Journal of Forecasting
J. Forecast. **17**, 369–388 (1998)

Improving the Pricing of Options: A Neural Network Approach

ULRICH ANDERS*, OLAF KORN AND CHRISTIAN SCHMITT
Centre for European Economic Research (ZEW), Mannheim, Germany

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 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. © 1998 John Wiley & Sons, Ltd.

KEY WORDS option pricing; neural networks; statistical inference; model selection

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 and 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 and Ross (1976), Geske (1979), Rubinstein (1983), Hull and 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 and Poggio (1994) and Malliaris and 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

* Correspondence to: Olaf korn, Centre for European Economic Research (ZEW), PO Box 103443, 68034 Mannheim, Germany. E-mail: korn@zew.de

Contract grant sponsor: German Research Foundation (DFG).

CCC 0277–6693/98/050369–20$17.50

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 non-linear 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 among others in Hornik, Stinchcombe and White (1989). First empirical results given in Hutchinson, Lo and Poggio (1994) and Malliaris and 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 (1989, a–c).

In this paper we adopt a model-selection strategy based on significance tests as suggested by Anders and 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 useful approach to model specification as used in Hutchinson, Lo and Poggio (1994), Lajbcygier *et al*. (1995) and Malliaris and Salchenberger (1993) is cross-validation. In cross-validation techniques, the in-sample data are 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 reduce in general 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 of theoretical pricing models.

The remainder of this paper is organized as follows: The next section briefly reviews some important results about statistical inference in neural networks and describes our architecture selection strategy. In the third section we introduce the option pricing models which are compared in this study. As a reference point we start with the Black/Scholes model, using both historical and implied volatility estimates. Then we consider pure neural networks chosen solely on statistical grounds. As a last specification the Black/Scholes model is nested in a neural network. The fourth section describes our data set while the fifth section 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.

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 includes 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 non-linear regression problems, such as

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

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

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 and are therefore well suited for modelling option prices, here we deal exclusively with the MLP type of neural networks. (Compare Poggio and Girosi, 1990, for RBF networks and Hornik, Stinchcombe and White, 1989, for MLP networks).

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:

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

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 non-linear transfer function attached to each hidden unit. Usually $g(\cdot)$ is either the logistic function or the tangent hyperbolic 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 and Poggio (1994) we focus exclusively on MLP networks for two reasons. First, it has been proved (Hornik, Stinchcombe and 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 with 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

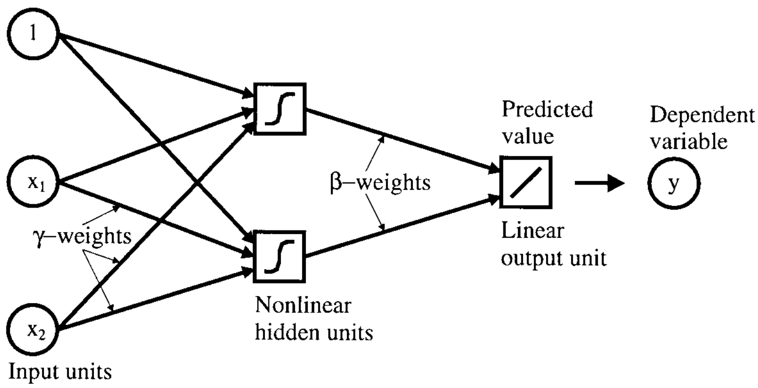


Figure 1. A multilayer perceptron neural network

to whether the estimated network pricing formula is consistent with some basic theoretical results. (For example, the option price should be a monotonically increasing function of the stock price).

Second, compared to the RBF network, the MLP network allows for the application of standard inference techniques known from parametric statistics. An application of such techniques may be possible for RBF networks as well. However, to our knowledge no work has 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 estimator of the network parameters follows an asymptotic normal distribution. 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 again is based on 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 $(1/T)C$ 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 $(1/T)C = (1/T)A^{-1}BA^{-1}$. 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 ∇ denotes the gradient, ∇^2 the hessian, and $\mathcal{L}_t$ the log-likelihood contribution of the tth 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 β_h equals zero, the corresponding weights γ_{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 and Granger (1993). The other was devised by Teräsvirta, Lin and 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 γ-weights of the additional hidden unit from a random distribution. This amounts to a random choice of the parameters in γ-space. The subsequent test is carried out conditional to the random values of γ. Teräsvirta, Lin and 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 and Korn (1996) which is based on the techniques of Teräsvirta, Lin and Granger (1993) and White (1989a–c).

