Option Pricing using Neural Networks

Technical Report by Norbert Fogarasi (Jan 2004)

1. Introduction

Among nonparametric option pricing techniques, probably the most fertile area for empirical research has been estimating option pricing formulas using neural networks.

The first attempt was made by Hutchinson, Lo and Poggio (1994) who used three different network architectures: Radial Basis Functions (RBF), Multi Layer Perceptron (MLP), Projection Pursuit Regression (PPR) to fit both Monte-Carlo simulated Brownian underlier and Black-Scholes option data, as well as S&P500 futures and options thereof. They used a minimalistic approach in their input selection, and limited the network inputs to time to maturity (T-t) and moneyness (ratio of underlying price to strike price, S/X) only, assuming interest rate and volatility to be constant. We also note that they used financial knowledge in this construction, namely the "homogeneity property" of the option pricing formula (Merton 1990), which justifies the use of the moneyness rather than the underlying price and strike price separately. Another important question this paper raises is that of measuring performance of the network. While the out-of-sample error, R^2 is a reasonable measure, they also examine the discrete delta-hedging performance of the neural network methods, which is of greater practical relevance. The result of these empirical tests is that these non-linear neural networks are able to recover the BS formula with remarkable accuracy, and even outperform BS in some cases in the discrete delta-hedging of simulated data. On the actual S&P500 data, the results are even more promising: all three of the neural network methods studied outperform the BS model both in terms of R^2 as well as delta-hedging performance, especially for longer term and out-of-the money options. The authors note that increasing the number of inputs, investigating the network architecture, re-examining the performance measures and establishing statistical significance of results are directions of further research. We examine the work done in each of these directions.

2. Selection of Input and Output Variables

Probably the most glaring shortcoming of Hutchinson et al (1994) is the severely limited number of input variables. Since they initially concentrate on recovering the BS formula, they argue that volatility and the risk-free interest rate are assumed to be constant across all training data, so there is no need to provide them as input to the network. In practice, however, we know that the reason why the BS formula has limited use is precisely this assumption, so having some measure of volatility and the level of interest rate as inputs to the model, makes intuitive sense.

Indeed, "the selection of input variables is a modelling decision and one which can greatly affect network performance. There is no well-defined theory to assist with the selection of input variables and heuristic methods are emplyed. One approach is to include all the variables in the network and performan analysis of the connection weights or a sensitivity analysis to determine which may be eliminated without reducing predictive accuracy. An alternative is to begin with a small number of variables and add new variables which improve network performance." (Malliaris, 1996) This study goes on to use the latter method and end up including no less than 13 input variables in their backpropagation network. While the specific results of this paper may be challenged due to the small amount of data and the basic performance measures (comparing MSE of implied volatility only, no hedge analysis) used, it is clear that addition of extra information in the form of input variables is an important factor in determining the effectiveness of neural networks. Amilon (2003) uses the first approach suggested above: this study starts out with 9 input variables, including lagged values of the underlying S(t-1), S(t-2), S(t-3) as well as 10-day and 30-day historical volitalities. Then, sensitivity analysis on each of the input variables is used to determine their relative importance.

Latter studies also spend ample time exploring this question. Qi and Maddala (1996) use historical volatility, interest rate and open interest as input variables, in addition to those used by Hutchinson et al (1994) and achieved better results than them. Lajbcygier (1995, 1996, 1997) advocates the use of implied volatility instead of historical volatility. He introduces the concept of Weighted Implied Standard Deviation (WISD), in which standard deviation implied by the Black model of at-the-money options is weighted more heavily than those that are out-of-money. DeWinne(2001) also experiments with various set of input variables and finds that the inclusion of both historical and implied volatility performs well, and liquidity information (such as trade size) improves the fit of intraday option pricing data.

As for the output variable, most studies use the option price, or ratio of option price to strike price. These are equivalent to outputting the implied volatility (as in Malliaris 1996, Carelli 1998). Some studies go further and attempt to model the bid-ask spread (Amilon 2003). An improvement was proposed by Lajbcygier (1995) who suggest that deviations from BS price be used as output, rather than the predicted option price itself. They call this a "hybrid network" and show that it has better performance than a network that tries to predict the option price only, without the BS information.

