# Improving the pricing of options: a neural network approach — Source link

Ulrich Anders, Olaf Korn, Christian Schmitt

**Published on:** 01 Sep 1998 - Journal of Forecasting (John Wiley & Sons, Ltd)

**Topics:** Valuation of options, Hedge (finance), Stock market index and Statistical inference

Anders, Ulrich; Korn, Olaf; Schmitt, Christian

**Working Paper — Digitized Version**

# Improving the pricing of options: a neural network approach

ZEW Discussion Papers, No. 96-04

**Provided in Cooperation with:**
ZEW - Leibniz Centre for European Economic Research

# Discussion Paper

Discussion Paper No. 96-04

# Improving the Pricing of Options – A Neural Network Approach –

Ulrich Anders
Olaf Korn
Christian Schmitt





ZEW

Zentrum für Europäische Wirtschaftsforschung GmbH

International Finance Series

Discussion Paper No. 96-04

# Improving the Pricing of Options
# – A Neural Network Approach –

Ulrich Anders
Olaf Korn
Christian Schmitt

# Improving the Pricing of Options
## — A Neural Network Approach —

Ulrich Anders, Olaf Korn, Christian Schmitt

Centre for European Economic Research (ZEW), Mannheim

May 1996

## Abstract

In this paper we apply statistical inference techniques to build neural network models which are able to explain the prices of call options written on the German stock index DAX. By testing for the explanatory power of several input variables serving as network inputs, some insight into the pricing process of the option market is obtained. The results indicate that statistical specification strategies lead to parsimonious networks which have a superior out-of-sample performance when compared to the Black/Scholes model. We further validate our results by providing plausible hedge parameters.

Key words: *Option Pricing, Neural Networks, Statistical Inference, Model Selection.*

### Acknowledgements

We thank Deutsche Börse AG, Frankfurt/Main and Deutsche Finanzdatenbank, Mannheim for providing the data. We are indebted to Daniel Schwamm for capable research assistance. Christian Schmitt gratefully acknowledges the financial support of the German Research Foundation (DFG), grant Ko 219/19-3.

# 1 Introduction

Most of the theoretical work on option pricing has focused on the idea of creating risk-free portfolios through dynamic hedging strategies, which should earn the risk-free rate of interest in the absence of arbitrage opportunities. This line of research follows the seminal papers of Black/Scholes (1973) and Merton (1973). The original model of Black and Scholes has since been refined in several directions. An important one of these is the derivation of pricing formulae which take into account some empirical characteristics of financial assets such as non-normal return distributions, stochastic volatilities or stochastic interest rates. See for example the models of Merton (1973, 1976), Cox/Ross (1976), Geske (1979), Rubinstein (1983), Hull/White (1987) and Duan (1995). A common feature of all these models is the assumption of a specific stochastic process driving the price dynamics of the underlying securities.

A different approach to option pricing was suggested by Hutchinson/Lo/Poggio (1994) and Malliaris/Salchenberger (1993). Rather than starting from a price process of the underlying security and subsequently deriving the corresponding option value, the option market's pricing mechanism is estimated from observed prices via a neural network. Thus both the implicit stochastic process of the underlying security and its relation to the option price are determined from observed data, i.e. from the market opinion. Once the network model has been estimated it can be used for out-of-sample pricing and the calculation of hedge parameters.

As option pricing theory typically derives nonlinear relations between an option price and the variables determining it, a highly flexible statistical model is required to capture the empirical pricing mechanism. Neural networks are well suited for this purpose due to their ability to approximate virtually any (measurable) function up to an arbitrary degree of accuracy, as was shown amongst others in Hornik/Stinchcombe/White (1989). First empirical results given in Hutchinson/Lo/Poggio (1994) and Malliaris/Salchenberger (1993) for S&P 500 futures options and in Lajbcygier et al. (1995) for Australian All Ordinary Share Price Index (SPI) futures options are promising for the network approach, though further research is needed.

In this study we apply neural networks to price call options on the leading German stock index, called the Deutscher Aktien Index (DAX). The main difference from previous work, however, is the use of statistical inference for neural networks as developed by White (1989a,b).

In this paper we adopt a model selection strategy based on significance tests, as suggested by Anders/Korn (1996). The application of this strategy leads to a network architecture which is particularly geared to the data set at hand. Moreover, as the resulting model contains only statistically significant terms, it will be protected against over-parameterization, and thus the out-of-sample performance of the network should improve.

The usual approach to model specification as used in Hutchinson/Lo/Poggio (1994), Lajbcygier et al. (1995a) and Malliaris/Salchenberger (1993) is cross-validation. In cross-validation techniques, the whole data sample is split into a training set and a validation set. Different networks are estimated from the training set and judged upon their performance on the validation set. This leads to a trial and error procedure which is usually quite time consuming. Moreover, as splitting the data set results in some loss of information, the out-of-sample pricing accuracy will in general reduce due to the less precisely estimated parameters.

By the help of statistical inference one can distinguish which input variables contribute significantly to the explanation of option prices. As theoretical pricing formulae are easily nested in a neural network, it is possible to investigate whether the relationships between each input variable and the observed option prices differ significantly from the propositions of the theory. The existence of such differences could suggest directions for further refinements to theoretical pricing models.

The remainder of this paper is organized as follows: Section 2 shortly reviews some important results about statistical inference in neural networks and describes our architecture selection strategy. In section 3 we introduce the option pricing models which are compared in this study. As a reference point we start with the Black/Scholes model and then consider pure neural networks chosen solely on statistical grounds. As a last specification we nest the Black/Scholes model in a neural network. This allows us to test whether single input variables influence the option price in addition to the contribution of the theoretical model. Section 4 describes our data set while section 5 provides the empirical results. We compare the out-of-sample pricing accuracy of different models and the behaviour of hedge parameters such as the option's delta and gamma, which are important for risk management. The results are summarized in section 6.

