Option Pricing Using Machine Learning Techniques

A thesis submitted in partial fulfilment of the requirements for the degree of Master of Science in Economics in the Department of Economics of the University of Konstanz

By: Pavel Podkovyrov (752453)

Rheingut Str. 30

78462 Konstanz

Period of completion:

19th October 2018 - 6th March 2019

1st assessor: Jun. Prof. Dr. Grigoryeva

2nd assessor: Prof. Dr. Jackwerth

Konstanz, 6th March 2019

Contents

1 Introduction **1**

1.1 Background of the Study . . . 1

1.2 Structure of the Study . . . 1

2 Literature Review **2**

2.1 Black-Scholes Model . . . 2

2.1.1 Shortcomings and Extensions of the Black-Scholes Model . . . 4

2.2 Stochastic Volatility . . . 5

2.2.1 Heston Model . . . 6

2.2.2 Shortcomings and Extensions of Heston Model . . . 7

2.2.3 Link with GARCH Models . . . 7

2.2.4 Estimation of Heston Model . . . 8

2.3 Neural Networks . . . 9

2.3.1 Previous Research . . . 12

2.3.2 Shortcomings of Neural Network . . . 13

2.4 Performance Measures . . . 14

3 Empirical Analysis **14**

3.1 Data Description . . . 14

3.2 Simulation Study . . . 17

3.3 Choosing Optimal Number of Nodes for Neural Network . . . 20

3.4 Comparison of Heston Model and Neural Network . . . 22

3.5 Forecast of the Option Price for the Next Day . . . 29

3.6 Futher Research Prospects . . . 35

4 Conclusion **35**

Appendices **37**

Appendix A **37**

Appendix B **38**

Appendix C **39**

List of Figures

1 Example of neural network with 1 hidden layer. 11
2 Evolution of S&P 500 in 2013-2014. 15
3 Daily returns of S&P 500 in 2014 16
4 Examples of simulated prices 18
5 Mean MSE for different number of nodes 21
6 Sorted prices for different models 26
7 Sample of implied volatility for different models 27
8 Histogram of RMSE extended neural network 1 29
9 One-step ahead forecast of realized volatility (2013-2014) 32
10 One-step ahead forecast of different models 34
11 Histogram of errors for different models (out of sample) 38
12 Errors 1 step ahead forecast 39

List of Tables

1 Decription statistics of daily returns S&P 500 in 2014. 16
2 Basic features of the options Wednesday 2013-2014 (Call) 17
3 Summary of the simmulation study 19
4 Evaluation of performance (out of sample) 19
5 Minimum mean of the performance measures 21
6 Summary of the models 22
7 Summary of the test sample 23
8 Estimated parameters of the Heston model 23
9 Performance measures of different models (out of sample) 24
10 Comparison of the models in terms of RMSE 25
11 Ratio of fulfilled condition across the models 28
12 Summary of the test sample (Forecast) 30
13 Models used in price forecasting 30
14 Comparison of 1 step ahead forecast 33
15 Performance forecast original dataset 34
16 P-value of the Diebold and Mariano test 35

1 Introduction

1.1 Background of the Study

The size of the derivative market is continuously growing and in order to increase investment gains and understand option market, researchers have spent years investigating this phenomenon. This study aims to investigate option pricing by employing the Black-Scholes model (Black and Scholes, 1973), the Heston model (Heston, 1993), and neural networks. This research focuses on the pricing of S&P 500 Call options. A rigorous comparison of estimation methods is not the primary goal of this thesis; rather, the objective is to see whether any of the methods can yield more useful results. There is an extensive number of articles that evaluate and compare different option pricing techniques; however, there has not yet been any research that compares stochastic volatility models and neural networks at the same time. Previously, the performance of the Black-Scholes model was compared either to neural networks or stochastic volatility models.

Another dilemma considered in the study is choosing between parametric and nonparametric methods. Parametric pricing methods are heavily dependent on capturing the dynamics of the underlying asset's price process and misspecification in the model will lead to systematic errors. On the contrary, nonparametric approaches like neural networks demand a large amount of data for precise estimation (Hutchinson et al., 1994). The main difference between an econometric approach and a neural network approach is that first one attempts to capture data generating process while the second focuses more on predictions without concern for the data generating process.

1.2 Structure of the Study

Before going into details of the study, I outline its structure. The second chapter provides a theoretical background of the research with a discussion of the Black-Scholes model, stochastic volatility, and neural network. This chapter also includes the justification of the methodology adopted for the study. The third chapter aims to analyse the obtained results, give comments with an emphasis on the implications of the findings, limitations and suggestions for further research. Finally, the last part offers concluding remarks.

2 Literature Review

2.1 Black-Scholes Model

An option is a contract that allows to buy or to sell a specified underlying asset at a fixed price at a particular time in the future (Jarrow and Rudd, 1983). There are different types of options; in order to be able to compare empirical performance with previous literature, this study will focus only on vanilla options. In addition, the study is focusing on European options that are the most traded in the world. Most of the option pricing models are derived under risk-neutral valuation, stating that investors are indifferent to risk and this proposition makes option pricing easier to evaluate as there is no risk premium (Hull, 2003).

The price of the option is called a premium. A premium of a European call option with the price of the underlying asset at time t (S_t) can be computed with Equation 1.

$$
\begin{aligned}
&(S_T - X)_+ = max(S_T - X, 0)\\
&Where:\\
&X -\ Strike\ of\ the\ option\\
&T -\ Expiration\ date\\
&S_T -\ Price\ of\ the\ underlying\ asset\ at\ the\ expiration
\end{aligned}
\tag{1}
$$

The question is how should one price the option at time t that worth $(S_T - X)_+$? To answer this question first of all absense of arbitrage opportunity on a market is assumed that leads to the call/put parity (Equation 2). In addition, a general assumption is that the market is liquid, and it is possible to borrow at the risk-free rate (Lamberton and Lapeyre, 2011).

$$
\begin{aligned}
&C_t - P_t = S_t - Xe^{-r(T-t)}\\
&Where:\\
&C_t -\ Price\ of\ the\ call\ option\\
&P_t -\ Price\ of\ the\ put\ option\\
&r -\ Risk\ free\ rate
\end{aligned}
\tag{2}
$$

Using Equation 2 and assuming that price movement of S_t follows Brownian motion (Equation 3) one can get closed form solution for premium, this was done by Fischer Black and Myron Scholes who introduced the Black-Scholes model.

In this case, only one price of the option is possible and can be found through partial differential equation (PDE) that describes price evolution of the option. Profound derivation of the Black-Scholes model can be found in Franke et al. (2015).

$$
\begin{aligned}
&dS_t = \mu S_t dt + \sigma S_t dW_t \\
&Where: \\
&\mu \ - \ Expected\,return \\
&\sigma \ - \ Volatility\,of\,return\,(constant), \\
&W_t \ - \ Brownian\,motion
\end{aligned}
\tag{3}
$$

Under the Black-Scholes model, price of the call option (C) is determined by Equation 4. C depends only on S and t; other parameters are assumed to be constant and known over time (Black and Scholes, 1973).

$$
\begin{aligned}
&\mathrm{C(S,t)} = \mathrm{N(d_1)S} - \mathrm{N(d_2)Xe}^{-rt} \\
&\mathrm{d}_1 = \frac{1}{\sigma\sqrt{\mathrm{t}}}\left[\ln\left(\frac{S}{X}\right) + t\left(r + \frac{\sigma^2}{2}\right)\right] \\
&\mathrm{d}_2 = d_1 - \sigma\sqrt{\mathrm{t}} \\
&Where: \\
&N(d_i) - Cumulative\,normal\,density\,function \\
&t - Time\,to\,expiration
\end{aligned}
\tag{4}
$$

This parametric model is the most well known and allows to capture specific relationship between parameters that traders concerned about. For example, the higher the ratio between S and X, the higher probability that option will be exercised; therefore price of the option increases. In addition, volatility is positively associated with C as d_1 will be higher if σ enlarges. At the same time, d_2 will be lower making the overall price of the option higher. A similar analysis can be done to conclude that t and r are positively associated with C. Volatility (σ) that is used in the Black-Scholes formula is called implied volatility and not directly observable. It is volatility measure of the underlying asset that is obtained through the Black-Scholes formula using option price of the market. It might be used as the forecast of future volatility; however, the investigation results of correlation between implied and future volatilities are controversial (Christensen and Prabhala, 1998).

Finally, the Black-Scholes model is generally not used in its original form; however, it is often used as a benchmark model. It also gives a delightful example of a parametric model whose assumptions are debatable in the context of financial data, therefore, considered in the current study.

2.1.1 Shortcomings and Extensions of the Black-Scholes Model

The initial version of this model has several strong assumptions that are usually violated on a financial market and these violations became even more significant after the crisis 1987 (Rubinstein, 1994). The most criticised drawback of Black-Scholes model is the constant variance assumption that was rejected by Akgiray (1989). In addition, **?** showed that on any given day, different market option prices had different values of σ. Moreover, differences in implied volatilities, computed through the Black-Scholes formula, vary across moneyness with the minimum of the volatility around at the money options. This feature of implied volatility having a convex shape with respect to the moneyness is called volatility smile. Another feature of implied volatility is that for at the money options implied volatility exhibits a steady rise with maturity (Backus et al., 2004). These patterns are referred to as the deviation from a normal distribution of returns (Baillie and Bollerslev, 1989). Nevertheless, the assumption of normality might have been relevant for the 1970s as Jackwerth and Rubinstein (1996) stated that before the crisis in 1987, price distribution was close to lognormal and only after the crisis it is consistently skewed.

Despite the criticism of the Black-Scholes model Yao et al. (2000) stated that the Black-Scholes model showed good results for pricing at the money options with maturity more than two months.

One of the ways to capture volatility smile might be done through deterministic volatility models. As stated in Buraschi and Jackwerth (2001), these models are useful for several reasons. They are able to capture observed patterns, can be well calibrated, and the market is complete under these models. Implied volatility might be obtained as a function of maturity and moneyness (Kuo and Wang, 2009). This dependency can be, for example, linear, square root and exponential. However, as was mentioned by Amin and Morton (1994), parameters of these models can be unstable. Another way to model deterministic volatility was suggested by Dumas et al. (1998) who modelled volatility as a function of maturities and asset price. The motivation behind this model is the possibility to value an option based on the

Black-Scholes partial differential equation. Dumas et al. (1998) also claimed that option pricing of S&P 500 with this approach was rather exact; however, the hedging and forecasting performance was worse than in models with smoothing volatilities.

Cox (1975) introduced a constant elasticity variance model that adds to the stochastic price process (Equation 3) additional elasticity parameter that determines the relationship between price and volatility. Further, Lauterbach and Schultz (1990) continued the study of Cox (1975) by applying a specific form of the constant elasticity variance model (CEV) in the context of the warrants pricing.

Backus et al. (2004) implemented Gram-Charlier expansion into option pricing to capture volatility smile. This approach takes into account skewness and kurtosis of the distribution of the underlying asset. That is a reasonable trial as observed returns of financial data have sample skewness and kurtosis significantly deviating from the normal with negative skewness and kurtosis larger than 3, representing asymmetric distribution and heavier tails (Cont, 2001). However, skewness obtained with a tailor series expansion that is used to calculate the call option price under the Gram-Charlier method is underestimated in comparison to a true one.

2.2 Stochastic Volatility

Another way of overcoming constant volatility assumed in the Black-Scholes model is introducing volatility as a stochastic process. The price for this more realistic model though is increased complexity of evaluation and methods of optimising become as important as the model itself (Cont and Ben Hamida, 2004).

Wiggins (1987) introduced stochastic volatility model. However, the performance of the model did not change drastically from the Black-Scholes model, and due to this fact model did not become very popular. Hull and White (1987) talked about stochastic volatility that obeys the same structure of the stochastic process as the price in the Black-Scholes model assuming that volatility and stock price are instantaneously uncorrelated.

Stein and Stein (1991) argue that numerical solution obtained using Taylor expansion suggested by Hull and White (1987) is not the most optimal and introduced their closed form solution to option pricing assuming that volatility is driven by mean reverting process. Nevertheless, the weakness of this model is the assumption of uncorrelatedness between stock price and volatility as it does not allow for the skewness effects between spot returns and volatility observed in data to be captured.

