Charles University in Prague

Faculty of Social Sciences

Institute of Economic Studies

MASTER'S THESIS

Neural network models for conditional quantiles of financial returns and volatility

Author: **Bc. Marek Hauzr**

Supervisor: **PhDr. Jozef Baruník, Ph.D.**

Academic Year: **2015/2016**

Declaration of Authorship

The author hereby declares that he compiled this thesis independently, using only the listed resources and literature, and the thesis has not been used to obtain a different or the same degree.

The author grants to Charles University permission to reproduce and to distribute copies of this thesis document in whole or in part.

Prague, July 27, 2016

Signature

Acknowledgments

I would like to thank to my supervisor PhDr. Jozef Baruník, Ph.D. for the topic suggestion, for all of the answers to the questions I had and all the comments he provided, for his willingness to help me and for providing me with the data.

Abstract

This thesis investigates forecasting performance of Quantile Regression Neural Networks in forecasting multiperiod quantiles of realized volatility and quantiles of returns. It relies on model-free measures of realized variance and its components (realized variance, median realized variance, integrated variance, jump variation and positive and negative semivariances). The data used are S&P 500 futures and WTI Crude Oil futures contracts. Resulting models of returns and volatility have good absolute performance and relative performance in comparison to the linear quantile regression models. In the case of in-sample the models estimated by Quantile Regression Neural Networks provide better estimates than linear quantile regression models and in the case of out-of-sample they are equally good.

JEL Classification C14, C45, C53, G17, G32
Keywords conditional quantiles, quantile regression neural networks, realized measures of volatility, value-at-risk

Author's e-mail marek.hauzr@gmail.com
Supervisor's e-mail barunik@fsv.cuni.cz

Abstrakt

V této práci se zkoumá chování kvantilové regrese neuronových sítí v odhadování kvantilů realizované volatility a kvantilů výnosů s výhledem více kroků. Závisí na realizované varianci a jejích komponentech (realizovaná variance, mediánová realizovaná variance, integrovaná variance, skoková variance a positivní a negativní semivariance). Použitá data jsou S&P 500 futures a WTI Crude Oil futures. Výsledné modely výnosů a volatility mají dobré absolutní chování a relativní chování v porovnání s modely ohodnocenými lineární kvantilovou regresí. V případě in-sample má lepší chování kvantilová regrese neuronových sítí a v případě out-of-sample mají chování stejně dobré.

Klasifikace JEL	C14, C45, C53, G17, G32
Klíčová slova	podmíněné kvantily, kvantilová regrese neuronových sítí, neparametrické odhady realizované volatility, VaR

E-mail autora	`marek.hauzr@gmail.com`
E-mail vedoucího práce	`barunik@fsv.cuni.cz`

Contents

List of Tables **ix**

List of Figures **xi**

Thesis Proposal **xii**

1 Introduction **1**

2 Previous literature **3**

2.1 Financial returns and volatility . . . 3
2.2 Žikeš & Baruník (2016) . . . 4
2.3 Taylor (2000) . . . 5
2.4 Other publications with QRNN . . . 6

3 About Neural Networks **7**

3.1 Single Neuron . . . 7
3.1.1 Linear Neuron . . . 7
3.1.2 Binary Threshold Neurons . . . 8
3.1.3 Rectified Linear Neuron . . . 9
3.1.4 Sigmoid Neurons . . . 9
3.1.5 Stochastic Binary Neurons . . . 10
3.2 Neural Network . . . 10
3.2.1 Neural network with one neuron . . . 11
3.2.2 Neural network with more neurons . . . 11
3.3 Learning . . . 12
3.4 Similarity between simple neural networks and classic estimators 12
3.4.1 Similarity between simple neural network and OLS . . . 12
3.4.2 Similarity between simple neural network and MLE . . . 13

4 Methodology 14
4.1 Quantile Regression 14
4.1.1 Quantiles 14
4.1.2 Quantile Regression 15
4.2 Value-at-Risk 15
4.3 Realized measures and forecasting Returns and Volatility 16
4.3.1 Realized Variance 17
4.4 Models for returns 18
4.5 Models for realized volatility 19
4.6 QRNN 20
4.6.1 Taylor (2000) 20
4.6.2 QRNN - implementation in R 21
4.7 How to compare results 22
4.7.1 Absolute performance of one-step-ahead forecasts 22
4.7.2 Absolute performance of multi-step-ahead forecasts . . . 24
4.7.3 Relative performance 25
4.8 Hypothesis 1 26
4.9 Hypothesis 2 26
4.10 Hypothesis 3 26

5 Data 28
5.1 WTI Crude Oil futures 29
5.2 S&P 500 30

6 Results 36
6.1 LQR and HARQ models of Žikeš & Baruník (2016) 37
6.1.1 LQR 37
6.1.2 HARQ 37
6.2 Absolute performance 40
6.2.1 LQR - Absolute performance 40
6.2.2 HARQ - Absolute performance 44
6.3 Relative performance 48
6.3.1 LQR - Relative performance 48
6.3.2 HARQ - Relative performance 52
6.4 Skewness 58

7 Conclusion 61

Bibliography **65**

List of Tables

5.1 Summary statistics for WTI crude oil futures 30
5.2 Summary statistics for S&P 500 futures 32

6.1 Conditional quantile models (LQR) of Žikeš & Baruník (2016) for WTI Crude Oil futures return quantiles. 38
6.2 Conditional quantile models (LQR) of Žikeš & Baruník (2016) for S&P 500 futures return quantiles. 39
6.3 Conditional quantile models (HARQ) of Žikeš & Baruník (2016) for WTI Crude Oil futures realized volatility quantiles. 41
6.4 Conditional quantile models (HARQ) of Žikeš & Baruník (2016) for S&P 500 futures realized volatility quantiles. 42
6.5 Absolute performance of in-sample and out-of-sample forecasts for WTI Crude Oil futures return quantiles. 43
6.6 Absolute performance of models on in-sample and out-of-sample forecasts for S&P 500 futures return quantiles. 44
6.7 Absolute performance of models on in-sample and out-of-sample forecasts for WTI Crude Oil futures realized volatility quantiles. 45
6.8 Absolute performance of models on in-sample and out-of-sample forecasts for S&P 500 futures realized volatility quantiles. 46
6.9 Relative performance of models for in-sample forecasts of WTI Crude Oil futures return quantiles. 49
6.10 Relative performance of models for out-of-sample forecasts of WTI Crude Oil futures return quantiles. 50
6.11 Relative performance of models for in-sample forecasts of S&P 500 futures return quantiles. 51
6.12 Relative performance of models for out-of-sample forecasts of S&P 500 futures return quantiles. 53

6.13 Relative performance of models for in-sample forecasts WTI Crude Oil futures realized volatility. 54
6.14 Relative performance of models for out-of-sample forecasts of WTI Crude Oil futures realized volatility. 55
6.15 Relative performance of models for in-sample forecasts of S&P 500 futures realized volatility. 56
6.16 Relative performance of models for out-of-sample forecasts of S&P 500 futures realized volatility. 57
6.17 Skewness of WTI Crude Oil futures forecasted distributions of returns and volatility 59
6.18 Skewness of S&P 500 futures forecasted distributions of returns and volatility 60

List of Figures

3.1 Artificial neuron 8
3.2 Biological neuron 8

4.1 Comparison of quantile forecast 23

5.1 WTI Crude Oil futures: time series of daily returns, realized volatility, jump variation and median realized volatility 31
5.2 WTI Crude Oil futures: ACF and PACF of returns and realized volatility 32
5.3 S&P 500 futures: time series of daily returns, realized volatility, jump variation and median realized volatility 34
5.4 S&P 500 futures: ACF and PACF of returns and realized volatility 35

Master's Thesis Proposal

Author	Bc. Marek Hauzr
Supervisor	PhDr. Jozef Baruník, Ph.D.
Proposed topic	Neural network models for conditional quantiles of financial returns and volatility

Motivation Forecasting financial volatility and financial returns is important for making rational decisions. Having the whole distribution of future changes in prices and volatility would be ideal, but having few quantiles may be sufficient. For example modeling 5% VaR, which is a threshold loss value, where there is 5% probability of a loss of given amount over some time period.

Linear models estimation of quantiles may be improved by using neural networks. The reason for that is that neural networks are used for estimation of non-linear models where the linear are a subset of more general non-linear models and they still can be estimated by using the neural networks.

In forecasting the use of neural networks provides generally better forecasts than linear models, for example Donaldson and Kamstra (1996) found that the use of artificial neural networks on the out of sample basis produces superior forecasts in comparison to linear procedures. Atiya (2001) found that neural networks in addition to classic indicators improve the prediction accuracy of corporate bankruptcies. QRNNs were used for example in civil engineering for estimating the distribution of concrete strength and Yeh (2014) found that QRNN can in this case build accurate quantile models. Xiong et al. (2013) shows that model based on neural networks is better in predicting accuracy in comparison to three commonly used prediction strategies in the case of WTI crude oil. Panella et al. (2012) found that their model based on neural networks may provide improvements in comparison to other well known models. Another paper that used neural networks successfully to predict oil prices is Jammazi and Aloui (2012).

Hypotheses

Hypothesis #1: Quantile regression neural networks provides better quantile estimates than standard linear quantile regression.

Hypothesis #2: Volatility is important for predicting quantiles of returns.

Hypothesis #3: Expected densities of returns are skewed.

Methodology To model non-linearity in quantiles, the quantile regression neural network approach (Taylor, 2000) will be used. The underlying instruments for comparing linear and non-linear models are S& P 500 and WTI Crude Oil futures and they are based on Žikeš & Baruník (2016). To compare relative performance Žikeš & Baruník (2016) suggest to follow Clements et al. (2008) and for absolute performance evaluation to use CAViaR test (Berkowitz, Christoffersen & Pelletier 2011).

Comparison of model including volatility and returns with a model including only returns will be used for determining the importance of volatility in predicting quantiles of returns. Same ways of relative and absolute performance as mentioned above will be used.

Distribution of expected returns may not be Gaussian, for example if the previous return is negative and too big then in some cases we can expect a correction in the market, we can expect that the return will fall, but with a limit to the fall which can result in skewed expected density of return. This will be tested for example by the comparison of mean and median.

Expected Contribution The thesis will contribute to the current academic literature by comparing the linear estimation of quantiles with potentially better approach of non-linear estimation. It should show that the usage of quantile regression neural network approach provides better forecasts for quantiles of return than the simple quantile regression.

Practical results are wide. For example option pricing (Hansen, 1994). Under the condition of having an estimated distribution of return and volatility we can calculate the expected return and expected volatility of assets, futures contracts and so on and in the case of this thesis S& P 500 and WTI Crude Oil futures. It can be also used to estimate the expected return and volatility in portfolio investments. Another contribution of applying quantile regression is it's ability to serve as a risk management tool for investors, since they can predict the risk of their losses (Žikeš & Baruník (2016)).

Outline

1. Introduction: Introduction to estimating quantiles of volatility and returns. How different authors use different approaches.
2. Methodology: Description of quantile regression and how it works with neural network.
3. Data description: Description of the dataset that is used.
4. Results: Discussion of results and comparison to different methods.
5. Summary.

Core bibliography

Taylor, J. W., 2000. A quantile regression neural network approach to estimating the conditional density of multiperiod returns. Journal of Forecasting. 19(4), 299-311.

Žikeš, F., Baruník, J., 2016. Semi-parametric Conditional Quantile Models for Financial Returns and Realized Volatility. Journal of Financial Econometrics. 14 (1): 185-226

Clements, M. P., Galväo, A.B., Kim, J. H., 2008. Quantile forecasts of daily exchange rate returns from forecasts of realized volatility. Journal of Empirical Finance. 15(4), 729-750.

Berkowitz, J., Christoffersen, P.,Pelletier, D., 2011. Evaluating value-at-risk models with desk-level data, Management Science. 52(12), 2213–2227.

Hansen, B.E., 1994. Autoregressive Conditional Density Estimation, International Economic Review. 35, 705-730.

Donaldson, R.G., Kamstra, M., 1996. Forecast combining with neural networks, Journal of Forecasting. 15(1), 49-61.

Atiya, A. F., 2001. Bankruptcy prediction for credit risk using neural networks: A survey and new results. IEEE Transactions on Neural Networks. 12 (4), 929-935.

Yeh, I.-C., 2014. Estimating distribution of concrete strength using quantile regression neural networks. Applied Mechanics and Materials 584-586, 1017-1025.

Xiong, T., Bao, Y., Hu, Z., 2013. Beyond one-step-ahead forecasting: Evaluation of alternative multi-step-ahead forecasting models for crude oil prices. Energy Economics. 40, 405–415.

Panella, M., Barcellona, F., D'Ecclesia, R. L., 2012. Forecasting energy commodity prices using neural networks. Advances in Decision Sciences 2012

Jammazi, R., Aloui, C., 2012. Crude oil price forecasting: Experimental evidence from wavelet decomposition and neural network modeling. Energy Economics. 34 (3), 828–841.

Author Supervisor

Chapter 1

Introduction

Understanding the volatility is important in the uncertainty of financial markets. Properly modeling its expected distribution together with the expected distribution of financial returns is important for managing risk. Some of the quantiles, usually the tails, of financial returns are already used for risk management in the Value-at-Risk models. Where Value-at-Risk can be computed directly through estimation of the corresponding quantile of returns or through assuming a distribution and forecasting volatility. It is important to forecast the distribution of volatility and returns not only one-step-ahead but also multi-step-ahead because not every investment position can be exited in one day. There are different measures of volatility, one of the few are option implied volatility and realized volatility. Here the realized volatility its components are studied.

Using non-linear model, specifically artificial neural network, for forecasting volatility and returns is not a new idea. There are academic papers on this topic and it is used in practice. What was done very little, in just few papers, is forecasting the distribution of returns and volatility using the artificial neural networks even though the idea is more than one and a half decades old. The non-linear models can improve the forecasting of the distribution and with the improved forecast we might be able to show that the distributions are skewed, where most of the models assume normal distribution or student's t-distribution which are symmetrical.

The thesis is structured as follows: Chapter 2 gives the introduction to the area of modeling quantiles of financial returns and volatility and to the Quantile Regression Neural Network. Chapter 3 covers the theoretical introduction to the area of artificial neural networks. Chapter 4 explains the methodology of realized measures, estimation techniques and performance evaluation. Chapter 5 describes the data. In Chapter 6 the results are reported. The final chapter, Chapter 7, concludes the thesis.

Chapter 2

Previous literature

2.1 Financial returns and volatility

Many papers were published on the topic of financial returns and volatility. In this thesis the focus will be on forecasting certain quantiles of returns which is mainly forecasting the Value-at-Risk and forecasting quantiles of volatility (realized volatility).

Value-at-Risk is complicated to forecast and it is complicated to evaluate the validity of forecasts due to the fact that it is not observable. What is usually done is that we evaluate the violaiton of the forecast and its behavior. The Value-at-Risk is violated when the actual loss is higher than the predicted loss. When the process of violations satisfies the martingale difference hypothesis then the model is considered valid (Dumitrescu *et al.* 2012). Value-at-Risk is probably the most used measure of portfolio risk in major commercial banks (Berkowitz *et al.* 2011). The popularity of this concept among financial practitioners is based on the simplicity of this concept (Engle & Manganelli 2004).

Dumitrescu *et al.* (2012) emphasize three main problems with evaluating Value-at-Risk forecasts where the first one is the most important one. The problem is related to the power of the backtesting test

$$Power = Pr(Reject\ H_0 |\ H_1\ is\ True)$$

where $H_0 :\ model\ is\ valid,\ H_1 :\ otherwise$ especially in small samples. These tests usually do not reject the model and its validity as often as it should, hence they have low power.

The test with one of the best finite-sample size and power properties is the CAViaR test of the Engle & Manganelli (2004), this was suggested by Berkowitz *et al.* (2011) in their paper where they assesed many different tests. The test is called DQ (Dynamic Quantile) test and is used in this thesis. Another test that arises from DQ test is suggested by Dumitrescu *et al.* (2012) and is based on Dynamic Binary (DB) regression model.

Measuring variation and prediction of financial returns is important for the pricing of securities. Volatility is an important input for option pricing and portfolio allocation (Andersen & Bollerslev 1998). They are also needed for performance evaluation, managerial decision making and understanding their distribution is important for the expectations of extreme shifts in portfolio (Andersen *et al.* 2003).

For measuring volatility the realized measures are used and the abstract concept of volatility is quantified through realized volatility. Papers such as Berkowitz *et al.* (2011), Andersen *et al.* (2003) and Comte & Renault (1998) are important stones in realized measures and specifically volatility. Barndorff-Nielsen & Shephard (2003) suggested a different way for measuring volatility and it is power variation, where instead of summing intraday squared returns to get the realized volatility, the intraday absolute returns are summed.

2.2 Žikeš & Baruník (2016)

Žikeš & Baruník (2016) found that for S&P 500 futures prices, both the realized and implied volatility possess significant predictive power when predicting quantiles of future returns. When they decomposed the realized volatility into realized downside and upside semivariance they found that the downside semivariance drives both the left and right tale quantiles and the upside semivariance does not have such influence. This means that the negative intraday returns possess more information than the positive intraday returns. In case of jumps they found that they play small role in forecasting quantiles and that across considered models they are not consistently significant. The data used are transaction prices from the front contract traded on the Chicago Mercantile Exchange (CME) between the main trading hours (9:30 - 16:00 EST), the data are high-frequency data of S&P 500 futures contract obtained from Tick Data ranging between January 1996 and December 2012 .

WTI Crude Oil futures are less well behaved (Žikeš & Baruník 2016) in sense that there is higher volatility and volatility of realized volatility than in the case of S&P 500. Žikeš & Baruník (2016) state that this provides an oportunity to test the methodology. They use intraday Tick Data with focus on the front contract traded on the New York Mercantile Exchange. The data range from September 2001 until December 2012 for each day between the main trading hours (9:00 - 15:00 EST). They found that their quantile models perform equally well in delivering quantile forecasts.

To compare relative performance Žikeš & Baruník (2016) use benchmarks such as CAViaR (Engle & Manganelli 2004) and ARFIMA-based lognormal-normal mixture (Andersen *et al.* 2003). To compare forecast accuracy they use tick loss function (Giacomini & Komunjer 2005). They found that no model outperforms other models across assets or quantiles.