## 2 Neural Network Models

Neural networks are a new, very flexible class of statistical models. Unfortunately, the term 'neural network' is not uniquely defined. Instead, it is comprised of many different network types. Since it is our goal to extract an alternative option pricing formula from market observations, we focus on those neural networks which are applicable to nonlinear regression problems, such as

$$y = F(X) + \varepsilon, \tag{1}$$

where $y$ is the dependent variable and where the columns of $X = [x_0, x_1, \dots, x_I]$ are the independent variables. The variable $x_0$ is defined to be constant and set to $x_0 \equiv 1$, while $\varepsilon$ stands for an *iid* error term with $E[\varepsilon\varepsilon'] = \sigma I$, $E[\varepsilon] = 0$ and $E[\varepsilon|X] = 0$.

The neural network literature knows basically two different types of regression networks, the so-called multilayer perceptron (MLP) and the so-called radial basis function (RBF) network. Although both network types have the universal approximation capability[1] and are therefore well suited to modelling option prices, here we deal exclusively with the MLP type of neural networks.

[1]Compare e.g. Poggio/Girosi (1990) for RBF–networks and Hornik/Stinchcombe/White (1989) for MLP–networks.

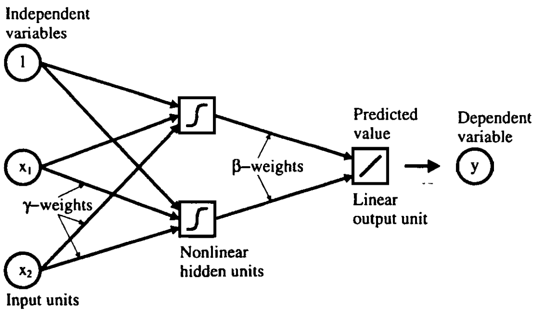


*Figure 1:* A multilayer perceptron neural network.

The network used in our study is a single hidden layer feedforward neural net, with a linear output unit as shown in figure 1. The output of this network is generated by the function:

$$f(X, w) = \sum_{h=1}^{H} \beta_h g \left( \sum_{i=0}^{I} \gamma_{hi} x_i \right) \tag{2}$$

with network weights $w = (\beta', \gamma')'$. The scalars $I$ and $H$ denote the number of input and hidden units in the net and $g(\cdot)$ is a nonlinear transfer function attached to each hidden unit. Usually $g(\cdot)$ is either the logistic function or the tangens hyperbolicus function. Apart from a monotonic transformation these transfer functions are identical. Due to its symmetry around the origin and its easily computable derivatives we prefer to use the tanh-function.

In contrast to Hutchinson/Lo/Poggio (1994) we focus exclusively on MLP–networks for two reasons. Firstly, it has been proven (Hornik/Stinchcombe/White, 1990) that feedforward networks with as little as one hidden layer and a linear output unit are able to approximate not only the unknown function, but simultaneously its unknown derivatives up to an arbitrary degree of accuracy. This characteristic is substantial since the partial derivatives of a pricing formula are needed for the hedging of option positions, a subject of similar importance as the pricing itself. Furthermore, the computation of the partial network derivatives provides a check as to whether the estimated network pricing formula is consistent with some basic theoretical results.[2]

Secondly, compared to the RBF–network, the MLP–network allows for the application of standard inference techniques known from parametric statistics. An application of such

[2]For example, the call option price should be a monotonically increasing function of the stock price.

techniques may be possible for RBF-networks as well. However, to our knowledge no work has as yet been published on this subject.

Statistical inference in MLP-networks was developed by White (1989a,b). He showed that — if the parameters of a neural network are identified — they can be consistently estimated by maximum likelihood methods. Moreover, the parameter estimates of a network are asymptotically normally distributed. This knowledge in principle allows the application of standard asymptotic hypotheses tests, such as Wald-tests or LM-tests.

However, as neural networks in general do not encompass the unknown functions but only approximate them, they are inherently misspecified models. The theory of statistical inference techniques for misspecified models is again based upon the work of White (1982, 1994). He proved that the application of standard asymptotic tests is valid even if the model is misspecified. One has though to take into account the misspecification when the covariance matrix $\frac{1}{T}C$ of the estimated parameters is computed. The estimated parameters $\hat{w}$ are asymptotically normally distributed around an optimum parameter vector $w^*$, which corresponds to the best projection of the misspecified model onto the true model $F$. In summary these results can be stated as

$$\sqrt{T} \cdot (\hat{w} - w^*) \sim N(0, C), \tag{3}$$

where $T$ is the number of observations. Due to the theory of misspecified models the covariance matrix of the parameters becomes $\frac{1}{T}C = \frac{1}{T}A^{-1}BA^{-1}$. The matrices $A$ and $B$ are defined as $A \equiv E[\nabla^2 \mathcal{L}_t]$ and $B \equiv E[\nabla \mathcal{L}_t \nabla \mathcal{L}_t']$ where $\nabla$ denotes the gradient and $\mathcal{L}_t$ the log-likelihood contribution of the $t$-th observation.

Unfortunately, we are left with the problem that the parameters of a neural network are not always identified, due to mutual dependencies between them. In such a case the parameters are no longer normally distributed and inference is cumbersome. To see the identification problem, consider equation (2). For instance, if a parameter $\beta_h$ equals zero, the corresponding weights $\gamma_{hi}$ can take any value without influencing the network's output, and are thus not identified. This situation occurs whenever the network is overparameterized in the sense that irrelevant hidden units exist.

Two techniques have been proposed in the literature to circumvent the identification problem. One was developed by White (1989c) and its properties investigated by Lee/White/Granger (1993). The other was devised by Teräsvirta/Lin/Granger (1993) and compared to the former. With these techniques we are able to perform an LM-test on whether or not an additional hidden unit is irrelevant.