Later, Hilliard and Schwartz (1996) implemented stochastic volatility approach combining it with the bivariate binomial model allowing non zero correlation and noted that obtained results are close to obtained by Hull and White (1987), but this model also did not gain popularity.

2.2.1 Heston Model

One of the most popular stochastic volatility models was introduced by Heston (1993), where assumptions of the Black-Scholes model of the volatility being constant and, at the same time, the absence of correlation between stock prices and volatility are relaxed. Equation 5 describes the option pricing under stochastic volatility model suggested by Heston (1993). The first part of the equation is identical to the Black-Sholes model with the main difference being that now volatility also described as a random process. Heston applied the mean reverting volatility process as was introduced by Cox et al. (1985) for the interest rate model and also adapted by Stein and Stein (1991) in option pricing.

Meaning of the parameters in Equation 5 are following: θ represents long-run mean of the volatility process, κ which is rate of reversion displaying how fast the variance converges to long run after the shock and also a degree of volatility clustering, σ stands for volatility of volatility and determines the magnitude of the shock that multiplied by the Brownian motion W_t^2 allowing the volatility to be stochastic. Finally, two Brownian motions are correlated with parameter ρ ensuring that processes are dependent.

$$\begin{aligned}
&dS_t = \mu S_t dt + \sqrt{V_t} S_t dW_t^1 \\
&dV_t = \kappa(\theta - V_t)dt + \sigma\sqrt{V_t} dW_t^2 \\
&dW_t^1 dW_t^2 = \rho dt \\
&Where: \\
&S_t - \ Price\, process \\
&V_t - \ Volatility\, process
\end{aligned} \tag{5}$$

Intuitively, the assumption that $\rho = 0$ is a limitation of the model. If $\rho \neq 0$, then it affects a skewness of the returns. Further, if we assume that $\rho < 0$ volatility will go up when the asset price/return goes down, thus enlarging the left tail and creating a bigger left-tailed distribution that is usually observed in the real financial data.

The σ adds kurtosis in the density function of spot returns. That results in a price decrease of near the money options and increase the far in and out of the money options. Finally, the higher σ the more noticeable is the smile (Heston, 1993).

2.2.2 Shortcomings and Extensions of Heston Model

Further research stated that the Heston model could not get enough skew in the implied volatility in comparison to the market volatility, especially for short term maturity (Mikhailov and Nögel, 2003). In addition, the model fails to acquire the kurtosis completely (Andersen et al., 2002). Christoffersen et al. (2009) introduced the model with the stochastic correlation between volatility and stock returns, and this proposition provides more flexibility of the time variation in the smirk. Performance of the model improved by 23% in comparison to the original Heston model in terms of RMSE. Other modifications have been carried out by Grzelak and Oosterlee (2011) who implemented stochastic interest rates in the Heston model.

The Brownian motion that is assumed in the models above considers continuity of the price path, but empirical analysis suggests that price evolution might not follow the Brownian motion as there are jumps observed in the price dynamics; therefore researchers added Poisson process to Equation 3. The modified process in the literature called Lévy process. Holding everything constant, the Poisson process will push the prices of out of the money options higher. Bates (1996) evaluated the stochastic volatility model with jumps and concluded that the major issue of the model is the instability of parameters due to large standard errors. Bates (1996) also noted that making models based only on more sophisticated descriptions of the underlying asset price process is not necessarily the right way to improve the performance of the option pricing models. Lastly, implementation and more detailed notes on option pricing with jumps can be found in Cont et al. (2004) and Lewis (2009).

2.2.3 Link with GARCH Models

In contrast to the stochastic volatility approach, General Autoregressive Conditional Heteroskedasticity GARCH models use squared returns as a proxy for volatility, where variance in the current period is determined as ARMA process depending on previous values of variance and innovations (Bollerslev, 1986). This idea was introduced by Duan (1995) and Heston and Nandi (2000) derived a closed form solution in GARCH framework for option pricing. Moyaert and Petitjean (2011) stated that Heston Nandi model displayed smaller hedging errors than the Heston model but the Heston model showed smaller in and out of sample MSEs in option pricing during the subprime crisis 2008. Nevertheless, both models performed better than the Black-Scholes model.

2.2.4 Estimation of Heston Model

The Heston model is used to compare its performance to the neural network as it is relatively simple to implement and captures the dynamics of implied volatility. In the analysis part, possible improvements of the Heston model would be taken into account. A model with jump diffusion is not applied because it requires additional parameters to estimate like frequency of jumps that is a computational burden. Finally, due to Engle and Patton (2007) Heston model corresponds to good volatility model as it is mean reverting, persistent, counts for asymmetry and exogenous shock influences volatility.

The original closed form solution of the Heston model is presented in Appendix A. The estimation by arbitrage shortens the pricing of the option to the computation of expectation of the discounted final payoff under probability measure Q that is equivalent to probability measure P under which the asset process is martingale (Fournié et al., 1997). For pricing an option by taking the discounted expected payoff density of the underlying asset is required. For the current model, PDF cannot be used but characteristic function that has one to one correspondence comes in handy. The approach used for derivation is beyond the focus of the thesis and detailed notes can be found in Heston (1993). However, the general idea is the following; first, a mapping of the characteristic function to the payoff is done and the next step is to apply the inversion formula to compute option prices.

In order to compute the option price, the optimal parameters of the Heston model are required. To achieve this the following minimisation problem is solved (Equation 6). The aim is to find the parameters that minimize the squared difference between the true price of the option and one obtained with the Heston model. Where $\Omega(V_0, \kappa, \theta, \sigma, \rho)$ is a set of parameters to be estimated where V_0 is an initial variance

(Mikhailov and Nögel, 2003). Equation 6 is solved with lsqnonlin Matlab function. It minimizes the function of initial parameters but result is dependent on the initial parameters as it is a local optimizer, therefore the chosen initial parameters are close to ones obtained by Mrázek and Pospíšil (2017). More on the nonlinear minimisation can be found in Coleman and Li (1996).

$$\begin{aligned}
&min_{\Omega} \textstyle\sum_1^N (C_i - C_i^{Heston})^2 \\
&subject\ to: \\
&2\kappa\theta > \sigma^2;\ -1 < \rho < 1 \\
&0 < \theta < 1;\ 0 < \sigma < 1
\end{aligned} \tag{6}$$

2.3 Neural Networks

There are different machine learning techniques such that support vector regression, gradient boosted trees, random forests and artificial neural network (Shalev-Shwartz and Ben-David, 2014). For the option pricing task neural networks are the most used, therefore also considered in this thesis (Amilon (2003), Lajbcygier and Connor (1997), Malliaris and Salchenberger (1993)).

A neural network is a nonparametric approach that is increasingly popular today. The reason for this is the availability of a massive amount of data required for robust estimation (Hu et al., 2018). Neural networks are very flexible and allow to discover patterns that are indistinguishable by standard statistical methods (Kaashoek and van Dijk, 2003). The name 'neural network' originates from the analogy to the human brain where neurons are connected, and signals are sent between each other (Simpson, 1990). This paper focuses on modelling the three-layer feedforward neural network with an input layer, one hidden layer, and the output, the justification of this choice is provided below. Feedforward network means that information goes only one direction from the input through the hidden layer and to the output. The current research adopts supervised learning during which outputs from the neural network are compared with the true ones and if they are not equal than weights are updated in accordance with learning rule (Franke et al., 2015). Neural network is a non-linear system that transforms input values over several intermediary steps (cells) to the output variable. The value of the cell in a hidden layer is a weighted sum of regressors from an input layer that is reshaped with the activation function. Activation function allows to construct a non-linear functional relationship between inputs and outputs; without this transformation neural network emerges into a linear regression. In the current study, I apply the sigmoid activation function that squeezes incoming values between zero and one (Equation 7).

$$g(x) = \frac{1}{1+e^{-x}} \tag{7}$$

A weighted sum of the transformed signals is then sent to the cells of the output layer. Weights in the neural network correspond to beta parameters in the linear regression. Equation 8 displays the mathematical formulation of the neural network. The notation to describe neural network is adopted from Kaashoek and van Dijk (2003).

$$\begin{aligned}
&y_j = \textstyle\sum_{h=1}^{H} c_{jh} g(\sum_{i=1}^{I} a_{ih} x_i + b_h) + d_j \\
&Where: \\
&i - Number\ of\ input\ cells,\ i = 1...I \\
&H - Number\ of\ cells\ in\ the\ hidden\ layer \\
&O - Number\ of\ the\ cells\ in\ the\ output \\
&j - Index\ of\ output\ cells\ , j = 1....O;\ for\ the\ current\ problem\ O = 1 \\
&h -\ Index\ of\ hidden\ layer\ cells,\ h = 1...H \\
&c_{jh} -\ Weight\ of\ the\ signal\ from\ hidden\ cell\ h\ to\ output\ cell\ j \\
&g -\ Actication\ function\ (sigmoid) \\
&a_{ih} -\ Weight\ of\ the\ signal\ from\ input\ cell\ i\ to\ hidden\ cell\ h \\
&b_h -\ Constant\ input\ weight\ for\ hidden\ cell\ h \\
&d_j\ - Constant\ weight\ for\ output\ cell\ j
\end{aligned} \tag{8}$$

Figure 1 represents the graphical representation of the one hidden layer neural network where the price of the option is computed through some of the Black-Scholes model inputs. It is worth noticing that all cells in the hidden layer should be connected with the input layer as a fully connected neural network is used in the study. Note that the absence of connections is made for the simplification of the graph.

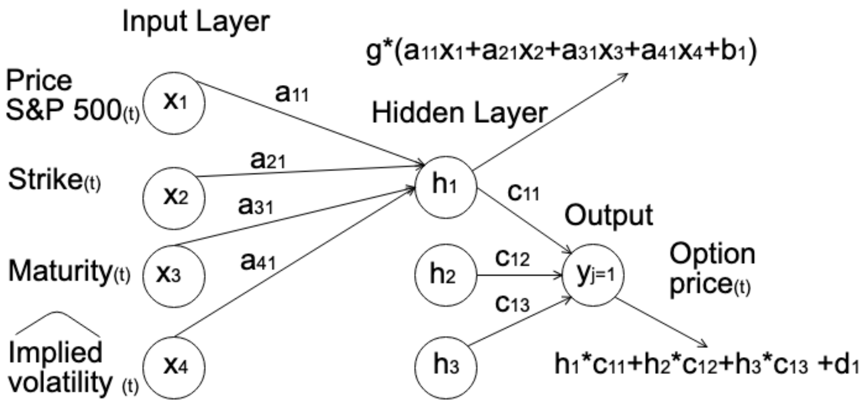


Figure 1: Example of neural network with 1 hidden layer.

The general principle of optimising the neural network is minimising the error that is usually done with a backpropagation method (Equation 9). This method consists of finding a derivative and changing the weights in accordance with the values of the gradient. To implement it I apply the Levenberg-Marquardt algorithm that is efficient and fast (Hagan and Menhaj, 1994). For furhter settings there is no unique rule, therefore default settings of Matlab are used. Derivation and more comprehensive notes can be found in Hagan and Menhaj (1994).

$$\begin{aligned}
&min_\theta \textstyle\sum_{n=1}^{N}(y_n - \hat{y}_n)^2 \\
&Where: \\
&\theta - \; All\, the\, weights\, from\, Equation\, 8 \\
&n - Number\, of\, observations,\; n = 1...N
\end{aligned} \tag{9}$$

In order to avoid overfitting, data is divided into training and testing sample, 80% and 20% respectively. In addition, 10% of the training sample is used as a validation set. The training is stopped when the validation error increases for six iterations in a row or the value of the gradient reaches $1e-07$. In addition, no cross-validation or permutation of the data is applied. Since the initial parameters are random, the network might stop at local minima. Kaastra and Boyd (1996) suggests five to ten random sets of starting weights to reach the global minimum, however, I apply 20 different starting initial values and use the best performing model, assuming that the global minimum is found.