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 overparameterized) neural net. To obtain statistically valid results, the strategy begins with the smallest model possible and successively adds hidden units. When the appropriate numbers of hidden units is determined, single input connections will successfully be removed in order to reach the optimal architecture.

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 procedure of Teräsvirta, Lin and Granger (1993) or White (1989c). If the hidden unit is not relevant the procedure stops. If the unit is relevant, it is included into the model. In this case the network is estimated and a further fully connected hidden unit tested for significance. The procedure continues until no further additional hidden unit shows relevance. After the number of hidden units is determined, Wald-tests are applied in a top-down approach to decide on the importance of single-input connections. If any connections are not significant on a 5% level, the one with the highest p-value is removed from the model and the reduced network retrained thereafter. This procedure is carried on until only significant connections remain in the model.

The proposed specification strategy is a kind of stepwise regression. As an alternative to hypothesis tests, specification methods based on information criteria, cross-validation or prediction errors could be applied. (Several criteria for regressor selection and their large-sample relations are discussed in Amemiya, 1980). These will in general be computationally more demanding. Moreover, the stepwise regression strategy outperformed AIC and cross-validation in the simultation study of Anders and Korn (1996).

One interesting relation between a stepwise regression strategy and an alternative criterion was established in Magee and Veall (1991). Using White's heteroscedasticity-consistent covariance matrix for the linear model they show that the exclusion of regressors with t-values smaller than 2 is asymptotically equivalent to the minimization of the prediction error sum of squares (PRESS) (see Allen, 1971). Once the number of hidden units is determined, our strategy coincides with the procedure of Magee and Veall (1991), except that it is applied to a non-linear network model.

It should be noted that stepwise regression techniques are subject to some criticism. Breiman (1995) reports a strong instability of models resulting from stepwise regressions in the sense that an observed in-sample performance will usually not be a good indicator of the out-of-sample performance. However, as opposed to Breiman's examples in this study the noise level is rather small and the data set quite large compared to the number of explanatory variables. This suggests that Breiman's criticism will be less important here. The results given later in this paper confirm this view.

In our model-selection strategy the specification of the final model depends on the output of preliminary tests of significance. Thus one has formed a pre-test estimator, which is discussed, for example, in Judge *et al.* (1988). Pre-test estimators are in general biased and in all but the simplest cases their sampling distribution can hardly be analysed analytically. However, this issue is less important in the context of inherently misspecified models like neural networks. The 'pseudo true' parameters can be estimated and asymptotically correct standard errors are available due to the work of White (1982, 1994). Thus, there exist valid standard errors for all network model specifications including the one finally chosen in our analysis.

OPTION PRICING MODELS

To judge the merits of network pricing, the performance of the neural network needs to be measured against an alternative model. Such a reference model should, as an estimated neural network does, produce a closed-form expression of the option price as to make it easily applicable in practice. In this study the Black/Scholes model (BS) is used for this purpose. Although several extensions and refinements of the model exist (see e.g. Hull, 1993, Chapter 17) 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 simplicity and robustness.

The derivation of the BS model 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 arbitrage 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

$$d_1 \equiv \frac{1\mathrm{n}(S/X) + (r + \sigma^2/2)(T-t)}{\sigma\sqrt{T-t}} \tag{5}$$

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

$S \equiv$ 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 option price C depends on five variables, namely the stock price S, the strike price X, the volatility σ, the interest rate r and the time to maturity $(T - t)$ of the contract. It is shown in 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 BS model, the number of input variables can be reduced to four by treating C/X as a function of S/X, σ, 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

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

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 BS 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 the previous section.

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

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

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 BS model 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 BS model.

An important task in practice is the hedging of option positions (see Hull, 1993, Chapter 13). 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:

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

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

THE DATASET

In our study we used transaction data on call options issued on the leading German stock index DAX. The index is composed of 30 major German stocks, selected with respect to market capitalization, turnover, and early availability of opening prices. The DAX is a capital-weighted performance index which is adjusted for stock splits, dividend markdowns,[1] 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 the Options Exchange (DTB). Since then it has developed into the most liquid option traded on the DTB.[2]

[1] In contrast to other indices, the adjustment for dividends is a particular feature of the DAX.
[2] The trading volume of DAX options is greater than that of all 20 DTB-traded stock options together.