To model the quantiles of volatility they use heterogenous autoregressive quantile model (HARQ) which is heterogenous autoregressive model (HAR) (Corsi 2009) estimated as a quantile regression. They specified and report three models, one with independent variables being previous day realized volatility, average realized volatility over the last 5 days and last 22 days. In the second model they split the last day volatility into positive and negative volatility based on positive and negative semivariances and they add the Volatility Index (VIX) calculated by the Chicago board of Exchange (CBOE) in case of S&P 500 and Crude Oil Volatility Index (OVX) introduced by CBOE which applies the same methodology as VIX. In the last (third) specification they use volatilities based on integrated variance and jump variation (square roots of them), VIX and 5 and 22 last days average of square root of jump variation.

For quantiles of returns (LQR) they use similar specifications. In the first specification the independent variable is only one and it is realized volatility. In the second case the independent variables are volatilites (square roots) based on integrated variance, jump variation and option implied volatility. In the last specification they use volatilities based on positive and negative semivariances with option implied volatility.

2.3 Taylor (2000)

Taylor (2000) uses artificial neural networks to estimate nonlinear quantile models. The cost function of the neural network corresponds to the cost function of linear quantile regression with added parameters to penalize complexity (overfitting) of the neural network.

The historical returns are used for multiperiod returns estimation. Suggested Quantile Regression Neural Network is compared to the GARCH(1,1) with empirical and gaussion distribution on Deutsche mark and Japanese yen, both quoted against U.S. dollar. The datasample starts by July 4^{th}, 1988 and ends by July 5^{th}, 1996. In case of Deutsche mark the GARCH(1,1) estimator of volatility with empirical distribution performs well for the $5^{th}, 25^{th}, 75^{th}$ and 95^{th} quantiles. The quantile regression neural network matches the first method for 5^{th} and 95^{th} quantile and is better for 99^{th} quantile. The GARCH(1,1) with gaussian distribution performs best for the 1^{st} quantile. The author states that interestingly for the 25^{th} quantile all three methods are severly underestimating it. In case of yen the quantile regression neural network is better than the other two in the case of 4 out 6 quantiles. GARCH(1,1) with empirical distribution performs best for 75^{th} quantile and GARCH(1,1) with gaussian distribution for 99^{th} quantile.

For comparison of relative performance of the methods Taylor (2000) calculated χ^2 goodness of fit statistics (Hull *et al.* 1998). He found that GARCH(1,1) with Gaussion distribution is best in 4 out of the 7 holding periods. The quantile regression neural network performs best in 2 out of the 7 cases and the GARCH(1,1) with empirical distribution performs worst. Another measure of relative performance was used by summing the number of times the method outperformed the other two. GARCH(1,1) with empirical distribution performed best in case of Deutsche mark and quantile regression neural network was the best in the case of yen and in total the quantile regression neural network performed best.

2.4 Other publications with QRNN

Quantile Regression Neural Networks were used on different types of data, ranging from the strength of concrete (Yeh 2014) to enterprise valuation (Liu & Yeh 2016) and forecasting of gas consumption in China (Zhu *et al.* 2014).

Yeh (2014) studies the compressive strength of high performance concrete and its distribution. 1030 observations were used for evaluation of Quantile Regression Neural Network and in those observations several variables were included such as amounts of cement, blast furnace slag, fly ash and so on. Several conclusions can be taken out of this study: Quantile Regression Neural Networks can build accurate models of the distribution of compressive strength of high performance concrete, variance of the error is not constant across the observations which implies heteroskedasticity in the predictions, normal distribution does not fit the empirical distribution as well as the logarithmic distribution.

Chapter 3

About Neural Networks

3.1 Single Neuron

Single neuron is supposed to mimic the biological neuron (see Figure 3.2). It consists of several parts. It starts with inputs, there can be one or more of them and they are supposed to mimic the dendrites in biological neuron (Figure 3.1). Next is a summing function that is linear and takes the inputs and creates a weighted sum of inputs with bias

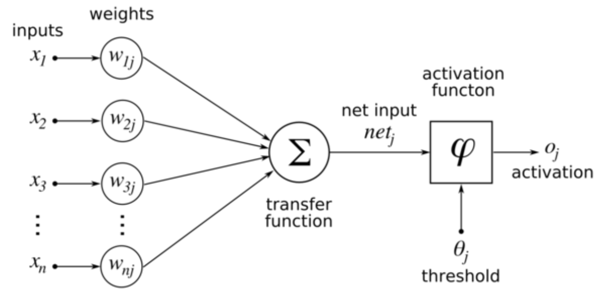
$$y = b + \sum(w_i * x_i) \tag{3.1}$$

where b is bias, x_i is i^{th} input, w_i is weight of the i^{th} input and y is the weighted sum. The sum is then passed through a function that is called activation function ($\varphi(y)$). Activation function is usually non-linear, but for example linear neuron has linear acivation function. Activation function is what differentiates neurons, it is usually non-linear, increasing, bounded and differentiable to make the learning easier and the optimisation faster, but it is not necessary to meet all of the criteria. This activation function mimics the axon in biological neuron. The last is the output that is supposed to mimic the terminal buttons of biological neuron.

3.1.1 Linear Neuron

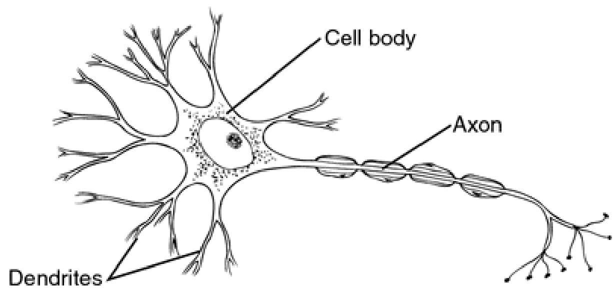
Linear neuron is a neuron where the output can be described by a linear function of input or inputs. In other words the activation function is linear and it can even be removed or equalled to identity (value of input is equal to the input). If it is not the case (the activation function is not identity) and the

Figure 3.1: Artificial neuron



Source: Figure taken from Chrislb (2005)

Figure 3.2: Biological neuron



Source: Figure taken from Dorland (2007)

activation function is still linear, lets say $\varphi(y) = \alpha + \beta * y$ where y is the output of the summing function (Equation 3.1) in the neuron. Then there exists a linear neuron with activation function equal to identity that has the same neuron output given the input. This neuron can be created by transforming the summing function (Subsection 3.1.1). The output of a neuron is equal to a function of inputs

$$output = \varphi(y) = \varphi(b + \sum_{i=1}^{n} w_i * x_i) = \alpha + \beta * (b + \sum_{i=1}^{n} w_i * x_i) = (\alpha + \beta * b) + \sum_{i=1}^{n} (\beta * w_i) * x_i$$

where n is the number of inputs, the $\alpha + \beta * b$ is the new bias in the summing function and $\beta * w_i$ is the new weight of i^{th} input.

3.1.2 Binary Threshold Neurons

The activation function of this neuron is equal to the Heaviside step function. Which is equal to 0 if the argument is negative and 1 otherwise.

$$H(n) = \begin{cases} 0 & \text{if } n \text{ is negative} \\ 1 & \text{otherwise} \end{cases}$$

The activation function can be specified as $\varphi(y) = H(y - \theta)$ where y is the result of the summing function of the neuron and θ is a threshold. The threshold is usually equal to 0, because the summing function can simly increase the bias (b in Equation 3.1) by the value of θ. It means that under some conditions the neuron is inactive and otherwise it sends signal that it is active.

3.1.3 Rectified Linear Neuron

This type of neuron is a combinaion of the previous two. It combines a linear activation function with a threshold value. When the argument of the activation function is negative then the function is equal to 0 and if not then it is equal to the argument. Where the 0 works as a threshold as in the previous type of neuron. To get to a threshold different from 0 we would have to change the bias in the summing function which is determined by the evaluation or instead of using $\varphi(y)$ use $\varphi(y + \delta)$

$$\varphi(y) = \begin{cases} 0 & \text{if } y \text{ is negative} \\ y & \text{otherwise} \end{cases}$$

We can also specify the activation function as $\varphi(y) = y * H(y)$ where the y is the sum from summing function and $H(y)$ is the Heaviside step function.

3.1.4 Sigmoid Neurons

Sigmoid neuron is a neuron whose activation function satisfies condition of non-linearity, it is increasing, bounded and differentiable. Sigmoid function typically refers to the special case of logistic function

$$\varphi(y) = \frac{1}{1 + e^{-t}} \tag{3.2}$$

where this function is differentiable, increasing, non-linear and bounded by 0 and 1 on the whole set of real numbers. Another possibility for sigmoid function can be hyperbolic tangent function

$$tanh(x) = \frac{e^x + e^{-x}}{e^x - e^{-x}} \tag{3.3}$$

which is also differentiable, non-linear, increasing and it is bounded by -1 and 1.

For the purposes of estimation in this thesis the second case will be used.

3.1.5 Stochastic Binary Neurons

Stochastic Binary Neuron is similar to the Sigmoid Neuron with logistic activation function. The difference between them is that the output of the logistic activation function is the probability of firing. The output of this neuron can be 0 (not fire) and 1 (fire). For example when the output of logistic activation function is equal to 0.74 then there is probability of 74% that the output of the neuron will be 1 and 26% that it will be 0.

3.2 Neural Network

Neural networks are models that are inspired by brain, its neurons and connections between the neurons. As neurons in models are similar to biological neurons the neural network models are similar to biological neural networks especially in the amount of connections between the neurons.

Neural networks can be considered (especially in the case of this thesis) a generalization of standard linear models. They are able to model non-linearities and they are especially used for modeling dependencies where there is no expectation on the type of dependency (the dependency is expected, but the form is unknown). They are used for modeling over big datasets such as recognition of what is on a picture, face recognition. They can also be used for recognition of speech and handwriting, for self driving cars and even for creating poetry.

Each neural network is typically divided into several layers of neurons. Into input layer, hidden layers and output layer. Input layer consists of neurons that send the data to the hidden layers. There can be only one, or many more hidden layers and they might have the same structure or they might not. These layers transform the input. Output layer gives the results of the transformation of the input.

For estimating the weights and biases in neurons in a neural network we use learning algorithms that are typically based on some form of gradient based non-linear optimization.

Another important part of the neural networks is the cost function which is minimized and through this minimization and learning we get the weights and biases in neurons.

3.2.1 Neural network with one neuron

Using neural networks with only one hidden neuron does not make much sense. There are better methods for estimation a model in econometrics than this. The non-linear activation function in neuron could be worked around by transforming the dependent variable of a linear model and the cost function only sets the apropriate econometric method.

3.2.2 Neural network with more neurons

Standard neural network contains three types of layers:

- Input layer (neurons in this layer are not connected between them selves but are only connected to all the neurons in the first hidden layer.
- Hidden layer (neurons in each layer are not connected between them selves but each neuron is connected to every neuron in previous (either input or hidden layer) and next layer (either hidden or output layer), the previous layer sends information to this neuron, the information gets transformed and is send to every neuron in the next layer)
- Output layer (neurons in this layer are not connected between them selves, but each neuron is connected to every neuron in the previous hidden layer and usually there is only one neuron in this layer)

Input Layer

Input neurons are used for transfering the input to the neural network and they are used for standardising the input. Easy way to think about them is to compare them to explanatory variables in a regression but the effect that single variable has on the output is more complex than in a simple regression.

Hidden Layers

These layers have a variable number of neurons based on the data, training process,... . Each neuron in each layer has a number of inputs that corresponds to the number of neurons in the previous layer (input layer or hidden layer) and its output works as an input for each neuron in the following layer (hidden or output layer).

Output Layer

The output layer can consist of many neurons, but usually there is only one. It adds outputs from the last hidden layer with assigned weights and then transforms it by the activation function.

3.3 Learning

There are three main learning ideas that are based on the problem needs, on the problem that is supposed to be solved.

- Supervised learning - for a given input vector and output, the supervised learning tries to find function that best fits the data, that minimizes the cost function.
- Reinforcement learning - data are usually given by the interaction with the problem that is supposed to be modeled, in each step the model performs action and the enviroment corresponding reaction and a cost. The goal is to minimize the sum of individual costs, the goal is to learn to select the best action for minimizing the cost.
- Unsupervised learning - it differs from supervised learning by not having an input, output and cost function, but only having input and cost function to be minimized. It is used for finding internal representation of the input.

3.4 Similarity between simple neural networks and classic estimators

The similarity of neural networks and classic estimators exists in some cases and should be examined. Classic estimators are in a way a subset of neural networs. They are not a perfect subset, but they are close. For example OLS is not a subset of neural networks, but we can get results that will converge to those of OLS.

3.4.1 Similarity between simple neural network and OLS

Linear neuron is build in following way: input, summing function, activation function and output (Figure 3.1). We can achieve a similar result as in OLS when we consider neural network that consists of one linear neuron with activation function being linear ($\psi(x) = x$) and cost function of the network being $C(x_i, y_i) = \sum_{i=1}^{n}(f(x_i) - y_i)^2$ where n is the number of observations, x_i is the i^{th} input, $f(x)$ is summing function and y_i is the i^{th} expected output. When we estimate this network, the parameters in the summing function should converge to the parametrs estimated by the OLS.

3.4.2 Similarity between simple neural network and MLE

As in the case of OLS, lets assume neural network with one linear neuron. The difference and what makes the network similar to a Maximum Likelihood Estimation is the neural network cost function which is

$$C(\boldsymbol{y}, \boldsymbol{z}) = -\Pi_{i=1}^{n} f(y_t|z_t) \tag{3.4}$$

where y_t is the dependent variable and z_t is the output of the neural network at time t, it corresponds to MLE where the $L(\boldsymbol{\theta}|\boldsymbol{y}, \boldsymbol{X}) = \Pi_{i=1}^{n} f(y_t|\boldsymbol{x_t}, \boldsymbol{\theta})$ is maximized. The z_t corresponds to $\boldsymbol{x}_t'\boldsymbol{\theta}$.

Chapter 4

Methodology

4.1 Quantile Regression

4.1.1 Quantiles

Quantiles are values that devide a set of values into equally sized subsets. Q-quantiles are values that devide a set into q subsets that are equal (or almost equal, depending on a particular case). k^{th} q-quantile of a set X is a value for which following applies:

$$P(y{<}z, y \in X) \leq \frac{k}{q} \tag{4.1}$$

$$P(y \leq z, y \in X) \geq \frac{k}{q} \tag{4.2}$$

where z is the value of the k^{th} q-quantile.

There are special quantiles that are used more often than others, for example median which is 1^{st} 2-quantile or 2^{nd} 4-quantile and so on. The rule that $k = \frac{q}{2}$ applies to median and in other words the median is the middle or central value.

Another often used quantile is k^{th} 100-quantile which can be called k^{th} percentile. In this thesis, only percentiles will be used and they will be denoted as quantiles. For example 95^{th} 100-quantile will be denoted as 95^{th} quantile.

Median is not the only quantile that is used often in economic literature. Confidence intervals are based on quantiles too, confidence interval at significance level of 5% (or 95% confidence interval) has a lower bound equal to 2.5^{th} quantile and upper bound equal to 97.5^{th} quantile. Quantiles that are important in this thesis are 5^{th} and 10^{th} quantiles of returns, which can be considered as 5% and 10% VaR, or 5% and 10% probability that the comodity (in the case of this thesis S&P 500 and WTI Crude Oil) will fall in value by more than the VaR.

4.1.2 Quantile Regression

Quantile regression is based on the idea of estimating the conditional quantile (for example median) in opposite to the OLS where the the conditional mean is estimated. While estimating the quantile regression we do not make any assumption about the distribution or about the conditional variance. To estimate (4.3) for q^{th} quantile we are looking for $\boldsymbol{\beta_q}$ such that the equation (4.4) is satisfied. Estimation is done through minimizining the function defined in (4.5)

$$y_i = \alpha + \boldsymbol{x_i'}\boldsymbol{\beta_q} \tag{4.3}$$

$$P(y \leq \boldsymbol{x'}\boldsymbol{\beta_q}|\boldsymbol{x}) = q \tag{4.4}$$

$$F_n(\boldsymbol{\beta_q}|\boldsymbol{y},\boldsymbol{X}) = \sum_{i:e_{i,q}\geq 0}^{n} q|y_i - \alpha - \boldsymbol{x_i'}\boldsymbol{\beta_q}| + \sum_{i:e_{i,q}<0}^{n} (1-q)|y_i - \alpha - \boldsymbol{x_i'}\boldsymbol{\beta_q}| \tag{4.5}$$

where $e_{i,q} = y_i - \alpha - \boldsymbol{x_i'}\boldsymbol{\beta_q}$.

4.2 Value-at-Risk

Value-at-Risk is an economic indicator that is used for estimation of the highest potential loss of a given instrument, investment or portfolio of financial instruments on a given significance level. VaR is basically a 95^{th} or sometimes 99^{th} quantile of maximum expected loss. Which means that it is expected that only in 5% or 1% of cases the loss will be higher than the value that VaR gives. For example in case of 95% VaR and the value of Value-at-Risk being \$1 means that there is 5% probability that there will be a loss of \$1 or more. Similarly the Value-at-Risk can be defined in an opposite way as a α-quantile of profits or better α-quantile of returns. The α that corresponds to the 95^{th} and 99^{th} quantile of maximum expected loss is 1^{st} and 5^{th} quantile of returns.

Lets define the Value-at-Risk, let $f(x_t)$ be a probability distribution function of a logarithmic price process in time t and $F(x_t)$ be its cumulative distribution function. We will consider the cumulative distribution function to be defined on the interval $(-\infty, \infty)$. It is defined on such interval because the logarithmic price process is not bounded, when the actual price approaches 0, then the logarithmic price approaches $-\infty$ and the return is this value minus some real value (logarithm of a real number - real price). It works similarly for the upper bound in case the real price starts from 0 (which can be for example start-up in its earliest moment). We will not require the cumulative distribution function to be strictly increasing, the non-decreasing restriction is enough. This means that there might be two or more different values x_l, x_m such that $F(x_l) = F(x_m) = \alpha$.