In a setting up neural network, two questions arise. Which structure of the model to choose and how complex the model should be? Leshno et al. (1993) stated that a neural network with one hidden layer and all continuous activation functions except polynomial could approximate any function accurately. However, approximation theory says nothing about the speed of convergence and Hinton et al. (2006) used more than one layer to make algorithm learn faster. Despite this finding, I rely on approximation theory and apply one hidden layer for option pricing as was done in the previous research.

Bailey and Thompson (1990) claimed that the optimal number of nodes in a neural network equals to three-fourths of the number of the input layer. Also, Katz (1992) stated that a number of hidden neurons depends on the number of inputs and should be between one-half to three times of the number of inputs. Kaastra and Boyd (1996) stated that selection of the number of hidden neurons involves investigation and one suggestion was to double the number of nodes until performance on the testing set declines. Methods suggested above are not stable, and Section 3.3 introduces more robust but computationally consuming approach of selecting an optimal structure of the neural network.

Before running the network, values of inputs are scaled to the range between 0 and 1 using Equation 10 to make input values consistent with the type of transfer function that ultimately lets neural network learn faster Kaastra and Boyd (1996).

$$\begin{aligned}
&X_{n,k} = \tfrac{\sigma_{max}-\sigma_{min}}{X^k_{max}-X^k_{min}}(X_{n,k} - X^k_{min}) + \sigma_{min} \\
&Where: \\
&\sigma_{max}\, and\, \sigma_{min} - \; Min\, and\, max\, values\, of\, activation\, function \\
&n = 1....N,\; Number\, of\, observation\, (rows) \\
&k = 1...K,\; Number\, of\, inputs\, (columns)
\end{aligned} \tag{10}$$

2.3.1 Previous Research

In recent years, neural networks gained significant popularity in different fields, the financial sector is no exception. Malliaris and Salchenberger (1993) studied option pricing using one layer neural network and compared it to the Black-Scholes model using MAPE and MSE. Authors used the Black-Scholes inputs and additional inputs: lagged price of the option and market return of the previous day. The proxy for volatility was annualized daily variance. Besides, the structure of the neural network included seven inputs nodes and four hidden nodes respectively. Estimators of the neural network were lower in comparison to the Black-Scholes model. However, both models underpriced in the money options and at the same time displayed the best performance estimating at the money options. Overall, the neural network outperformed the Black-Scholes model in about 50 % of the cases. The criticism of this study can be the size of the dataset used. The authors used only options from January 1, 1990, to June 30, 1990, and with maturities between 30 and 60 days having overall 224 observations that make the results doubtful as neural network requires many observations for decent performance.

Same results were obtained by Anders et al. (1996) who applied one layer feed forward neural network and found that the neural network performed better than the Black-Scholes model in option pricing of the DAX index with the dataset of 13.676 observations. Anders et al. (1996) used four inputs: moneyness, interest rate, volatility and maturity and three hidden nodes in their settings.

Similar results were obtained by Amilon (2003) who concluded that neural networks outperform the Black-Scholes model in estimating Swedish stock index call options from 1997 to 1999.

Gradojevic et al. (2009) applied ensemble method for option pricing by adopting several neural networks. Selection of the neural network was based on 3 groups of moneyness and 3 groups of maturity. Authors concluded that this type of model outperformed standard neural network and the Black-Scholes model as input authors used Moneyness and time to maturity and used ratio of premium/strike as an output variable.

Finally, Kohzadi et al. (1996) forecasted monthly wheat prices with neural network and concluded that neural network performed 56% better than ARIMA model in terms of MSE . Similiar results were obtained by Chakraborty et al. (1992) and Schöneburg (1990) authors concluded that non-linear behaviour could be better captured with neural networks.

2.3.2 Shortcomings of Neural Network

Despite the gains in terms of forecasting, neural networks do have their imperfections. First of all, it is model interpretability as it as not easy to get marginal effects as in OLS. Improved predictions are only measurements that do not tell anything about the underlying process between asset prices and independent variables (Anders et al., 1996).

Another criticism was by Kaastra and Boyd (1996) that claimed that neural networks struggle with replicating stable solutions, require a large number of parameters that must be manually and empirically selected to achieve a reasonable forecast. Moreover, adding complexity of the neural network also increases computation time and might lead to overfitting that results in low out of sample performance.

Lastly, Krauss et al. (2017) used machine learning techniques like deep neural networks, gradient boosted trees and random forest for return predictions, however, they noted that neural networks are more difficult to tune precisely and gave worse results in comparison to other techniques.

2.4 Performance Measures

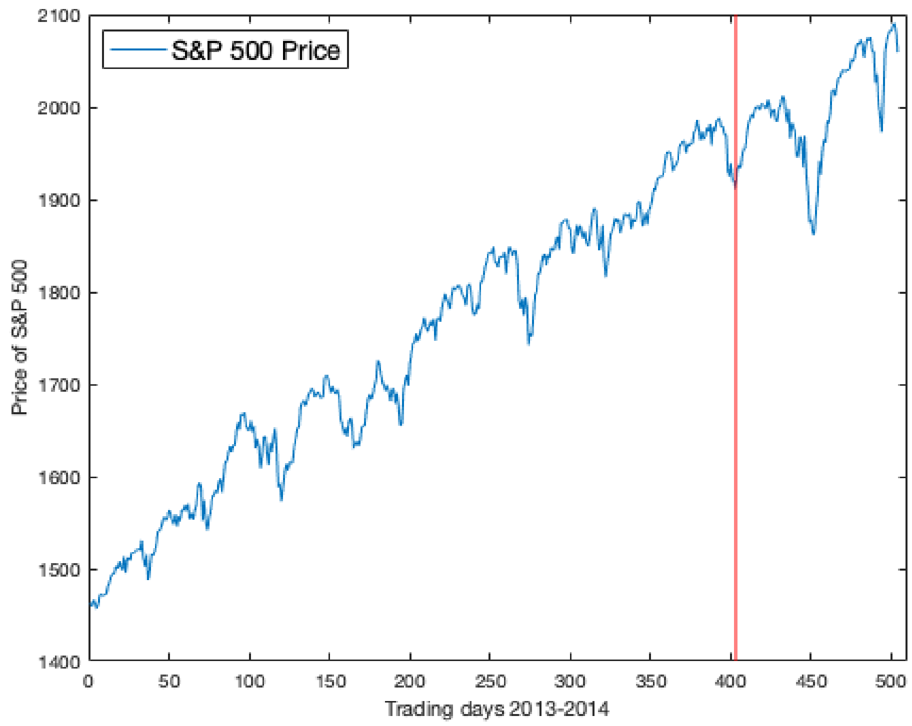
Most of the papers use the following performance measures (Equation 11) to be able to compare obtained results and previous they are applied. RMSE is a standard measure that is used in almost every research as an objective function. %RMSE is essential as well as for cheap options even small difference in absolute terms may become crucial. Finally, IVRMSE focuses on the difference between the market implied volatility and model implied volatility.

$$
\begin{aligned}
&RMSE = \sqrt{\tfrac{1}{N}\textstyle\sum_{i=1}^{N}(P_i^{market} - P_i^{model})^2} \\
&\%RMSE = \sqrt{\tfrac{1}{N}\textstyle\sum_{i=1}^{N}((P_i^{market} - P_i^{model})^2/P_i^{market})} \\
&MAE = \tfrac{1}{N}\textstyle\sum_{1}^{N}|P_i^{market} - P_i^{model}| \\
&IVRMSE = \sqrt{\tfrac{1}{N}\textstyle\sum_{i=1}^{N}(IV_i^{market} - IV_i^{model})^2} * 100
\end{aligned}
\tag{11}
$$

3 Empirical Analysis

3.1 Data Description

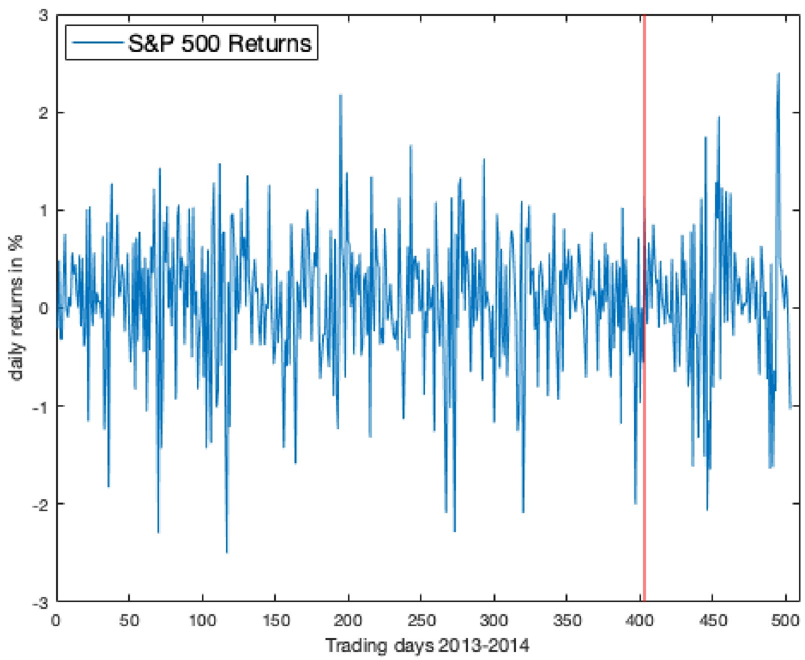
Taking into account that options are very tightly connected with the underlying asset, examing of S&P 500 dynamics is considered. Figure 2 represents the evolution of S&P 500 in 2013-2014. The red line splits data in the training sample and testing sample respectively. As it can be noticed in general there was an upward trend, however, after the splitting date, there is a 10% drop and then fast sharp growth around 13%.



Red line splits the data in training and testing samples

Figure 2: Evolution of S&P 500 in 2013-2014.

From Figure 3 which shows daily returns, one can see volatility clustering and that returns fluctuate around 0. Table 1 reflects the descriptive statistics of daily returns that is in line with stylized facts about financial time series (Cont, 2001). It is worth noticing that H_0 of mean returns being equal to zero can be rejected at 5% level but not at 1%. The skew is negative, which means that distribution is asymmetric around the mean and kurtosis is greater than three indicating that distribution has more massive tails than the normal one, in the literature it is often called leptokurtic property. Skewness and kurtosis are tested against 0 and 3 respectively. Looking at the p-values conclusion would be that assuming a normal distribution for S&P 500 returns is inappropriate.



Red line splits the data in training and testing samples

Figure 3: Daily returns of S&P 500 in 2014

Table 1: Decription statistics of daily returns S&P 500 in 2014.

Measure	Value	P-value
Mean	0.0705	0.0237
Std	0.6989	
Skewness	-0.4459	0
Kurtosis	4.2171	0

Dataset used in the study consists of S&P 500 European call options from January 1st, 2013 till December 31st, 2014. The dataset is reduced using the maturity of the options from between 15 and 250 days and moneyness between 0.9 and 1.1 in a similar manner that was done by Badescu et al. (2016). This dataset is used to estimate the price of the option quoted every Wednesday. In order to prevent predicted prices of the neural network from being negative, the dataset is further reduced, and only observations with price larger than three are used. The features of the dataset are: price of the S&P 500 index, strike, time to maturity, implied volatility, volume, return S&P 500 of this date and open interest of the option.

Following the methodology of Bakshi et al. (1997) options are grouped within the following categories at the money if moneyness $\in [0.97, 1.03]$; out of the money if moneyness $\in [0.9, 0.97)$; in the money $\in (1.03, 1.1]$. By the term of the expiration short term $\in [15, 60]$; medium term $\in [61, 180]$; long term $\in [181, 250]$. The summary of the characteristics is presented in Table 2. Most of the contracts are short term and they are either at the money or out of the money. Consistent with the previous foundings of Rosenberg (1998) implied volatility differs across moneyness and maturity. In addition, implied volatility for at the money options increases with the increase in maturity (Backus et al., 2004). Note that there are in total 6166 observations.