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 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 to be considered for the calculation of option prices. Furthermore, as the DAX option is of European style, the standard BS model provides a suitable pricing formula.

Our dataset contains intraday time-stamped data on DAX call options traded on the DTB from January 1992 to the end of 1994.[3] Since this dataset 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 Rubenstein (1985), Sheikh (1991), Resnick, Sheikh and Song (1993) and Xu and Taylor (1994):

1. The call option is traded at less than 10 points.[4]
2. The option has less than 15 days to maturity.
3. The lower boundary condition for the value of European call options is violated:

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

4. The option is extremely deep-in- or deep-out-of-the-money:

$$\frac{S}{X} < 0{\cdot}85 \text{ or } \frac{S}{X} > 1{\cdot}15$$

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 (see Hull, 1993, p. 156). With criterion 4, extremely deep-in-the-money and deep-out-of-the-money options are excluded, as these options

[3] The dataset was provided by the Deutsche Börse AG, Frankfurt/Main.
[4] The value of 10 points leads to an exclusion of options which are traded at a price of less than 5% of the average DAX in 1994.

are traded roughly at their intrinsic value and have almost no informational content. Furthermore, the trading volume is very low for these options. Our resulting dataset consists of 13,676 observations.

To obtain a theoretical price according to the BS formula, we had to tie our option prices to an appropriate level of the DAX (S), the risk-free interest rate (r) and the return volatility (σ). In this respect, every transaction was linked with the current intraday DAX index level.[5] This means that each transaction between, say, 11:20 and 11:21 was combined with the DAX level of 11:20.[6]

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.[7] 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.

A crucial point is the volatility estimation since this is the only input variable of the BS model which cannot be observed directly. In accord with previous studies on network option pricing (see Hutchinson, Lo and Poggio, 1994; Lajbeygier *et al*., 1995), our first estimate is a historical volatility, computed according to

$$\sigma_{30} \equiv s \cdot \sqrt{252} \tag{10}$$

where s is the standard deviation of the returns for the close-to-close DAX levels of the most recent 30 days. The factor 252 corresponds to the number of trading days in 1994.

A poor volatility estimator can have particularly strong effects on the pricing accuracy of the BS model, as in contrast to the network no adjustment of the functional form is possible. There are two approaches to improve the volatility estimates in equation (10).

The first would be to stay with historical estimates while using more sophisticated methods based, for example, on GARCH or EGARCH models. This approach has the major disadvantage that it is not consistent with BS valuation. Instead the corresponding GARCH and EGARCH pricing models developed by Duan (1995) and Schmitt (1996) should be applied. However, in our view these are not well suited as reference models for this study as option prices have to be obtained via costly Monte Carlo simulations. Also such models rely on assumptions about investors' preferences.

The second approach is to use an implied volatility estimate deduced from observed option prices, which summarizes the market's view of future volatility. We follow this line here as the current market opinion should incorporate almost all information present in the historical return series. Our second volatility estimate is the DAX Volatility Index (VDAX), which was constructed by the DTB and reaches back to 1992.[8]

$$\sigma_{\text{VDAX}} \equiv \text{VDAX}. \tag{11}$$

The VDAX is a weighted average of volatilities implied by different DAX options traded at the DTB. It is published once a day at 1:30 p.m. Since the option prices used in this study refer to the time window 11:00 a.m. to 11:30 a.m., the VDAX of the previous day serves as the volatility estimate.

[5] The DAX data also stem from the Deutsche Börse AG, Frankfurt, a.M.
[6] 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.
[7] The data was supplied by the Deutsche Finanzdatenbank, Mannheim.
[8] Details on the construction of the VDAX can be found in Deutsche Börse (1995).

