expected shortcomings in certain parametric approaches, in particular when applying normal distribution. It was therefore, highly unexpected that the GARCH(1,1) model with normal distribution produced an acceptable result in the Christoffersen test. Especially since the null-hypothesis of the Jarque-Bera test for normality had been rejected over the whole sample period together with Barone-Adesi's claim that the normal distribution together with a GARCH setup produced poor results; therefore, further strengthening our expectation that the normal distribution was a poor choice for WTI. Even so, this result

[121] For graph of violations see figure 9.

shows that the GARCH(1,1) model with normal distributions works fairly well for estimating VaR with a high confidence level for WTI, which is in line with Xiliang and Xi's conclusion, but contradiction to Hung. Though it can only be speculated as to why, simple models such as the GARCH(1,1) model have proven to perform better than more advanced models despite its simple nature,[122] which is certainly true in this case for the normal distribution. Thus, it may not have been unexpected that the Asymmetric EGARCH(1,1) and TGARCH(1,1) models perform poor estimates of VaR for WTI, knowing that the histogram showed no signs of skewedness rendering the models useless.

In conjunction with Hung's results for the t-distribution, our results also concluded that it was a poor estimate for evaluating VaR for WTI, as the risk was overestimated. However, though the asymmetric models with normal distribution failed the Christoffersen test, this was not the case for the t-distribution as both TGARCH(1,1), TGARCH(1,1) AR(1), and EGARCH(1,1) all failed to reject the null-hypothesis of the Christoffersen test. Additionally, all models with an underlying t-distribution performed exceedingly well on the Basel test, even though this is largely down to the model overestimating the risk; as can be seen in figure 10. A result, that becomes clear when comparing with the VaR forecasts with GED in figure 11.

The results for GED was expected, given that both Fan and Aghayev together with Rizvanoghlu established that GARCH(1,1) with GED outperformed both GARCH-N and HSAF in their studies. A result, which was in line with our hypothesis that a parametric model with a fat tailed distribution, would prove superior in estimating VaR for WTI at a high confidence level (99%). Although, this result was contradictory to Xiliang and Xi who found that GARCH(1,1) with GED was the best model for WTI Crude Oil at a low confidence level (95%) but not at a high confidence level. Further on the results, the difference between GARCH and the asymmetric GARCH models was not large in terms of VaR accuracy. This was to be expected, since the histogram was not skewed, meaning that EGARCH and TGARCH models have no asymmetries effects to capture.
In conclusion, the model that best captures VaR for WTI Crude Oil is a standard GARCH(1,1)-GED model.

[122] Hansen, Peter R., and Asger Lunde. "*A forecast comparison of volatility models: does anything beat a GARCH (1, 1)?*." *Journal of applied econometrics* 20.7 (2005): 873-889.

WTI compared with S&P 500

For S&P 500, all parametric methods with a normal distribution were rejected in the Christoffersen backtest. This is coherent with Angelidis as he also rejected the normal distribution for VaR estimates on S&P 500. This result is not entirely unexpected as the null hypothesis for the Jarque-Bera test was rejected, leading us to conclude that S&P 500 doesn't follow a normal distribution and that the normal distribution produces poor VaR estimation. The t-distribution however, seems to better capture the risk of S&P 500 while GED systematically underestimated the risk. This is also similar to Angelidis, who finds that EGARCH(1,1) with student's t-distribution provides the best results. However, though our EGARCH(1,1)-t produces reasonable results, our standard GARCH(1,1)-t produces marginally better results. Nevertheless, these results stand in contrast with WTI, where the t-distribution overestimated the risk, but GED accurately captured it, which would indicate that the t-distribution is generally more parsimonious while the normal distribution is the least parsimonious and the GED in-between. This is strengthened by appendix 8.2 since it indicates that the critical value for the t-distribution is the most parsimonious. However, it is worth noting, that even though we fail to measure VaR for S&P 500 with GED, Angelidis found that GED actually produces acceptable results.

