Using Neural Networks to Forecast the Implied Volatility: the Case of S&P100XEO

Mehmet Can and Šadi Fadda
International University of Sarajevo, Faculty of Engineering and Natural Sciences, Hrasnicka Cesta 15, Ilidža 71210 Sarajevo, Bosnia and Herzegovina

Abstract
Currently the most popular method of estimating volatility is the implied volatility. It is calculated using the Black-Scholes option price formula, and is considered by traders to be a significant factor in signaling price movements in the underlying market. A trader is able to establish the proper strategic position in anticipation of changes in market trends if she/he could accurately forecast future volatility. There is an abundance of ways to compute the volatility. For two decades neural networks has been developed to forecast future volatility, using past volatilities and other options market factors. In this article a network is created for this purpose whose performance demonstrates the value of neural networks as a predictive tool in volatility analysis.

1. INTRODUCTION

During the past three decades (Engle and Rothschild 1992), the increasing volume of the financial markets, also boomed the desire to forecast volatility of markets. This trend has motivated a large body of research. Volatility is the size of the price movement in the underlying markets, and is often used to predict the risk and forecast large moves in those markets (Caplan and Stein 1993; Black and Scholes 1973).

To understand the underlying process, relationships between volatility and numerous other variables such that market price of the underlying asset, strike price, risk free interest rate, time to maturity, option price have been studied in an attempt to make accurate predictions (Haugen, et al, 1985; Duan 1995; Hull and White 1987; Choi, and Shastri 1989; Merville and Piętea ,1989; Lockwood, and Linn 1990; Dubofsky 1991; Amin and Ng 1993; Scott 1997; Duan and Zhang 2001). The issue of testing the relationship between the implied volatility and future realized volatility has been the subjects for a number of studies (Canina and Figlewski, 1993; Christensen and Prabhala, 1998; Fleming, 1998). The test between implied and realized volatility also forms a formal test of information efficiency of the option market. Hull and White (1987) show that when volatility is constant, the Black-Scholes implied volatility of an at-the-money option approximately equals to the expected future realized volatility during the option life. They also showed that by trading on market information based on changes in volatility, profits can be earned. Hence the predictability of market volatility is important for accurately determining expected market return to properly price stocks (Merton 1980). To forecast closing prices, and to determine the optimal position to take early in the day, options traders use predictions of volatility. A correct prediction of volatility is also critical in designing optimal dynamic

Traders mostly use implied volatility calculated with the Black-Scholes formula. In this work, it will be shown that Neural networks, which have been shown to effectively model nonlinear relationships, accurately forecasts the volatility, and can thus be used to compute reliable forecasts for market volatility.

The paper is organized as follows. In Section 2, the techniques of computation of implied and historical volatility is shown. In Section 2, a general discussion of neural networks as prediction models, and for volatility prediction is presented. Section 3 provides the results of the networks and compares the predictions to the implied volatility estimates. In Section 4, along with suggestions for future research, a discussion of the results is presented.

2. CALCULATING IMPLIED AND HISTORICAL VOLATILITIES

In their famous paper on pricing options, Black and Scholes (1973) assumed that the price of the asset underlying the option follows an Ito process

\[ \frac{dS}{S} = \mu dt + \sigma dZ \]  

(2.1)

where \( \frac{dS}{S} \) denotes the rate of return, \( \sigma \) is the expected instantaneous volatility, \( \mu \) is the instantaneous expected rate of return, and \( Z \) is a standardized Wiener process, or \( dZ \) is a continuous-time random walk.

The Black-Scholes option pricing formula for calculating the equilibrium price of call and put options are

\[ c = S N(d_1) - X e^{-rT} N(d_2) \]  

(2.2)

\[ p = -S N(-d_1) + X e^{-rT} N(-d_2) \]  

(2.3)

where

\[ d_1 = \left( \frac{\ln(S_0/X) + (r + \sigma^2/2)T}{\sigma \sqrt{T}} \right)/\sigma \sqrt{T}, \]

and

\[ d_2 = d_1 - \sigma \sqrt{T} \]  

(2.4)

In (2.2-4) \( c \) and \( p \) are the market price to be charged for the call and put options, \( N \) is the cumulative normal distribution, \( T \) is the number of days remaining until expiration of the option expressed as a fraction of a year, \( S \) is the price of the underlying asset, \( r \) is the risk-free interest rate prevailing at period \( t \), \( X \) is the exercise price of the option, \( \sigma^2 \) is the variance rate of return for the underlying asset. For any time interval \([0, t]\) of length \( t \), the return on the underlying asset is assumed to be normally distributed with variance \( \sigma^2 t \).