White (1989c) suggests drawing the $\gamma$-weights of the additional hidden unit from a random distribution. This amounts to a random choice of the parameters in $\gamma$-space. The subsequent test is carried out conditional to the random values of $\gamma$. Teräsvirta/Lin/Granger (1993) propose the application of a third order taylor expansion to the additional hidden unit which equally leads to an avoidance of the identification problem.

In order to specify a network architecture we have to choose both the relevant input variables and the appropriate number of hidden units, i.e. the complexity of the functional form. For this purpose, we apply one of the model selection strategies suggested by Anders/Korn (1996) which is based on the techniques of White (1989c) or Teräsvirta/Lin/Granger (1993).

In the process of network architecture selection we have to ensure the identification of our model whenever inference techniques are used. Consequently, the strategy cannot adopt a top down approach which starts with a large (and probably over-parameterized) neural net. To obtain statistically valid results, the strategy begins with the smallest model possible and successively adds more complexity.

The strategy runs as follows: in the first step, all $I$ input variables are combined with one hidden unit and the relevance of the hidden unit is tested by the LM–test procedures of White (1989b) or Teräsvirta/Lin/Granger (1993). If the hidden unit is not relevant the procedure stops. If the unit is relevant, it is included in the model and the identification of the extended network follows. This allows the application of standard Wald–tests to decide the significance of each input unit connection. Only the significant connections remain in the model. In the next step the significant input units are connected with a second hidden unit and the whole procedure is repeated. The procedure stops if no further hidden unit shows relevance.

Since this model selection strategy is built upon inference techniques the resulting network leaves us with some insight into the statistical significance of the inputs fed into the network. In the simplest case, inputs which have no connection to hidden units show no relevance in explaining the observed call option prices.

# 3 Option Pricing Models

In this study we use the Black/Scholes model (1973) as a reference point. Although several extensions and refinements of this model exist,[3] which might give superior results for specific data sets, we believe that the basic model is still the most relevant in practice due to its simple closed-form solution and its robustness.

The derivation of the Black/Scholes model (BS) relies on the following assumptions: Asset prices follow a geometric Brownian motion; mean returns and volatilities are constant over time; interest rates are both constant over time and equal for all maturities; trading occurs continuously on frictionless markets and no arbitage opportunities exist. From these premises Black and Scholes derived the following formula for the price of a European call option written on a non-dividend paying stock:

$$C_{\mathrm{BS}}(t) = S\,\mathcal{N}(d_1) - X\,\mathrm{e}^{-r(T-t)}\,\mathcal{N}(d_2) \tag{4}$$

where

[3] See e.g. Hull (1993), Chapter 17.

$$d_1 \equiv \frac{\ln(\frac{S}{X}) + (r + \sigma^2/2)(T-t)}{\sigma\sqrt{T-t}} \tag{5}$$

$$d_2 \equiv d_1 - \sigma\sqrt{T-t} \tag{6}$$

| | | |
|---|---|---|
| $S$ | $\equiv$ | Price of the underlying stock |
| $X$ | $\equiv$ | Strike price of the option |
| $\sigma$ | $\equiv$ | Volatility of the continuously compounded stock returns |
| $r$ | $\equiv$ | Continuously compounded interest rate |
| $T-t$ | $\equiv$ | Time to maturity of the option contract |

and $\mathcal{N}(x)$ is the cumulative distribution function of the standard normal distribution.

Following equation (4) the call option price $C$ depends on five variables, namely the stock price $S$, the strike price $X$, the volatility $\sigma$, the interest rate $r$ and the time to maturity $(T-t)$ of the contract. It was shown by Merton (1973, theorem 9) that the option price is linear homogeneous of order one in $X$ and $S$ for every 'rational' pricing model, if the return distribution of the underlying stock does not depend on the stock price level. As this condition is valid for the Black/Scholes model, the number of input variables can be reduced to four by treating $C/X$ as a function of $S/X$, $\sigma$, $r$ and $(T-t)$. The corresponding pricing formula becomes:

$$\frac{C_{\mathrm{BS}}(t)}{X} = \frac{S}{X}\mathcal{N}(d_1) - \mathrm{e}^{-r(T-t)}\mathcal{N}(d_2). \tag{7}$$

Our second option pricing formula relies exclusively on an estimated neural network (NN). The formula takes the form given in equation (8):

$$\frac{C_{\mathrm{NN}}(t)}{X} = \sum_{h=1}^{H} \hat{\beta}_h \cdot g\left(\hat{\gamma}_{h0} + \hat{\gamma}_{h1} \cdot \frac{S}{X} + \hat{\gamma}_{h2} \cdot r + \hat{\gamma}_{h3} \cdot (T-t) + \hat{\gamma}_{h4} \cdot \sigma\right), \tag{8}$$

where $\hat{\beta}$ and $\hat{\gamma}$ are the parameter values estimated from a regression of the observed prices on the neural network. As input units we choose the same four variables as those contained in the reduced Black/Scholes model, though stock prices are added in a second step to test for level effects. The model is estimated by least squares and the network architecture results from the selection strategy outlined in section 2.

In a third pricing formula the Black/Scholes model is nested in a neural network (BS+NN). This leads to the following pricing equation (9):

$$\frac{C_{\mathrm{BS+NN}}(t)}{X} = \frac{C_{\mathrm{BS}}(t)}{X} + \sum_{j=1}^{J} \hat{\beta}_j \cdot g\left(\hat{\gamma}_{j0} + \hat{\gamma}_{j1} \cdot \frac{S}{X} + \hat{\gamma}_{j2} \cdot r + \hat{\gamma}_{j3} \cdot (T-t) + \hat{\gamma}_{j4} \cdot \sigma\right). \tag{9}$$

It is an advantage of the nested model that those parts of the pricing mechanism which are already explained by the theoretical formula need not be approximated by the network. When the Black/Scholes model already provides reasonable results the network can concentrate on the differences between theoretical and observed prices. If estimation errors are reduced, the out-of-sample accuracy of the pricing formula should improve.