In addition, though the Asymmetric GARCH models under t-distribution produce good results, they do not necessarily improve the VaR estimate which is a contrast to Awartani, since they found that asymmetric GARCH models do improve the VaR estimate. Their sample period 1987-2002 was not the same as this papers which might explain the contrasting results. When examining the histogram over the whole data set the series does not seem to show any noticeable skewedness in the returns. This symmetry helps explaining why the asymmetric GARCH models do not improve the VaR estimates. The non-parametric models perform very badly for S&P 500, since we reject all the models in our backtest. This is a similar result as with WTI, since we rejected the non-parametric methods there as well.

Further Research

This study can be extended in order to broaden the scope of understanding VaR for WTI or crude oil in general. Most obvious, would be by adding a two tailed VaR approach, since the tail of interest for a market participant depends on where in the production line they are located. However, a more difficult and probably more rewarding extension would be to check VaR for different holding periods and confidence intervals. This would broaden the understanding of the optimal VaR method for the returns series. Additionally, another possible extension would be to focus on the nonparametric methods by perform sensitivity analysis on different assumptions and sample lengths.

6 Conclusion

When estimating VaR for the P/L series for West Texas Intermediate, the parametric methods produced superior results in comparison to all non-parametric methods. We also found that a heavy tailed distribution produced the best VaR estimates, which concurs with **H1**. For **H2**, the results for S&P 500 were comparable to that of WTI, since the optimal method for S&P 500 was a parametric method with fat tail distribution. However, for WTI the most accurate method in estimating VaR was GARCH(1,1)-GED, while for S&P 500 it was GARCH(1,1)-t.

7 Bibliography

Books

Dowd, Kevin. “*Measuring market risk.”* John Wiley & Sons, (2005).

Edwards, Davis W. "*Energy Trading and Investing."* (2010).

Enders, Walter. *“Applied econometric time series.”* John Wiley & Sons, 2010.

Fattouh, Bassam. “*An anatomy of the crude oil pricing system*.” Oxford, England: Oxford Institute for Energy Studies, 2011.

Hamilton, James Douglas*. “Time series analysis*.” Vol. 2. Princeton: Princeton university press, 1994.

Jorion, Philippe. “ *Value at risk: the new benchmark for managing financial risk*.” Vol. 2. New York: McGraw-Hill, 2007.

Saunders, Anthony, and Linda Allen. "*Credit risk measurement*." 2nd John Wiley & Sons Inc. New York (2002).

Verbeek, Marno. “*A guide to modern econometrics*.” John Wiley & Sons, 2012 4:th ed.

Academic Articles

Aghayeva, Huseyn, and Islam Rizvanoghlub*. "Understanding the crude oil price Value at Risk: the Case of Azeri Light."* Available at SSRN 2402622 (2014).

Aloui, Chaker, and Samir Mabrouk. " *Value-at-risk estimations of energy commodities via long-memory, asymmetry and fat-tailed GARCH models."* Energy Policy 38.5 (2010): 2326-2339.

Angelidis, Timotheos, Alexandros Benos, and Stavros Degiannakis. *"The use of GARCH models in VaR estimation."* Statistical Methodology 1.1 (2004): 105-128.

Awartani, Basel, and Valentina Corradi. *"Predicting the volatility of the S&P-500 stock index via GARCH models: the role of asymmetries."* International Journal of Forecasting 21.1 (2005): 167-183.

Balke, Nathan S., Stephen PA Brown, and Mine K. Yücel. " *Oil Price Shocks and US Economic Activity."* Resources of The Future (2010): 10-37.

Barone-Adesi, Giovanni, Kostas Giannopoulos, and Les Vosper. " *VaR without correlations for portfolios of derivative securities."* Journal of Futures Markets 19.5 (1999): 583-602.

Bollerslev, Tim. "*Generalized autoregressive conditional heteroskedasticity."* Journal of econometrics 31.3 (1986): 307-327.

Boudoukh, Jacob, Matthew Richardson, and Robert Whitelaw. *"The best of both worlds."* Risk 11.5 (1998): 64-67.

Cabedo, David J., Moya, Ismael. "*Estimating oil price- Value at Risk using the historical simulation approach".* Energy Economics 25.3 (2003): 239-253.