Table 2: Basic features of the options Wednesday 2013-2014 (Call)

		Moneyness			
	Maturity	[0.900, 0.970)	[0.97,1.03]	(1.03, 1.1]	Across moneyness
Number of contracts	[15,60]	670	3061	203	3936
	[61,180]	795	881	89	1765
	[181,250]	178	248	39	465
Across maturities		1643	4190	331	6166
Averge price	[15,60]	6.6565	22.3408	88.4145	23.0709
	[61,180]	17.3827	56.5902	118.2388	42.0388
	[181,250]	36.3357	92.1393	133.7256	74.2658
Across maturities		15.0620	33.6734	101.7725	32.4033
Average IV	[15,60]	0.1153	0.1182	0.1614	0.1199
	[61,180]	0.1210	0.1398	0.1640	0.1326
	[181,250]	0.1337	0.1550	0.1687	0.1480
Across maturities		0.1201	0.1249	0.1630	0.1257

3.2 Simulation Study

The simulation study is conducted in order to check the validity of the approximation theory. This analysis aims to answer the question, if we imagine that prices follow the Black-Scholes model, can the neural network learn it? The simulation study is similar to one conducted by Hutchinson et al. (1994). Taking into account the observed dynamics of S&P 500 index, I try to mimic it, prices of underlying asset are simulated by adding the standard random number with mean 1 and standard deviation 2 to the previous value. The simulated path contains 100 days that equal to 4.5 trading months. With the initial price of 1500, hundred of different paths with the following data generating process ($P_t = P_{t-1} + normrnd(1, 2)$) are computed. Some examples of the simulated paths are presented in Figure 4.

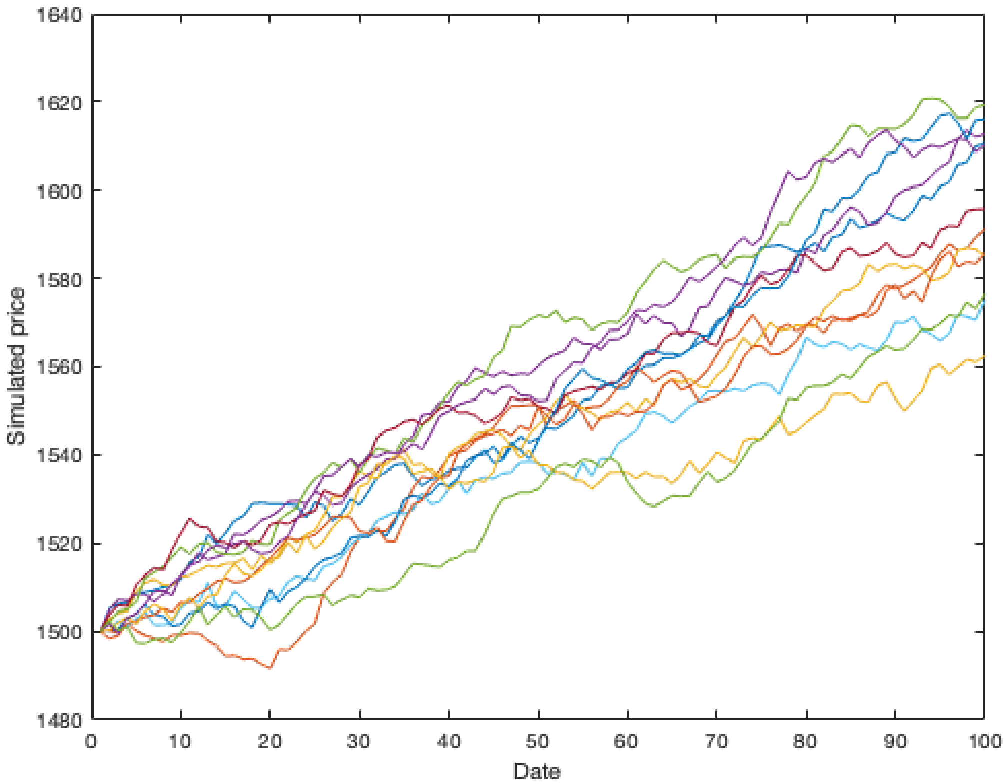


Figure 4: Examples of simulated prices

Around prices of underlying asset strikes are created with the values starting from 1455 till 1650 with the step of 10. At the first day for each strike options with initial maturity (15, 30, 60, 90, 180) are assigned. With each day the maturity of the option decreases by one day. If the option is close to expiration, it is replaced by a new one with the same initial maturity. This procedure is done in two days before the exercise date. Overall, each daily observation has 100 different options (20 strikes and 5 maturities). Moneyness is between 0.92 and 1.06 with mean 0.99 that is similar to the original dataset. Then for each strike and maturity option prices are calculated using the Black-Scholes formula with constant annual risk-free rate 0.0007 provided with the option dataset and constant implied volatility (σ) 0.13

that is mean implied volatility obtained from the data.

The summary, of the simulation study is presented in Table 3. Inputs of the neural network are price of S&P500, time to maturity, strike and the output variable is the price of the call option obtained with the Black-Scholes formula. Implied volatility and interest rate are not included as they are assumed to be constant. Later the whole sample is split into training, testing and validation sets with the principle stated in the previous section. During the next stage, one layer neural network is used to approximate the Black-Scholes formula. The size of the hidden layer is a random integer between 1 and 100 that is different for each day. Random size of the neural network is used as approximation theory states that neural network can approximate any function no matter the complexity of the neural network.

Table 3: Summary of the simmulation study

Evolution of the price - $P_t = P_{t-1} + normrnd(1, 2)$

N - Total number of observations for each price path

IV - Implied volatility

Strikes	Maturities in Days	Starting Price	Moneyness	N	Interest rate	IV
1755:10:1950	15,30,60,90,180	1500	[0.92,1.06]	10000	0.0007	0.13

Results of the simulation study are presented in Table 4. It can be seen that one-layer neural network can recover option prices very accurately no matter the size of the hidden layer, RMSE is 0.2228 and MAE is 0.1512 that is very small taking into account average prices of the option. The conclusion of this part is that neural network can approximate nonlinear function and results of Hutchinson et al. (1994) and Leshno et al. (1993) approved.

Table 4: Evaluation of performance (out of sample)

Measure	Value
RMSE	0.2228
MAE	0.1512
RMSE %	0.0266
Mean Price of the option	67.5827

3.3 Choosing Optimal Number of Nodes for Neural Network

The next step is to choose the optimal number of nodes for the one layer neural network for the real data. Before choosing the number of nodes, I create a proxy for implied volatility. It appears not to be fair to use an input that was obtained by using the actual price of the option that is an output variable. Initially, realized volatility is considered as a proxy and historical volatility as was used in previous research by Malliaris and Salchenberger (1993) and Anders et al. (1996), however, these proxies do not create a volatility smile observed in the real data as it would not vary across moneyness. To overcome it, I use constant volatility approach. From section 2.1.1 there are two suggestions to model volatility as a function of moneyness and maturity or as a function of underlying asset price and maturity. Moreover, I do not want to impose a specific linear structure between regressors and dependent variable; therefore I run a one hidden layer neural network to estimate implied volatility. The neural network is trained on a similar dataset but from 2012 that is cleaned in the same manner as in section 3.1. Surprisingly, the model that uses moneyness and maturity as inputs performs worse than one suggested by Dumas et al. (1998) with price and maturity as inputs. The difference in terms of percentage accuracy is 16%. It is worth mentioning that best estimated implied volatility ($\hat{IV}$) differs on average by 10% in absolute terms from the true one in the testing sample from 2012 that is rather precise estimate. Finally, obtained weights from the neural network are used to determine estimated implied volatility in 2013-2014.

The inputs used to determine the optimal number of nodes are: price of S&P 500 index, strike, maturity, estimated implied volatility and the output variable is the actual price of the option. 100*100 neural networks are estimated with the number of nodes from one to hundred. Neural network with each node estimated a hundred times with different initial parameters in order to get consistent results.

Lastly, to select the optimal number of nodes several criteria are used. For each number of nodes mean testing error, mean training error and mean network information criteria (NIC) suggested by Murata et al. (1994) are computed. NIC is a standard information criterion but applied to the neural network. Performance in terms of mean MSE for each number of nodes is presented in Figure 5. The training error continuously decreasing with the increase in the number of nodes. However, the testing error at first decreases and then slightly increases with a minimum at 67. This is a standard pattern of overfitting the data.

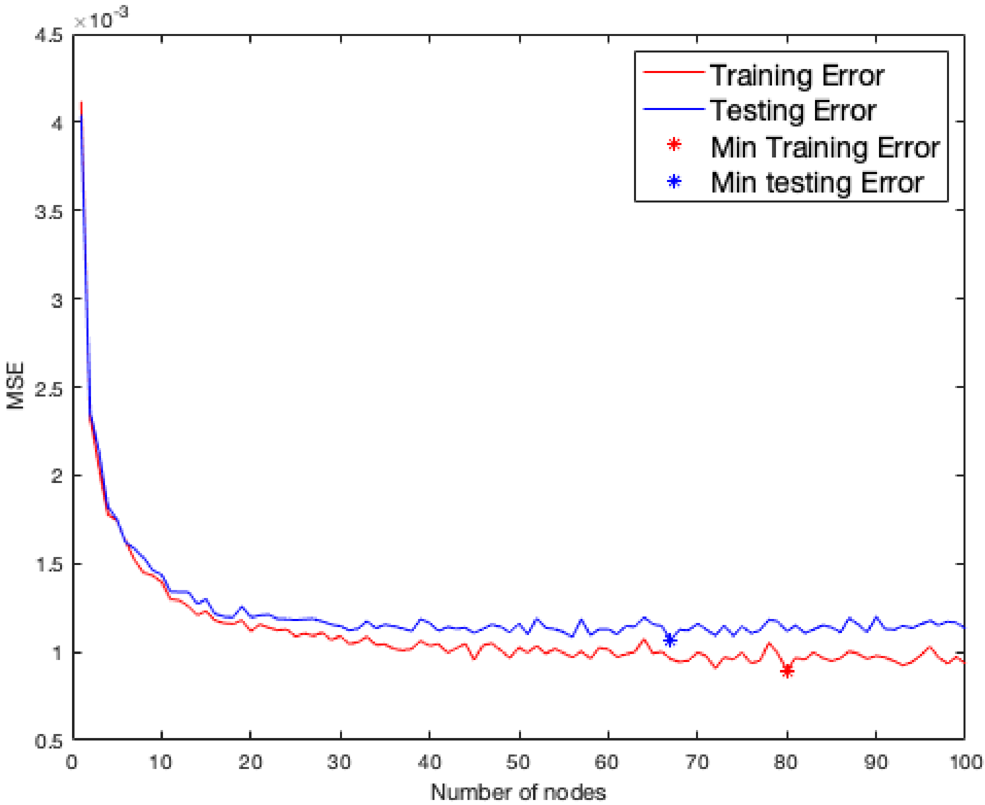


Figure 5: Mean MSE for different number of nodes

Performance measures are shown in Table 5. Despite the fact that AIC information criteria and minimal testing error suggest more complicated structure, for further analysis neural network with 20 nodes is used as 67 nodes might cause overfitting and performance improvement between 20 and 67 nodes is negligible.

Table 5: Minimum mean of the performance measures

Measure	Value	Number of nodes
Testing error MSE	0.0011	67
Training error MSE	0.00009	80
SBC	-2.6361	20
AIC	-2.3385	45

3.4 Comparison of Heston Model and Neural Network

The task of the models is to estimate the price of the option in period t. In this part I run four neural networks. First one is the neural network with the price of S&P 500, strike and maturity served as an input. The second neural network has an additional $\hat{IV}$ input. The third neural network has all the available input parameters. This is done in order to determine whether the additional parameters add valuable information to the option pricing. Last neural network has all available inputs but $\hat{IV}$ is excluded. Finally, the Heston model with optimal parameters is considered. Full description of the models can be found in Table 6.