A pressing question is which variables should enter the network part of pricing equations (8) and (9). This problem has not been addressed in previous work as it deserves the application of statistical inference. In this study, we test for both the significance of single input variables and the number of necessary hidden units, i.e. the degree of additional functional complexity needed to improve the Black/Scholes model.

An important task in practice is the hedging of option positions.[4] The chosen pricing model provides important information about the appropriate strategies. Of primary interest are the hedging parameters or 'greek' letters resulting from the pricing model. They are defined as follows:

$$\Delta \equiv \frac{\partial C}{\partial S}, \; \Theta \equiv -\frac{\partial C}{\partial (T-t)}, \; \Gamma \equiv \frac{\partial \Delta}{\partial S},$$

where delta ($\Delta$) and theta ($\Theta$) are the partial derivatives of the option price with respect to changes in the stock price and the time to maturity, while gamma ($\Gamma$) gives the sensitivity of delta with respect to changes in the stock price. We calculate hedge parameters in order to further validate our models.

# 4 The Dataset

In our study, we used transaction data on call options issued on the leading German stock index, called DAX. The index is comprised of 30 major German stocks, selected with respect to market capitilization, turnover, and early availability of opening prices. The DAX is a capital-weighted performance index which is adjusted for stock splits, dividend markdowns[5], and capital changes. It is calculated by the minute during trading hours at an accuracy of 0.01 index points.

In August 1991, the DAX option was introduced at the German Futures and Options Exchange (DTB). Since then it has developed into the most liquid option traded on the DTB.[6] The value of an option contract is the current index level multiplied by 10 German Marks (DM). Option prices are quoted in points where each point represents DM 10,– of contract value. The tick size is 0.1 points which corresponds to a tick value of DM 1,–.

The option's exercise prices have fixed increments of 25 index points, e.g. 2050, 2075, 2100. For each contract month there are at least five option series: two in-the-money, one

[4]For a discussion of this topic see Hull (1993), Chapter 13.

[5]In contrast to other indices, the adjustment for dividends is a particular feature of the DAX.

[6]The trading volume of DAX options is greater than that of all 20 DTB-traded stock options together.

at-the-money, and two out-of-the-money. If the DAX falls below (rises above) the average of the second- and third-lowest (highest) exercise price, option series with new exercise prices are introduced. At all times, there are options with five different expiration months available. The maximum time to maturity of an option contract is nine months.

Since the adjustment for dividends is carried out by reinvesting the total amount of dividend payments into the dividend-paying stock, a stock's value in the DAX portfolio remains unchanged. Consequently, the dividend payments of the 30 DAX–shares need not be considered for the calculation of option prices. Furthermore, as the DAX option is of European style, the standard Black/Scholes model provides a suitable pricing formula.

Our data set contains intraday time-stamped data on DAX call options traded on the DTB from January 1992 to the end of 1994.[7] Since this data set consists of more than half a million transaction data records, it had to be restricted.

For the empirical investigation we chose the most recent one year period, covering the whole of 1994. Within each trading day we selected all transactions that took place between 11:00 a.m. and 11:30 a.m. Each transaction record contains the option price ($C$), the exercise price ($X$) and the time to maturity ($T - t$).

In order to remove uninformative and non-representative option records we employed exclusion criteria similar to those of Rubinstein (1985), Sheikh (1991), Resnick/Sheikh/Song (1993) and Xu/Taylor (1994):

1. The call option is traded at less than 10 points.[8]

2. The option has less than 15 days to maturity.

3. The lower boundary condition for the value of European call options is violated:

$$C < S - X \cdot e^{-r \cdot \tau}.$$

4. The option is deep-in- or deep-out-of-the-money: $\frac{S}{X} < 0.85$ or $\frac{S}{X} > 1.15$.

Despite the tick size of 0.1 points, a preliminary analysis of the data showed that there is a tendency for options to be traded at integer values. This leads to high percentage deviations between observed and theoretical prices when the option value is very low. Thus, criterion 1 excludes options with low prices. Criterion 2 is used to eliminate options with a short time to maturity, as these options have only a small time-value and the integer pricing behaviour leads to severe deviations when calculating theoretical option prices. The third criterion excludes options whose prices are not consistent with a no-arbitrage condition which is binding for all European-style options independent of a specific option pricing model.[9] With criterion 4, deep-in-the-money and deep-out-of-the-money options are excluded, as these options are traded roughly at their intrinsic value and have almost

[7]The data set was provided by the Deutsche Börse AG, Frankfurt/Main.

[8]The value of 10 points leads to an exclusion of options which are traded at a price of less then 5% of the average DAX in 1994.

[9]See Hull (1993), page 156.

no informational content. Furthermore, the trading volume is very low for these options. Our resulting data set consists of 13,676 observations.

To obtain a theoretical price according to the Black/Scholes formula, we had to tie our option prices to an appropriate level of the DAX ($S$), the riskfree interest rate ($r$) and the return volatility ($\sigma$). In this respect, every transaction was linked with the current intraday DAX index level.[10] This means that each transaction between say 11:20 and 11:21 was combined with the DAX index level of 11:20.[11]

Our interest rate data consist of averaged daily bid and ask interbank rates for overnight, one month, three month, six month and twelve month money.[12] In order to calculate an adequate interest rate which matches the time to maturity for each option, we linearly interpolated the neighbouring interest rates and transformed the resulting values into compounded rates.

As an estimate of the volatility ($\sigma$) we calculated the historical 30-day-volatility using

$$\sigma = s \cdot \sqrt{252}, \tag{10}$$

where $s$ is the standard deviation of the returns for the close-to-close DAX levels of the most recent 30 days. We chose the 30-day-volatility since it showed the highest similarity to the German volatility index VDAX.[13] The factor 252 corresponds to the number of trading days in 1994.