Christoffersen, P. F. *"Backtesting, Prepared for the Encyclopedia of Quantitative Finance, R. Cont.*" (2008).

Costello, Alexandra, Ebenezer Asem, and Eldon Gardner. *"Comparison of historically simulated VaR: Evidence from oil prices."* Energy economics 30.5 (2008): 2154-2166.

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

Fallon, William. Calculating value-at-risk. Wharton School, University of Pennsylvania, 1996.

Fan, Ying, et al. "Estimating *"'Value at Risk'of crude oil price and its spillover effect using the GED-GARCH approach."* Energy Economics 30.6 (2008): 3156-3171.

Hansen, Peter R., and Asger Lunde*. "A forecast comparison of volatility models: does anything beat a GARCH (1, 1)?." Journal of applied econometrics* 20.7 (2005): 873-889.

Giot Pierre, and Sébastien Laurent*. "Market risk in commodity markets: a VaR approach."* Energy Economics 25.5 (2003): 435-457.

Giot, Pierre, and Sébastien Laurent. *"Modelling daily value-at-risk using realized volatility and ARCH type models."* journal of Empirical Finance 11.3 (2004): 379-398.

Glosten, Lawrence R., Ravi Jagannathan, and David E. Runkle. *"On the relation between the expected value and the volatility of the nominal excess return on stocks."* The journal of finance 48.5 (1993): 1779-1801.

Hamilton, James D. *Understanding crude oil prices*. No. w14492. National Bureau of Economic Research, 2008.

Hendricks, Darryll. *"Evaluation of value-at-risk models using historical data."* Federal Reserve Bank of New York Economic Policy Review 2.1 (1996): 39-69.

Hull, John, and Alan White. *"Incorporating volatility updating into the historical simulation method for value-at-risk."* Journal of Risk 1.1 (1998): 5-19.

Kilian, Lutz, and Cheolbeom Park. *"The impact of oil price shocks on the us stock market*.*" International Economic Review 50.4 (2009): 1267-1287.

Kilian, Lutz. *"Not all oil price shocks are alike: Disentangling demand and supply shocks in the crude oil market."* The American Economic Review (2009): 1053-1069

Longin, Francois M. *"From value at risk to stress testing: The extreme value approach."* Journal of Banking & Finance 24.7 (2000): 1097-1130.

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

Pérignon, Christophe, and Daniel R. Smith. *"Diversification and value-at-risk."* Journal of Banking & Finance 34.1 (2010): 55-66.

Politis, Dimitris N. *"A heavy-tailed distribution for ARCH residuals with application to volatility prediction."* (2004).

Pritsker, M*. "Evaluating Value-at-Risk Methodologies: Accuracy versus Computational Time in 'Model Risk: concepts, calibration and pricing'."* (2000).

Riskmetrics, T. M. "JP Morgan Technical Document." (1996).

Sadeghi, Mehdi, and Saeed Shavvalpour. "*Energy risk management and value at risk modeling."* Energy policy 34.18 (2006): 3367-3373.

Subbotin, M. Th. *"On the law of frequency of error."* Математический сборник 31.2 (1923): 296-301.

Sfetsos, A., and L. Kalyvas. *"Are conditional Value-at-Risk models justifiable?."* Applied Financial Economics Letters 3.2 (2007): 129-132.

Xiliang, Zhao, and Zhu Xi. *"Estimation of Value-at-Risk for Energy Commodities via CAViaR Model."* Cutting-Edge Research Topics on Multiple Criteria Decision Making. Springer Berlin Heidelberg, 2009. 429-437.

Zakoian, Jean-Michel. *"Threshold heteroskedastic models."* Journal of Economic Dynamics and control 18.5 (1994): 931-955.

Regulatory articles

Supervisory Framework for the use of "Backtesting" in conjunction with The Internal Models Approach to Market Risk Capital Requirements Basle Committee on Banking Supervision 1996.