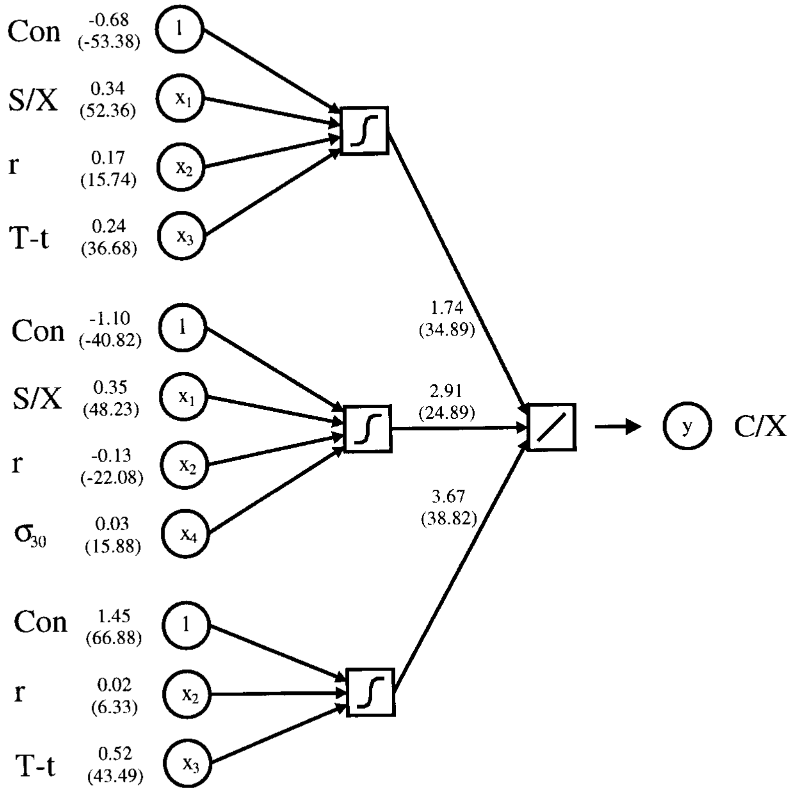


Figure 2. Optimal network architecture (NN1) with four input variables, three hidden units and historical volatility σ_{30}. The numbers are the estimated weight values with corresponding pseudo-*t*-values in brackets

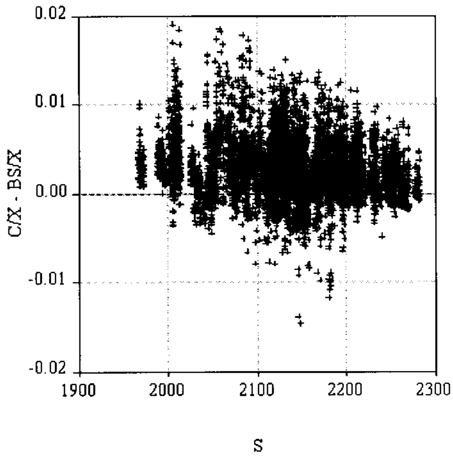
RESULTS

Optimal network architectures

We will now present the network architectures which arose from our specification strategies outlined in the second section. 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 architecture of a pure network model as it was defined in equation (8). This model employs the volatility estimates σ_{30}. The architecture results independently of which additional hidden unit LM-test was applied, the one of White (1989c) or of Teräsvirta, Lin and Granger (1993). The network consists of three hidden units of which none is fully connected.[9] We

[9] The computation time to specify this neural network was less than three hours on a personal computer with Pentium 133 processor.



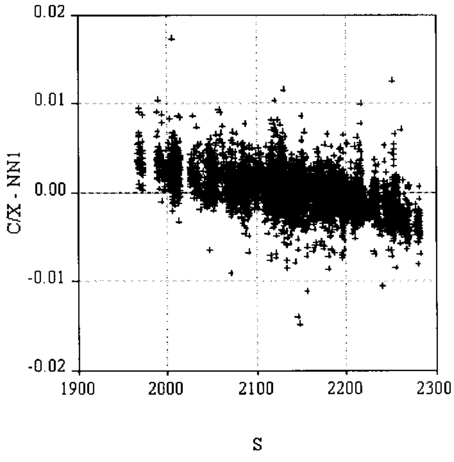


Figure 3. (a) Pricing error of the Black/Scholes model using historic volatilities plotted against the index level *S*. (b) Pricing error of the network NN1 using historical volatilities plotted against the index level *S*

further provide network weights and pseudo[10] *t*-values, the latter in parentheses. As one would expect from theory, all input variables significantly contribute to the pricing mechanism.

An important question is whether further input variables improve the pricing accuracy. As mentioned earlier, the index level *S* should have no explanatory power if the return distribution of the index is independent of its level. Nonetheless some structure can be found. As shown in Figures 3(a) and 3(b), there exists a negative relation between *S* and the pricing errors of both the BS model and our first neural network model (NN1).