# 5 Results

## 5.1 Optimal Network Architectures

We will now present the network architectures which arose from our specification strategies outlined in section 2. Model selection and estimation were carried out on a subsample consisting of the observations in the first nine months of 1994, a total of 10,848 data records. The remaining 2,828 observations — corresponding to the last three months of 1994 — were held back in order to evaluate the out-of-sample performance of the competing models. During the selection process all tests were run on a 5% significance level. For estimation purposes, we scaled our data to a mean of zero and a variance of one and then rescaled them for comparison of the different models.

Figure 2 shows the resulting architecture of a pure network model as it was defined in equation (8). This architecture results independently of which additional hidden unit LM-test was applied, the one of White (1989c) or the one of Teräsvirta/Lin/Granger (1993). The network consists of three hidden units, of which none is fully connected.

[10] The DAX data also stem from the Deutsche Börse AG, Frankfurt a.M.

[11] Since the DAX is calculated every minute, but updated only when there are changes in the level, we used the last published value before the transaction took place.

[12] The data was supplied by the Deutsche Finanzdatenbank, Mannheim.

[13] The VDAX is a volatility index which represents the average implied volatility of the DAX options.

We further provide network weights and pseudo[14] $t$-values, the latter in brackets. As one would expect from theory, all input variables significantly contribute to the pricing mechanism.

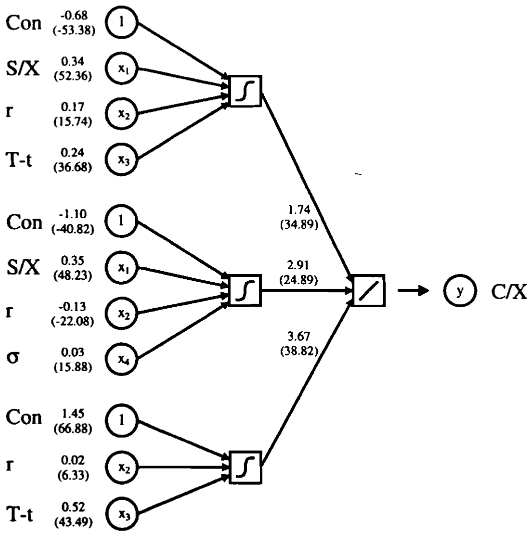


*Figure 2:* Optimal network architecture with four input variables and three hidden units (NN43). The numbers are the estimated weight values with corresponding $t$-values in brackets.

It is interesting to note, that the network model selected by statistical tests is markedly more parsimonious than the ones of Hutchinson/Lo/Poggio (1994) and Lajbcygier et al. (1995), who chose four to ten fully connected hidden units. In particular, our network architectures were not restricted to be fully connected, since the selection strategy tests for both the significance of the hidden units and the significance of single input variables.

An important question is whether further input variables improve the pricing accuracy significantly. As mentioned in section 3 the index level $S$ should have no explanatory power if the index's return distribution is independent of its level.

[14]The term "pseudo" accounts for the fact that the $t$-values do not actually obey a $t$-distribution. Inference relies on the asymptotic normality of the network weights.

Nonetheless some structure can be found. As shown in figures 3 and 4 there exists a negative relation between $S$ and the pricing errors of both the Black/Scholes model and our first neural network model (NN43).

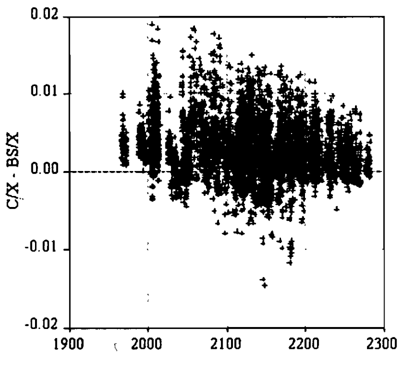


*Figure 3:* Pricing error of the Black/Scholes model plotted against the index level $S$.

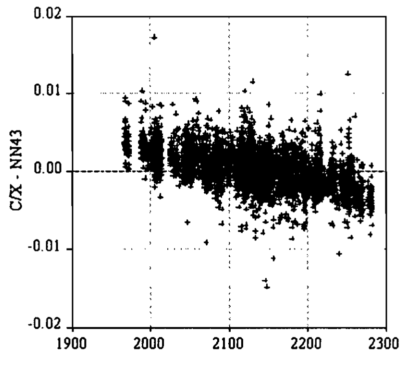


*Figure 4:* Pricing error of the Network NN43 plotted against the index level $S$.

When $S$ is considered as a further input variable, the selection strategy chooses the network shown in figure 5.[15] The index level turns out to have significant connections with two hidden units. Otherwise the architecture is very similar to that of figure 2.

An explanation for the significance of $S$ is difficult to provide, though the observed relation may indicate that expectations concerning the trend or the volatility of the stock market are influenced by "relatively" high or low index levels, resulting in some risk premium in the option prices. In any case, further research is needed into this subject, in particular into the question as to whether the same level effect can be found for other option markets and different time periods.

The third pricing model introduced in section 3 is the Black/Scholes model nested in a neural network. The specification of the network part provides information on which input variables can improve the explanation of observed prices in addition to the theoretical formula. Although in the first step of our specification strategy the LM-tests showed a significant hidden unit, the optimization algorithm did not converge when all five inputs ($S/X$, $r$, $T-t$, $\sigma$, $S$) were included. Thus we tested each of the input variables separately and excluded those which showed little or no significance. The resulting network is shown in figure 6. It contains one hidden unit and the three variables $r$, $\sigma$ and $S$. A second hidden unit was not accepted by the model selection strategy.