Electronic resources

[1] Investopia "Standard & Poor's 500 index– S&P 500". Retrieved from <http://www.investopedia.com/terms/s/sp500.asp > Accessed: 2014-08-13

Buckley, Neil (2012, 20 june)" *Economy: Oil dependency remains a fundamental weakness".* Retrieved from < http://www.ft.com/cms/s/0/438712b2-b497-11e1-bb2e-00144feabdc0.html#axzz381n6dNp6 > Accessed: 2014-08-17

Brown, Stan (2012-12-27) "Measures of Shape: Skewness and Kurtosis" retrieved from http://www.tc3.edu/instruct/sbrown/stat/shape.htm Accessed: (2014-08-10)

Normal Distribution. Retrieved from <http://en.wikipedia.org/wiki/File:Normal_Distribution_PDF.svg> Accessed: 2014-08-01

Student's t-DistributionRetrieved from <http://en.wikipedia.org/wiki/File:T_distribution_1df_enhanced.svg> Accessed: 2014-08-01

Generalized Error Distribution Retrieved from <http://en.wikipedia.org/wiki/File:Generalized_normal_densities.svg> Accessed: 2014-08-01

Lecture notes

Nilsson, Birger(2014) "*Value-at risk*" lecture notes in. *NEKN83/TEK180 spring 2014. Lund University*

Nilsson, Birger(2014) "*Extreme value theory for VaR estimation*" lecture notes in. *NEKN83/TEK180 spring 2014. Lund University*

8 Appendix

8.1 Distribution density functions[123]

GED distributions

$$f(\varepsilon) = \frac{k^* \exp(\frac{1}{2}\ [\varepsilon/\lambda]^k}{\lambda^* 2^{[k+1/k]} \Gamma(1/k)} (0 \le k \le \infty)$$

$$\lambda = \left[\frac{2^{(-1/k)} \Gamma(1/k)}{\Gamma(3/k)}\right]^{1/2}$$

Note, in the text the shape parameter *k*, .is denoted *β*.

Student's t-distribution

Density function for the t- distribution is as follows.[124]

$$f^*(x) = \frac{\Gamma[v + 1/2]}{\sigma^* \sqrt{v\pi}\, \Gamma(v/2)} \left[1 + \frac{1}{v} \left(\frac{x - \mu}{\sigma^*}\right)^2\right]^{-(v+1)/2} \quad for\ x\ \in (-\infty, \infty)$$

Normal distribution

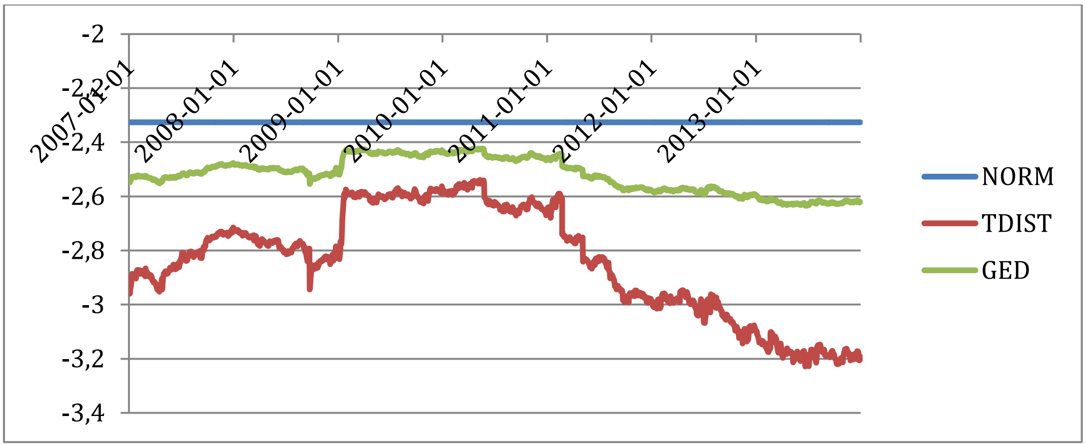
Density function for the normal distribution is as follows.[125]

$$f(y) = \frac{1}{\sqrt{2\pi\sigma^2}} exp \left\{\frac{-1}{2} \frac{(y - \mu)^2}{\sigma^2}\right\}$$

[123] Hamilton (1994).
[124] Hamilton (1994)
[125] Verbeek (2012)

8.2 Critical values GARCH(1,1)

Illustration of how the critical values of the VaR changes with the different distributions over time. This example is changes in the critical value when estimating VaR on WTI with GARCH(1,1) and a constant in the mean equation.