Lets define function $G(x_t)$

$$G(x_t) = \begin{cases} F(x_t) & \text{if } \forall x_m, x_m > x_t : F(x_t) \neq F(x_m) \\ \text{not defined} & \text{otherwise} \end{cases}$$

by the definition, the $G(x_t)$ is invertible and its functional values are in interval [0, 1]. Now we can finally define the α-VaR

$$\alpha\text{-VaR} = \begin{cases} G^{-1}(1\text{-}\alpha) & \text{if } G^{-1}(1\text{-}\alpha) \text{ is defined} \\ & \text{where } \alpha_u \text{ is such value that } G^{-1}(1\text{-}\alpha_u) \text{ is defined,} \\ G^{-1}(1\text{-}\alpha_u) & \text{and } \nexists\ \alpha_l,\ \alpha > \alpha_l > \alpha_u \text{ such that } G^{-1}(1\text{-}\alpha_l) \\ & \text{is defined.} \end{cases}$$

4.3 Realized measures and forecasting Returns and Volatility

As in Žikeš & Baruník (2016) the logarithmic price process (X_t) is used. Quadratic variation is used as a measure of volatility in the logarithmic price process. It can be split into the integrated variance and jump variation. Where the jump variation is there due to the discontinuos part of the logarithmic price process (X_t) and the integrated variance is there due to the continuous part of the X_t.

$$QV_t = IV_t + JV_t$$

To split the quadratic variation into integrated variance and jump variation lets assume data sample of size T*(M+1), which corresponds to T days and M+1 observations in each day (intraday observations). Lets define i^{th} return observation in day t as:

$$\Delta_i X_t = X_{t-1+\frac{i+1}{M}} - X_{t-1+\frac{i}{M}}$$

Return can be defined this way due the usage of logarithmic price process, $X_t = log(p_t)$ where p_t is the price process, then

$$\Delta_i X_t = X_{t-1+\frac{i+1}{M}} - X_{t-1+\frac{i}{M}} = log(p_{t-1+\frac{i+1}{M}}) - log(p_{t-1+\frac{i}{M}}) = log(\frac{p_{t-1+\frac{i+1}{M}}}{p_{t-1+\frac{i}{M}}})$$

hence $\Delta_i X_t$ can be understood as logarithmic return.

4.3.1 Realized Variance

A consistent estimator of the overall quadratic variation is realized variance (Andersen & Bollerslev 1998):

$$RV_{t,M} = \sum_{i=0}^{M-1} (\Delta_i X_t)^2$$

where $RV_{t,M} \xrightarrow{p} QV_t$ with $M \to \infty$. To split the quadratic variation into integrated variance and jump variation we filter the process X_t by using the median realized variance (Andersen *et al.* 2012; Žikeš & Baruník 2016)

$$MedRV_{t,M} = \frac{\pi}{6 - 4\sqrt{3} + \pi} \frac{M}{M-2} \sum_{i=0}^{M-3} med(|\Delta_i X_t|, |\Delta_{i+1} X_t|, |\Delta_{i+2} X_t|)^2$$

The median realized variance is a consistent estimator of the integrated variance and the difference between the realized variance and the median realized variance is a consistent estimator of the jump variation.

$$IV_{t,M} = MedRV_{t,M}$$

$$JV_{t,M} = RV_{t,M} - IV_{t,M}$$

The realized variance can be also split into two parts where one is based on the positive intraday returns and the second one on the negative intraday returns (Barndorff-Nielsen *et al.* 2010; Sévi 2014). Žikeš & Baruník (2016) found that the negative semivariance possesses much more information for the forecasting of future volatility than the negative semivariance.

$$RS^{-}_{t,M} = \sum_{i=0}^{M-1} (\Delta_i X_t)^2 1_{\{\Delta_i X_t < 0\}} \xrightarrow{\text{p}} 0.5 IV_t + \sum_{t-1 \le s \le t} 1_{\{\Delta J_s < 0\}} (\Delta J_s)^2$$

$$RS^{+}_{t,M} = \sum_{i=0}^{M-1} (\Delta_i X_t)^2 1_{\{\Delta_i X_t > 0\}} \xrightarrow{\text{p}} 0.5 IV_t + \sum_{t-1 \le s \le t} 1_{\{\Delta J_s > 0\}} (\Delta J_s)^2$$

4.4 Models for returns

The specification of the following models is the same as Žikeš & Baruník (2016) used. These three models will be used for comparison of the estimation techniques of Žikeš & Baruník (2016) and techniques suggested by this thesis, in the sense that the same dependent and independent variables will be used for linear quantile regression estimation and quantile regression neural network estimation approach.

These models forecast quantiles of returns, specifically they are used to forecast 5%, 10%, 50%, 90% and 95% quantiles of returns. The first two quantiles can considered as Value-at-Risk. Generally for α-quantile the model can be specified as follows

$$q_\alpha(r_{t+1}|\Omega_t) = \beta_0(\alpha) + \boldsymbol{\beta_v(\alpha)' v_{t,M}} + \boldsymbol{\beta_z(\alpha)' z_t} \tag{4.6}$$

where

$$r_{t+1} = X_{t+1} - X_t$$

,

$$\boldsymbol{v_{t,M}} = (QV^{1/2}_{t,M}, QV^{1/2}_{t-1,M}, ..., IV^{1/2}_{t,M}, IV^{1/2}_{t-1,M}, ..., JV^{1/2}_{t,M}, JV^{1/2}_{t-1,M}, ...)$$

$\boldsymbol{z_t}$ is a vector of other exogenous variables and $q_\alpha(r_{t+1}|\Omega_t)$ is the α-quantile of returns conditional on the information set Ω_t.

More specifically, Žikeš & Baruník (2016) define three models

LQR1:

$$q_\alpha(r_{t+1}|\Omega_t) = \beta_0(\alpha) + \beta_1 RV_t^{1/2} \tag{4.7}$$

LQR2:

$$q_\alpha(r_{t+1}|\Omega_t) = \beta_0(\alpha) + \beta_1 IV_t^{1/2} + \beta_2 JV_t^{1/2} + \beta_3 ImV_t \tag{4.8}$$

LQR3:

$$q_\alpha(r_{t+1}|\Omega_t) = \beta_0(\alpha) + \beta_1 RS_t^{+1/2} + \beta_2 RS_t^{-1/2} + \beta_3 ImV_t \tag{4.9}$$

4.5 Models for realized volatility

As in Section 4.4 Žikeš & Baruník (2016) specify three models for quantiles of realized volatility dependent on the realized measures. The quantiles of realized volatility that will be used are 50%, 75%, 90% and 95% quantiles. The models will be estimated with linear quantile regression and corresponding models will be estimated with quantile regression neural network and then compared. General α-quantile model can be specified as follows

$$\begin{aligned} q_\alpha(RV_{t+1,M}|\Omega_t) = \beta_0(\alpha) + \boldsymbol{\beta_{v1}(\alpha)' v_{t,M}} + \boldsymbol{\beta_{v5}(\alpha)' v_{t,t-5,M}} + \\ \boldsymbol{\beta_{v22}(\alpha)' v_{t,t-22,M}} + \boldsymbol{\beta_z(\alpha)' z_t} \end{aligned} \tag{4.10}$$

where

$$v_{t,t-k,M} = \frac{1}{k}\sum_{j=0}^{k-1} v_{t-j,M}$$

and $v_{t,t-k,M}$ is the average of $v_{t,M}$ over the past k days, $\boldsymbol{z_t}$ is a vector of other exogenous variables and $q_\alpha(r_{t+1}|\Omega_t)$ is the α-quantile of returns conditional on the information set Ω_t. In case of S&P 500 there will be no other exogenous variables and in the case of WTI Crude Oil there will be dummy for wednesday ($D_t^W = 1$). The model Equation 4.10 is called heterogenous autoregressive quantile model (HARQ) and the following specifications suggested by Žikeš & Baruník (2016) will be used (without the wednesday dummy variable for the WTI case)

HARQ1:

$$q_\alpha(RV_{t+1}^{1/2}|\Omega_t) = \beta_0(\alpha) + \beta_1 RV_t^{1/2} + \beta_2 RV_{t,t-5}^{1/2} + \beta_3 RV_{t,t-22}^{1/2} \tag{4.11}$$

HARQ2:

$$q_\alpha(RV_{t+1}^{1/2}|\Omega_t) = \beta_0(\alpha) + \beta_1 RS_t^{+1/2} + \beta_2 RS_t^{-1/2} + \beta_3 RV_{t,t-5}^{1/2} + \beta_3 RV_{t,t-22}^{1/2} + \beta_4 ImV_t \tag{4.12}$$

HARQ3:

$$q_\alpha(RV_{t+1}^{1/2}|\Omega_t) = \beta_0(\alpha) + \beta_1 IV_t^{1/2} + \beta_2 IV_{t,t-5}^{1/2} + \beta_3 IV_{t,t-22}^{1/2} + \beta_4 JV_t^{1/2} + \beta_5 ImV_t \tag{4.13}$$

4.6 QRNN

4.6.1 Taylor (2000)

Selection of appropriate explanatory variables might not be simple task, so Taylor (2000) uses artificial neural network for estimation of nonlinear quantile models. Author uses historical returns from different periods and proposes a quantile regression approach to estimate the multiperiod returns distribution. He estimates quantile regression neural network model (with one hidden layer)

$$f(\mathbf{x_t}, \mathbf{v}, \mathbf{w}) = g_2(\sum_{j=0}^{m} v_j g_1(\sum_{i=0}^{n} w_{ji} x_{it})) \tag{4.14}$$

of the q^{th} quantile by minimizing the following expression:

$$\min_{\mathbf{v},\ \mathbf{w}} \Bigg(\sum_{t|y_t \geq f(\boldsymbol{x_t}, \mathbf{v}, \mathbf{w})} q|y_t - f(\boldsymbol{x_t}, \mathbf{v}, \mathbf{w})| + \sum_{t|y_t < f(\boldsymbol{x_t}, \mathbf{v}, \mathbf{w})} (1-q)|y_t - f(\boldsymbol{x_t}, \mathbf{v}, \mathbf{w})| + \lambda_1 \sum_{j,i} w_{ji}^2 + \lambda_2 \sum_i v_i^2 \Bigg) \tag{4.15}$$

λ_1 and λ_2 are parameters that penalize the complexity (overfitting) of the neural network. Taylor (2000) suggests to select the optimal values of λ_1, λ_2 and the number of neurons in the hidden layer m by cross-validation. Another suggestion, based on Tang & Fishwick (1993), is to use n neurons, where n is number or inputs (variables).

4.6.2 QRNN - implementation in R

The quantile regression neural network estimation used in this thesis was implemented in R by Cannon (2011). The model is based on the standard multilayer perceptron artificial neural network. The outputs from the model are calculated in a following way. Each hidden layer node output is calculated by using hyperbolic tangent function (sigmoidal function) on a weighted sum of inputs

$$g_j(t) = tanh(\sum_{i=1}^{l} x_i(t) w_{ij}^h + b_j^h)$$

where the b_j^h is the j^{th} hidden neuron bias, $x_i(t)$ is the i^{th} input, l is the number of inputs and w_{ij}^h is the weight that corresponds to a given neuron and input.

The output $y_\tau(t)$ of the neural network (conditional α-quantile) is calculated as

$$y_\tau(t) = f(\sum_{j=1}^{J} g_j(t) w_j^{(0)} + c)$$

where J is number of neurons in hidden layer, $g_j(t)$ is the output of j^{th} neuron, $w_j^{(0)}$ is the j^{th} output layer weight, $b^{(0)}$ is the output layer bias and $f(.)$ is the output layer transfer function that is used for left censoring (it is either identity or ramp function). When the left censoring is applied then the Huber norm is used to for construction of smooth approximation cost function then standard gradient-based optimization algorithms are applied. The cost function is constructed as follows, starting with the Huber norm

$$h(u) = \begin{cases} \frac{u^2}{2\epsilon} & \text{if } 0 \le |u| \le \epsilon \\ |u| - \frac{\epsilon}{2} & \text{otherwise} \end{cases}$$

tilted absolute value function

$$\rho_\tau^{(a)} = \begin{cases} \tau h(u) & \text{if } 0 \le u \\ (\tau - 1) h(u) & \text{otherwise} \end{cases}$$

ramp function

$$r^{(a)}(u) = \begin{cases} h(u) & \text{if } l \le u \\ l & \text{otherwise} \end{cases}$$

and the cost function (in Taylor (2000) error function)

$$E_{\tau}^{(a)} = \frac{1}{N} \sum_{t=1}^{N} \rho_{\tau}^{(a)} (y(t) - \hat{y}_{\tau}(t))$$

The problem with overfitting is tackled through weight decay regularization with the addition of penalty to the cost function

$$E_{\tau}^{(a)} = \frac{1}{N} \sum_{t=1}^{N} \rho_{\tau}^{(a)} (y(t) - \hat{y}_{\tau}(t)) - \lambda \frac{1}{IJ} \sum_{i=1}^{I} \sum_{j=1}^{J} (w_{ij}^{(h)})^2$$

where lambda controls the relative contribution of the penalty and is positive. As Taylor (2000) says this approach reduces the non-linearity in the model. The bagging is also used to reduce the overfitting, the model is trained on resampled datasets and then the result is the average of these models (same models trained on different subsamples of the dataset).

4.7 How to compare results

As a forecasting measure of accuracy in forecasting quantiles we use percentage of observations that fall bellow the estimator. The reason for not using conventional measures is the unobservable nature of quantiles (Taylor 2000).

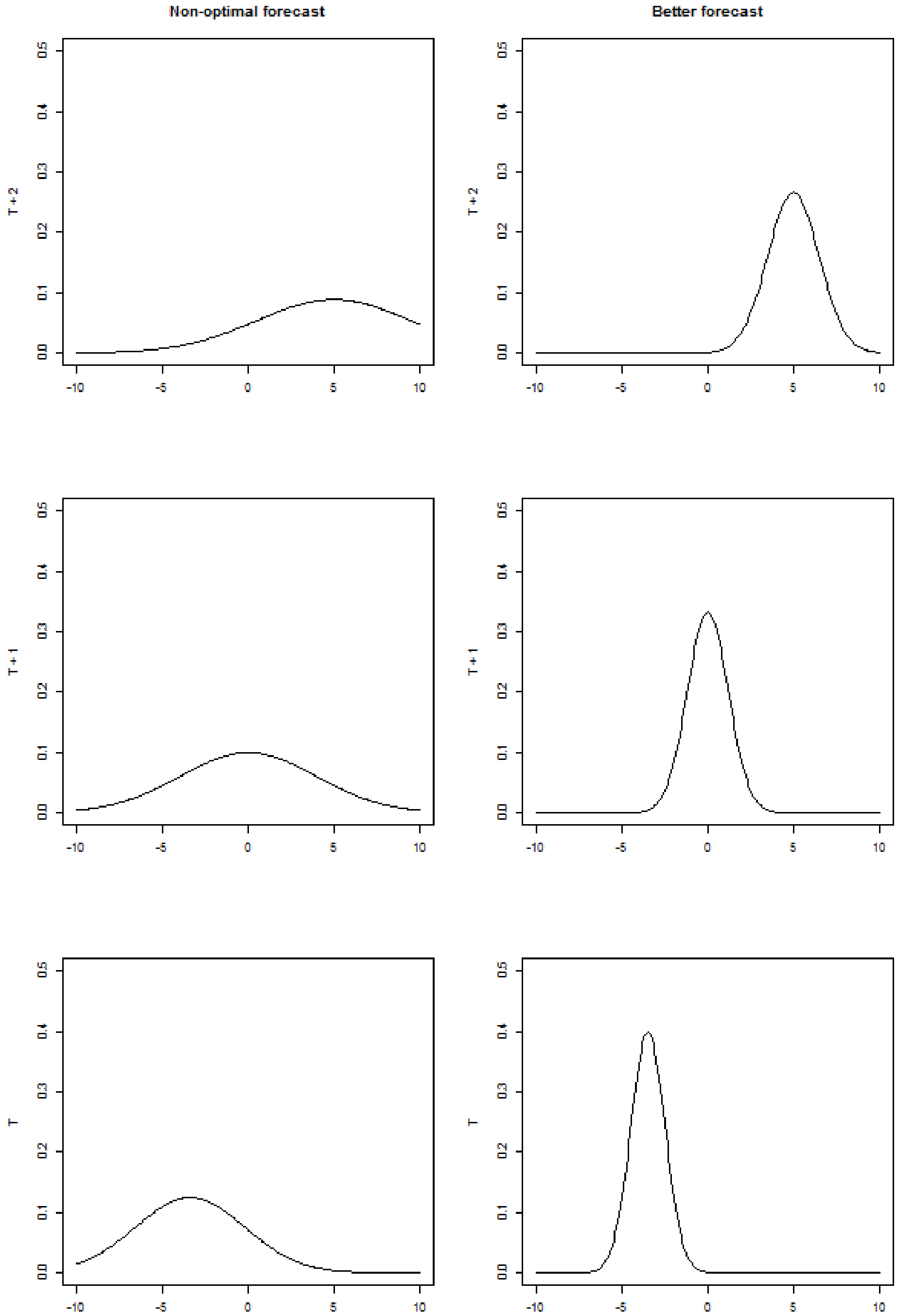
To get a good estimate of the q^{th} quantile we need to get the forecasting measure of accuracy equal to q which is just saying that it is the q^{th} quantile, this is called the *absolute* performance. If the model for quantile forecasts is correctly dynamicaly specified this should be always true.

Second important requirement from quantile models is that the forecasted distribution should be as narrow as possible. Forecasts of distribution in Figure 4.1 might fulfil the first requirement (on average) but having the *better forecast* gives more precise information. Comparison of this is called comparison of the *relative* performance.

4.7.1 Absolute performance of one-step-ahead forecasts

The approach taken here is only suited for the one-step-ahead forecasts. The absolute performance of the models is evaluated by the CAViar test (Berkowitz *et al.* 2011) (version of DQ test of Engle & Manganelli (2004)). The test is based on *Hit* variable

Figure 4.1: Comparison of quantile forecast



$$Hit_{t+1} = 1\{r_{t+1} \leq q_\alpha(r_{t+1}|\Omega_t)\}$$

which is a binary variable that is equal to 1 if the r_{t+1} is below the quantile prediction and zero when it is above. If the quantiles are correctly specified they should be i.i.d. with Bernoulli distribution with parameter α. Berkowitz *et al.* (2011) suggest to test this by estimating following logistic regression:

$$Hit_t = c + \sum_{k=1}^{n} \beta_{1,k} Hit_{t-k} + \sum_{k=1}^{n} \beta_{2,k} q_\alpha(r_{t+1-k}|\Omega_{t-k}) + u_t \tag{4.16}$$

To show that α-quantile is correctly specified is to have the β coefficients insignificant and $P(Hit_t = 1) = \frac{e^c}{e^c+1} = \alpha$. The critical values for the test are obtained through Monte Carlo simulation. This test tests that there is no autocorrelation in *Hit* and that it is not dependent on lagged α-quantile forecasts.