[15]Note that the model selection strategy again led to identical specifications independent of which LM-test we applied.

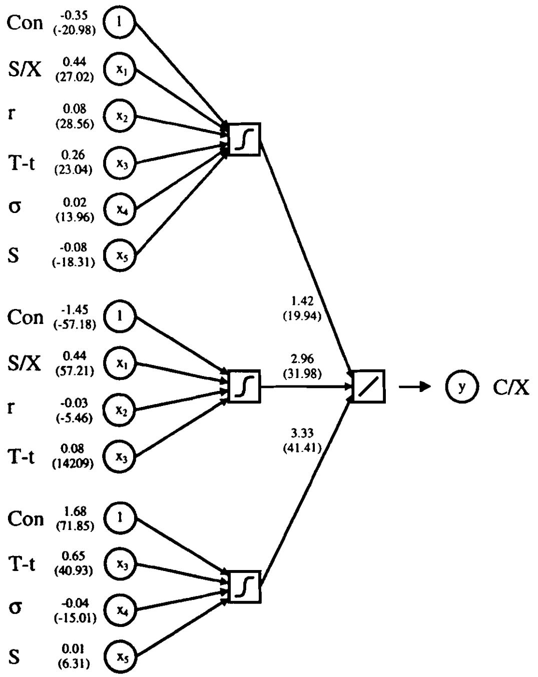


*Figure 5:* Optimal network architecture with $S$ as an additional input variable (NN53).

From this model we can draw the following conclusion. The Black/Scholes model matches the functional relationship between the call price and $S/X$ as well as $(T - t)$ up to very small deviations. This seems reasonable, as these inputs are readily available when calculating call prices. Thus, much of the deviation between observed and Black/Scholes prices seems to stem from a wrong assessment of the remaining variables $r$ and $\sigma$.

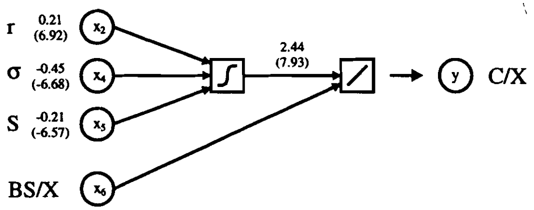


*Figure 6:* Optimal architecture of a network nesting the Black/Scholes model (BS+NN31).

## 5.2 Pricing Accuracy

To compare observed prices[16] with those obtained from the different models, the following measures of fit were computed:

$$R^2 = \frac{\sum_{t=1}^{T}[(\widehat{C/X})_t - (\overline{C/X})]^2}{\sum_{t=1}^{T}[(C/X)_t - (\overline{C/X})]^2}$$

$$\text{RMSE} = \sqrt{\frac{1}{T}\sum_{t=1}^{T}[(C/X)_t - (\widehat{C/X})_t]^2}$$

$$\text{ME} = \frac{1}{T}\sum_{t=1}^{T}[(C/X)_t - (\widehat{C/X})_t]$$

$$\text{MAE} = \frac{1}{T}\sum_{t=1}^{T}|(C/X)_t - (\widehat{C/X})_t|$$

$$\text{MAPE} = \frac{1}{T}\sum_{t=1}^{T}\frac{|(C/X)_t - (\widehat{C/X})_t|}{|(C/X)_t|}$$

The measures aim to highlight different aspects of the pricing accuracy. While the $R^2$ provides a measure of correlation between observed and fitted option prices, the mean error (ME) indicates a pricing bias. The root mean squared error (RMSE) and the mean absolute error (MAE) give absolute measures of price discrepancy while the mean absolute percentage error (MAPE) judges the price differences relative to the price level. Table 1 shows the performance measures for both the estimation period January to September 1994 and the out-of-sample evaluation period October to December 1994.

Table 1 yields some general results. Firstly, all neural network models possess a better pricing accuracy in-sample than for the out-of-sample period. However, this can not be explained by overfitting as the same result holds for the Black/Scholes model as well. On the contrary, the superior performance of the neural network models, in-sample as well as out-of-sample, indicate the quality of the statistical model selection approach. Secondly, the mean error is in general quite small. As none of the values is significantly different from zero, the models show no pricing bias.

[16]The prices refer to $C/X$, the call price divided by the corresponding strike price.

For two of the four models we obtain a clear ranking. Black/Scholes prices are in general the least accurate as they show the highest RMSE, MAE, MAPE and the lowest $R^2$ for both periods. The values obtained from the five input neural network (NN53) on the other hand, are closest to the observed prices with regard to all performance measures.

| in-sample | RMSE | ME | MAE | MAPE | $R^2$ |
|---|---|---|---|---|---|
| BS: | 0.0040 | 0.0025 | 0.0031 | 0.1764 | 0.9661 |
| NN43: | 0.0021 | 0.0000 | 0.0017 | 0.1096 | 0.9901 |
| NN53: | 0.0016 | 0.0000 | 0.0012 | 0.0754 | 0.9948 |
| BS+NN31: | 0.0020 | 0.0000 | 0.0015 | 0.1011 | 0.9912 |
| out-of-sample | RMSE | ME | MAE | MAPE | $R^2$ |
| BS: | 0.0049 | 0.0023 | 0.0037 | 0.2200 | 0.9302 |
| NN43: | 0.0025 | 0.0012 | 0.0018 | 0.1154 | 0.9823 |
| NN53: | 0.0022 | 0.0002 | 0.0018 | 0.1217 | 0.9853 |
| BS+NN31: | 0.0028 | 0.0000 | 0.0021 | 0.1493 | 0.9774 |

*Table 1:* Performance measures of competing models.

It is an interesting question why the combined network showed a worse performance than the pure network models. In our view, this may stem from effects in the variables $S/X$ and $(T-t)$ such as the smile–effect or the volatility–skew.[17] These effects are present in our data, but apparently not strong enough to require a further hidden unit in the combined network. In the more complex network structures, however, they might be implicitely modelled.