Their formula expresses the call price \( c \), and put price \( p \) as functions of five inputs

\[ c = c(s, x, T, \sigma, r), \quad p = p(s, x, T, \sigma, r). \]  

(2.5)

Observe that the \( \mu \) of Eq. (2.1) does not appear in (2.5). If there is any arbitrage opportunity exists, market participants will execute arbitrage strategies in order to obtain riskless profits. As a result, biased prices will be influenced by investors and no more arbitrage opportunities exist. Thus a fair value must promise no arbitrage opportunities that are widely known as put-call parity (Stoll 1969).

The mathematical derivation of the call option pricing formula as shown in (Malliaris 1982) or (Lee, 1990) shows that arbitrage requires that the per unit of risk excess returns between two appropriately designed portfolios must be equal. Making the necessary substitutions in this put-call parity relationship, the term containing \( \mu \) drops out. With \( \mu \) now out of the picture and with four of the five remaining variables directly observable, an estimate of the asset’s volatility \( \sigma \) in (2.5) becomes the focal point of attention for both theorists and traders.

In the above it is seen that predictability of market volatility is important for accurately determining expected market return to properly price stocks, and traders use predictions of volatility to forecast closing prices.

There are two main approaches to estimating and predicting variable non constant volatility \( \sigma \); the historical approach and the implied volatility approach.

2.1 Calculating historical volatility

Calculation of historical volatility bases on the statistical definition of volatility. It is the simplest way of determination of the volatility because tomorrow’s volatility \( \sigma_{t+1} \) is an estimate obtained from a sample, of past prices of the underlying asset. Suppose that the sample size is \( n \) and let

\[ S_{t-n+1}, \ldots, S_{t-1}, S_t \]  

(2.6)

denote daily historical prices for the underlying asset. To get an estimate for \( \sigma_{t+1} \) first compute daily returns, \( \rho_{t-p} = 0, \ldots, n - 2 \), where

\[ \rho_{t-i} = \ln(S_{t-i}) - \ln(S_{t-i-1}). \]  

(2.7)

For a sample of \( n \) historical prices, we obtain \( n - 1 \) rates of daily return. The annualized standard deviation of these rates of return is defined as the volatility and called historical volatility and is used as an estimate of \( \sigma_{t+1} \). The nearby historical volatility uses 30 days of data, the middle historical volatility uses 45, and the distant historical volatility has 60 daily prices.
An obvious problem with the historical approach is that it assumes that future volatility will not change and that history will exactly repeat itself. Markets, however, are forward looking and numerous illustrations can be presented to show that historical volatility does not always anticipate future volatility. A better estimate, the one most used by traders to price options, comes from the Black–Scholes option pricing model itself (Choi and Wohar 1992).

2.2 Calculating implied volatility

Simply stated, supporters of implied volatility claim that tomorrow’s volatility \( \sigma_{t+1} \) can only be estimated during trading tomorrow, i.e. in real time. As option prices are being formed by supply and demand considerations, each trader assesses the asset’s volatility prior to making his or her bid or ask prices and, accepting the consensus price of a call as a true market price reflecting the corporate opinions of the trading participants, one solves the Black–Scholes model for the volatility that yields the observed call price. When volatility is calculated in this way, it is called the implied volatility, with the adjective ‘implied’ referring to the volatility estimate obtained from the Black–Scholes pricing formula. Unlike historical volatility, which uses past returns, the implied volatility is forward-looking to the stock’s future returns from now to the time of the expiration of the option. This implied volatility technique has become the standard method of estimating volatility at the moment of trading.

However, the parameter \( \sigma \) in the Black–Scholes–Merton pricing formulas (2.2-2.4) cannot be solved from these expressions. To illustrate how implied volatilities are calculated, suppose that the value of a European call option on a non-dividend-paying stock is 1.875 when \( S_0 = 21; \ K = 20, \ r = 0.1, \) and \( T = 0.25. \) The implied volatility is the value of \( \sigma \) that, when substituted into equation (2.2), gives \( c = 1.875. \) Unfortunately, it is not possible to invert equation (2.2) so that \( \sigma \) is expressed as a function of \( S_0, K, r, T, \) and \( c. \) However, an iterative search procedure can be used to find the implied \( \sigma. \) For example, we can start by trying \( \sigma = 0.20. \) This gives a value of \( c \) equal to 1.76, which is too low.