Table 6: Summary of the models

	Model				
	Neural Network 1	Neural Network 2	Extended NN 1	Extended NN 2	Heston
Inputs	Price of SP 500 t	Price of SP 500 t	Price of SP 500 t	Price of SP 500 t	Price of SP 500 t
	Strike	Strike	Strike	Strike	Strike
	Maturity	Maturity	Maturity	Maturity	Maturity
		$\hat{IV}$	$\hat{IV}$		Interest Rate
			Realized Volatility	Realized Volatility	V(0)*
			Volume	Volume	κ*
			Open Interest	Open Interest	θ*
			SP Return	SP Return	σ*
			Moneyness	Moneyness	ρ*
Output	Mean price of the option for the current day				

* Stands for optimal parameter obtained from minimisation problem (Equation 6)
NN - Neural Network

The comparison of the models is made out of the sample. The characteristics of the testing sample are presented in Table 7. It is somewhat similar to the full dataset as the prevailing part of the options are at the money. In addition, the volatilities features are the same as in the overall sample.

Table 7: Summary of the test sample

	Maturity	[0.900, 0.970)	[0.97,1.03]	(1.03, 1.1]	Across moneyness
		Moneyness			
Number of contracts	[15,60]	192	649	28	869
	[61,180]	129	182	14	325
	[181,250]	17	30	3	50
Across maturities		338	861	45	1244
Averge price	[15,60]	7.8186	25.4634	97.7679	23.8946
	[61,180]	26.4500	81.6027	146.3679	62.5012
	[181,250]	49.5029	127.8733	161.3000	103.2330
Across maturities		17.0028	41.3307	117.1233	37.4056
Average IV	[15,60]	0.1240	0.1242	0.1746	0.1258
	[61,180]	0.1295	0.1509	0.1760	0.1435
	[181,250]	0.1386	0.1621	0.1794	0.1551
Across maturities		0.1269	0.1315	0.1754	0.1318

Training sample is used to estimate the parameters of the Heston model. Table 8 presents the estimated parameters of the Heston model. As it can be seen Feller condition is fulfilled as $2\kappa\theta > \sigma^2$. κ coefficient displays the shock persistence in volatility and it is rather small in comparison to one observed in GARCH models where the decaying rate of volatility is closer to one (Andersen and Bollerslev (1997) and Ding and Granger (1996)). This might be due to the additional restriction of the Heston model (Feller condition) and inability of the model to capture the long memory process of volatility completely. The correlation coefficient is negative meaning that it reduces prices of out of the money options relative to in the money options that correspond to previous research of Mrázek and Pospíšil (2017). σ is not equal to zero and it partially captures the kurtosis of spot returns. Lastly, having the essential parameters option prices with the closed solution from Appendix A are calculated.

Table 8: Estimated parameters of the Heston model

V(0)	κ	θ	σ	ρ
0.0156	0.0519	0.8353	0.2450	-0.6843

Using the inputs of Table 6 4 neural networks with 20 different starting weights are estimated. The obtained weights of the neural networks cannot be easily interpreted, therefore only performance measures are discussed.

Table 9 presents an overall comparison of the models in terms of the different performance measures. As it can be seen extended neural network 1 with additional inputs performs better in comparison to the Heston model, in particular, RMSE of the neural network is more than two times lower than the Heston model. It is worth noticing that even by incorporating stochastic correlation conducted by Christoffersen et al. (2009) the performance of the neural network would still be superior, assuming that stochastic correlation gives the same improvement as in Christoffersen et al. (2009) dataset. However, in terms of IVRMSE, the difference between these models is not that large and extended neural network 1 shows around 30% improvement. Another interesting observation is that $\hat{IV}$ increased the performance of the neural networks; this finding is expected as implied volatility should be beneficial for option pricing. However, the mean percentage error of the estimated implied volatility in 2013-2014 dataset increased up to 20%, while the 2012 test sample was only 10%.

Table 9: Performance measures of different models (out of sample)

Criteria	Neural network 1	Neural network 2	Extended NN 1	Extended NN 2	Heston model
IVRMSE	4.2596	3.6536	2.7825	3.6251	3.9513
MAE	10.7087	9.1291	4.8081	5.7162	9.6435
RMSE	19.1977	15.6764	7.1929	10.4825	18.6612

Looking deeper into the performance of price valuation the estimated price values are divided into groups and RMSEs for each group are calculated, results are presented in Table 10. The neural network 1 and 2 display similar performance, but $\hat{IV}$ significantly improved the performance of medium term options across all types of moneyness. In addition, extended neural network 1 outperformed other neural networks and the Heston model across all moneyness. Heston model showed results close to the neural network 2 that has Black-Scholes inputs. Also, the Heston model estimated relatively good short term options. Finally, extended neural network 1 displayed superior results in estimating long term option across the moneyness showing almost four times lower RMSE in comparison to the Heston model.

Poor performance of the Heston model might be caused by getting trapped in a local minimum. Some other initial values were implemented in the estimation of the Heston model but results did not change much.

Table 10: Comparison of the models in terms of RMSE

		Moneyness			
	Maturity	[0.900, 0.970)	[0.97,1.03]	(1.03, 1.1]	Across moneyness
Neural network 1	[15,60]	9.5841	8.4733	6.0420	8.0099
	[61,180]	22.6850	36.7475	21.9292	30.9648
	[181,250]	37.1251	42.7946	30.6391	38.7388
Neural network 2	[15,60]	15.6843	7.5479	5.6045	7.5538
	[61,180]	15.7173	27.5450	15.6593	22.9498
	[181,250]	34.4013	41.4366	30.0376	37.5230
Extended neural Network 1	[15,60]	9.5035	4.7285	3.0259	4.6313
	[61,180]	14.0032	12.6103	7.7016	10.9490
	[181,250]	5.9046	12.9817	7.5527	11.0726
Extended neural Network 2	[15,60]	4.9728	4.5427	5.1552	4.7077
	[61,180]	12.1363	20.9448	10.8371	17.1690
	[181,250]	12.1917	23.0239	13.0255	19.6128
Heston model	[15,60]	8.3597	7.6003	5.5771	7.2039
	[61,180]	19.6112	35.3488	21.4959	29.8368
	[181,250]	36.2172	45.8961	32.1255	41.1519

Appendix B represents the error histogram of the models ($Error = \hat{Y} - Y$). From the Figure 11, it can be concluded that all the models except extended neural networks tend to significantly undervalue the option prices and this is the reason for large RMSEs across the other models in Table 10. Besides, extended neural network 1 displays less diverse errors and they are more centred around zero.

During the next step, I inspect the estimated option prices by grouping and sorting original prices with predicted prices that are presented in Figure 6. Neural network 1 estimated relatively good cheap options while significantly depreciating more expensive options. Adding $\hat{IV}$ to neural network 1 as an input slightly improved estimates by reducing the error among the expensive options and at the same adding some noise in evaluating cheap options.

Better improvements of adding $\hat{IV}$ might be observed between extended neural networks. The absence of $\hat{IV}$ caused mispricing of expensive options (purple line). Extended neural network 1 provided the most precise estimates over the whole testing sample. Finally, the Heston model performed comparably to the neural network 1 displaying an increase of error with the price increase.

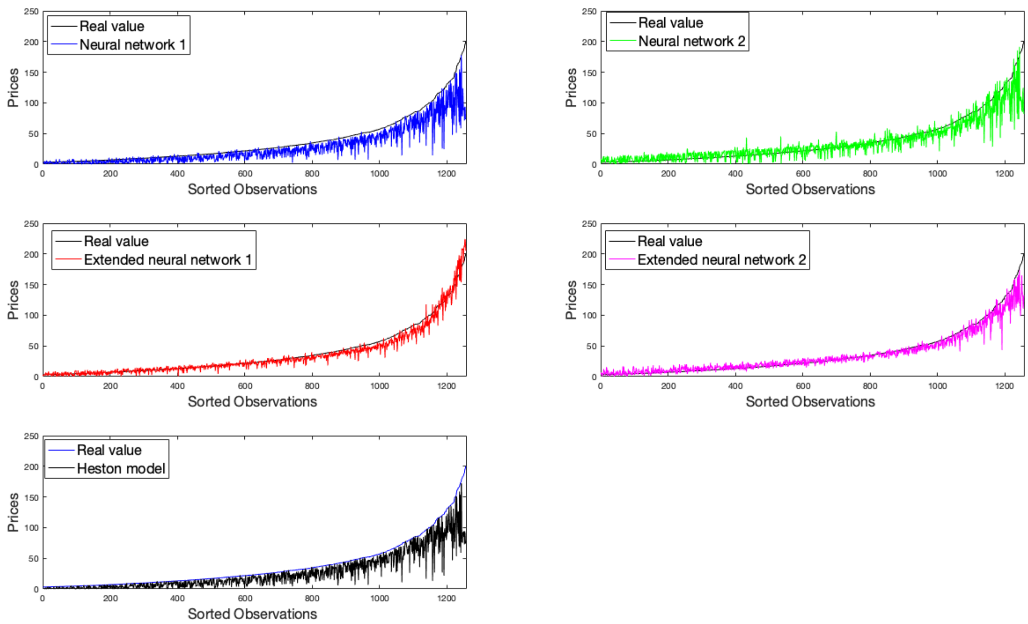


Figure 6: Sorted prices for different models

Further, estimated option prices are used to calculate implied volatility. It is hard to plot all the examples of estimated implied volatilities for all the possible maturities, therefore, Figure 7 displays a sample of estimated implied volatilities for a particular day in testing sample. During this particular day, neural network 1 and 2 underestimate the implied volatility. Extended neural network 1 has the closest estimates to the market implied volatility. It is worth mentioning that extended neural network 1 sometimes provided extreme outliers for in the money options across different maturities by producing to steep curve. Finally, the Heston model is slightly better than neural networks 1 that uses only S%P 500 price and maturity as an input but worse than neural network 2 with estimated implied volatility served as input.

As was noted by Mikhailov and Nögel (2003) Heston model cannot get enough skew in the implied volatility. However, for the current dataset, in general, the

Heston model more or less is able to replicate the shape of the volatility smile but the scale is not precise. The similar tendency of estimated implied volatility is observed among the overall testing sample.

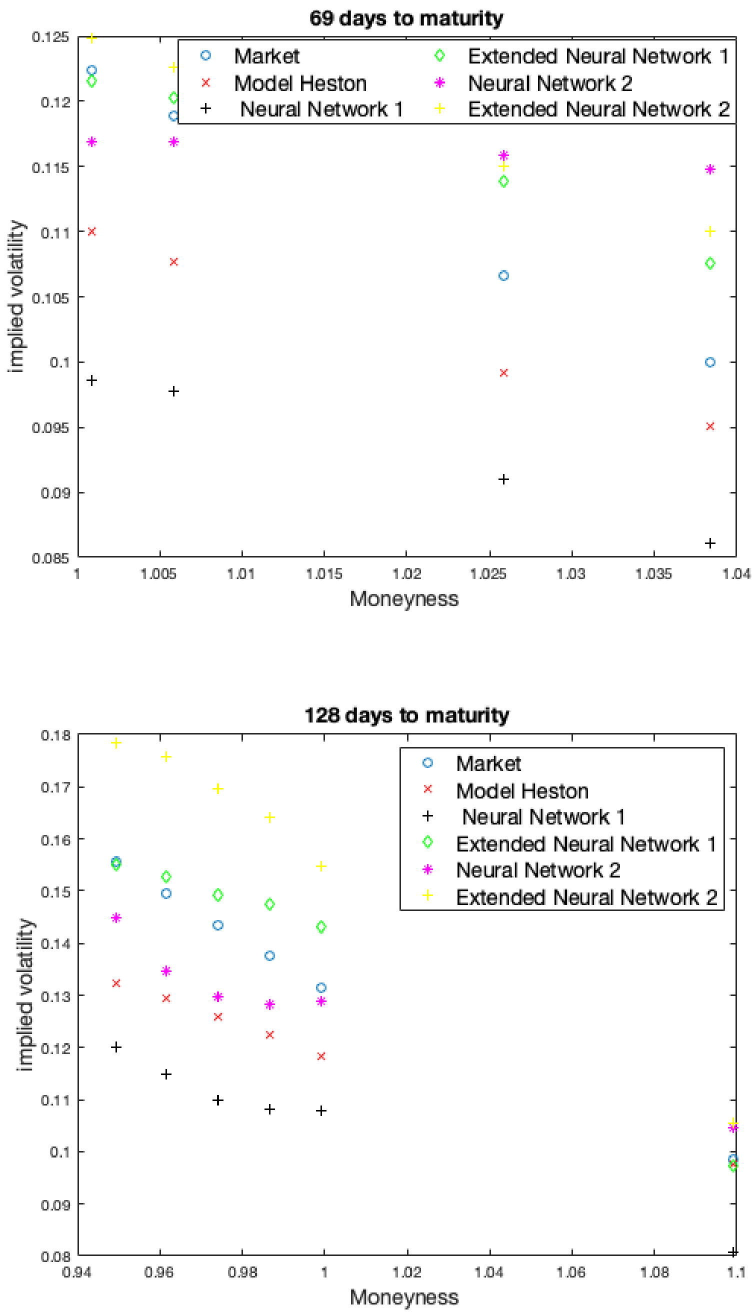


Figure 7: Sample of implied volatility for different models

Further, to estimate the reasonability of results, I examine the following Equation 12 that should hold for two European call options with strikes $X_1 \leq X_2$ and same maturity date T.