Nevertheless, all network models clearly dominate the Black/Scholes results. As this is true to the same extent for both time periods, the estimated relations seem to be stable over time. When looking at the magnitude of the improvement over the Black/Scholes model, the gain through the neural networks is considerable, for example the MAPE reduces from 22% to 12% for the best network.

## 5.3 Hedge Parameters

As option pricing models are frequently used to calculate hedge parameters, it is necessary to check whether the parameters obtained from the neural networks are reliable insofar as they follow certain patterns suggested by theory. The hedge parameters delta ($\Delta$), gamma ($\Gamma$) and theta ($\Theta$) of the neural network model (NN43) are shown in the figures 7 to 9.[18] For the computation of the derivatives, the volatility and interest rate were kept constant at $\sigma = 15\%$ and $r = 5\%$.

According to Cox/Rubinstein (1985) the value of a call is an increasing convex function of the stock price. Although this is not enforced by arbitrage, it is "true as an empirical

[17]The smile–effect is for example described in Tompkins (1994), pages 153–172 and the volatility–skew in Natenberg (1994), pages 405–418.

[18]Note that the derivatives were taken with respect to the normalized index value $S/X$.

fact".[19] Consequently, delta and gamma must always be positive, whereas delta should also be non-decreasing and only take values less than or equal to one.

As shown in figure 7 the delta-values fulfill these conditions for a large range of $S/X$ values. An exception are deep-in-the-money options ($S/X > 1.10$), where the deltas decrease with growing $S/X$. As gamma is the sensitivity of delta to changes in the stock price, it takes negative values in this region.

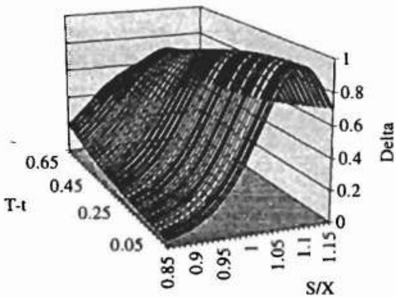


*Figure 7:* Network Delta

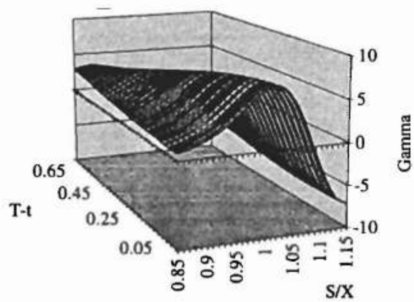


*Figure 8:* Network Gamma

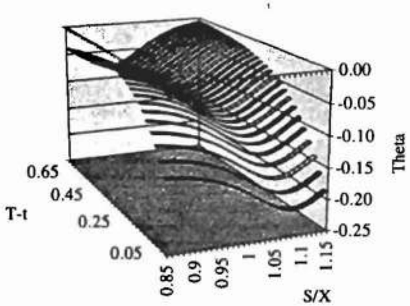


*Figure 9:* Network Theta

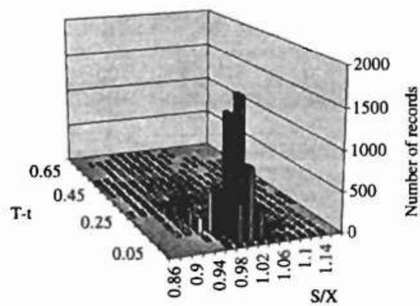


*Figure 10:* Distribution of records

In order to investigate whether this inconsistency comes from the data or from the network being unable to reproduce the derivatives appropriately, we estimated networks of similar complexity to model NN43 with simulated Black/Scholes prices.[20] As a result we obtained

[19] See Cox/Rubinstein(1985), page 156f.

[20] We used 21,150 Black/Scholes prices uniformly covering the area from $S/X = 0.85$, $r = 0\%$, $\sigma = 5\%$, $(T-t) = 0$ to $S/X = 1.15$, $r = 10\%$, $\sigma = 35\%$, $(T-t) = 0.75$.

hedge parameters similar to those of the Black/Scholes model that met all conditions mentioned above.

A plausible explanation for the delta- and gamma-deviations is provided by the distribution of our data with respect to $S/X$. Deep-in-the-money options are thinly traded even if time-to-maturity is short. Our data set thus contains very few observations in this region, which can be seen in figure 10. As the most liquid options are those at-the-money with a short time-to-maturity they weigh heavily when the networks are fitted.[21]

The theta must always be negative, since the value of an option decreases with diminishing time-to-maturity while keeping the other variables constant. Figure 9 confirms this for the thetas of the neural network model. Due to their high time value, at-the-money options correctly show the most negative thetas for the range of different maturities.[22]

In summary the hedge parameters of the network model follow the patterns suggested by theory, which provides a further check for the validity of the network approach. The performance of actual hedging strategies based on neural network hedge parameters, however, needs to be investigated in further research.

# 6 Summary and Conclusions

This paper shows that statistical inference techniques can successfully be applied to improve the pricing of options via neural networks. Networks are fitted to the normalized prices $C/X$ of call options written on the German stock index DAX. We adopt a network model selection strategy that is based on significance tests developed by White (1989b,c) and Teräsvirta/Lin/Granger (1993). This strategy leads to rather parsimonious networks which consist of only three hidden units that are not fully connected. Though all of the considered input variables $S/X$, $r$, $\sigma$ and $(T-t)$ show statistical significance.

Our statistical approach allows one to test for additional input variables in the network. It turns out that the index level $S$ has some additional explanatory power. As this finding can not be explained by traditional pricing models, further research is needed. In particular, it has to be investigated whether similar level effects are found in other data sets and what kind of economic explanation may stand behind it.