Because \( c \) is an increasing function of \( \sigma, \) a higher value of \( \sigma \) is required. We can next try a value of 0.30 for \( \sigma. \) This gives a value of \( c \) equal to 2.10, which is too high and means that \( \sigma \) must lie between 0.20 and 0.30. Next, a value of 0.25 can be tried for \( \sigma. \) This also proves to be too high, showing that \( \sigma \) lies between 0.20 and 0.25. Proceeding in this way, we can halve the range for \( \sigma \) at each iteration and the correct value of \( \sigma \) can be calculated to any required accuracy.

Using interval halving one has:

\[
T=0.25; \\
r=0.1; \\
S_0=21; \\
K=20; \\
a=.0; \\
b=.4; \\
c_0=1.875; \\
\epsilon_0=0.000001; \\
\]

\[
\text{Do}\{s=(a+b)/2, \\
d_1=(\log[S_0/K]+(r+s^2/2)T)/(s \sqrt{T}), \\
d_2=d_1-s \sqrt{T}, \\
p_1=\text{CDF[NormalDistribution[],d_1]}, \\
p_2=\text{CDF[NormalDistribution[],d_2]}, \\
call=S_0*p_1-K*\text{Exp[-r T]*p}_2, \\
\text{If[Abs[call-c]<\epsilon_0,} \{\text{Print[s],Break[]}\}, \\
\text{If[call<c_0,a=s,b=s}\}]\},\{i,1,1000\}\}
\]

\[
\sigma = 0.234513
\]
2.3. Comparison of implied and historical volatilities

Data have been collected for the most successful options market: the S&P 100 (OEX), traded at the Chicago Board Options Exchange. Daily closing call and put prices and the associated exercise prices closest to at-the-money, S&P 100 Index prices, call volume, put volume, call open interest and put open interest were collected from the Wall Street Journal for five days in April in the year 2014.

The historical and implied volatility for April 4 through April 9, 2014 are shown in Fig. 3.

It is seen that the historical estimate significantly underestimates the implied volatility. Since the historical volatility is an average based on returns from 30 preceding days, it is not surprising that the estimate smooths out the peaks, giving a value for each day which is less variable, and thus less sensitive to daily market fluctuations, while the implied volatility for any given day uses only trading information from that day, not a previous time period, to generate a value. Mean square of differences are shown in Table 1.

### Table 1. Mean square of differences between historical, and implied volatilities.

<table>
<thead>
<tr>
<th></th>
<th>Put</th>
<th>Call</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSE</td>
<td>0.063224</td>
<td>0.031132</td>
</tr>
</tbody>
</table>

3. NEURAL NETWORKS FOR A BETTER PREDICTION

Neural networks are an information processing technology which model mathematical relationships between inputs and outputs. Based on the architecture of the human brain, a set of processing elements or neurons (nodes) are interconnected and organized in layers. These layers of nodes can be structured hierarchically, consisting of an input layer, an output layer, and middle (hidden) layers. Each connection between neurons has a numerical weight associated with it which models the influence of an input cell on an output cell. Positive weights indicate reinforcement; negative weights are associated with inhibition. Connection weights are ‘learned’ by the network through a training process, as examples from a training set are presented repeatedly to the network. Each processing element has an activation level, specified by continuous or discrete values. If the neuron is in the input layer, its activation level is determined in response to input signals it receives from the environment. For cells in the middle or output layers, the activation level is computed as a function of the activation levels on the cells connected to it and the associated connection weights. This function is called the transfer function or activation function and may be a linear discriminant function, i.e. a positive signal is output if the value of this function exceeds a threshold level, and 0 otherwise. It may also be a continuous, nondecreasing function. Feedforward networks map inputs into outputs with signals flowing in one direction only, from the input layer to the output layer.

While there are dozens of network paradigms, the backpropagation network has frequently been applied to classification, prediction, and pattern recognition problems. Financial applications of neural networks
include underwriting (Collins 1988), bond-rating (Dutta and Shekhar 1988), predicting thrift institute failure (Salchenberger et al. 1992), and estimating options prices (Malliaris and Salchenberger 1993).

3.1. Neural Networks for a Better Prediction in Option Pricing

The term backpropagation technically refers to the method used to train the network, although it is commonly used to characterize the network architecture. In this learning algorithm, mean squared error and gradient descent are employed to determine a set of weights for the trained network. At each iteration, current weights are updated by minimizing the mean squared differences between the actual response of the system to a given example and the desired response. The nonlinear response functions generate gradients of the error function with respect to the weights and the chain rule is used to determine the appropriate weight changes which propagate back through the layers of the network. For more details of this method, see D. E. Rumelhart and J.L. McClelland (Rumelhart and McClelland 1986). Currently, a number of variations on this method exist which overcome some of its limitations.