8.3 Stationary tests.

ADF test for Stationarity WTI

Null Hypothesis: R has a unit root
Exogenous: Constant
Lag Length: 0 (Automatic - based on SIC, maxlag=27)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-56.14385	0.0001
Test critical values:	1% level	-3.432385	
	5% level	-2.862325	
	10% level	-2.567232	

*MacKinnon (1996) one-sided p-values.

ADF test for Stationarity S&P

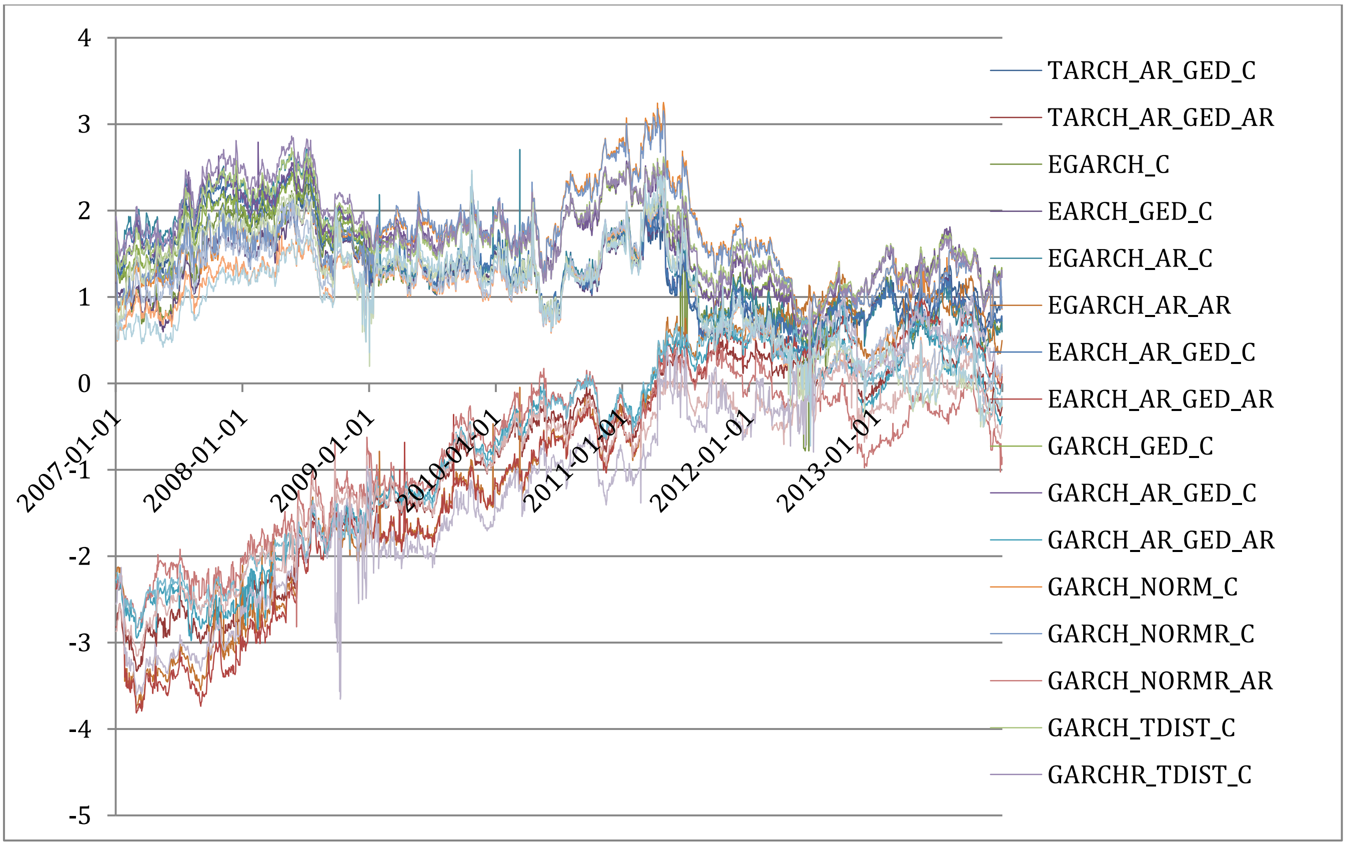
Null Hypothesis: R has a unit root
Exogenous: Constant
Lag Length: 1 (Automatic - based on SIC, maxlag=27)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-42.73203	0.0000
Test critical values:	1% level	-3.432385	
	5% level	-2.862325	
	10% level	-2.567232	