4.7.2 Absolute performance of multi-step-ahead forecasts

In the case of multi-step-ahead forecast we have to redefine the *Hit* variable as

$$Hit_{t|t+h} = 1\{r_{t+1} + r_{t+2} + ... + r_{t+h} \leq q_\alpha(r_{t+1} + r_{t+2} + ... + r_{t+h}|\Omega_t)\}$$

where the $q_\alpha(r_{t+1} + r_{t+2} + ... + r_{t+h}|\Omega_t)\}$ is the h-step-ahead α-quantile forecast for cumulative h-period return under the information Ω_t that is available at time t. The cumulative h-period return is a sum of returns due to the fact that the logarithmic returns are used.

Unlike the one-step-ahead forecast we can not use the Equation 4.16 for testing correct dynamic specification of quantiles, due to the fact that the sequence is h-dependent, as is suggested by Žikeš & Baruník (2016). They also suggested a way around this problem, but they reported that the test has poor performance on finite samples. Also they stated that to the best of their knowledge there is no reliable test for correct dynamic specification of multi-step-ahead quantile forecasts. Similarly to the best knowledge of the author of this thesis, there is still no reliable test for absolute performance.

4.7.3 Relative performance

The relative performance of quantile models is based on the Clements *et al.* (2008) and the core idea is comparison of values from a tick loss function

$$L_\alpha(e^m_{t+1}) = (\alpha - 1\{e^m_{t+1} < 0\})e^m_{t+1} \tag{4.17}$$

where we look at the expected tick loss:

$$L_{\alpha,m} = E((\alpha - 1\{e^m_{t+1} < 0\})e^m_{t+1})$$

where $e^m_{t+1} = r_{t+1} - q^m_\alpha(r_{t+1}|\Omega_t)$ and $q^m_\alpha(r_{t+1}|\Omega_t)$ is the fitted value of α-quantile at time $t+1$ of model m. If the return is higher than the forecasted quantile (in case of 50^{th} or higher quantiles) then the penalization is higher than if the return was below the forecasted quantile. In case of smaller than 50^{th} quantile the situations that are more penalized are those when the returns are below the forecasted quantiles.

The relative performance compares two models and their tick loss. With the null hypothesis $H_0 : L_{\alpha,m} = L_{\alpha,n}$ against a general alternative, where we test that the two models have equal expected tick loss. The Diebold & Mariano (1995) test is used.

The test works in a following way. We transform the e^m_{t+1} by the tick loss function (4.17) of both models that we want to compare. Lets say that the model n is the model we compare to and m is the model we compare. We take their differences

$$d_t = L_\alpha(e^m_{t+1}) - L_\alpha(e^n_{t+1}),\ for\ each\ t\ in\ 1, ...T$$

and we calculate the following statistics

$$DM = \frac{\bar{d}}{\sqrt{\frac{1}{T}var(d)}} \ \sim\ N(0,1) \tag{4.18}$$

where $\bar{d}$ is the mean of d_t, T is the number of observations (predictions, length of the d) and *var(d)* is in the case of one-step-ahead forecast the variance of d and in case of multi-step-ahead forecasts the Newey-West variance of d and DM is distributed by N(0, 1) (Diebold 2015). The Newey-West variance (Newey & West 1987) for h-step-ahead forecast can be computed by:

$$NWvar(x) = var(x) + 2\sum_{i=1}^{h-1} cov(x_t, x_{t-i})$$

where *var(x)* is the usual variance and $cov(x_t, x_{t-i})$ is covariance.

4.8 Hypothesis 1

First hypothesis states that QRNN provides better quantile estimates than standard linear quantile regression. This will be tested by comparing the LQR1, LRQ2, LQR3 (Section 4.4), HARQ1, HARQ2 and HARQ3 (Section 4.5) of Žikeš & Baruník (2016) to their generalization through neural networks, in other words for each of these models the QRNN will be estimated with the same input data and the same input specification. The predictions of two corresponding models will be compared. This will be done for both in-sample and out-of-sample data.

4.9 Hypothesis 2

In the second hypothesis the volatility is expected to be important for predicting quantile returns. This will be evaluated by comparing models with and without volatility.

4.10 Hypothesis 3

The hypothesis 3 states that the expected densities of returns are skewed. Simple example of skewed expected density of returns can be in a situation when there is the same probability of a good and bad result but one (lets assume the good result) has limitations and the other one does not. This is an example of negative skew.

Before constructing the test lets define the skewness. There are different measures of sample skewness and we will base ours on Equation 4.19 (Cramer 1946; Joanes & Gill 1998).

$$g_1 = \frac{m_3}{m_2^{3/2}} \tag{4.19}$$

$$m_r = \frac{1}{n}\sum(x_i - \bar{x})^r$$

where m_r are sample moments, $\bar{x}$ is the sample mean. But Joanes & Gill (1998) show that they are biased. Based on their study, we will use two measures for skewness that work better in small samples. Where first one is G_1. Bias in the sample moments are biased (estimates of the population moments μ_r)

$$E(m_2) = \frac{(n-1)}{n}\mu_2$$

$$E(m_3) = \frac{(n-1)(n-2)}{n^2}\mu_3$$

by making the correction we get

$$K_2 = \frac{n}{n-1}m_2$$

$$K_3 = \frac{n^2}{(n-1)(n-2)}m_3$$

and we can define the G_1 as

$$G_1 = \frac{K_3}{K_2^{3/2}} = \frac{\sqrt{n(n-1)}}{n-2}g_1 \tag{4.20}$$

Second measure that we will use is b_1

$$b_1 = \left(\frac{n-1}{n}\right)^{3/2}\frac{m_3}{m_2^{3/2}} \tag{4.21}$$

Both parameters will be tested with null hypothesis of them being equal to 0 against general alternative. It will be done by estimating skewness (b_1 and G_1) for each time t from its prediction of distribution and then testing series of b_1 and G_1 that they are equal to 0.

Chapter 5

Data

As in Žikeš & Baruník (2016), data from S&P 500 futures and WTI Crude Oil futures contracts are used.[1]

Both S&P 500 futures and WTI Crude Oil futures use tick-data that are condensed into daily observations of date, logarithmic return, option implied volatility and components of quadratic variation - realized measures: realized variance, positive and negative semivariances, median realized variance. Realized measures are obtained from 5-min logarithmic returns where the returns are based on last tick method. The sampling frequency is the same as in Žikeš & Baruník (2016) so corresponding results can be obtained.

The two variables of high importance are returns and realized volatility which will be used as dependent variables in the models. Explanatory variables are realized volatility, option implied volatility, square root of integrated variance and averages of their lags, square root of jump variation and square root of both positive and negative semivariances.

The option implied volatilities are the volatility indecies calculated by the Chicago Board of Exchange (CBOE). In the case of WTI Crude Oil futures it is the crude oil volatility index (OVX) which is the 30-day volatility implied by oil futures options, but it goes back only to May 2007, so as in Žikeš & Baruník (2016) the model free implied volatility index suggested by Carr & Wu (2009) and Trolle & Schwartz (2010) is used, using the American style futures options settlement prices of oil traded on the Chicago Mercantile Exchange (CME). The option implied volatility (September 2001 - August 2008) and the OVX (September 2008 - December 2012) was spliced together as in Žikeš & Baruník (2016) to get the final implied volatility.[2]

[1] Data were provided by Josef Baruník.

[2] Already calculated implied volatility was provided by Josef Baruník in the dataset.

In the case of S&P 500 the problem with option implied volatility is simplier, the Volatility Index (VIX) index is used (again calculated by the CBOE). This index (as OVX) measures the one month expected volatility which is implied by a portfolio of put and call options.

Jumps (jump variation) are detected from the median realized variance and the realized variance. When their difference for a certain day is significantly different from 0 (at 0.1% significance level), then we detected a jump. When a jump is detected for day t, the integrated variance is set to be equal to the median realized variance $IV_{t,M} = MedRV_{t,M}$ and the jump variantion is set as a difference between realized variance and median realized variance $JV_{t,M} = RV_{t,M} - MedRV_{t,M}$. When a jump for day t was not detected, then the integrated variance is equal to the realized variance and the jump variation is equal to 0 $IV_{t,M} = RV_{t,M}$, $JV_{t,M} = 0$.

5.1 WTI Crude Oil futures

The data for WTI Crude Oil futures start by September 4^{th}, 2001 and end with December 31^{th}, 2012. All together they contain 2829 observations. Basic statistics are provided in Table 5.1. The average return is 0.04 with standard deviation 1.87 which gives (under normal distribution) 95% confidence interval for the return (-0.03; 0.11), the median is above the mean approximately equal to the upper bound of the confidence interval which suggests slightly skewed probability density function which is supported by the sample skewness. The second main variable that will be forecasted is the realized volatility. The realized volatility mean is equal to 1.73 with standard deviation 0.73 which gives (under normal distribution) 95% confidence interval (1.7; 1.76). Other variables that will be used as explanatory variables are positive and negative semivariance with means 1.8 and 1.71, median realized variance with mean 3.25 and option implied volatility with mean 40.91. The option implied volatility is higher than the realized volatility which shows the existence of a negative variance risk premium (Bollerslev *et al.* 2009). The positive and negative semivariances account both for approximately half of the realized variance(48.7% and 51.3%). Jump Variation is present in 71 days (2.5% of days) and the amount of it accounts for 0.7% of the realized variance.

Figure 5.1 suggests high volatility clustering as can be seen especially for the period between 2008 and 2010, whilst other periods experience very little volatility (after the year 2006). Similar observation can be made for jump

Table 5.1: Summary statistics for WTI crude oil futures

	Mean	Std. D.	Median	Min	Max	Skewness	Kurtosis
r_t	0.04	1.87	0.11	-12.54	14.43	-0.15	3.67
RV_t	3.51	3.75	2.46	0.23	37.59	3.93	20.65
$RVol_t$	1.73	0.73	1.57	0.48	6.13	2.04	6.02
$log(RV_t)$	0.95	0.73	0.90	-1.45	3.63	0.54	0.74
RSV_t^-	1.80	2.09	1.18	0.09	21.43	3.86	20.15
RSV_t^+	1.71	1.96	1.17	0.10	23.62	4.82	33.76
$MedRV_t$	3.25	3.50	2.29	0.16	43.79	4.04	23.45
VIX_t	40.91	12.20	37.42	24.63	106.50	2.09	5.41

All realized measures are calculated from 5-minute prices. The sample period starts at September 4, 2001 and ends with December 31, 2012, with 2829 observations.

variation which is the highest and most pronounced in 2009, another spike is in 2001. Almost no variation is around the year 2006 and in 2002. The negative semivariance has higher extreme values prior to the 2008. The returns of WTI suggest that the process is MA(1) as can be seen in Figure 5.2 in the ACF of returns with the one significant lag. The realized variance shows signs of long-memory which is supported by the Ljun-Box test of no autocorrelation with 20 lags.

5.2 S&P 500

S&P 500 futures data range from January 2^{nd}, 1996 and end with December 31^{st}, 2012. The sample length is higher than in the case of WTI Crude Oil futures, alltogether it is 4265 observations. The returns have negative mean of -0.01 with standard deviation of 1.11 which makes it statistically indistinguishable from 0 (under assumption of normality) with 95% confidence interval (-0.04; 0.02). The absolute value of skewness is lower than in the case of WTI Crude Oil futures returns and is equal to -0.04. The volatility measured with the realized volatility is also smaller than in the case of WTI Crude Oil futures and is equal to 0.93 with standard deviation of 0.56 which gives 95% confidence interval of (0.91; 0.95). The realized variance is split in half into the positive and negative semivariances (0.60 and 0.59), the median realized variance mean is equal to 1.10 with 2.22 standard deviation and (1.03; 1.17) 95% confidence interval.

S&P 500 experiences similar behavior as WTI Crude Oil futures with re-

Figure 5.1: WTI Crude Oil futures: time series of daily returns, realized volatility, jump variation and median realized volatility

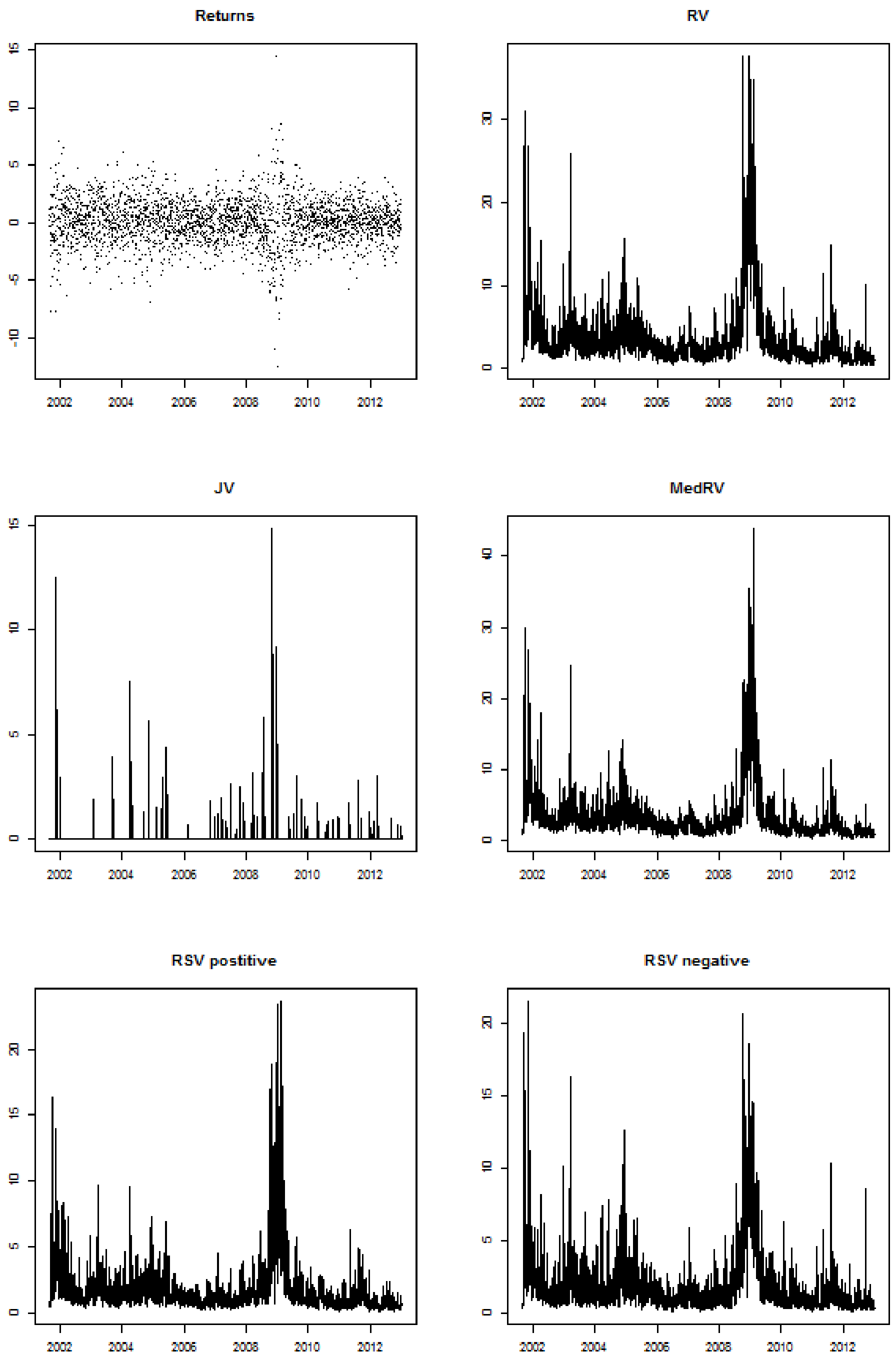


Figure 5.2: WTI Crude Oil futures: ACF and PACF of returns and realized volatility

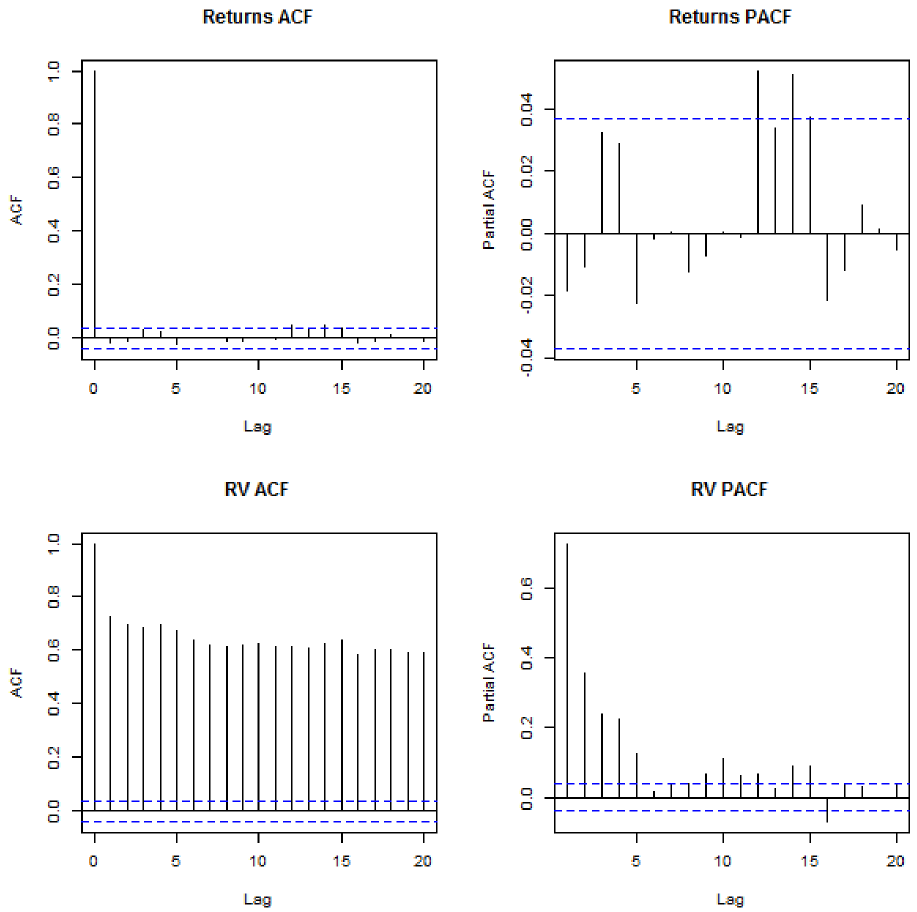


Table 5.2: Summary statistics for S&P 500 futures

	Mean	Std. D.	Median	Min	Max	Skewness	Kurtosis
r_t	-0.01	1.11	0.04	-8.02	8.38	-0.04	7.11
RV_t	1.19	2.33	0.65	0.04	61.35	10.20	167.92
$RVol_t$	0.93	0.56	0.80	0.21	7.83	3.28	19.33
$log(RV_t)$	-0.39	0.95	-0.44	-3.14	4.12	0.55	0.65
RSV_t^-	0.59	1.10	0.32	0.01	19.13	8.08	92.84
RSV_t^+	0.60	1.34	0.31	0.02	42.23	13.54	299.2
$MedRV_t$	1.10	2.22	0.58	0.03	55.88	9.96	155.62
VIX_t	1.15	0.45	1.08	0.52	4.23	1.93	6.60

All realized measures are calculated from 5-minute prices. The sample period starts at January 2, 1996 and ends with December 31, 2012, with 4265 observations.

alized variance, only here the magnitude is higher (Figure 5.3) and it applies similarly to the median realized variance. The highest peaks of positive semivariance are higher than the peaks of negative semivariance in 2009 and before 2000, whereas the WTI Crude Oil experiences opposite effect in the peaks of semivariances where in 2009 they have similar magnitude and in the beggining of the sample (year 2001) the peaks of negative semivariance are higher than the positive ones. In the case of jumps (jump variation) there are very few jumps in 2009 and they have limited magnitude which suggests that the volatility was consistently high, whereas in the case of WTI Crude Oil the jumps where more pronounced and even. Jumps in 2001 are proportionally similar in both cases of S&P 500 and WTI. Overall jumps are present in 103 days (2.4% of days) and they account for 2% of realized variance which higher than in the case of WTI.