When *S* is considered as a further input variable, the selection strategy chooses the network shown in Figure 4.[11] The index level turns out to have significant connections with two hidden units. On the one hand, this result could be viewed as an indication for a return distribution which depends on the index level. On the other hand, the significance of *S* can be due to a poor volatility estimate. In this case the network model may try to compensate for the poor estimate by including an additional input variable and by changing the functional form.

Figure 5 shows the pure network model where the VDAX is used as a volatility estimator. As in the models in Figures 2 and 4 the network architecture consists of three hidden units, where only one is fully connected. While all other explanatory variables remain in the model the index level is no longer statistically significant.

Figures 6(a) and 6(b) show the pricing errors of the BS model and the pure network plotted against *S*. The relations seen in Figures 3(a) and 3(b) have disappeared. This result highlights the importance of reliable volatility estimates.

It is interesting to note that the network models selected by statistical tests are markedly more parsimonious than the ones of Hutchinson, Lo and Poggio (1994) and Lajbcygier *et al.* (1995), who chose four to ten fully connected hidden units. In particular, our network architectures were

[10] 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.

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

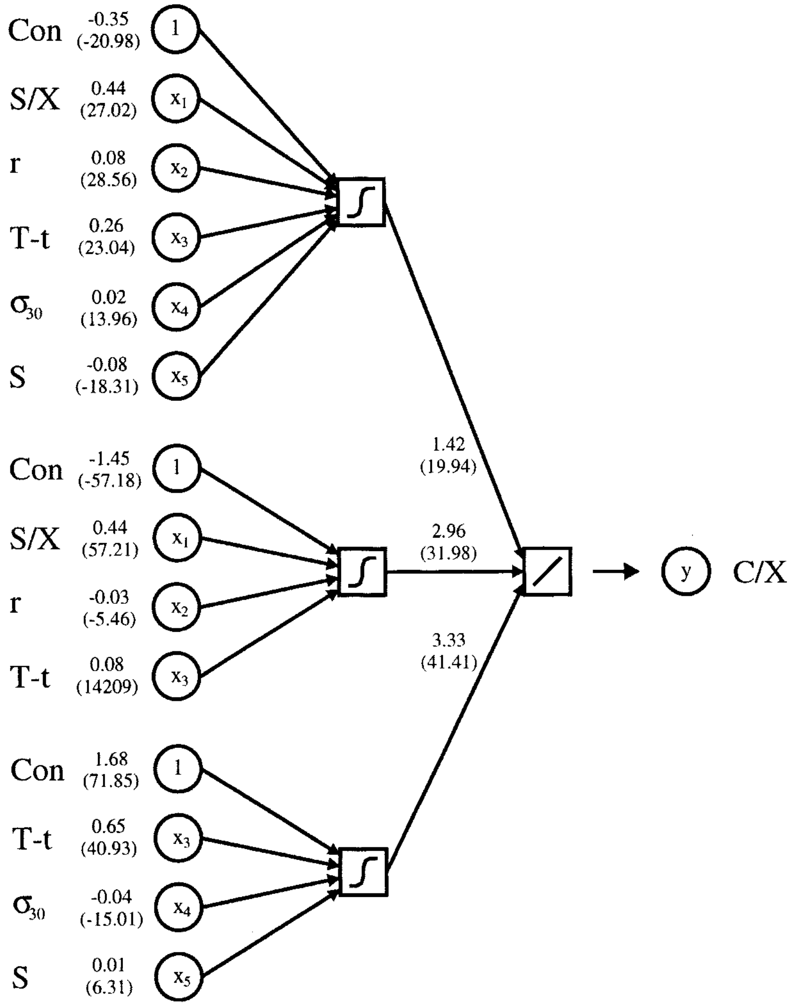


Figure 4. Optimal network architecture (NN2) with S as an additional input variable using historical volatility σ_{30}

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.

The third pricing model introduced above is the BS 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. The model resulting from our specification strategy is shown in Figure 7. Its network part contains only one hidden unit and two explanatory variables S/X and $T-t$. Thus much of the deviation between the observed and the BS prices is due to a wrong assessment of the variables S/X and $T-t$. This result can be interpreted as an attempt of the network model to account for some structures