$$
\begin{aligned}
&0 \leq C_{X_1,T}(S_t,\tau) - C_{X_2,T}(S_t,\tau) \leq (X_2 - X_1)e^{-r\tau} \\
&Where: \\
&\tau = T - t \\
&{}^{*}Source:\ Franke\,et\,al.\,(2015)
\end{aligned}
\tag{12}
$$

I use a ratio that is computed as a fraction between the number of fulfiled conditions and the overall number of cases of options with different strikes but the same maturity. Results of the examination are presented in Table 11, across all the models the rate of fulfilment is close to 1. It is a desirable result in a sense that no-arbitrage assumption holds, however, if two options with the same maturity but different strikes are simultaneously undervalued or overpriced with the same scale the assumption would still hold. Therefore this measure is useful as a robustness check, but the overall conclusion about the performance of the models cannot be done using only this criterion.

Table 11: Ratio of fulfilled condition across the models

	Model				
	Neural Network 1	Neural Network 2	Extended NN 1	Extended NN 2	Heston
Ratio of fulfilled condition	0.9825	0.9893	0.9708	0.9766	0.9805

Finally, I examine the validity of Kaastra and Boyd (1996) claim who stated that neural networks do not provide stable results. Extended neural network 1 is estimated one thousand times with different starting values on the original training and testing samples. Figure 8 presents the distribution of out of sample RMSE. The distribution is rather dispersed with some extreme outliers, meaning that results of neural network dependent on starting values and multiple runs of the models is required. In addition, 5-10 runs of the neural network might not be enough to find a global minimum and more runs are required.

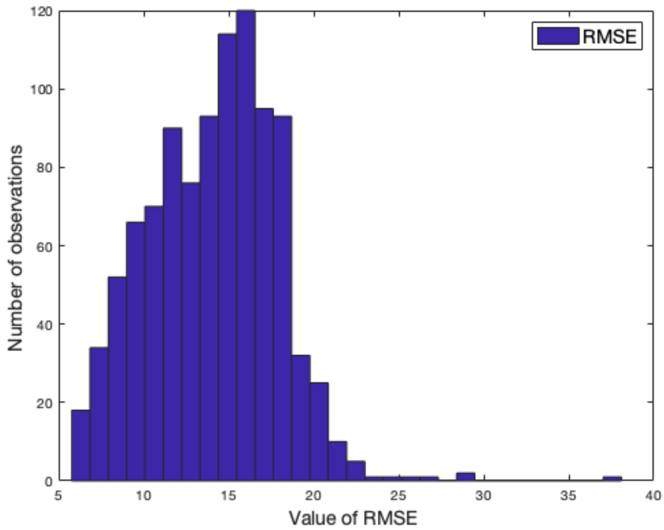


Figure 8: Histogram of RMSE extended neural network 1

3.5 Forecast of the Option Price for the Next Day

Taking into account the success in monthly price forecasting with a neural network that was achieved in the previous research; this section of the study considers forecasting option prices for the next day as the lifetime of the option is relatively short and building model on monthly prices looks infeasible. It is expected that option prices are martingale but it is still interesting to compare the models in terms of forecast accuracy.

First of all, the training sample is reduced until 4161 observations, options that are traded on Wednesday and Thursday are used (Same strikes but with maturity lower by one day). In the current settings data from Wednesday is used to forecast option price on Thursday. Then the reduced dataset is divided into training and testing sets. Summary of the testing sample is presented in Table 12. Prevailing part of the options is at the money in a short and medium-term group. The structure is similar to the initial set.

Table 12: Summary of the test sample (Forecast)

	Moneyness			
Maturity	[0.900, 0.970)	[0.970, 1.03]	(1.03, 1.1]	Across moneyness
[15,60]	63	463	6	532
[61,180]	65	127	2	194
[181,250]	26	75	1	102
Across maturity	154	665	9	828

I employ several models for price forecast. The first model states that the best prediction for tomorrow is the option price today. The second model is the Black-Scholes model with maturity lowered by one day, and the other inputs from today's price. Finally, the neural network with all available inputs is implemented. During this step of the study, log prices of the neural network are forecasted. Heston model is not included in the list as it is aimed to estimate prices in time t. Summary of the models presented in Table 13.

Table 13: Models used in price forecasting

	Model		
Inputs	Guess	Black-Scholes	Neural network
	Price of the option in t	Price of SP 500 t	Price of SP 500 t
		Strike	Strike
		Maturity t+1 (reduced by 1)	Realized volatility t
		Interest rate	Forecasted realized volatility t+1
		Implied volatility t	Implied volatility t
			Maturity t
			Volume t
			Open interest t
			SP return in t
			Moneyness t
Output	Price of the option t+1		

Taking into account that volatility plays a significant role in option pricing, forecasted volatility of the next day is used as an input for the neural network to improve the performance. To get an estimate of the volatility the HAR model is employed. This model is used because Vortelinos (2017) stated that HAR model performed better in comparison to GARCH(1,1) and neural network with one hidden layer in predicting volatility. HAR model was introduced by Corsi (2009) and this is an approximation of a long memory process that is a legitimate approach as autocorrelations of the square returns display very strong persistence. The core idea of this model is that volatility consists of several components. Agents cause various price movements, short term traders create daily activity, midterm traders weekly and long-term investors monthly movements. The HAR model is presented in Equation 13.

$$\begin{aligned}
&RV_{\text{t}} = a + \beta_1 RV_{\text{t-1}} + \beta_2 RV^5{}_{\text{t-1}} + \beta_3 RV^{22}{}_{\text{t-1}} + u_t \\
&Where: \\
&RV^{(h)}{}_{\text{t-1}} - \; h\, period\, realized\, volatility \\
&RV^5{}_{\text{t-1}} - \; Corresponding\, to\, 1\, week\, average \\
&RV^{22}{}_{\text{t-1}} - \; Corresponding\, to\, 1\, month\, average
\end{aligned} \tag{13}$$

Daily realized volatility is obtained from 5 minutes observation and data is collected from the Oxford library (Lunde and Shephard, 2018), 5-minute observation is used as Liu and Patton (2015) stated that it is hard to outperform 5 minute RV measure significantly. For realized volatility prediction I use 252 previous observations to compute 1 step ahead forecast with a rolling window approach. For example, for the first Thursday in 2013, 252 previous observations including the first Wednesdays are used to estimate the volatility process and get a forecast for Thursday. In addition, values are scaled by 10.000, and the logarithm transformation is applied to get more precise estimation and avoid negatively predicted volatility.

Predicted values of the realized volatility are presented in Figure 9. Forecasted values are rather accurate and model can replicate the innovations in the volatility process. Therefore it might make sense to include it as an input variable in the neural network.

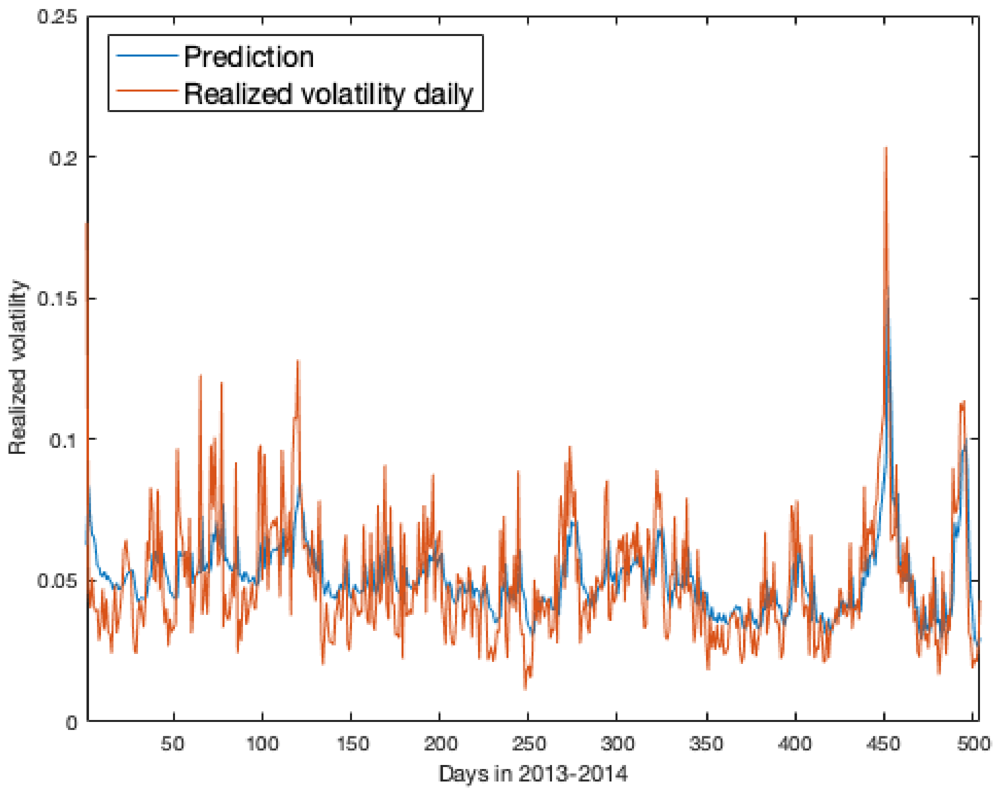


Figure 9: One-step ahead forecast of realized volatility (2013-2014)

Now with additional input neural network has overall 10 inputs. In order to choose the structure of the neural network for prediction exercise, the same procedure as in Section 3.3 implemented. Mean minimum testing MSE is observed for six nodes and information criterion suggests 8 nodes. Following the same logic as above, model with the simpler structure is used for further estimation.

Table 14 presents the overall comparison of 1 step ahead forecasts. At first sight, the performance of the neural network is inferior and the Black-Scholes model performs approximately like a guess. The reason for such imperfections might be the sharp drop of S%P 500 price in the testing sample and neural network had no information to learn it. From figure 12 Appendix C that presents the error histograms of all models it is clear that all models fail to predict extreme changes in the option prices that were probably caused by a sharp drop and then rise of S%P 500 in the testing sample. In addition, neural network has more spread errors in comparison to other models.

Table 14: Comparison of 1 step ahead forecast

	Measure	Value
Neural Network	RMSE	11.9916
	MAE	8.8396
	AP % Err	28.2785
	IVRMSE	2.6566
Black-Scholes with reduced maturity	RMSE	8.7263
	MAE	5.4087
	AP % Err	19.5315
	IVRMSE	1.9422
Guess yt=yt+1	RMSE	8.7010
	MAE	5.3606
	AP % Err	18.8799
	IVRMSE	2.0338
Mean absolute change in Price %	Test sample	18.8799

Table 15 presents the performance across moneyness and maturity. Here I would like to note that the neural network performs worse than other models almost across all of the cases, the exceptions are mid term out of the money and long term at the money options; long term out of the money group displays the worst result.

In some cases, the Black-Scholes model displays slightly smaller RMSEs in comparison to the guess but the differences are negligible.