The behavior of ACF (Figure 5.4) and PACF of returns is similar to the WTI Crude Oil, it suggests the MA process with the first lag in ACF being highly significant and the second and third being small but significant. The behavior of realized variance suggests long-memory as in the case of WTI which is supported by the Ljun-Box test of no autocorrelation with 20 lags.

Figure 5.3: S&P 500 futures: time series of daily returns, realized volatility, jump variation and median realized volatility

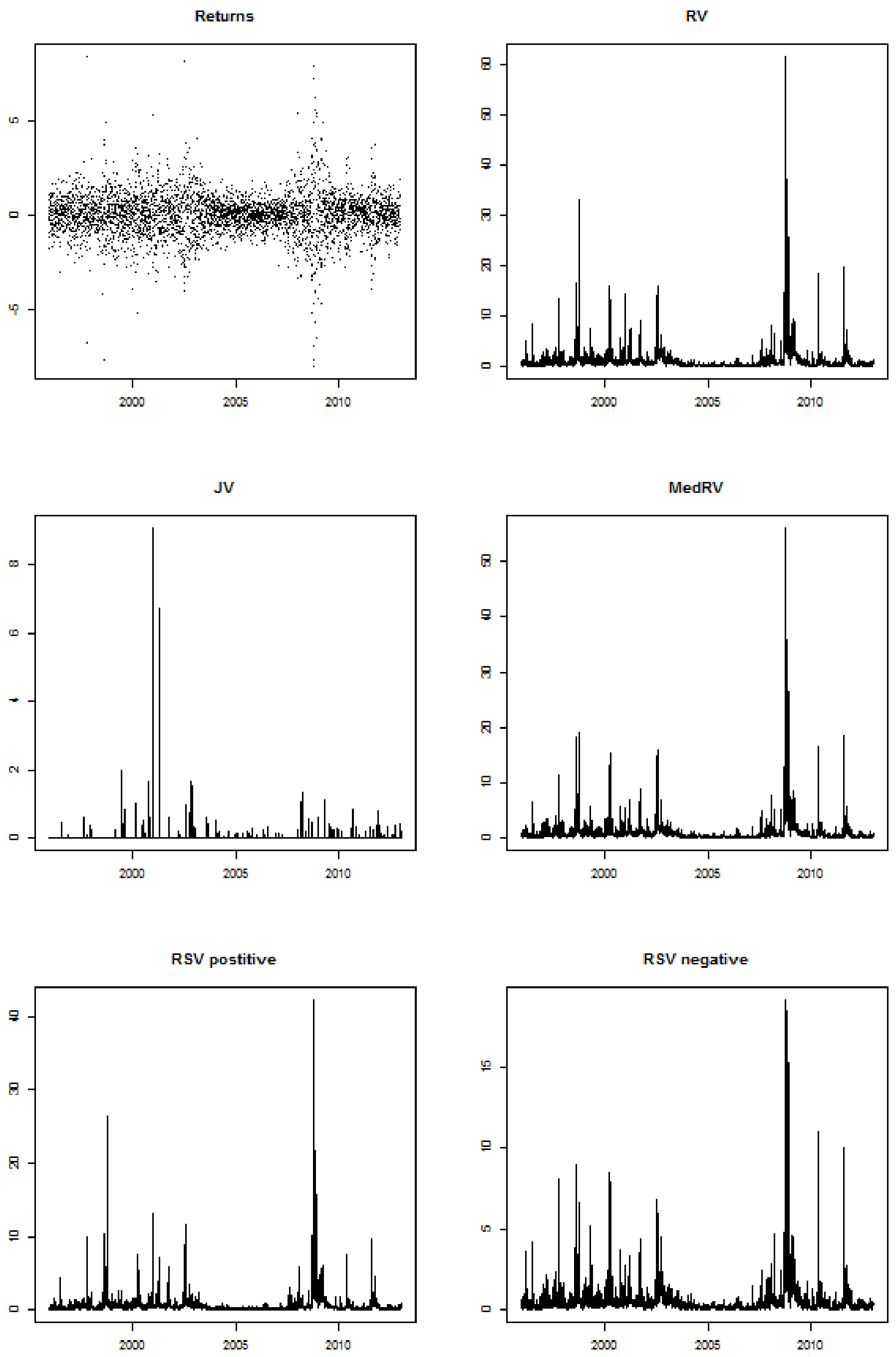
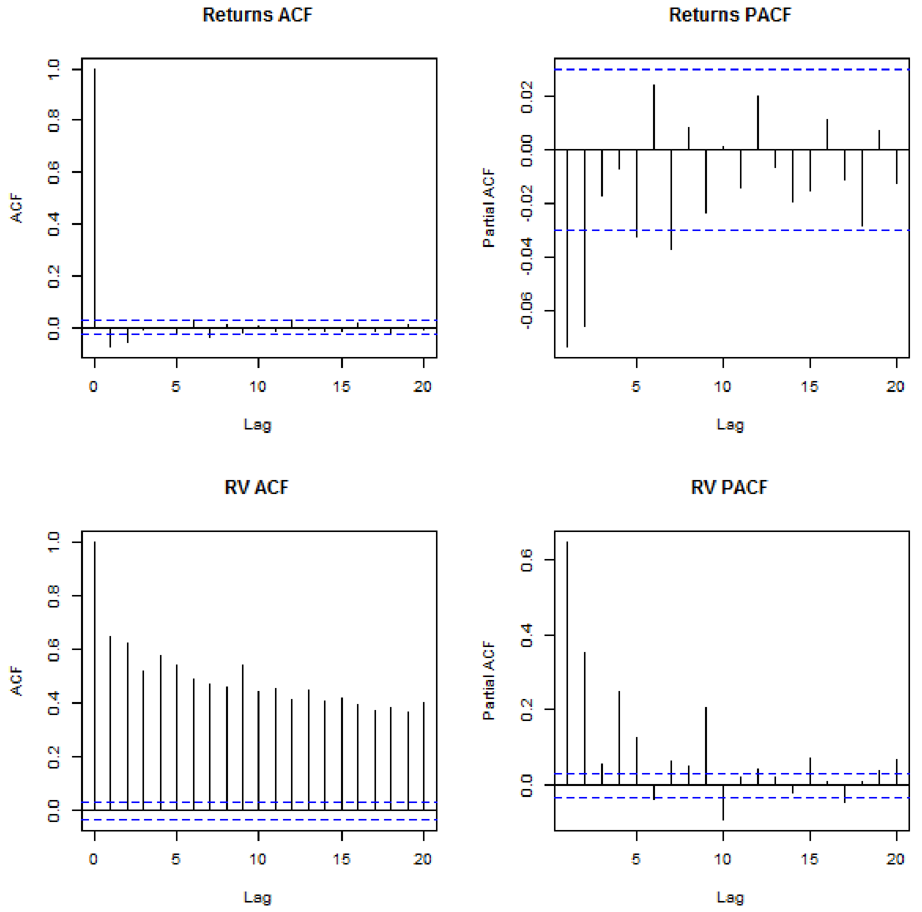


Figure 5.4: S&P 500 futures: ACF and PACF of returns and realized volatility



Chapter 6

Results

Here we compare models suggested by Žikeš & Baruník (2016) with the corresponding models estimated by Quantile Regression Neural Network. The neural networks specification is that all of the neurons have sigmoid activation function, more specifically the activation function is hyperbolic tangent. The number of neurons is n, where n corresponds to the number of independent variables with the exception of the model that has the specification as the LQR1 which has only one independent variable but two neurons. The n number of neurons is suggested by Tang & Fishwick (1993).

- LQR1 - 2 neurons
- LQR2 - 3 neurons
- LQR3 - 3 neurons
- HARQ1 - 3 neurons (4 neurons in the case of WTI)
- HARQ2 - 5 neurons (6 neurons in the case of WTI)
- HARQ3 - 5 neurons (6 neurons in the case of WTI)

The LQR (linear quantile regression) models are models of quantiles of returns with explanatory variables being mainly realized measures. The HARQ (heterogenous quantile autoregression) models, model quantiles of realized volatility with similar explanatory variables as LQR models. In the case of HARQ models and WTI Crude Oil data there is added dummy variable for wednesday.

Out-of-sample forecasts start at July 1^{st}, 2010 and end by December 31^{st}, 2012 which gives 2.5 years for forecasting and almost 10 years for training. The models are always estimated on the entire previous dataset. The linear models are reestimated (in the case of out-of-sample forecasts) for each time t (for each day) which is over 600 reestimations, the quantile regression neural networks are reestimated only 5 times on the out-of-sample due to the fact that it is more time demanding.

6.1 LQR and HARQ models of Žikeš & Baruník (2016)

The models of Žikeš & Baruník (2016) are estimated. These models will be used as bases of comparison to the neural network models.

6.1.1 LQR

Firstly the models suggested by Žikeš & Baruník (2016) are estimated to act as a comparison models for the models suggested by this thesis. The 5%, 10%, 50%, 90% and 95% quantiles of daily returns are estimated, where the first two quantiles can be interpreted as Value-at-Risk at corresponding levels for holding long position and the last two for holding short position. The estimated models are reported in Table 6.1 for WTI Crude Oil futures and in Table 6.2 for S&P 500 futures. The WTI Crude Oil futures starts by September 4, 2001 and ends by December 31, 2012. The S&P 500 starts by January 2, 1996 and ends by December 31, 2012.

In both cases (WTI and S&P 500) the realized volatility is statistically significant and as expected in lower quantiles it is negative and in upper quantiles positive. In LQR2 the jump variation is statistically insignificant as expected and volatility based on integrated variance is significant with signs that were expected, but it is lower in magnitude than the RV in LQR1 which was not anticipated, but could be due to the implied volatility in the model. In LQR3 the semivariances have expeced signs and in most cases are significant.

6.1.2 HARQ

Similarly as in Subsection 6.1.1 we estimate the HARQ models of Žikeš & Baruník (2016) for 50%, 75%, 90% and 95% quantiles where the last two quantiles can be used by traders that are under higher volatility risk. All three HARQ models were estimated both for WTI Crude Oil futures starting by

Table 6.1: Conditional quantile models (LQR) of Žikeš & Baruník (2016) for WTI Crude Oil futures return quantiles.

	LQR1					LQR2					LQR3				
α	0.05	0.1	0.5	0.9	0.95	0.05	0.1	0.5	0.9	0.95	0.05	0.1	0.5	0.9	0.95
$const$	-0.789	-0.587	0.178	0.732	0.864	-0.093	0.350	0.363	0.223	0.537	-0.073	0.329	0.432	0.316	0.393
	(0.003)	(0.003)	(0.109)	(0.000)	(0.000)	(0.793)	(0.216)	(0.028)	(0.385)	(0.028)	(0.837)	(0.253)	(0.008)	(0.184)	(0.089)
$RV_t^{1/2}$	-1.246	-0.915	-0.045	0.883	1.144										
	(0.000)	(0.000)	(0.539)	(0.000)	(0.000)										
$RS_t^{+1/2}$											-0.362	-0.122	-0.187	0.097	0.208
											(0.234)	(0.591)	(0.159)	(0.635)	(0.066)
$RS_t^{-1/2}$											-0.263	-0.225	0.306	0.684	0.754
											(0.348)	(0.284)	(0.010)	(0.000)	(0.000)
$IV_t^{1/2}$						-0.420	-0.274	0.091	0.692	0.636					
						(0.048)	(0.077)	(0.325)	(0.000)	(0.000)					
$JV_t^{1/2}$						-0.287	-0.348	-0.087	0.057	0.275					
						(0.112)	(0.396)	(0.778)	(0.879)	(0.473)					
ImV_t						-0.935	-0.961	-0.193	0.401	0.568	-0.941	-0.982	-0.221	0.457	0.598
						(0.000)	(0.000)	(0.055)	(0.011)	(0.000)	(0.000)	(0.000)	(0.027)	(0.002)	(0.000)

Reported are estimation results for linear quantile regression models of Žikeš & Baruník (2016) with p-values in parenthesis.
The sample period starts by September 4, 2001 and ends by December 31, 2012.

Table 6.2: Conditional quantile models (LQR) of Žikeš & Baruník (2016) for S&P 500 futures return quantiles.

	LQR1					LQR2					LQR3				
α	0.05	0.1	0.5	0.9	0.95	0.05	0.1	0.5	0.9	0.95	0.05	0.1	0.5	0.9	0.95
$const$	-0.341	-0.224	0.005	0.104	0.194	-0.103	-0.029	-0.057	-0.279	-0.212	-0.123	-0.027	-0.050	-0.259	-0.221
	(0.000)	(0.002)	(0.908)	(0.163)	(0.012)	(0.448)	(0.739)	(0.269)	(0.001)	(0.003)	(0.336)	(0.758)	(0.350)	(0.002)	(0.002)
$RV_t^{1/2}$	-1.454	-1.106	0.039	1.156	1.475										
	(0.000)	(0.000)	(0.471)	(0.000)	(0.000)										
$RS_t^{+1/2}$											-0.708	-0.372	-0.114	0.054	0.248
											(0.000)	(0.092)	(0.408)	(0.790)	(0.103)
$RS_t^{-1/2}$											-0.701	-0.551	0.080	0.822	0.817
											(0.014)	(0.005)	(0.534)	(0.000)	(0.000)
$IV_t^{1/2}$						-0.953	-0.619	-0.025	0.578	0.74					
						(0.000)	(0.000)	(0.745)	(0.000)	(0.000)					
$JV_t^{1/2}$						-0.784	0.435	0.244	0.012	-0.132					
						(0.383)	(0.262)	(0.008)	(0.914)	(0.688)					
VIX_t						-0.591	-0.560	0.109	0.816	0.933	-0.549	-0.539	0.108	0.754	0.935
						(0.003)	(0.000)	(0.197)	(0.000)	(0.000)	(0.002)	(0.000)	(0.232)	(0.000)	(0.000)

Reported are estimation results for linear quantile regression models of Žikeš & Baruník (2016) with p-values in parenthesis.
The sample period starts by January 2, 1996 and ends by December 31, 2012.

September 4, 2001 and ending by December 31, 2012 in Table 6.3 and the S&P 500 starting by January 2, 1996 and ending by December 31, 2012 in Table 6.4.

In both the WTI and S&P 500 the realized volatility and its lags are highly significant for all of the quantiles. The negative semivariances are high and highly significant close to the value of RV, whereas the positive semivariance are close to 0 and insignificant, which suggests that the negative semivariances possess most of the explanatory power of the realized variance. In LQR3 the volatility based on the integrated variance is statistically significant and the jump variation is significant only in few cases (in S&P 500 only for 90% quantile and in WTI for 50% and 75% quantiles). The option implied volatility is significant across models and in both instruments.

6.2 Absolute performance

Here the absolute performance of all of the models is reported for one-step-ahead forecasts. The statistics of unconditional coverage, DQ test statistic (Berkowitz *et al.* 2011) of correct dynamic specification with corresponding (Monte-Carlo based) p-value for the null hypothesis that all of the beta coefficients in the logistic regression Equation 4.16 are equal to 0. The regression is estimated with 5 lags.

6.2.1 LQR - Absolute performance

The absolute performance of linear quantile regression models (LQR) with dependent variable being the returns is reported here for both the linear case and the neural network case. The dynamic specification is tested for both S&P 500 and WTI Crude Oil, for forecasting horizon being equal to 1 and for both in-sample and out-of-sample. The forecasted quantiles are 5%, 10%, 50%, 90% and 95%.

Absolute performance of WTI Crude Oil futures return quantiles is reported in Table 6.5. In case of in-sample the unconditional coverage ($\hat{\alpha}$) is close to perfect in all of the cases, in out-of sample case it is shifted from the α values, it is shifted away from the center (5% α-quantile has lower unconditional coverage than it should have and 95% α-quantile has higher unconditional coverage than it should have). All of the models in in-sample case except one (QRNN LQR1 90% quantile) are correctly dynamically specified. In case of out-of-sample the models for 5%, 10% and 50% are correctly dynamically specified, in case of

Table 6.3: Conditional quantile models (HARQ) of Žikeš & Baruník (2016) for WTI Crude Oil futures realized volatility quantiles.