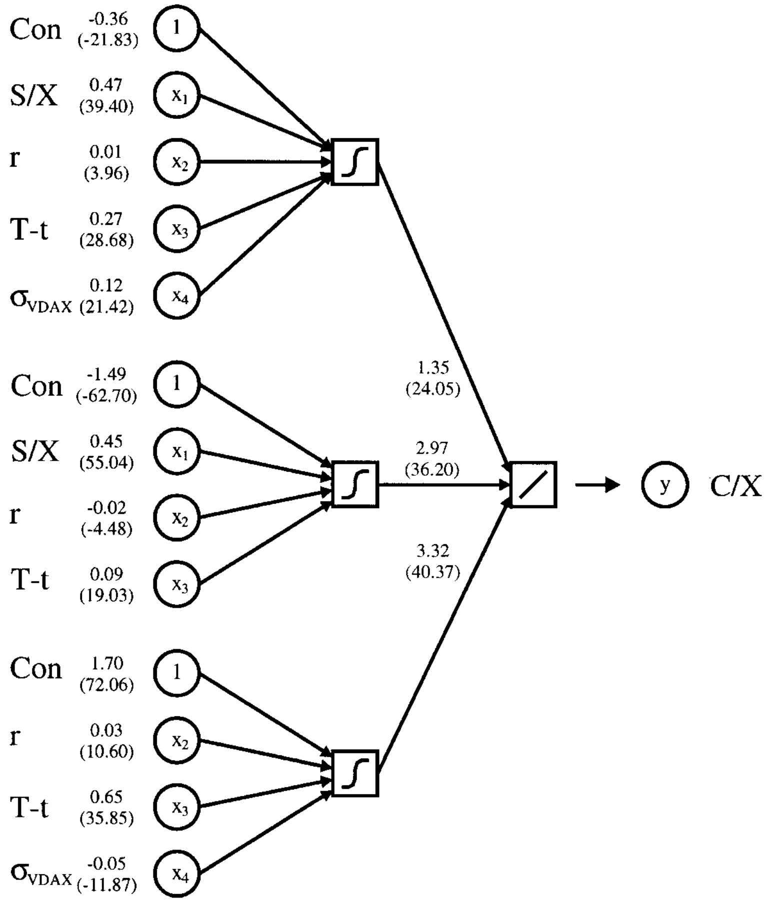


Figure 5. Optimal network architecture (NN3) using the VDAX as volatility estimator σ_{VDAX}

as the smile effect or the volatility skew[12] which are present in our data (compare Schmitt, 1996).

Pricing accuracy

It still remains to be examined how far the different network specifications and volatility estimates lead to a good out-of-sample pricing performance. To compare the observed prices with those obtained from the models, the following measures of fit were computed:

$$\bar{R}^2 = 1 - \left(\frac{T-1}{T-K}\right)\frac{\sum_{t=1}^{T}(C_t - \hat{C}_t)^2}{\sum_{t=1}^{T}(C_t - \bar{C}_t)^2}$$

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

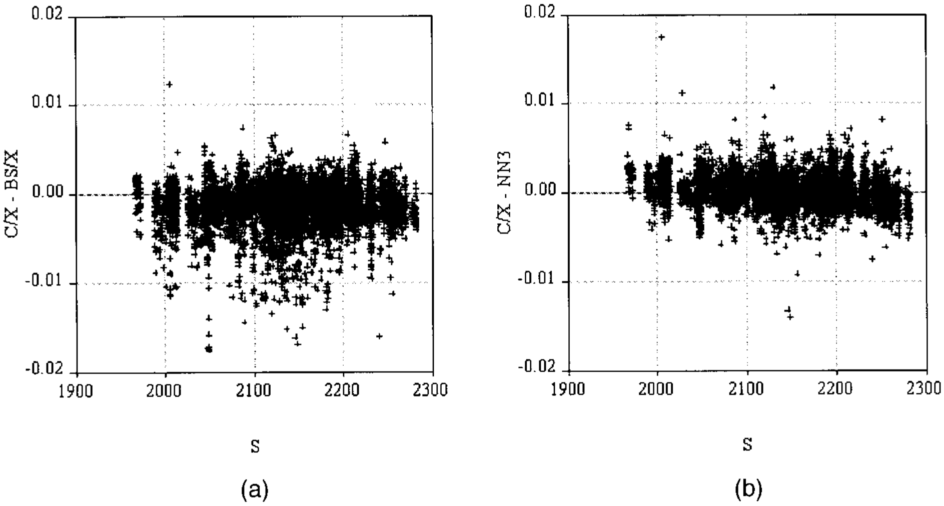


Figure 6. (a) Pricing error of Black/Scholes using the VDAX plotted against the index level S. (b) Pricing error of the network NN3 using the VDAX plotted against the index level S