*MacKinnon (1996) one-sided p-values.

Since we reject the null hypothesis of unit root we conclude that the converted series is stationary.

8.4 t-value evolution of the mean equation parameters for GARCH estimations WTI



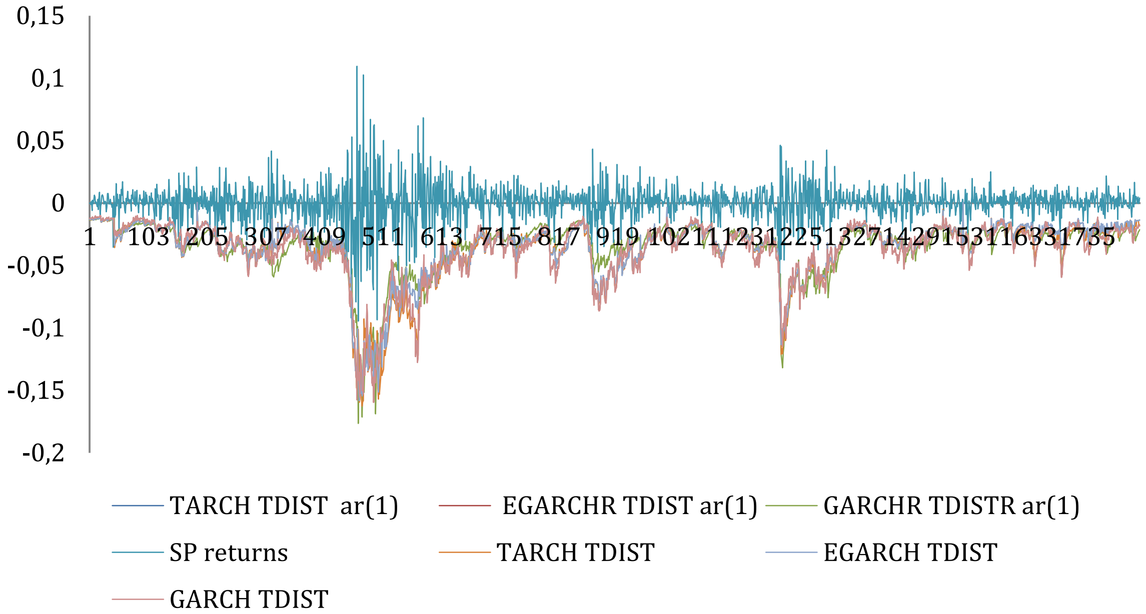
8.4 Parametric results for S&P 500 and GARCH (1, 1) model

	GARCH n-dist.	GARCH n-dist. AR(1)	TGARCH n-dist.	TGARCH n-dist. AR(1)	EGARCH n-dist.	EGARCH n-dist. AR(1)
Violations	49	49	54	55	55	56
Independence	2.703	2.703	3.292	3.417	3.417	3.544
Uncond. Convergence	35.817	35.817	46.373	48.601	48.601	50.867
Cond. Convergence	38.519	38.519	49.664	52.018	52.018	54.411
Basel	16	16	14	15	17	17