	HARQ1				HARQ2				HARQ3			
α	0.5	0.75	0.9	0.95	0.5	0.75	0.9	0.95	0.5	0.75	0.9	0.95
$const$	0.036	0.038	0.096	0.091	-0.040	-0.046	-0.039	-0.059	-0.051	-0.066	-0.065	-0.079
	(0.196)	(0.367)	(0.107)	(0.322)	(0.230)	(0.362)	(0.532)	(0.624)	(0.138)	(0.195)	(0.383)	(0.558)
D_t^W	0.180	0.235	0.256	0.186	0.155	0.225	0.271	0.174	0.167	0.234	0.280	0.163
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)
$RV_t^{1/2}$	0.210	0.235	0.270	0.274								
	(0.000)	(0.000)	(0.000)	(0.004)								
$RS_t^{+1/2}$					0.015	0.087	0.083	-0.012				
					(0.600)	(0.056)	(0.038)	(0.914)				
$RS_t^{-1/2}$					0.202	0.218	0.240	0.177				
					(0.000)	(0.000)	(0.000)	(0.011)				
$RV_{t,t-5}^{1/2}$	0.408	0.449	0.576	0.692	0.399	0.393	0.570	0.750				
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)				
$RV_{t,t-22}^{1/2}$	0.290	0.360	0.327	0.343	0.266	0.277	0.162	0.175				
	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)	(0.000)	(0.023)	(0.118)				
$IV_t^{1/2}$									0.188	0.232	0.238	0.165
									(0.000)	(0.000)	(0.000)	(0.001)
$JV_t^{1/2}$									0.118	0.072	0.054	-0.034
									(0.000)	(0.000)	(0.114)	(0.480)
$IV_{t,t-5}^{1/2}$									0.383	0.392	0.583	0.731
									(0.000)	(0.000)	(0.000)	(0.097)
$IV_{t,t-22}^{1/2}$									0.244	0.248	0.134	0.098
									(0.000)	(0.000)	(0.000)	(0.000)
ImV_t					0.113	0.171	0.239	0.294	0.124	0.193	0.257	0.354
					(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)

Reported are estimation results for linear quantile regression models of Žikeš & Baruník (2016) with p-values in parenthesis. The sample period starts by September 4, 2001 and ends by December 31, 2012.

Table 6.4: Conditional quantile models (HARQ) of Žikeš & Baruník (2016) for S&P 500 futures realized volatility quantiles.

	HARQ1				HARQ2				HARQ3			
α	0.5	0.75	0.9	0.95	0.5	0.75	0.9	0.95	0.5	0.75	0.9	0.95
$const$	0.058	0.066	0.077	0.044	-0.017	-0.010	-0.044	-0.095	-0.034	-0.029	-0.068	-0.119
	(0.000)	(0.000)	(0.001)	(0.277)	(0.135)	(0.482)	(0.060)	(0.016)	(0.002)	(0.050)	(0.007)	(0.000)
$RV_t^{1/2}$	0.432	0.485	0.683	0.729								
	(0.000)	(0.000)	(0.000)	(0.000)								
$RS_t^{+1/2}$					-0.008	-0.025	-0.066	0.015				
					(0.815)	(0.548)	(0.335)	(0.888)				
$RS_t^{-1/2}$					0.461	0.594	0.691	0.755				
					(0.000)	(0.000)	(0.000)	(0.000)				
$RV_{t,t-5}^{1/2}$	0.278	0.430	0.546	0.675	0.275	0.387	0.636	0.557				
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)				
$RV_{t,t-22}^{1/2}$	0.175	0.126	0.007	0.006	0.029	-0.029	-0.230	-0.253				
	(0.000)	(0.000)	(0.871)	(0.958)	(0.261)	(0.194)	(0.000)	(0.010)				
$IV_t^{1/2}$									0.341	0.397	0.530	0.615
									(0.000)	(0.000)	(0.000)	(0.000)
$JV_t^{1/2}$									0.024	0.080	0.202	0.385
									(0.379)	(0.361)	(0.001)	(0.186)
$IV_{t,t-5}^{1/2}$									0.239	0.371	0.506	0.512
									(0.000)	(0.000)	(0.000)	(0.000)
$IV_{t,t-22}^{1/2}$									0.009	-0.078	-0.234	-0.332
									(0.703)	(0.001)	(0.000)	(0.000)
VIX_t					0.282	0.301	0.420	0.571	0.325	0.370	0.481	0.647
					(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)

Reported are estimation results for linear quantile regression models of Žikeš & Baruník (2016) with p-values in parenthesis.
The sample period starts by January 2, 1996 and ends by December 31, 2012.

Table 6.5: Absolute performance of in-sample and out-of-sample forecasts for WTI Crude Oil futures return quantiles.

		In-sample					Out of-sample				
	α	0.05	0.1	0.5	0.9	0.95	0.05	0.1	0.5	0.9	0.95
	$\hat{\alpha}$	0.050	0.100	0.500	0.900	0.950	0.056	0.091	0.515	0.928	0.978
LQR1	DQ	2.522	2.104	7.651	11.037	2.357	6.647	5.305	4.021	7.579	16.124
	p-value	0.773	0.835	0.177	0.051	0.798	0.355	0.505	0.674	0.271	0.013
	$\hat{\alpha}$	0.050	0.100	0.500	0.900	0.950	0.038	0.072	0.499	0.946	0.981
LQR2	DQ	1.246	2.160	7.974	10.382	2.920	3.756	7.861	2.443	19.018	20.724
	p-value	0.940	0.827	0.158	0.065	0.712	0.710	0.248	0.875	0.004	0.002
	$\hat{\alpha}$	0.050	0.100	0.500	0.900	0.950	0.042	0.070	0.506	0.946	0.981
LQR3	DQ	2.022	1.705	3.206	3.291	6.192	3.864	8.273	3.428	18.097	20.724
	p-value	0.846	0.888	0.668	0.655	0.288	0.695	0.219	0.753	0.006	0.002
	$\hat{\alpha}$	0.050	0.100	0.500	0.900	0.950	0.067	0.080	0.506	0.938	0.976
QRNN LQR1	DQ	4.103	4.079	8.544	14.397	5.849	5.465	5.962	5.013	12.391	14.319
	p-value	0.535	0.538	0.129	0.013	0.321	0.486	0.427	0.542	0.054	0.026
	$\hat{\alpha}$	0.051	0.101	0.501	0.900	0.950	0.042	0.078	0.490	0.930	0.976
QRNN LQR2	DQ	1.317	1.752	7.502	7.501	2.054	5.981	6.355	5.243	10.912	14.824
	p-value	0.933	0.882	0.186	0.186	0.842	0.425	0.385	0.513	0.091	0.022
	$\hat{\alpha}$	0.050	0.101	0.501	0.900	0.950	0.043	0.075	0.501	0.944	0.978
QRNN LQR3	DQ	1.093	7.114	3.378	5.975	5.824	4.725	7.379	4.137	20.426	16.480
	p-value	0.955	0.212	0.642	0.309	0.324	0.580	0.287	0.658	0.002	0.011

Where α is the quantile, $\hat{\alpha}$ is the unconditional coverage, DQ is the Berkowitz *et al.* (2011) test statistic for correct dynamic specification, p-value is the corresponding p-value to the DQ.

90% half of them is well specified and half is not and in case of 95% quantile the DQ shows that all of the models are not well specified.

Table 6.6: Absolute performance of models on in-sample and out-of-sample forecasts for S&P 500 futures return quantiles.

		In-sample					Out-of-sample				
	α	0.05	0.1	0.5	0.9	0.95	0.05	0.1	0.5	0.9	0.95
	$\hat{\alpha}$	0.050	0.100	0.500	0.900	0.950	0.034	0.077	0.473	0.929	0.963
LQR1	DQ	1.734	2.907	15.814	19.869	15.147	11.495	19.082	3.405	17.680	10.775
	p-value	0.885	0.714	0.007	0.001	0.010	0.074	0.004	0.757	0.007	0.096
	$\hat{\alpha}$	0.050	0.100	0.500	0.900	0.950	0.029	0.077	0.482	0.936	0.977
LQR2	DQ	3.801	3.340	13.985	8.351	4.965	9.741	14.517	1.462	24.458	17.088
	p-value	0.578	0.648	0.016	0.138	0.420	0.136	0.024	0.962	0.000	0.009
	$\hat{\alpha}$	0.050	0.100	0.500	0.900	0.950	0.029	0.079	0.474	0.937	0.976
LQR3	DQ	3.815	2.371	13.033	11.101	3.764	12.683	13.395	3.488	23.083	15.524
	p-value	0.576	0.796	0.023	0.049	0.584	0.048	0.037	0.746	0.001	0.017
	$\hat{\alpha}$	0.050	0.100	0.500	0.900	0.950	0.032	0.080	0.474	0.928	0.963
QRNN LQR1	DQ	1.986	3.845	17.538	17.154	15.394	12.548	17.624	2.666	17.840	10.775
	p-value	0.851	0.572	0.004	0.004	0.009	0.051	0.007	0.850	0.007	0.096
	$\hat{\alpha}$	0.050	0.101	0.502	0.901	0.950	0.031	0.080	0.471	0.936	0.973
QRNN LQR2	DQ	3.053	2.198	15.471	23.234	4.128	12.141	15.876	3.343	22.944	13.008
	p-value	0.692	0.821	0.009	0.000	0.531	0.059	0.014	0.765	0.001	0.043
	$\hat{\alpha}$	0.051	0.101	0.500	0.900	0.950	0.031	0.079	0.474	0.941	0.979
QRNN LQR3	DQ	3.129	1.984	22.252	7.677	3.871	13.569	19.454	2.351	23.726	18.875
	p-value	0.680	0.851	0.000	0.175	0.568	0.035	0.003	0.885	0.001	0.004

Where α is the quantile, $\hat{\alpha}$ is the unconditional coverage, DQ is the Berkowitz *et al.* (2011) test statistic for correct dynamic specification, p-value is the corresponding p-value to the DQ.

The situation of unconditional coverage for S&P 500 futures (Table 6.6) is similar to WTI. In in-sample the unconditional coverage is close to perfect, in out-of-sample it is shifted away from 50% quantile. What changes is the DQ in in-sample for 50%, 90% and 95% quantiles where the models (in most cases) are not well dynamically specified. At 5% significance level and in the case of out-of-sample only the 5% and 50% quantiles seem to be well specified.

6.2.2 HARQ - Absolute performance

The absolute performance of heterogenous quantile autoregression models (HARQ) where the dependent variable is the realized volatility is reported here for both the linear case and the neural network. Again the DQ specification is tested for both S&P 500 and WTI Crude Oil, for only one foresting horizon being

Table 6.7: Absolute performance of models on in-sample and out-of-sample forecasts for WTI Crude Oil futures realized volatility quantiles.

		In-sample				Out-of-sample			
	α	0.5	0.75	0.9	0.95	0.5	0.75	0.9	0.95
	$\hat{\alpha}$	0.501	0.750	0.900	0.950	0.539	0.776	0.906	0.949
HARQ1	DQ	12.464	2.567	2.929	8.581	8.030	7.045	9.894	13.361
	p-value	0.029	0.766	0.711	0.127	0.236	0.317	0.129	0.038
	$\hat{\alpha}$	0.500	0.750	0.900	0.949	0.606	0.838	0.947	0.971
HARQ2	DQ	13.594	3.769	3.149	3.009	36.352	34.684	28.780	11.372
	p-value	0.018	0.583	0.677	0.699	0.000	0.000	0.000	0.078
	$\hat{\alpha}$	0.500	0.750	0.900	0.950	0.616	0.832	0.944	0.976
HARQ3	DQ	17.258	5.628	3.199	2.394	38.996	33.548	27.263	14.681
	p-value	0.004	0.344	0.669	0.792	0.000	0.000	0.000	0.023
	$\hat{\alpha}$	0.501	0.750	0.900	0.950	0.565	0.784	0.914	0.958
QRNN HARQ1	DQ	17.137	6.979	1.642	2.109	16.991	11.574	12.574	14.936
	p-value	0.004	0.222	0.896	0.834	0.009	0.072	0.050	0.021
	$\hat{\alpha}$	0.501	0.751	0.900	0.949	0.621	0.848	0.944	0.963
QRNN HARQ2	DQ	13.340	5.602	1.157	3.597	42.191	40.763	27.643	8.807
	p-value	0.020	0.347	0.949	0.609	0.000	0.000	0.000	0.185
	$\hat{\alpha}$	0.500	0.751	0.900	0.950	0.635	0.832	0.936	0.958
QRNN HARQ3	DQ	15.379	3.147	0.830	4.208	49.169	31.682	13.693	16.146
	p-value	0.009	0.677	0.975	0.520	0.000	0.000	0.033	0.013

Where α is the quantile, $\hat{\alpha}$ is the unconditional coverage, DQ is the Berkowitz *et al.* (2011) test statistic for correct dynamic specification, p-value is the corresponding p-value to the DQ.

equal to 1 and for in-sample and out-of-sample. The forecasted quantiles are 50%, 75%, 90% and 95%.

The unconditional coverage in in-sample is again close to perfect and in the case of out-of-sample it suffers form the same problem as LQR models do. The dynamic specification of realized volatility models for WTI (Table 6.7) seem to be incorrectly specified for 50% quantile in both the in-sample and most of the out-of-sample models - only the HARQ1 is well specified. In in-sample the 75%, 90% and 95% quantiles are well specified. The out-of-sample is in most cases incorrectly specified, only the HARQ1 seems to be ok.

The S&P 500 (Table 6.8) has close to perfect unconditional coverage in in-sample and in out-of-sample the unconditional coverages are higher than the true value. In in-sample the 50% and 75% quantiles are mostly incorrectly specified and the 90% and 95% are correctly specified. In out-of-sample the HARQ2,

Table 6.8: Absolute performance of models on in-sample and out-of-sample forecasts for S&P 500 futures realized volatility quantiles.

		In-sample				Out-of-sample			
	α	0.5	0.75	0.9	0.95	0.5	0.75	0.9	0.95
	$\hat{\alpha}$	0.500	0.750	0.900	0.950	0.529	0.749	0.907	0.953
HARQ1	DQ	14.682	11.208	3.430	7.667	4.320	1.391	7.070	2.450
	p-value	0.012	0.047	0.634	0.176	0.633	0.966	0.314	0.874
	$\hat{\alpha}$	0.500	0.750	0.900	0.950	0.603	0.790	0.931	0.973
HARQ2	DQ	35.401	13.297	6.692	1.393	31.797	6.355	20.539	11.278
	p-value	0.000	0.021	0.245	0.925	0.000	0.385	0.002	0.080
	$\hat{\alpha}$	0.500	0.750	0.900	0.950	0.601	0.804	0.941	0.973
HARQ3	DQ	39.315	21.389	6.652	5.416	35.077	13.155	19.886	11.278
	p-value	0.000	0.001	0.248	0.367	0.000	0.041	0.003	0.080
	$\hat{\alpha}$	0.500	0.750	0.900	0.951	0.526	0.744	0.905	0.953
QRNN HARQ1	DQ	7.107	12.514	1.814	4.563	5.518	4.703	3.671	5.127
	p-value	0.213	0.028	0.874	0.472	0.479	0.582	0.721	0.528
	$\hat{\alpha}$	0.501	0.750	0.900	0.950	0.572	0.785	0.924	0.974
QRNN HARQ2	DQ	10.276	18.826	2.834	2.249	16.661	6.843	9.274	12.904
	p-value	0.068	0.002	0.726	0.814	0.011	0.336	0.159	0.045
	$\hat{\alpha}$	0.500	0.750	0.900	0.950	0.590	0.794	0.939	0.982
QRNN HARQ3	DQ	15.531	19.540	6.051	1.241	28.763	12.556	18.245	21.029
	p-value	0.008	0.002	0.301	0.941	0.000	0.051	0.006	0.002

Where α is the quantile, $\hat{\alpha}$ is the unconditional coverage, DQ is the Berkowitz *et al.* (2011) test statistic for correct dynamic specification, p-value is the corresponding p-value to the DQ.

HARQ3 and their corresponding QRNN models are in most cases incorrectly specified. The HARQ1 and QRNN HARQ1 are specified correctly, where these models are the models with dependent variables being lagged volatility, 5-day average volatility and 22-day average volatility.

6.3 Relative performance

Relative performance of two models is compared through the DM test statistic Equation 4.18 which is based on the comparison of the tick-loss function (Diebold 2015; Clements *et al.* 2008). The reported values are the unconditional coverage, mean tick-loss value and the DM test statistic which is reported only for the models that are compared to the benchmark models, to the corresponding linear models.

6.3.1 LQR - Relative performance

Generally we can say that in in-sample, the QRNN performed better (on 5% significance level) and on out-of-sample the methods were equivalent, in some cases linear model was better and in some cases the neural network was better.

For WTI and return quantiles (Table 6.9) there is no case in which the linear method would outperform the neural net. In most cases the methods are statistically equivalent, the methods were statistically indistinguishable on 5% significance level. There is no case where the QRNN has positive DM, which means that the QRNN are better performing, but the majority is not on the 5% significance level.

In the case of out-of-sample forecasts of WTI return quantiles (Table 6.10), most of the methods were similarly good. In few cases the linear method was better and in few cases the neural network method performed better. The QRNN performed generally better on 5-step-ahead and 10-step-ahead forecasts, but not significantly enough in most cases. The unconditional coverages are close to their true values, but not as close as in the case of in-sample.

The case of S%P 500 futures in-sample return quantiles (Table 6.11) work similarly as WTI, in some cases the neural networks perform statistically better (at 5%), but there is no situation where they would perform worse, where the DM is positive. When LQR3 and QRNN LQR3 are compared then the neural net works in 7 out of 15 cases better then the linear model and in the rest they are not statistically distinguishable. The unconditional coverage values are close to perfect.

In S&P 500 return quantiles on out-of-sample (Table 6.12), we can say that only in 4 cases the neural net is worse than the linear model, but in the rest of cases (most of them), their performance is similar (on 5% significance level).