Table 15: Performance forecast original dataset

	Moneyness				
	Maturity	[0.900, 0.970)	[0.970, 1.03]	(1.03, 1.1]	Across moneyness
Neural Network	[15,60]	15.7015	11.7071	8.2490	11.3614
	[61,180]	12.6691	14.7684	11.2895	13.6113
	[181,250]	28.4736	11.8262	10.9451	11.8879
Black-Scholes	[15,60]	11.8295	7.6632	4.7686	7.4053
	[61,180]	19.7672	9.0059	5.6256	8.1674
	[181,250]	2.9326	15.7207	9.8732	14.3754
Guess tom=today	[15,60]	12.0983	7.6223	4.6292	7.3621
	[61,180]	20.0306	8.9911	5.5239	8.1390
	[181,250]	3.3000	15.7870	9.7157	14.4023

The same pattern might be observed in Figure 10 that represents sorted actual and forecasted prices. The behaviour of the models is somewhat similar but the neural network tends to underestimate the prices more.

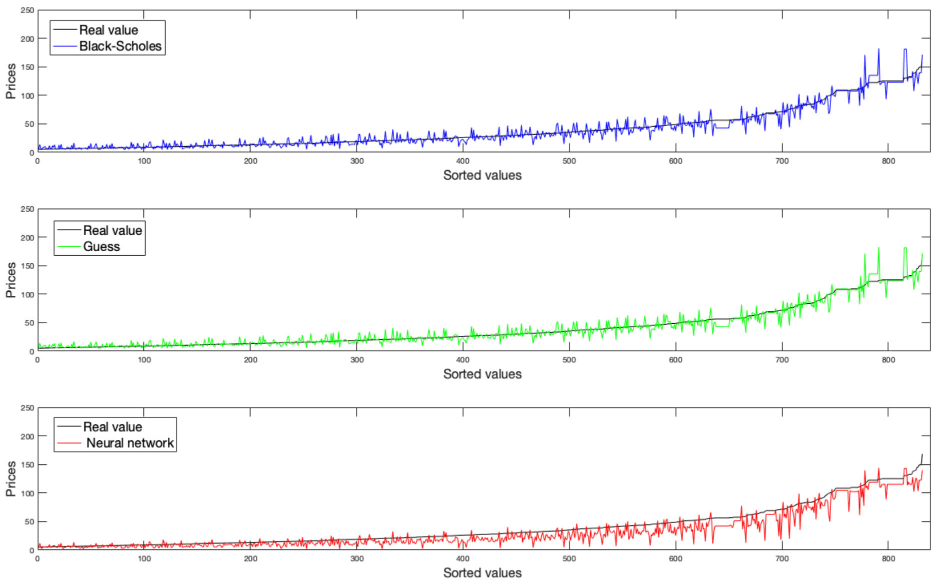


Figure 10: One-step ahead forecast of different models

Further, to analyse the statistical significance between predictions Diebold and Mariano test is applied (Diebold and Mariano, 2002). Under the null hypothesis two forecasts provide the same accuracy and under the alternative hypothesis, forecasts are different from each other. Table 16 presents the p-values of the Diebold and Mariano test. Forecasts of the neural network are significantly different from the Black-Scholes model and a guess. Aside from this, guess and the Black-Scholes model have the same accuracy in predictions. That means that the neural network performs worse in particular dataset.

Table 16: P-value of the Diebold and Mariano test

		Model	
		Neural Network	Black-Scholes
Model	Guess	0.00	0.12
	Neural Network		0.00

3.6 Futher Research Prospects

Limitation of the study is that data is treated as panel data and including time series dependency might lead to better results, this can be done by implementing the recurrent neural network that can use internal memory in order to generate the output. Zhang and Xiao (2000) stated that this type of architecture is a has a great potential for time-series analysis. Further, the option pricing in period t with the neural network was rather precise and additional improvements could be achieved by the ensemble method similar to one applied by Gradojevic et al. (2009). On another hand, it might worth implementing deeper neural networks as Eldan and Shamir (2016) stated that an increasing number of layers even by one could give a substantial improvement in the performance of the neural network. To set a deep structure one may use geometric pyramid rule Masters (1993).

4 Conclusion

In this paper, I compare the performance of the neural network with the Heston model in the context of option pricing. Nonparametric method namely neural network with one hidden layer and twenty nodes significantly outperformed the Heston model in pricing options out of sample, on the dataset from January 2013 until December 2014 from S&P 500 index. The significant improvement of the neural network performance was achieved by its flexibility and capability to incorporate different inputs.

Further, one layer neural network is not able to outperform the Black-Scholes model and guess in one-step ahead prediction. The problem with the predictions might be that the model tries to fit the noise while there is no signal and system always changes nonlinearly, therefore a guess turned out to be the best forecast. Finally, the findings of the thesis help to justify the growing role of machine learning in the financial industry.

Appendices

Appendix A

$$
\begin{aligned}
&C(S,v,t,T) = SP_1 - Ke^{-r(T-t)}P_2\\
&Where:\ for\ j=1,2\\
&P_j(x_t,v_t;x_T,lnK) = \tfrac{1}{2}+\tfrac{1}{\pi}\int_0^{\infty}(Re(\tfrac{e^{-i\phi LnK}f_j(\phi;t,x,v)}{i\phi})d\phi)\\
&f_j(\phi;t,x,v) = exp[C_j(\tau,\phi)+D_j(\tau,\phi)v_t+i\phi x_t]\\
&C_j(\tau,\phi) = (r-q)i\phi\tau+\tfrac{a}{\sigma^2}((b_j-\rho\sigma i\phi+d_j)\tau-2ln\tfrac{1-g_je^{d_j\tau}}{1-g_j})\\
&D_j(\tau,\phi) = \tfrac{b_j-\rho\sigma i\phi+d_j}{\sigma^2}(\tfrac{1-e^{d_j\tau}}{1-g_je^{d_j\tau}})\\
&Where:\\
&g_j = \tfrac{b_j-\rho\sigma i\phi+d_j}{b_j-\rho\sigma i\phi-d_j}\\
&d_j = \sqrt{(b_j-\rho\sigma i\phi)^2-\sigma^2(2iu_j\phi-\phi^2)}\\
&u_1=\tfrac{1}{2};u_2=-\tfrac{1}{2};a=\kappa\theta;b_1=k-\rho\sigma;b_2=\kappa;i^2=-1
\end{aligned}
\tag{14}
$$

Appendix B

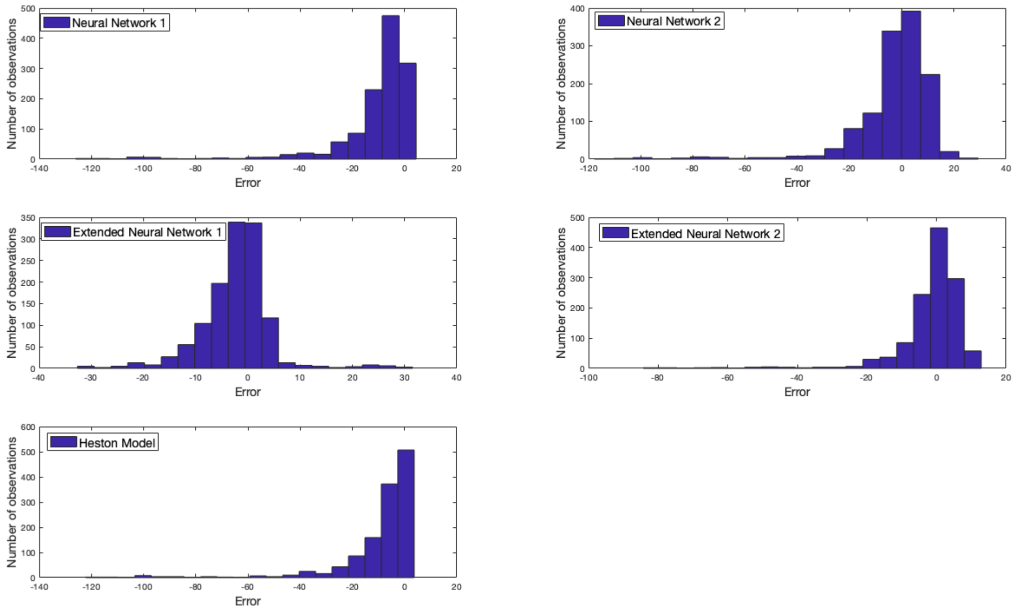


Figure 11: Histogram of errors for different models (out of sample)

Appendix C

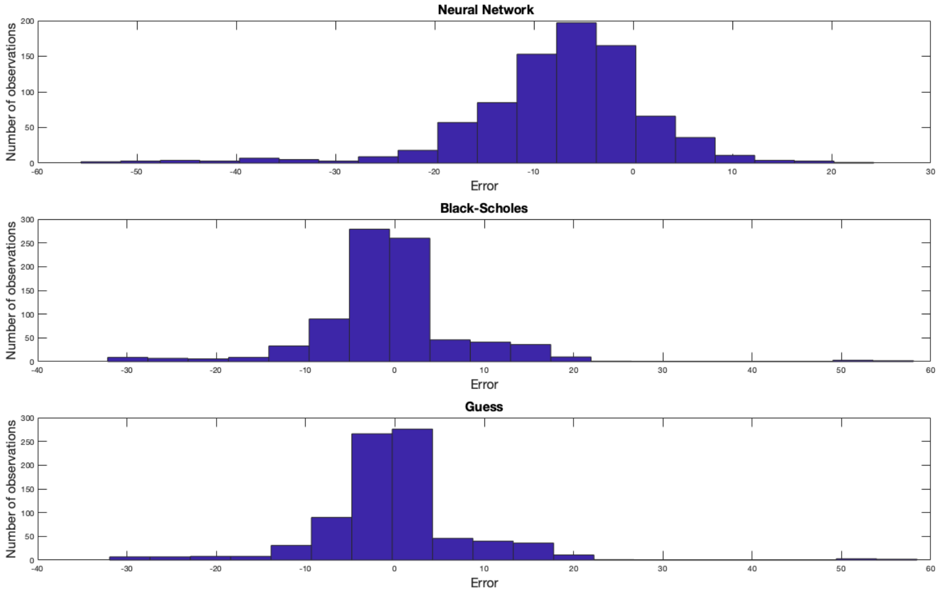


Figure 12: Errors 1 step ahead forecast

References

Akgiray, V., 1989, "Conditional heteroscedasticity in time series of stock returns: Evidence and forecasts," *Journal of business*, pp. 55–80.

Amilon, H., 2003, "A neural network versus Black–Scholes: a comparison of pricing and hedging performances," *Journal of Forecasting*, volume 22, no. 4, 317–335.

Amin, K. and Morton, A., 1994, "Implied volatility functions in arbitrage-free term structure models," *Journal of Financial economics*, volume 35, no. 2, 141–180.

Anders, U., Korn, O., and Schmitt, C., 1996, "Improving the pricing of options: A neural network approach," *Journal of Forecasting*, volume 17, no. 5-6, 369–388.

Andersen, T., Benzoni, L., and Lund, J., 2002, "An empirical investigation of continuous-time equity return models," *The Journal of Finance*, volume 57, no. 3, 1239–1284.

Andersen, T. and Bollerslev, T., 1997, "Intraday periodicity and volatility persistence in financial markets," *Journal of empirical finance*, volume 4, no. 2-3, 115–158.

Backus, D., Foresi, S., and Wu, L., 2004, "Accounting for biases in Black-Scholes," *Available at SSRN 585623*, pp. 1–46.

Badescu, A., Elliott, R., Grigoryeva, L., and Ortega, J.-P., 2016, "Option pricing and hedging under non-affine autoregressive stochastic volatility models," .

Bailey, D. and Thompson, D., 1990, "Developing neural-network applications," *AI expert*, volume 5, no. 9, 34–41.

Baillie, R. and Bollerslev, T., 1989, "The message in daily exchange rates: a conditional-variance tale," *Journal of Business & Economic Statistics*, volume 20, no. 1, 60–68.

Bakshi, G., Cao, C., and Chen, Z., 1997, "Empirical performance of alternative option pricing models," *The Journal of finance*, volume 52, no. 5, 2003–2049.