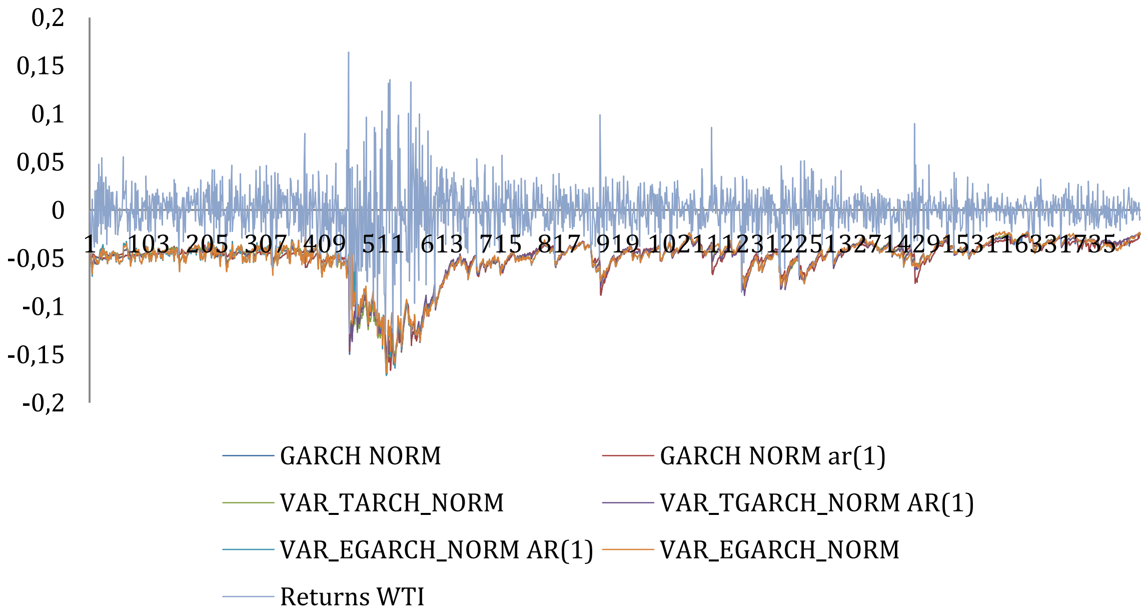
	GARCH t-dist.	GARCH t-dist. AR(1)	TARCH t-dist.	TGARCH t-dist. AR(1)	EGARCH t-dist.	EGARCH t-dist. AR(1)
Violations	20	10	21	16	16	20
Independence	0.443	0.110	0.489	0.283	0.283	0.443
Uncond. Convergence	0.164	4.506	0.399	0.292	0.292	0.164
Cond. Convergence	0.607	4.616	0.888	0.575	0.575	0.607
Basel	12	6	8	8	8	12

	GARCH GED	GARCH GED AR(1)	TGARCH GED	TGARCH GED AR(1)	EGARCH GED	EGARCH GED AR(1)
Violations	29	28	33	35	37	38
Independence	0.936	0.872	1.215	1.368	1.531	1.615
Uncond. Convergence	5.426	4.523	9.715	12.239	14.995	16.457
Cond. Convergence	6.362	5.395	10.930	13.607	16.526	18.072
Basel	11	11	12	13	14	14