The estimated networks show a higher pricing accuracy with respect to the performance measures $R^2$, RMSE, MAE and MAPE than the theoretical model of Black and Scholes both in-sample and out-of-sample. This result indicates that restriction to significant hidden units and input connections helps both to avoid overfitting and to approximate stable functional relationships. Fitting a network to the residuals of the Black/Scholes model leads to additional significant contributions of $r$, $\sigma$ and $S$, which increase the pricing accuracy, though no further improvement to the pure network models is achieved.

As a final observation, the hedge parameters estimated from the networks turn out to be consistent with theory. This is promising for the performance of hedging strategies,

[21]The same effect can be observed in Hutchinson/Lo/Poggio (1994), figure 4, page 865 and figure 5, page 867, where the authors also obtained a decreasing delta for high $S/X$ values.

[22]Deep in-the-money options with short time-to-maturity are again an exception.

whose evaluation is a topic for future research. In summary the results are encouraging. In our view, the use of statistical methods for model specification and inference in neural networks is to be highly recommended when the aim of analysis is both to obtain an accurate description of the data and to learn about the underlying economic processes.

# References


Anders U., Korn O. (1996): *Model Selection in Neural Networks.* ZEW Discussion Paper 96-21.

Black F., Scholes M. (1973): *The Pricing of Options and Corporate Liabilities.* Journal of Political Economy, Vol 81, 673–654.

Cox J.C., Ross S.A. (1976): *The Valuation of Options for Alternative Stochastic Processes.* Journal of Financial Economics, Vol 3, 145–166.

Cox J.C., Rubinstein M. (1985): *Options Markets.* Prenctice-Hall, Englewood Cliffs, 1985.

Duan J.C. (1995): *The GARCH Option Pricing Model.* Mathematical Finance, Vol 5, 13–32.

Geske R. (1979): *The Valuation of Compound Options.* Journal of Financial Economics, Vol 7, 63–81.

Hornik K., Stinchcombe M., White H. (1989): *Multilayer Feedforward Networks are Universal Approximators.* Neural Networks, Vol 2, 359–366.

Hornik K., Stinchcombe M., White H. (1990): *Universal Approximation of an Unknown Mapping and Its Derivatives Using Multilayer Feedforward Networks.* Neural Networks, Vol 3, 551–560.

Hull J.C. (1993): *Options, Futures, and other Derivative Securities.* Prentice Hall.

Hull J.C., White A. (1987): *The Pricing of Options on Assets with Stochastic Volatilities.* The Journal of Finance, Vol 38, 281–300.

Hutchinson J.M., Lo A.W., Poggio T. (1994): *A Nonparametric Approach to Pricing and Hedging Derivative Securities Via Learning Networks.* The Journal of Finance, Vol 49, 851–889.

Lajbcygier P., Boek C., Palaniswami M., Flitman A. (1995): *Neural Network Pricing of all Ordinaries SPI Options on Futures.* Presented at the 3rd Conference on Neural Networks in the Capital Markets.

Lee T.H., White H., Granger C.W.J. (1993): *Testing for Neglected Nonlinearity in Time Series Models.* Journal of Econometrics, Vol 56, 269–290.

Malliaris M., Salchenberger L. (1993): *A Neural Network Model for Estimating Option Prices.* Journal of Applied Intelligence, Vol 3, 193–206.

Merton R.C. (1973): *Theory of Rational Option Pricing.* Bell Journal of Economics and Management Science, Vol 4, 141–183.

Merton R.C. (1976): *Option Pricing When Underlying Stock Returns are Discontinuous.* Journal of Financial Economics, Vol 3, 125–144.

Natenberg S. (1995): *Option Volatility & Pricing: Advanced Trading Strategies and Techniques.* Probus.

Poggio T., Girosi F. (1990): *Networks for Approximation and Learning.* Proceedings of the IEEE, Vol 78, No 9, 91–106.

Resnick B., Sheikh A.M., Song Y.-S. (1993): *Time Varying Volatilities and Calculation of the Weighted Implied Standard Deviation.* Journal of Financial and Quantitative Analysis, Vol 28, No 3, 417–430.

Rubinstein, M. (1983): *Displaced Diffusion Option Pricing.* The Journal of Finance, Vol 38, 213–217.

Rubinstein, M. (1985): *Nonparametric Tests of Alternative Option Pricing Models Using All Reported Trades and Quotes on the 30 Most Active CBOE Options Classes from August 23, 1976 through August 21, 1978.* The Journal of Finance, Vol 40, No 2, 455–480.

Sheikh A.M. (1991): *Transaction Data Tests of S&P 100 Call Option Pricing.* Journal of Financial and Quantitative Analysis, Vol 26, No 4, 459–475.

Teräsvirta T., Lin C.-F., Granger C.W. (1993): *Power of the Neural Network Linearity Test.* Journal of Time Series Analysis, Vol 14, No 2, 209–220.

Tompkins R (1994): *Options Explained* [2]. Macmillan.

White H. (1982): *Maximum Likelihood Estimation of Misspecified Models.* Econometrica, Vol 50, No 1, 1-25.

White H. (1989a): *Some Asymptotic Results for Learning in Single Hidden–Layer Feedforward Network Models.* Journal of the American Statistical Association, Vol 84, No 404, 1003–1013.

White H. (1989b): *Learning in Neural Networks: A Statistical Perspective.* Neural Computation, Vol 1, 425–464.

White H. (1989c): *An Additional Hidden Unit Test for Neglected Non-linearity in Multilayer Feedforward Networks.* Proceedings of the International Joint Conference on Neural Networks, Washington, DC. San Diego: SOS Printing, II, 451–455.

White H. (1994): *Estimation, Inference and Specification Analysis.* Cambridge University Press.

Xu X., Taylor S.J. (1994): *The Term Structure of Volatility Implied by Foreign Exchange Options.* Journal of Financial and Quantitative Analysis, Vol 29, No 1, 57–74.