Table 6.9: Relative performance of models for in-sample forecasts of WTI Crude Oil futures return quantiles.

		h=1					h=5					h=10				
	α	0.05	0.1	0.5	0.9	0.95	0.05	0.1	0.5	0.9	0.95	0.05	0.1	0.5	0.9	0.95
LQR1	$\hat{\alpha}$	0.050	0.100	0.500	0.900	0.950	0.050	0.100	0.500	0.900	0.950	0.050	0.100	0.500	0.900	0.950
	$\hat{L}$	0.197	0.325	0.687	0.308	0.184	0.197	0.323	0.686	0.308	0.185	0.200	0.328	0.686	0.310	0.185
QRNN LQR1	$\hat{\alpha}$	0.050	0.100	0.500	0.900	0.950	0.050	0.101	0.501	0.900	0.950	0.051	0.100	0.500	0.900	0.950
	$\hat{L}$	0.195	0.323	0.685	0.308	0.182	0.195	0.322	0.684	0.306	0.183	0.198	0.325	0.685	0.309	0.185
	DM	-1.553	-1.148	-1.022	-0.725	-2.027^{+}	-1.421	-0.831	-0.834	-1.054	-0.920	-1.022	-0.979	-0.881	-0.447	-0.439
LQR2	$\hat{\alpha}$	0.050	0.100	0.500	0.900	0.950	0.050	0.100	0.500	0.900	0.950	0.050	0.100	0.500	0.900	0.950
	$\hat{L}$	0.192	0.319	0.686	0.307	0.183	0.193	0.318	0.685	0.305	0.183	0.197	0.325	0.685	0.307	0.183
QRNN LQR2	$\hat{\alpha}$	0.051	0.101	0.501	0.900	0.950	0.051	0.100	0.500	0.900	0.950	0.051	0.100	0.501	0.900	0.950
	$\hat{L}$	0.190	0.315	0.680	0.303	0.179	0.191	0.315	0.680	0.303	0.180	0.194	0.319	0.681	0.304	0.181
	DM	-1.689^{+}	-2.576^{+}	-2.753^{+}	-2.480^{+}	-1.076	-1.176	-1.083	-1.648^{+}	-2.024^{+}	-1.321	-1.779^{+}	-1.416	-1.038	-1.360	-1.173
LQR3	$\hat{\alpha}$	0.050	0.100	0.500	0.900	0.950	0.050	0.100	0.500	0.900	0.950	0.050	0.100	0.500	0.900	0.950
	$\hat{L}$	0.193	0.319	0.685	0.306	0.182	0.193	0.319	0.685	0.306	0.183	0.196	0.324	0.686	0.308	0.183
QRNN LQR3	$\hat{\alpha}$	0.050	0.101	0.501	0.900	0.950	0.051	0.101	0.500	0.900	0.950	0.050	0.101	0.500	0.900	0.950
	$\hat{L}$	0.190	0.314	0.680	0.303	0.179	0.188	0.315	0.678	0.303	0.180	0.191	0.316	0.680	0.303	0.179
	DM	-1.559	-2.322^{+}	-1.792^{+}	-2.069^{+}	-1.737^{+}	-1.352	-1.740^{+}	-1.655^{+}	-1.250	-1.409	-1.997^{+}	-1.364	-1.309	-1.513	-1.139

Where h is the forecast horizon, α is the quantile, $\hat{\alpha}$ is the unconditional coverage, $\hat{L}$ the value of tick-loss function and DM the Diebold-Mariano test statistic for comparison of predictive accuracy between QR and QRNN, with + denoting significantly more accurate QRNN and * denoting that linear models are significantly more accurate.

Table 6.10: Relative performance of models for out-of-sample forecasts of WTI Crude Oil futures return quantiles.

		h=1					h=5					h=10				
	α	0.05	0.1	0.5	0.9	0.95	0.05	0.1	0.5	0.9	0.95	0.05	0.1	0.5	0.9	0.95
LQR1	$\hat{\alpha}$	0.056	0.091	0.515	0.928	0.978	0.060	0.089	0.506	0.921	0.963	0.049	0.091	0.507	0.924	0.964
	$\hat{L}$	0.158	0.258	0.508	0.223	0.132	0.160	0.260	0.508	0.235	0.141	0.160	0.262	0.508	0.235	0.143
QRNN LQR1	$\hat{\alpha}$	0.067	0.080	0.506	0.938	0.976	0.060	0.084	0.503	0.924	0.969	0.046	0.073	0.501	0.927	0.969
	$\hat{L}$	0.181	0.259	0.508	0.225	0.132	0.157	0.261	0.507	0.234	0.140	0.159	0.263	0.509	0.235	0.143
	DM	1.785*	0.555	-0.172	1.852*	-1.047	-1.652+	-0.078	-0.241	-0.330	-1.505	-0.841	0.521	0.890	0.348	-0.494
LQR2	$\hat{\alpha}$	0.038	0.072	0.499	0.946	0.981	0.035	0.084	0.497	0.935	0.976	0.031	0.075	0.488	0.941	0.979
	$\hat{L}$	0.156	0.261	0.509	0.229	0.137	0.160	0.262	0.507	0.237	0.142	0.162	0.264	0.511	0.240	0.145
QRNN LQR2	$\hat{\alpha}$	0.042	0.078	0.490	0.930	0.976	0.050	0.084	0.489	0.935	0.971	0.036	0.078	0.494	0.940	0.976
	$\hat{L}$	0.158	0.265	0.516	0.247	0.146	0.157	0.257	0.505	0.242	0.147	0.162	0.264	0.515	0.238	0.145
	DM	1.345	1.072	1.519	1.150	1.220	-1.490	-1.486	-0.503	0.485	0.916	-0.241	-0.129	0.662	-0.430	0.033
LQR3	$\hat{\alpha}$	0.042	0.070	0.506	0.946	0.981	0.035	0.079	0.498	0.932	0.977	0.039	0.075	0.494	0.943	0.974
	$\hat{L}$	0.158	0.262	0.508	0.228	0.137	0.159	0.261	0.507	0.237	0.142	0.162	0.264	0.510	0.239	0.144
QRNN LQR3	$\hat{\alpha}$	0.043	0.075	0.501	0.944	0.978	0.044	0.076	0.495	0.945	0.974	0.037	0.076	0.491	0.943	0.976
	$\hat{L}$	0.161	0.262	0.510	0.239	0.138	0.164	0.262	0.506	0.238	0.146	0.166	0.264	0.518	0.241	0.147
	DM	1.716*	-0.070	0.584	1.941*	0.773	1.120	0.057	-0.284	0.651	1.791*	1.209	0.030	1.277	1.425	1.735*

Where h is the forecast horizon, α is the quantile, $\hat{\alpha}$ is the unconditional coverage, $\hat{L}$ the value of tick-loss function and DM the Diebhold-Mariano test statistic for comparison of predictive accuracy between QR and QRNN, with + denoting significantly more accurate QRNN and * denoting that linear models are significantly more accurate.

Table 6.11: Relative performance of models for in-sample forecasts of S&P 500 futures return quantiles.

		h=1					h=5					h=10				
	α	0.05	0.1	0.5	0.9	0.95	0.05	0.1	0.5	0.9	0.95	0.05	0.1	0.5	0.9	0.95
LQR1	$\hat{\alpha}$	0.050	0.100	0.500	0.900	0.950	0.050	0.100	0.500	0.900	0.950	0.050	0.100	0.500	0.900	0.950
	$\hat{L}$	0.111	0.184	0.384	0.170	0.101	0.116	0.188	0.384	0.177	0.107	0.119	0.191	0.384	0.177	0.108
QRNN LQR1	$\hat{\alpha}$	0.050	0.100	0.500	0.900	0.950	0.050	0.100	0.500	0.900	0.950	0.050	0.100	0.500	0.900	0.950
	$\hat{L}$	0.111	0.183	0.384	0.169	0.100	0.116	0.186	0.384	0.176	0.106	0.119	0.190	0.384	0.176	0.107
	DM	-1.194	-0.802	-0.227	-1.457	-1.235	-0.210	-1.777^{+}	-0.635	-1.612	-1.559	-0.922	-0.938	-0.743	-0.912	-0.641
LQR2	$\hat{\alpha}$	0.050	0.100	0.500	0.900	0.950	0.050	0.100	0.500	0.900	0.950	0.050	0.100	0.500	0.900	0.950
	$\hat{L}$	0.111	0.183	0.384	0.167	0.098	0.114	0.186	0.384	0.174	0.105	0.118	0.189	0.384	0.175	0.107
QRNN LQR2	$\hat{\alpha}$	0.050	0.100	0.500	0.900	0.950	0.051	0.101	0.502	0.900	0.950	0.050	0.100	0.500	0.900	0.950
	$\hat{L}$	0.108	0.180	0.382	0.165	0.097	0.113	0.185	0.382	0.171	0.103	0.117	0.188	0.383	0.174	0.105
	DM	-2.653^{+}	-1.529	-1.617	-1.723^{+}	-1.754^{+}	-1.371	-1.286	-1.578	-1.323	-1.449	-0.902	-0.984	-1.178	-1.034	-1.121
LQR3	$\hat{\alpha}$	0.050	0.101	0.502	0.901	0.950	0.050	0.100	0.500	0.900	0.950	0.050	0.100	0.500	0.900	0.950
	$\hat{L}$	0.111	0.183	0.384	0.166	0.098	0.114	0.187	0.384	0.174	0.105	0.118	0.189	0.384	0.175	0.107
QRNN LQR3	$\hat{\alpha}$	0.051	0.101	0.500	0.900	0.950	0.050	0.100	0.500	0.900	0.950	0.050	0.100	0.500	0.900	0.950
	$\hat{L}$	0.108	0.181	0.381	0.165	0.096	0.112	0.184	0.381	0.171	0.103	0.117	0.188	0.382	0.173	0.104
	DM	-2.396^{+}	-1.525	-2.367^{+}	-1.711^{+}	-1.550	-1.684^{+}	-1.682^{+}	-1.667^{+}	-1.481	-1.560	-1.416	-1.407	-1.392	-1.232	-1.655^{+}

Where h is the forecast horizon, α is the quantile, $\hat{\alpha}$ is the unconditional coverage, $\hat{L}$ the value of tick-loss function and DM the Diebhold-Mariano test statistic for comparison of predictive accuracy between QR and QRNN, with + denoting significantly more accurate QRNN and * denoting that linear models are significantly more accurate.

The unconditional coverages follow the true values but they are not as well behaved as in the in-sample case.

6.3.2 HARQ - Relative performance

For the forecasting of realized volatility we can say that in the in-sample the neural networks dominate the linear models. 68 out of 72 pairwise comparisons shows better performance of QRNN, in the rest the difference in performance is statistically insignificant. In the case of out-of-sample they seem to perform similarly, in most cases the models do not show statistical difference in relative performance. Only in few cases QRNN or QR is better.

In the case of WTI Crude Oil in-sample (Table 6.13), the unconditional coverage is close to perfect and in all cases the QRNN outperforms the linear models. Whereas for out-of-sample (Table 6.14) the unconditional coverage is not well behaved especially for the 50% and 75% quantiles of HARQ2, HARQ3 and their corresponding neural network models, across all the forecasting horizons, in the rest of the cases the behavior is better.

S&P 500 realized volatility models in-sample (Table 6.15) has conditional coverage as every in-sample estimation close to perfect, the out-of-sample is not well behaved, but it can be tolerated. The relative performance is in favour of neural network models with 32 out of 36 cases being statistically significant and 2 out of 36 just not significant at 5%. Three of the cases where the QRNN is not significantly better is in the case HARQ1 and forecasting horizon being 10 days and one is also in the HARQ1 but for forecasting horizon being 5 days. In the case of out-of-sample the models perform equally well with one case where the linear model is better than the neural network model (in HARQ2, 10-step-ahead and 95% quantile).

Table 6.12: Relative performance of models for out-of-sample forecasts of S&P 500 futures return quantiles.

		h=1					h=5					h=10				
	α	0.05	0.1	0.5	0.9	0.95	0.05	0.1	0.5	0.9	0.95	0.05	0.1	0.5	0.9	0.95
LQR1	$\hat{\alpha}$	0.034	0.077	0.473	0.929	0.963	0.037	0.079	0.475	0.925	0.956	0.039	0.077	0.474	0.922	0.962
	$\hat{L}$	0.089	0.147	0.301	0.126	0.078	0.095	0.154	0.298	0.134	0.083	0.098	0.156	0.298	0.134	0.082
QRNN LQR1	$\hat{\alpha}$	0.032	0.080	0.474	0.928	0.963	0.037	0.079	0.475	0.929	0.956	0.039	0.077	0.474	0.923	0.966
	$\hat{L}$	0.089	0.148	0.301	0.127	0.078	0.094	0.154	0.299	0.134	0.083	0.099	0.156	0.299	0.134	0.082
	DM	0.342	1.239	-0.137	0.819	0.302	-0.768	-0.694	1.164	-0.241	-0.359	1.006	0.392	0.301	0.570	-0.494
LQR2	$\hat{\alpha}$	0.029	0.077	0.482	0.936	0.977	0.037	0.068	0.473	0.938	0.969	0.034	0.064	0.467	0.936	0.977
	$\hat{L}$	0.091	0.147	0.300	0.127	0.076	0.096	0.154	0.299	0.134	0.082	0.100	0.157	0.299	0.134	0.083
QRNN LQR2	$\hat{\alpha}$	0.031	0.080	0.471	0.936	0.973	0.037	0.070	0.476	0.930	0.961	0.034	0.064	0.471	0.931	0.974
	$\hat{L}$	0.092	0.153	0.298	0.126	0.077	0.096	0.155	0.298	0.134	0.081	0.102	0.157	0.300	0.134	0.083
	DM	0.594	2.035*	-1.025	-1.349	2.603*	-0.023	0.914	-0.525	-0.050	-1.084	1.233	0.917	0.927	-0.035	-0.232
LQR3	$\hat{\alpha}$	0.029	0.079	0.474	0.937	0.976	0.037	0.066	0.473	0.940	0.969	0.033	0.064	0.475	0.935	0.977
	$\hat{L}$	0.091	0.148	0.300	0.127	0.076	0.096	0.154	0.298	0.134	0.082	0.100	0.157	0.299	0.134	0.083
QRNN LQR3	$\hat{\alpha}$	0.031	0.079	0.474	0.941	0.979	0.037	0.070	0.470	0.935	0.966	0.031	0.075	0.479	0.925	0.969
	$\hat{L}$	0.093	0.154	0.301	0.126	0.076	0.098	0.154	0.301	0.133	0.082	0.100	0.157	0.299	0.134	0.084
	DM	2.874*	2.020*	0.893	0.319	0.568	1.626	0.529	1.230	-0.686	-0.152	-0.426	0.258	0.253	0.303	0.883

Where h is the forecast horizon, α is the quantile, $\hat{\alpha}$ is the unconditional coverage, $\hat{L}$ the value of tick-loss function and DM the Diebhold-Mariano test statistic for comparison of predictive accuracy between QR and QRNN, with + denoting significantly more accurate QRNN and * denoting that linear models are significantly more accurate.

Table 6.13: Relative performance of models for in-sample forecasts WTI Crude Oil futures realized volatility.

		h=1				h=5				h=10			
	α	0.5	0.75	0.9	0.95	0.5	0.75	0.9	0.95	0.5	0.75	0.9	0.95
HARQ1	$\hat{\alpha}$	0.501	0.750	0.900	0.950	0.500	0.750	0.900	0.950	0.500	0.750	0.900	0.950
	$\hat{L}$	0.142	0.128	0.080	0.052	0.159	0.143	0.092	0.060	0.170	0.155	0.098	0.063
QRNN HARQ1	$\hat{\alpha}$	0.501	0.750	0.900	0.950	0.500	0.750	0.900	0.950	0.500	0.751	0.901	0.950
	$\hat{L}$	0.140	0.125	0.077	0.050	0.157	0.141	0.089	0.058	0.167	0.152	0.096	0.061
	DM	-3.466^{+}	-3.026^{+}	-3.684^{+}	-2.714^{+}	-2.009^{+}	-1.911^{+}	-2.611^{+}	-2.117^{+}	-1.842^{+}	-1.846^{+}	-1.849^{+}	-1.812^{+}
HARQ2	$\hat{\alpha}$	0.500	0.750	0.900	0.950	0.500	0.751	0.900	0.951	0.500	0.750	0.901	0.950
	$\hat{L}$	0.140	0.126	0.079	0.051	0.158	0.142	0.091	0.058	0.169	0.154	0.097	0.062
QRNN HARQ2	$\hat{\alpha}$	0.501	0.751	0.900	0.950	0.502	0.750	0.901	0.950	0.501	0.751	0.900	0.951
	$\hat{L}$	0.136	0.120	0.073	0.045	0.152	0.135	0.086	0.054	0.162	0.146	0.090	0.057
	DM	-5.128^{+}	-5.840^{+}	-4.842^{+}	-5.479^{+}	-2.816^{+}	-2.766^{+}	-3.465^{+}	-2.597^{+}	-3.169^{+}	-2.894^{+}	-3.317^{+}	-2.271^{+}
HARQ3	$\hat{\alpha}$	0.500	0.750	0.900	0.950	0.500	0.750	0.900	0.950	0.500	0.750	0.900	0.950
	$\hat{L}$	0.141	0.126	0.079	0.051	0.158	0.142	0.091	0.058	0.170	0.154	0.097	0.062
QRNN HARQ3	$\hat{\alpha}$	0.500	0.751	0.900	0.950	0.500	0.750	0.901	0.950	0.501	0.750	0.900	0.951
	$\hat{L}$	0.136	0.119	0.074	0.047	0.152	0.137	0.085	0.054	0.163	0.147	0.090	0.056
	DM	-4.653^{+}	-5.576^{+}	-5.098^{+}	-5.072^{+}	-2.835^{+}	-2.399^{+}	-3.123^{+}	-3.056^{+}	-3.173^{+}	-2.867^{+}	-3.208^{+}	-2.663^{+}

Where h is the forecast horizon, α is the quantile, $\hat{\alpha}$ is the unconditional coverage, $\hat{L}$ the value of tick-loss function and DM the Diebold-Mariano test statistic for comparison of predictive accuracy between QR and QRNN, with + denoting significantly more accurate QRNN and * denoting that linear models are significantly more accurate.