Nonlinear, multilayer, feedforward networks differ from traditional modelling techniques in several ways. Relationships between inputs and outputs are learned during a training process in which the network is repeatedly presented with historical examples. Neural networks possess the ability to approximate arbitrary mappings with no a priori assumptions about the nature of the underlying model required. Also, no assumptions about the distributions of the variables are required and the variables may be highly correlated.

In the option pricing implied volatility is better than the historical volatility, but for any given day it uses only trading information from that day, to generate a value. Therefore it has a limited value to predict the future. Artificial Neural Network (ANN) technique on the other hand has an online adaptation, and traders can benefit from it in the option pricing. Many researches that analyze ANN performance from different perspectives. Malliaris and Salchenberger examined the performance of neural network option prices with the Black-Scholes prices by using S&P 100 index options (Malliaris and Salchenberger 1993). Approximately for half of the cases that they examined, mean squared error for the neural network is smaller than that of Black-Scholes, which implies the good performance of ANN relative to Black and Scholes. Hutchinson, Lo and Poggio studied whether artificial neural network can be used for pricing option in replace of Black and Scholes model with S&P 500 index options (Hutchinson et al. 1994). He reported that when parametric methods fail, nonparametric learning network alternatives can be useful substitutes, but they emphasized that the study did not claim that the learning network alternatives would be successful in general.

Anders, Korn and Schmitt examine the artificial neural network model to call options written on the German stock index DAX in order to determine the right combination of input for the best out-of-sample performance by applying the statistical inferences methods (Anders et al. 1998). Yao, Li and Tan reported the forecasting performance of back propagation neural network with Nikkei 225 Index futures data (Yao et al. 1999) that neural network option pricing outperforms the Black-Scholes for high volatile markets. They grouped the data differently to feed the neural network analysis in order to find best combination of input. They also do not take the volatility as an input in the neural network model, but they provide the volatility to be captured by the neural network. They concluded that the grouping data differently creates varying degree of accuracy.

Garcia and Gençay also work on the how pricing accuracy can be improved by homogeneity hint (Garcia and Gençay 2000). Instead of setting up a learning network mapping moneyness and maturity directly into the derivative pricing, they break down the pricing function two parts: one with moneyness, the other with the time to maturity. The results of their study showed that the homogeneity hint always reduces the out-of-sample performance. Amilon studied whether Multi Layer Perceptron (MLP) neural network can be used to find a call option pricing formula better than Black Scholes option pricing formula (Amilon 2003). He extended the Hutchinson, Lo and Poggio’s nonparametric approach to model the spread between bid and ask price by neural network, instead of taking simply the average of bid and ask price. Amilon made the performance comparison against two benchmarks, Black-Scholes prices with historical volatility and implied volatility. He reported that neural network models outperform against either benchmark in both pricing and hedging performances. By working on the Australian Stock Price Index, which also includes American Type option, Daglish reported that neural network analysis showed superior performance for in-sample pricing, however, the parametric methods showed a better performance in explaining the future prices and showed higher hedging performance (Daglish 2003).

Z. İltüzür Samur and G. Tekin Temur (İltüzür Samur and Tekin Temur 2003) used ANN to study European put options and American call and put options. The performance of ANN in put/call, American-European, moneyness dimensions are analyzed. They checked whether the contribution of the volatility in neural network analysis make improvement in prediction performance. They showed that ANN over performs in pricing call options compared to put options; and the use of volatility parameter as an input does not improve the performance. Bennell and Sutcliffe also compared the Black-Scholes performance with artificial neural network (ANN) in pricing European type FTSE 100 call options (Bennell and Sutcliffe 2004). They reported that for the out of money options, ANN have arguable superior performance over Black and Scholes model, but when moving to in-the-money options, performance of Black-Scholes is much better than ANN. However, they also reported that if input data exclude the options with moneyness greater than 1.15
and smaller than 0.9; and maturity greater than 200 days and smaller than 14, then both ANN and Black-Scholes show the same performance.

3.2 Architecture of an Artificial Neural Network to be used

ANNs are information processing models that are developed by inspiring from the working principles of human brain. The most essential property of an ANN is its ability of learning from sample sets. There are three different kinds of layers in a typical architecture of an ANN model. Input layer contain neurons that serve as input terminals of the data. Hidden layer contain neurons that receives information from input neurons, process this information and transfers to the output neurons. All neurons of a layer in and ANN architecture are internally in connection with neurons from subsequent layers. The ability of ANNs to process depends on these connections which are named as weights. The weights give the abilities of prediction or classification to the system (Haykin 1999; Gupta and Sinha 1999; Kahraman et al. 2009). The weights are iteratively changed according to learning rules’ results. The weights are modified until the output fits the expected values best (Johnson and Picton 1995).