3. Selection of Network Architecture

Another potential point of improvement over the initial results of Hutchinson et al (1994) is the neural network architecture used. They used a single hidden layer in their study, which, according to the theoretical results of Cybenko (1989) and Hornik (1989) is sufficient to universally approximate any continuous function. The selection of the number of hidden neurons is also quite important: too few neurons would result in a poor fit, whilst the use of too many is likely to result in overfitting. There are numerous methods to address this problem, including the validation-set-method of Peterson et al (1991), Lagrange-multiplier test of White (1989) and Terasvirta (1993), cross-validation techniques of Wahba (1990) as well as various heuristic and comparative methods.

Carelli et al (1998) propose the use of a non-linear regression technique called "profiling" to select the optimal network structure specifically for the problem of option pricing using feedforward neural networks.

The most recent strand of literature on this topic introduces various *constraints* on the architecture, in an attempt to incorporate theoretical knowledge into the network. Garcia et al (2001) introduce the concept of "homogeneity hint" which is a constraint on the set of allowable solutions to which the learning process may converge, thereby avoiding overfitting and improving the out-of-sample performance. They show that further enforcing the homogeneity property (ie. that the option pricing function has homogeneity of degree one in the asset price and the strike price) in the structure of the network improves out-of-sample pricing as well as dynamic hedging performance. Lajbcygier (2001) also used constraints in a similar fashion: the "maturity constraint" is introduced to enforce the knowledge that when an option reaches maturity, the pay-off is certain, other "model constraints" are used to keep the hybrid model (see above) consistent with some of the extremes of the underlying modified BS model. Finally, we note that in other areas of nonparametric option pricing research (i.e. those which do not user neural networks for estimation), pricing under various theoretical constraints is an active area (see for example Ghysels 1997, Ait-Sahalia 2003).

4. Selection of Performance Measures

There are two major groups of performance tests that have been conducted on neural networks used for option pricing. The first group of measures indicates the difference between the network-predicted out-of-sample option prices and the realized market prices. R^2, mean error (ME), residual mean squared error (RMSE), mean absolute error (MAE), mean absolute percentage error (MAPE) can all be used to quantify this difference and thus compare the pricing accuracy of various methods.

The second group of measures concerns the hedging performance of the option pricing model. "The ability to hedge an option can be used as a measure-of-fit in model comparisons, and, in many cases, it is a more accurate way to proceed. The market option prices reflect the anticipation of the future distribution of the underlying asset. If these expectations are not fulfilled on average, then the market option prices are wrong and, consequently, measures-of-fit based on price accuracy are of little use." (Amilon 2003) Based on these considerations, Hutchinson et al (1994) introduce the concept of "tracking error" of replicating portfolios designed to delta-hedge an option position. They use the mean and variance of these tracking errors to show that the discreet (daily) hedging performance of neural networks is superior to that of BS. Taking this concept one step further, based on the ideas of Chesney and Scott (1989), we can define a trading policy based on a particular pricing model. Assuming we buy options which are underpriced in the market according to our model and then delta-hedge it until it becomes overpriced or matures (whichever occurs first), and vice-versa for overpriced ones, we can compute the discounted trading profit/loss associated with a particular pricing model. This is a particularly intuitive and practical measure, which is used in Lajbcygier (2001) and in Amilon (2003).

5. Selection of Benchmarks

Given its theoretical foundations and wide-spread use, the BS model, using historical volatilities is an ideal benchmark for any option-pricing model. Most studies, including Hutchinson et al (1994) use this as the basis of comparison. However, given that BS prices have been shown to be quite far from market prices, especially for longer term and out-of-money options, recent studies have started using "more competitive" benchmarks. DeWinne (2001) uses the discrete binomial tree model of Cox, Ross and Rubinstein to compare the performance of neural networks in pricing American-style options. Amilon (2003) uses the BS model with both historical estimates of volatility (the standard method) and also with implicit volatility estimates (extracting the volatility implied by the market price of at-the-money options).

6. Establishing Statistical Significance of Results

Hutchinson et al (1994) note that statistical tests to determine whether the results are statistically significant, are difficult to formulate because the errors in fitting the option prices are likely to be correlated across options and over time. Amilon (2003) uses the block bootstrap method of Kunsch (1989) and Fitzenberger (1998) to establish a confidence interval for the RMSE to establish that neural network generated prices fit market prices better than BS prices do, at the 95% confidence level.

Lajbcygier (1997) establishes confidence intervals around the model-generated prices using bootstrap methods (Tibshirani 1996, LeBaron and Weigend 1998), and shows that trading policies which only only deal in "badly mispriced" (ie. market price 2-3 sigmas away from model price) options generate more profits (and better Sharpe ratio) than those without regard to confidence intervals.