Table 6.14: Relative performance of models for out-of-sample forecasts of WTI Crude Oil futures realized volatility.

		h=1				h=5				h=10			
	α	0.5	0.75	0.9	0.95	0.5	0.75	0.9	0.95	0.5	0.75	0.9	0.95
HARQ1	$\hat{\alpha}$	0.539	0.776	0.906	0.949	0.576	0.787	0.921	0.944	0.577	0.802	0.914	0.953
	$\hat{L}$	0.116	0.106	0.069	0.046	0.136	0.122	0.083	0.058	0.144	0.136	0.092	0.062
QRNN HARQ1	$\hat{\alpha}$	0.565	0.784	0.914	0.958	0.623	0.823	0.926	0.956	0.657	0.875	0.935	0.956
	$\hat{L}$	0.118	0.106	0.068	0.045	0.142	0.126	0.085	0.062	0.150	0.144	0.097	0.069
	DM	1.736*	-0.014	-0.215	-0.673	2.305*	1.387	0.772	0.843	2.069*	1.942*	2.344*	2.411*
HARQ2	$\hat{\alpha}$	0.606	0.838	0.947	0.971	0.627	0.845	0.940	0.971	0.663	0.857	0.948	0.971
	$\hat{L}$	0.119	0.109	0.069	0.045	0.141	0.128	0.088	0.060	0.153	0.144	0.098	0.065
QRNN HARQ2	$\hat{\alpha}$	0.621	0.848	0.944	0.963	0.656	0.840	0.934	0.965	0.641	0.833	0.946	0.954
	$\hat{L}$	0.122	0.111	0.068	0.046	0.148	0.132	0.092	0.063	0.152	0.150	0.101	0.071
	DM	2.082*	1.170	-0.852	0.534	2.068*	1.520	1.627	1.897*	-0.292	1.357	0.690	1.922
HARQ3	$\hat{\alpha}$	0.616	0.832	0.944	0.976	0.629	0.844	0.940	0.971	0.667	0.855	0.948	0.969
	$\hat{L}$	0.119	0.109	0.070	0.046	0.141	0.128	0.088	0.060	0.153	0.145	0.099	0.065
QRNN HARQ3	$\hat{\alpha}$	0.635	0.832	0.936	0.959	0.656	0.824	0.940	0.969	0.667	0.857	0.945	0.946
	$\hat{L}$	0.122	0.111	0.069	0.048	0.144	0.132	0.090	0.064	0.157	0.155	0.107	0.076
	DM	1.920*	1.151	-0.170	0.940	0.760	1.121	0.515	1.105	0.920	2.143*	1.992*	3.083*

Where h is the forecast horizon, α is the quantile, $\hat{\alpha}$ is the unconditional coverage, $\hat{L}$ the value of tick-loss function and DM the Diebhold-Mariano test statistic for comparison of predictive accuracy between QR and QRNN, with + denoting significantly more accurate QRNN and * denoting that linear models are significantly more accurate.

Table 6.15: Relative performance of models for in-sample forecasts of S&P 500 futures realized volatility.

		h=1				h=5				h=10			
	α	0.5	0.75	0.9	0.95	0.5	0.75	0.9	0.95	0.5	0.75	0.9	0.95
HARQ1	$\hat{\alpha}$	0.500	0.750	0.900	0.950	0.500	0.750	0.900	0.950	0.500	0.750	0.900	0.950
	$\hat{L}$	0.093	0.085	0.055	0.037	0.114	0.109	0.073	0.049	0.125	0.121	0.084	0.058
QRNN HARQ1	$\hat{\alpha}$	0.500	0.750	0.900	0.951	0.500	0.750	0.900	0.950	0.501	0.750	0.900	0.950
	$\hat{L}$	0.091	0.084	0.053	0.035	0.112	0.106	0.072	0.048	0.123	0.119	0.082	0.057
	DM	-3.498^{+}	-2.339^{+}	-2.735^{+}	-2.711^{+}	-2.498^{+}	-1.839^{+}	-1.250	-1.934^{+}	-1.807^{+}	-1.638	-1.641	-1.462
HARQ2	$\hat{\alpha}$	0.500	0.750	0.900	0.950	0.500	0.750	0.900	0.950	0.500	0.750	0.900	0.950
	$\hat{L}$	0.089	0.081	0.052	0.035	0.112	0.107	0.072	0.048	0.123	0.120	0.083	0.057
QRNN HARQ2	$\hat{\alpha}$	0.501	0.750	0.900	0.950	0.501	0.751	0.900	0.950	0.500	0.751	0.900	0.950
	$\hat{L}$	0.086	0.078	0.050	0.032	0.109	0.104	0.069	0.045	0.120	0.115	0.079	0.054
	DM	-4.399^{+}	-3.965^{+}	-3.617^{+}	-3.225^{+}	-2.975^{+}	-2.040^{+}	-2.585^{+}	-2.688^{+}	-1.915^{+}	-2.109^{+}	-2.120^{+}	-2.359^{+}
HARQ3	$\hat{\alpha}$	0.500	0.750	0.900	0.950	0.500	0.751	0.900	0.950	0.500	0.750	0.900	0.950
	$\hat{L}$	0.090	0.083	0.053	0.035	0.112	0.107	0.072	0.049	0.123	0.120	0.083	0.057
QRNN HARQ3	$\hat{\alpha}$	0.500	0.750	0.900	0.950	0.500	0.750	0.900	0.950	0.500	0.750	0.900	0.950
	$\hat{L}$	0.087	0.080	0.050	0.033	0.110	0.104	0.070	0.045	0.120	0.116	0.080	0.055
	DM	-3.949^{+}	-3.186^{+}	-3.300^{+}	-3.804^{+}	-2.806^{+}	-2.117^{+}	-1.846^{+}	-2.764^{+}	-1.922^{+}	-2.177^{+}	-2.160^{+}	-1.976^{+}

Where h is the forecast horizon, α is the quantile, $\hat{\alpha}$ is the unconditional coverage, $\hat{L}$ the value of tick-loss function and DM the Diebhold-Mariano test statistic for comparison of predictive accuracy between QR and QRNN, with + denoting significantly more accurate QRNN and * denoting that linear models are significantly more accurate.

Table 6.16: Relative performance of models for out-of-sample forecasts of S&P 500 futures realized volatility.

		h=1				h=5				h=10			
	α	0.5	0.75	0.9	0.95	0.5	0.75	0.9	0.95	0.5	0.75	0.9	0.95
HARQ1	$\hat{\alpha}$	0.529	0.749	0.907	0.953	0.525	0.781	0.927	0.963	0.560	0.792	0.923	0.974
	$\hat{L}$	0.076	0.067	0.042	0.027	0.093	0.086	0.060	0.043	0.098	0.096	0.069	0.051
QRNN HARQ1	$\hat{\alpha}$	0.526	0.744	0.905	0.953	0.512	0.775	0.927	0.959	0.557	0.788	0.926	0.971
	$\hat{L}$	0.075	0.067	0.041	0.025	0.093	0.086	0.061	0.043	0.099	0.097	0.069	0.051
	DM	-0.090	0.430	-0.910	-1.338	-0.044	-0.098	1.601	0.595	1.008	1.518	-0.072	-0.029
HARQ2	$\hat{\alpha}$	0.603	0.789	0.931	0.973	0.593	0.807	0.942	0.968	0.631	0.824	0.951	0.975
	$\hat{L}$	0.076	0.067	0.040	0.026	0.095	0.087	0.061	0.043	0.103	0.099	0.071	0.051
QRNN HARQ2	$\hat{\alpha}$	0.572	0.785	0.924	0.974	0.575	0.802	0.934	0.966	0.634	0.837	0.961	0.977
	$\hat{L}$	0.075	0.066	0.040	0.025	0.095	0.087	0.062	0.043	0.103	0.102	0.073	0.054
	DM	-1.287	-0.673	-0.341	-0.813	0.047	0.461	0.653	0.704	0.017	1.184	1.281	2.719*
HARQ3	$\hat{\alpha}$	0.601	0.804	0.941	0.973	0.588	0.812	0.947	0.968	0.636	0.825	0.949	0.971
	$\hat{L}$	0.076	0.067	0.041	0.027	0.094	0.086	0.061	0.043	0.103	0.098	0.071	0.051
QRNN HARQ3	$\hat{\alpha}$	0.590	0.794	0.939	0.982	0.558	0.801	0.940	0.966	0.632	0.846	0.959	0.977
	$\hat{L}$	0.075	0.067	0.041	0.027	0.094	0.088	0.063	0.046	0.103	0.102	0.073	0.054
	DM	-0.674	0.190	-0.286	0.122	-0.490	0.870	1.795*	1.495	0.221	1.503	1.047	1.318

Where h is the forecast horizon, α is the quantile, $\hat{\alpha}$ is the unconditional coverage, $\hat{L}$ the value of tick-loss function and DM the Diebhold-Mariano test statistic for comparison of predictive accuracy between QR and QRNN, with + denoting significantly more accurate QRNN and * denoting that linear models are significantly more accurate.

6.4 Skewness

Calculation of skewness is based on the Section 4.10 and the estimators G_1 and b_1(Equation 4.20 and Equation 4.21). Neural networks for out-of-sample were trained only once, not 5 times like in the cases above, because of the fact that the estimation is done for each quantile and it is computationally demanding, but with the same number of neurons. The calculated skewnesses for S&P 500 futures and WTI Crude Oil futures both for in-sample and out-of-sample are in Table 6.18 and Table 6.17 (all the results are statistically significantly different from 0, at 0.1%). The one-step-ahead returns and volatility were estimated.

The distributions of expected returns are negatively skewed for both instruments and both in-sample and out-of-sample with the S&P 500 having higher magnitude of skewness than the WTI. In case of WTI the out-of-sample cases are skewed less than the in-sample. The distributions of realized volatility are positively skewed with WTI having slightly lower magnitude of skewness than the S&P. In the case of S&P 500 the out-of-sample skewnesses are little higher in magnitude than the in-sample. The difference between G_1 and b_1 is positive across the models (first is higher than the second) but the differences are negligible.

Table 6.17: Skewness of WTI Crude Oil futures forecasted distributions of returns and volatility

	In-sample		Out-of-sample	
	b_1	G_1	b_1	G_1
QRNN LQR1	-0.130	-0.134	-0.045	-0.046
QRNN LQR2	-0.090	-0.092	-0.067	-0.069
QRNN LQR3	-0.104	-0.108	-0.090	-0.093
QRNN HARQ1	0.905	0.933	0.907	0.935
QRNN HARQ2	0.796	0.820	0.934	0.963
QRNN HARQ3	0.880	0.907	0.851	0.877

All of the estimates of skewness are statistically different from 0 on significance level 0.1%

Table 6.18: Skewness of S&P 500 futures forecasted distributions of returns and volatility

	In-sample		Out-of-sample	
	b_1	G_1	b_1	G_1
QRNN LQR1	-0.213	-0.219	-0.227	-0.234
QRNN LQR2	-0.235	-0.242	-0.233	-0.240
QRNN LQR3	-0.245	-0.253	-0.217	-0.224
QRNN HARQ1	0.897	0.924	1.013	1.045
QRNN HARQ2	0.951	0.980	1.006	1.038
QRNN HARQ3	0.924	0.952	1.074	1.107

All of the estimates of skewness are statistically different from 0 on significance level 0.1%

Chapter 7

Conclusion

This thesis applies new approach to the estimation and forecasting of volatility and return quantiles of financial instruments. Volatility is measured through realized measures and these measures are also used as explanatory variables. The approach newly used for modeling volatility and return quantiles is called Quantile Regression Neural Network, it combines the linear quantile regression with a feedforward neural network to output the quantiles. Having some quantiles of expected volatility is important for risk management, where we can now understand the probability distribution of volatility and not only the expected value. The quantiles of returns are important for risk management from a different perspective, some of the quantiles of returns are used in Value-at-Risk models and this thesis provides different approach to their estimation.

The models provide a good fit on S&P 500 futures and WTI Crude Oil futures both in in-sample and out-of-sample. The realized measures and specifically volatilities that are based on realized measures such as realized volatility, realized semivariance, median realized variance, jump variation and integrated variance, with option implied volatility were used for modeling the returns and volatility. It was showed that volatility and its components are important for predicting the quantiles of returns. Models perform well even though the option implied volatility and especially realized volatility is not consistent over time and even though the clustering behavior of jump variation.

The models were compared to the models suggested by Žikeš & Baruník (2016) in three different forecasting time horizons - one-step-ahead, five-steps-ahead and 10-steps-ahead. For both the WTI Crude Oil and S&P 500 and all forecasting horizons following applies: in the case of returns the QRNN works as good or better than the linear models on in-sample and in the case of out-of-sample it works as good as linear models, but in few cases the QRNN have inferior performace. For quantiles of realized volatility the models on in-sample have better performance than the linear models (in few cases their performance is statistically indistinguishable) on out-of-sample in the case of S&P 500 they are equally good and for WTI Crude Oil the linear models and neural network models are in most cases equally good and few cases the QRNN have lower relative performance.

The distribution skewness of expected (one-step-ahead) returns and volatility was estimated with two estimators, over both the S&P 500 and WTI Crude Oil and in-sample and out-of-sample. The skewness of expected distribution of returns is negative for both instruments and both in-sample and out-of-sample. The WTI has lower magnitude of the skewness than S&P 500 and the skewness of out-of-sample is lower than the in-sample. The realized volatility experiences positive skewness with the WTI being slightly lower in magnitude than the S&P 500 and in the case of S&P 500 the out-of-sample being little higher in magnitude.

Overall the models perform well. The linear models and the neural network models have equal performance, only in the case of realized volatility quantiles and for in-sample the neural network models dominate the linear models. Not all techniques and all model specifications were tried so there may be estimation techniques and specifications that could improve the performance and provide more precise forecasts.

Bibliography

ANDERSEN, T. G. & T. BOLLERSLEV (1998): "Answering the skeptics: Yes, standard volatility models do provide accurate forecasts." *International Economic Review* **39(4)**: pp. 885–905.

ANDERSEN, T. G., T. BOLLERSLEV, F. X. DIEBOLD, & P. LABYS (2003): "Modeling and forecasting realized volatility." *Econometrica* **71(2)**: pp. 579–625.

ANDERSEN, T. G., D. DOBREV, & E. SCHAUMBURG (2012): "Jump-robust volatility estimation using nearest neighbor truncation." *Journal of Econometrics* **169(1)**: pp. 75–93.

BARNDORFF-NIELSEN, O. E., S. KINNEBROCK, & N. SHEPHARD (2010): *Measuring Downside Risk - Realized Semivariance.* Volatility and Time Series Econometrics: Essays in Honor of Robert Engle.

BARNDORFF-NIELSEN, O. E. & N. SHEPHARD (2003): "Realized power variation and stochastic volatility models." *Bernoulli* **9(2)**: pp. 243–265.

BERKOWITZ, J., P. CHRISTOFFERSEN, & D. PELLETIER (2011): "Evaluating value-at-risk models with desk-level data." *Management Science* **57(12)**: pp. 2213–2227.

BOLLERSLEV, T., G. TAUCHEN, & H. ZHOU (2009): "Expected stock returns and variance risk premia." *Review of Financial Studies* **22(11)**: pp. 4463–4492.

CANNON, A. J. (2011): "Quantile regression neural networks: Implementation in r and application to precipitation downscaling." *Computers and Geosciences* **37(9)**: pp. 1277–1284.

CARR, P. & L. WU (2009): "Variance risk premiums." *Review of Financial Studies* **22(3)**: pp. 1311–1341.

Chrislb (2005): "Diagram of an artificial neuron." [Online; accessed July 25, 2016].

Clements, M. P., A. B. GalvĂŁo, & J. H. Kim (2008): "Quantile forecasts of daily exchange rate returns from forecasts of realized volatility." *Journal of Empirical Finance* **15(4)**: pp. 729–750.

Comte, F. & E. Renault (1998): "Long memory in continuous-time stochastic volatility models." *Mathematical Finance* **8(4)**: pp. 291–323.

Corsi, F. (2009): "A simple approximate long-memory model of realized volatility." *Journal of Financial Econometrics* **7(2)**: pp. 174–196.

Cramer, H. (1946): *Mathematical Methods of Statistics*. Princeton: Princeton University Press.

Diebold, F. X. (2015): "Comparing predictive accuracy, twenty years later: A personal perspective on the use and abuse of dieboldâ€"mariano tests." *Journal of Business and Economic Statistics* **33(1)**.

Diebold, F. X. & R. S. Mariano (1995): "Comparing predictive accuracy." *Journal of Business and Economic Statistics* **13(3)**: pp. 253–263.

Dorland, W. A. N. (2007): *Dorland's illustrated medical dictionary*. Saunders.

Dumitrescu, E.-I., C. Hurlin, & V. Pham (2012): "Backtesting value-at-risk: From dynamic quantile to dynamic binary tests." Working paper or preprint.

Engle, R. F. & S. Manganelli (2004): "Caviar: Conditional autoregressive value at risk by regression quantiles." *Journal of Business and Economic Statistics* **22(4)**: pp. 367–381.

Giacomini, R. & I. Komunjer (2005): "Evaluation and combination of conditional quantile forecasts." *Journal of Business and Economic Statistics* **23(4)**: pp. 416–431.

Hull, J., , & A. White (1998): "Value at risk when daily changes in market variables are not normally distributed." *Journal of Derivatives* **5(Spring)**: pp. 9–19.

Joanes, D. N. & C. A. Gill (1998): "Comparing measures of sample skewness and kurtosis." *Journal of the Royal Statistical Society Series D: The Statistician* **47(1)**: pp. 183–189.

Liu, Y. . & I. . Yeh (2016): "Building valuation model of enterprise values for construction enterprise with quantile neural networks." *Journal of Construction Engineering and Management* **142(2)**.

Newey, W. K. & K. D. West (1987): "A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix." *Econometrica* **55(3)**: pp. 703–08.

Sévi, B. (2014): "Forecasting the volatility of crude oil futures using intraday data." *European Journal of Operational Research* **235(3)**: pp. 643–659.

Tang, Z. & P. Fishwick (1993): "Feedforward neural nets as models for time series forecasting." *ORSA Journal on Computing* **5**: pp. 374–385.

Taylor, J. W. (2000): "A Quantile Regression Neural Network Approach to Estimating the Conditional Density of Multiperiod Returns." *Journal of Forecasting* **19**: pp. 299–311.

Trolle, A. B. & E. S. Schwartz (2010): "Variance risk premia in energy commodities." *Journal of Derivatives* **17(3)**: pp. 15–32.

Žikeš, F. & J. Baruník (2016): "Semi-parametric Conditional Quantile Models for Financial Returns and Realized Volatility." *Journal of Financial Econometrics* **14(1)**: pp. 185–226.

Yeh, I. . (2014): "Estimating distribution of concrete strength using quantile regression neural networks." *Applied Mechanics and Materials* **584-586**: pp. 1017–1025.

Zhu, Q., Q. . Lu, X. . Zhou, & K. K. Lai (2014): "A driving force analysis and forecast for gas consumption demand in china." *Mathematical Problems in Engineering* **2014**.