Bates, D., 1996, "Jumps and stochastic volatility: Exchange rate processes implicit in deutsche mark options," *The Review of Financial Studies*, volume 9, no. 1, 69–107.

Black, F. and Scholes, M., 1973, "The pricing of options and corporate liabilities," *Journal of political economy*, volume 81, no. 3, 637–654.

Bollerslev, T., 1986, "Generalized autoregressive conditional heteroskedasticity," *Journal of Econometrics*, volume 31, no. 3, 307–327.

Buraschi, A. and Jackwerth, J., 2001, "The price of a smile: Hedging and spanning in option markets," *The Review of Financial Studies*, volume 14, no. 2, 495–527.

Chakraborty, K., Mehrotra, K., Mohan, C. K., and Ranka, S., 1992, "Forecasting the behavior of multivariate time series using neural networks," *Neural networks*, volume 5, no. 6, 961–970.

Christensen, B. and Prabhala, N., 1998, "The relation between implied and realized volatility," *Journal of financial economics*, volume 50, no. 2, 125–150.

Christoffersen, P., Heston, S., and Jacobs, K., 2009, "The shape and term structure of the index option smirk: Why multifactor stochastic volatility models work so well," *Management Science*, volume 55, no. 12, 1914–1932.

Coleman, T. and Li, Y., 1996, "An interior trust region approach for nonlinear minimization subject to bounds," *SIAM Journal on optimization*, volume 6, no. 2, 418–445.

Cont, R., 2001, "Empirical properties of asset returns: stylized facts and statistical issues," *Quantitative Finance*, volume 1, no. 2, 223–236.

Cont, R. and Ben Hamida, S., 2004, "Recovering volatility from option prices by evolutionary optimization," .

Cont, R., Tankov, P., and Voltchkova, E., 2004, "Option pricing models with jumps: integro-differential equations and inverse problems," .

Corsi, F., 2009, "A simple approximate long-memory model of realized volatility," *Journal of Financial Econometrics*, volume 7, no. 2, 174–196.

Cox, J., 1975, "Notes on option pricing I: Constant elasticity of variance diffusions," *Unpublished note, Stanford University, Graduate School of Business.*

Cox, J., Ingersoll, J., and Ross, S., 1985, "A theory of the term structure of interest rates," in "Theory of Valuation," pp. 129–164, World Scientific.

Diebold, F. and Mariano, R., 2002, "Comparing predictive accuracy," *Journal of Business & economic statistics*, volume 20, no. 1, 134–144.

Ding, Z. and Granger, C., 1996, "Modeling volatility persistence of speculative returns: a new approach," *Journal of econometrics*, volume 73, no. 1, 185–215.

Duan, J., 1995, "The GARCH option pricing model," *Mathematical finance*, volume 5, no. 1, 13–32.

Dumas, B., Fleming, J., and Whaley, R. E., 1998, "Implied volatility functions: Empirical tests," *The Journal of Finance*, volume 53, no. 6, 2059–2106.

Eldan, R. and Shamir, O., 2016, "The power of depth for feedforward neural networks," in "Conference on Learning Theory," pp. 907–940.

Engle, R. and Patton, A., 2007, "What good is a volatility model?" in "Forecasting Volatility in the Financial Markets (Third Edition)," pp. 47–63, Elsevier.

Fournié, E., Lasry, J., and Touzi, N., 1997, "Monte Carlo methods for stochastic volatility models," *Numerical Methods in Finance*, volume 146, 164.

Franke, J., Härdle, W. K., and Hafner, C., 2015, *Statistics of financial markets*, volume 4, Springer.

Gradojevic, N., Gençay, R., and Kukolj, D., 2009, "Option pricing with modular neural networks," *IEEE transactions on neural networks*, volume 20, no. 4, 626–637.

Grzelak, L. and Oosterlee, C., 2011, "On the Heston model with stochastic interest rates," *SIAM Journal on Financial Mathematics*, volume 2, no. 1, 255–286.

Hagan, M. and Menhaj, M., 1994, "Training feedforward networks with the Marquardt algorithm," *IEEE transactions on Neural Networks*, volume 5, no. 6, 989–993.

Heston, S., 1993, "A closed-form solution for options with stochastic volatility with applications to bond and currency options," *The review of financial studies*, volume 6, no. 2, 327–343.

Heston, S. and Nandi, S., 2000, "A closed-form GARCH option valuation model," *The review of financial studies*, volume 13, no. 3, 585–625.

Hilliard, J. and Schwartz, A., 1996, "Binomial option pricing under stochastic volatility and correlated state variables," *Journal of Derivatives*, volume 4, no. 1, 23–39.

Hinton, G., Osindero, S., and Teh, Y.-W., 2006, "A fast learning algorithm for deep belief nets," *Neural computation*, volume 18, no. 7, 1527–1554.

Hu, G., Peng, X., Yang, Y., Hospedales, T., and Verbeek, J., 2018, "Frankenstein: Learning deep face representations using small data," *IEEE Transactions on Image Processing*, volume 27, no. 1, 293–303.

Hull, J., 2003, *Options futures and other derivatives*, Pearson Education India.

Hull, J. and White, A., 1987, "The pricing of options on assets with stochastic volatilities," *The journal of finance*, volume 42, no. 2, 281–300.

Hutchinson, J., Lo, A., and Poggio, T., 1994, "A nonparametric approach to pricing and hedging derivative securities via learning networks," *The Journal of Finance*, volume 49, no. 3, 851–889.

Jackwerth, J. C. and Rubinstein, M., 1996, "Recovering probability distributions from option prices," *The Journal of Finance*, volume 51, no. 5, 1611–1631.

Jarrow, R. and Rudd, A., 1983, *Option pricing*, Richard d Irwin.

Kaashoek, J. and van Dijk, H., 2003, "Neural networks: an econometric tool," *Statistics Textbooks and Monographs*, volume 169, 351–384.

Kaastra, I. and Boyd, M., 1996, "Designing a neural network for forecasting financial and economic time series," *Neurocomputing*, volume 10, no. 3, 215–236.

Katz, J. O., 1992, "Developing neural network forecasters for trading," *Technical Analysis of Stocks and Commodities*, volume 10, no. 4, 160–168.

Kohzadi, N., Boyd, M., Kermanshahi, B., and Kaastra, I., 1996, "A comparison of artificial neural network and time series models for forecasting commodity prices," *Neurocomputing*, volume 10, no. 2, 169–181.

Krauss, C., Do, X. A., and Huck, N., 2017, "Deep neural networks, gradient-boosted trees, random forests: Statistical arbitrage on the S&P 500," *European Journal of Operational Research*, volume 259, no. 2, 689–702.

Kuo, I.-D. and Wang, K.-L., 2009, “Implied deterministic volatility functions: an empirical test for Euribor options,” *Journal of Futures Markets: Futures, Options, and Other Derivative Products*, volume 29, no. 4, 319–347.

Lajbcygier, P. and Connor, J., 1997, “Improved option pricing using artificial neural networks and bootstrap methods,” *International journal of neural systems*, volume 8, no. 04, 457–471.

Lamberton, D. and Lapeyre, B., 2011, *Introduction to stochastic calculus applied to finance*, Chapman and Hall/CRC.

Lauterbach, B. and Schultz, P., 1990, “Pricing warrants: An empirical study of the Black-Scholes model and its alternatives,” *The Journal of Finance*, volume 45, no. 4, 1181–1209.

Leshno, M., Lin, V. Y., Pinkus, A., and Schocken, S., 1993, “Multilayer feedforward networks with a nonpolynomial activation function can approximate any function,” *Neural networks*, volume 6, no. 6, 861–867.

Lewis, A., 2009, *Option Valuation Under Stochastic Volatility Ii*, Finance Press, Newport Beach, CA.

Liu, L. and Patton, K., Andrew Jand Sheppard, 2015, “Does anything beat 5-minute RV? A comparison of realized measures across multiple asset classes,” *Journal of Econometrics*, volume 187, no. 1, 293–311.

Lunde, A. and Shephard, K., 2018, “Data on Realized volatility,” .
URL `https://realized.oxford-man.ox.ac.uk/data`

Malliaris, M. and Salchenberger, L., 1993, “A neural network model for estimating option prices,” *Applied Intelligence*, volume 3, no. 3, 193–206.

Masters, T., 1993, *Practical neural network recipes in C++*, Morgan Kaufmann.

Mikhailov, S. and Nögel, U., 2003, “Heston?s Stochastic Volatility Model Implementation,” *Fraunhofer Institute for Industrial Mathematics,*.

Moyaert, T. and Petitjean, M., 2011, “The performance of popular stochastic volatility option pricing models during the subprime crisis,” *Applied Financial Economics*, volume 21, no. 14, 1059–1068.

Mrázek, M. and Pospíšil, J., 2017, "Calibration and simulation of Heston model," *Open Mathematics*, volume 15, no. 1, 679–704.

Murata, N., Yoshizawa, S., and Amari, S.-i., 1994, "Network information criterion-determining the number of hidden units for an artificial neural network model," *IEEE Transactions on Neural Networks*, volume 5, no. 6, 865–872.

Rosenberg, J., 1998, "Pricing multivariate contingent claims using estimated risk–neutral density functions," *Journal of International Money and Finance*, volume 17, no. 2, 229–247.

Rubinstein, M., 1994, "Implied binomial trees," *The Journal of Finance*, volume 49, no. 3, 771–818.

Schöneburg, E., 1990, "Stock price prediction using neural networks: A project report," *Neurocomputing*, volume 2, no. 1, 17–27.

Shalev-Shwartz, S. and Ben-David, S., 2014, *Understanding machine learning: From theory to algorithms*, Cambridge university press.

Simpson, P., 1990, *Artificial neural systems: foundations, paradigms, applications, and implementations*, Pergamon.

Stein, E. and Stein, J., 1991, "Stock price distributions with stochastic volatility: an analytic approach," *The review of financial studies*, volume 4, no. 4, 727–752.

Vortelinos, D., 2017, "Forecasting realized volatility: HAR against Principal Components Combining, neural networks and GARCH," *Research in international business and finance*, volume 39, 824–839.

Wiggins, J., 1987, "Option values under stochastic volatility: Theory and empirical estimates," *Journal of financial economics*, volume 19, no. 2, 351–372.

Yao, J., Li, Y., and Tan, C. L., 2000, "Option price forecasting using neural networks," *Omega*, volume 28, no. 4, 455–466.

Zhang, J.-S. and Xiao, X.-C., 2000, "Predicting chaotic time series using recurrent neural network," *Chinese Physics Letters*, volume 17, no. 2, 88.

Declaration

1. I hereby declare that this thesis, entitled ”Option Pricing Using Machine Learning Techniques” is a result of my own work and that no other than the indicated aids have been used for its completion. Material borrowed directly or indirectly from the works of others is indicated in each individual case by acknowledgement of the source and also the secondary literature used.
This work has not previously been submitted to any other examining authority and has not yet been published.
2. After completion of the examining process, this work will be given to the library of the University of Konstanz, where it will be accessible to the public for viewing and borrowing. As author of this work, I agree to this procedure.

Konstanz, 6. March, 2019 ....

Pavel Podkovyrov

Erklärung

1. Ich versichere hiermit, dass ich die vorliegende Arbeit mit dem Thema "Option Pricing Using Machine Learning Techniques" selbständig verfasst und keine anderen Hilfsmittel als die angegebenen benutzt habe. Die Stellen, die anderen Werken dem Wortlaut oder dem Sinne nach entnommen sind, habe ich in jedem einzelnen Falle durch Angaben der Quelle, auch der benutzten Sekundärliteratur, als Entlehnung kenntlich gemacht. Die Arbeit wurde bisher keiner anderen Prüfungsbehörde vorgelegt und auch noch nicht verffentlicht.
2. Diese Arbeit wird nach Abschluss des Prüfungsverfahrens der Universitätsbibliothek Konstanz übergeben und ist durch Einsicht und Ausleihe somit der Öffentlichkeit zugänglich. Als Urheber der anliegenden Arbeit stimme ich diesem Verfahren zu.

Konstanz, 6. March 2019 ....

Pavel Podkovyrov