7. Directions for Further Research

Each of the above areas still has open questions, but the following have direct consequences, in terms of the practical applicability of neural networks in option pricing:

- A well-defined quantitative selection method to decide on the input variables and on the network structure.
- Inclusion of further financial axioms and rules, as constraints to the learning network.
- Inclusion of the practical applicability (hedging performance, trading performance etc) of the network in the learning rules/objective function.

8. References

Ait-Sahalia, Y and Duarte, J. Nonparametric option pricing under shape restrictions. *Journal of Econometrics.* 116:9-47, 2003.

Amilon, H. A Neural Network Versus Black-Scholes: A Comparison of Pricing and Hedging Performances. *Journal of Forecasting,* 2003.

Carelli, A, Silani S. and Stella, F. Profiling Neural Networks for Option Pricing. *International Journal of Theoretical and Applied Finance,* 2:183-204, 2000.

Chesney, M and Scott, L. Pricing European Currency Options: A Copmarison of the Modified Black-Scholes Model and the Random Variance Model. *Journal of Financial and Quantitative Analysis.* 24:267-284, 1989.

Cybenko, G. Approximation by SUperpositions of a Sigmoidal Function. *Mathematics of Control, Signals and Systems.* 2, 303-314, 1989

DeWinne R, Meurisse B and Francois-Heude A. Market Microstructure and Option Pricing: A Neural Network Approach. Working paper, 2001.

Fitzenberger, B. The Moving Blcoks Bootstrap and Robust Inference for Linear Least Squares and Quantile Regressions. *Journal of Econometrics.* 82:235-287, 1998.

Garcia, R. and Gencay, R. Pricing and Hedging Derivative Securities With Neural Networks and a Homogeneity Hint. *Journal of Econometrics.* 93-115, 2000.

Ghysels, E. Patilea, V. Renault, E. and Torres, O. Nonparametric Methods and Option Pricing. Discussion Paper, CIRANO, Montreal, 1997.

Hornik, K. Multilayer Feedforward Networks are Universal Approximators. *Neural Networks.* 2:359-366, 1989.

Hutchinson, JM., Lo, AW. and Poggio, T. A Nonparametric approach to pricing and hedging derivative securities via learning network. *Journal of Finance.* 3:851-889, 1994.

Kunsch, H. The Jackknife and Bootstrap for General Stationary Observations. *Annals of Statistics.* 17:1217-1241.

Lajbcygier, P, Boek C., Palaniswami M. and Flitman, A. A Hybrid Neural Network Approach to the Pricing of Options. Working paper, 1995.

Lajbcygier, P, Boek C., Palaniswami M. and Flitman, A. Comparing Conventional and Artificial Neural Network Models For the Pricing of Options on Futures. *Neurovest Journal.* 5:16-24, 1996.

Lajbcygier, P and Connor, J. Improved Option Pricing Using Artificial Neural Networks and Bootstrap Methods. *International Journal of Neural Systems*, 8:457-471, 1997

Le Baron and Weigend, AS. A Bootstrap evaluation of the effect of data splitting on financial time series. *IEEE Trans. Neural Networks.* 9:213-220, 1998.

Malliaris, M. and Salchenberger, L. Using Neural Networks to Forecast the S&P 100 Implied Volatility. *Neurocomputing,* 183-194, 1996.

Peterson, C. and Rognvaldsson, T. An Introduction to Artificial Neural Networks, Department of Theoretical Physics, Lund University, Lund, 1991.

Qi, M and Maddala, G.S. Option Pricing Using Artificial Neural Netoworks: The Case of S&P 500 Index Call Options, *Neural Networks in Financial Engineering, Proceedings of 3rd International Conference on Neural Networks in the Capital Markets,* London, World Sceintific, 78-91, 1995.

Terasvirta T, Lin C-F and Granger, C.W. Power of the Neural Network Linearity Test. *Journal of Time Series Analysis.* 14:209-220, 1993.

Tibshirani, R. A comparison of some error estimates for neural network models. *Neural Comput.* 8:152-163, 1996.

Wahba, G. Spline Models for Observational Data, volume 59 of *Regional Converence Series in Applied Mathematics.* SIAM Press, Philadelphia, 1990.

White, H. An Additional Unit Test for Neglected Non-linearity in Multilayer Feedforward Networks. *Proceedings of the International Joint Converence on Neural Networks,* San Diego 451-455, 1989.