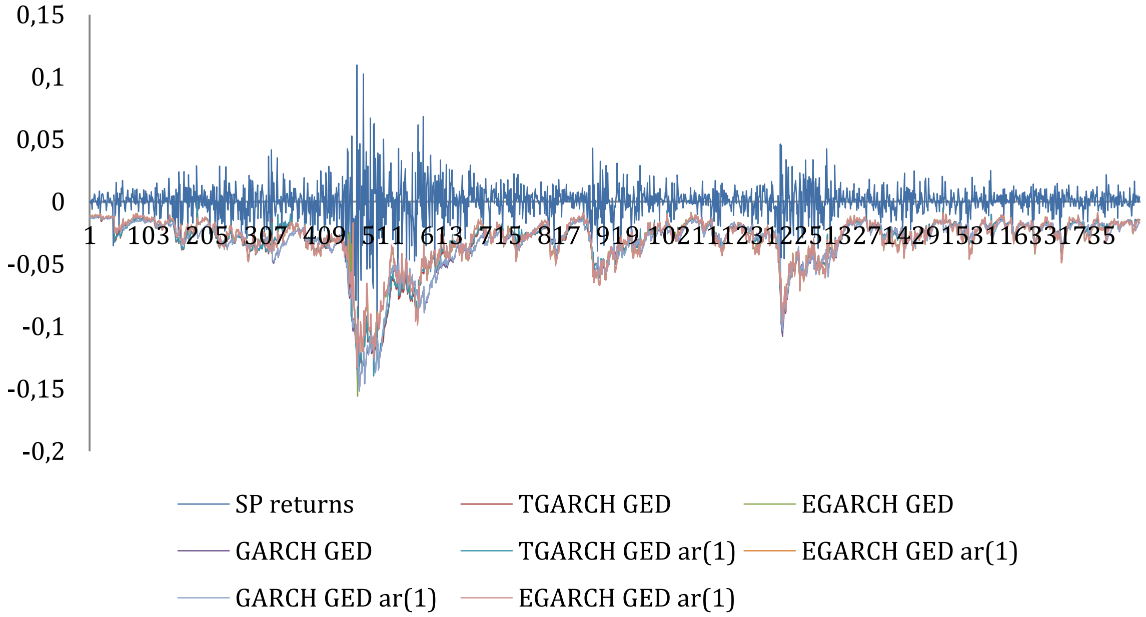
8.5 Value-at-Risk graphs



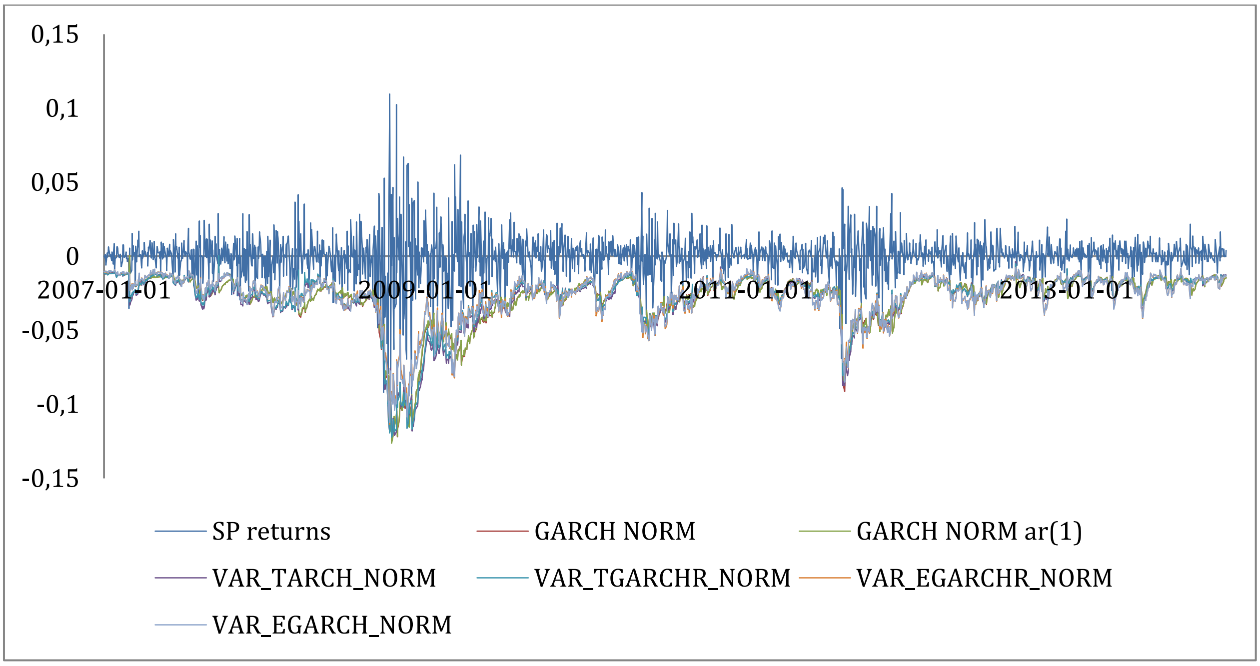
Figur 1 Var estimates for S&P 500 with t-distribution



Figur 2 Var estimates for WTI with normal distribution



Figur 3 Var estimates for S&P 500 with GED



Figur 4 Var estimates for S&P 500 with normal distribution