To deduce volatility from data the most appropriate type of ANN’s is multi layered perception (MLP) with an output layer that consists of only one neuron which transmits the output to the decision makers. Input layer is the input receiver from external environment of the network and has neurons whose number is equal the dimension of the input data. Between input and output layers, there is an intermediate layer which is called the hidden layer. The basic architecture of a MLP network model is showed in Fig. 4.

Many trials are required for deciding on the best loads, best numbers of hidden layers and best numbers of neurons in each layer. However, in literature some rules are also defined to find the best numbers of hidden layers, such as \( \frac{n}{2} \), \( \frac{2n}{3} \), \( n+1 \) and \( 2n+1 \) (\( n \) is the number of input nodes). Previous researches also show that an ANN model with one hidden layer is sufficient for many complex systems. Larger number of neurons makes the ANN memorize the training data; hence ANN has very poor generalization power.

3.3 Implementation

In this section, ANN’s is used to predict the volatility that fits best to the market behavior. Then the volatility produced by ANN will be used to compute option prices via Black-Scholes formula, and these prices will be compared with prices in the market.

3.3.1 Definition of Data Set and Variable Set

There are 314 data for put and, 196 data for call options. The data cover the quotes covering S&P 100 European index options which are traded on the Chicago Board Options Exchange. The data are obtained from the Wall Street Journal web site. The data includes the options with varied strike prices and maturities on different days, i.e. 4th April to 9 April 2014.

For ANN implementation, spot price/strike price \( S/K \), time to maturity \( T \), risk free interest rate \( r \), call option price (c), put option price (p) are chosen as free variables, volatility \( \sigma \) is the dependent variable. ANN will perform as a multivariate regression function which connects regressors \( (S/K, T, r, c \text{ or } p) \) to the dependent variable volatility \( \sigma \).

3.3.2 Network Implementation

Input data is a four dimensional vector \( (S/K, T, r, c \text{ or } p) \). Therefore the input layer has \( 4+1 \) input neurons. A hidden layer with 10 neurons is found to be sufficiently complex for this job. Output layer has only one neuron that will give the predicted value volatility. For both call and put options data, half of the data is used for training and the other half is used for testing.

<table>
<thead>
<tr>
<th>Put Data</th>
<th>Call Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training</td>
<td>157</td>
</tr>
<tr>
<td>Testing</td>
<td>157</td>
</tr>
</tbody>
</table>

Implied volatilities computed using Black-Scholes formula and neural networks are compared. Mean square errors of 5 runs are as follows:

<table>
<thead>
<tr>
<th>MSE</th>
<th>Puts</th>
<th>Calls</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.018532</td>
<td>0.028747</td>
<td></td>
</tr>
</tbody>
</table>
4. RESULTS

We evaluated the performance of the neural network by measuring MSE of the differences between implied volatilities computed using Black-Scholes formula and neural networks. These results are shown in Fig. 5 and Table 3. After several runs, the smallest MSE is obtained is 0.0116.

5. DISCUSSION

The results of this study of neural networks for forecasting volatility are encouraging. Because historical estimates are traditionally poor predictors, traders have been forced to rely on formulas like the Black-Scholes which can be solved implicitly for the real-time volatility. However, these models can only provide real-time estimates to the traders. Furthermore, they fail to incorporate knowledge of the history of volatility. The neural network model, on the other hand, employs both short-term historical data and contemporaneous variables to forecast future implied volatility, enabling the trader to take a position when the market opens which will provide a strategic advantage. For example, high implied volatility often indicates the market is about to consolidate while low volatility often signals that the market is preparing for a breakout.

The neural network approach has two advantages which make it more useable as a forecasting tool. First, daily predictions can be made using data from previous trading cycles, thus providing a trading advantage. Secondly, in the cases we tested, the network forecasts were very accurate estimates of the volatility preferred by traders.

There are several ways to extend this research. Improvement may be possible through experimentation with other variables and network architectures. Radial basis networks offer another avenue of investigation, since these have been used for financial forecasting. In this research, we have predicted nearby volatility.

However, networks for predicting middle and distant volatility may be developed, as well, using different variables and different network architectures.

REFERENCES


Duan, J. and H. Zhang, “Pricing Hang Seng Index options around the Asian financial crisis- a Garch approach,”


Iltüzer Samur, Z. and G. Tekin Temur, The Use of Artificial Neural Network in Option Pricing: The Case of S&P 100 Index Options, World Academy of Science, Engineering and Technology 54 2009.


Samad, T., Backpropagation is significantly faster if expected value of source unit is used for update, INNS Conf. Abstracts, 1988.