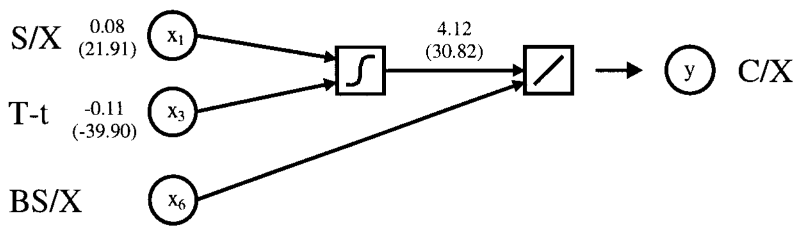


Figure 7. Optimal architecture of a network nesting the Black/Scholes model BS + NN4 using the VDAX as volatility estimator σ_{VDAX}

$$\mathrm{RMSE} = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(C_t - \hat{C}_t)^2}$$

$$\mathrm{ME} = \frac{1}{T}\sum_{t=1}^{T}(C_t - \hat{C}_t)$$

$$\mathrm{MAE} = \frac{1}{T}\sum_{t=1}^{T} | C_t - \hat{C}_t |$$

$$\mathrm{MAPE} = \frac{\frac{1}{T}\sum_{t=1}^{T} | C_t - \hat{C}_t |}{| C_t |}$$

The measures aim to highlight different aspects of pricing accuracy. They were computed for call prices C instead of standardized prices C/X to facilitate the economic interpretation.

Table I. Performance measures of competing models

	RMSE	ME	MAE	MAPE	R^2
In-sample					
BS (σ_{30})	8.5942	−5.3060	6.5940	0.1764	0.9616
BS(σ_{VDAX})	5.6591	−3.1074	4.0928	0.1200	0.9833
NN1 (σ_{30})	4.6549	−0.0918	3.6807	0.1096	0.9887
NN2 (σ_{30})	3.3512	−0.0029	2.5854	0.0754	0.9942
NN3 (σ_{VDAX})	3.4698	−0.0301	2.6917	0.0809	0.9937
BS + NN4 (σ_{VDAX})	3.6931	−0.0410	2.7378	0.0760	0.9929
Out-of-sample					
BS (σ_{30})	10.2375	4.7466	7.7875	0.2110	0.9217
BS (σVDAX)	5.7610	−3.3600	3.7811	0.1216	0.9752
NN1 (σ_{30})	5.1562	2.4421	3.7781	0.1154	0.9800
NN2 (σ_{30})	4.7493	−0.3882	3.7246	0.1218	0.9831
NN3 (σ_{VDAX})	2.7558	0.5973	2.0437	0.0654	0.9943
BS + NN4 (σ_{VDAX})	3.6835	−0.0369	2.2458	0.0636	0.9899

The R^2 provides a measure of correlation between observed and fitted option prices. We chose the adjusted version here to take into account the different numbers K of parameters estimated in different models. Especially for the in-sample-period, this allows a more appropriate comparison between the BS and the network model than the ordinary R^2. Some indication of pricing bias is given by the mean error (ME). The root mean square 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 I shows the performance measures for both the estimation period January to September 1994 and the out-of-sample evaluation period October to December 1994.

First, we will look at the in-sample results. Concerning all performance measures the network models achieve a better pricing accuracy than the BS models. This holds irrespective of the volatility estimate. Nevertheless, the choice of the volatility estimate plays an important role. In the BS model the VDAX leads to smaller pricing errors than the historical volatility and thus is to be preferred. When σ_{30} is used, the corresponding network allocates the additional input variable S with which it reaches a similar in-sample performance as the models NN3(σ_{VDAX}) and BS + NN4 (σ_{VDAX}). However, such additional complexity bears the danger of data fitting and could worsen the out-of-sample results.

The out-of-sample performance is the relevant yardstick for a comparison of different models. Here the models NN3(σ_{VDAX}) and BS + NN4 (σ_{VDAX}) are clearly the most accurate as they show the lowest RMSE, MAE, MAPE and the highest R^2. Their errors are even smaller out-of-sample than in-sample, which indicates that the statistical model-selection approach successfully avoided overfitting. The performance of the most complex neural network NN2 (σ_{30}) markedly depreciated from in-sample to out-of-sample and produces an even higher MAPE than the network NN1 (σ_{30}) with only four inputs. This suggests that the additional input S does not actually contribute to the explanation of call prices. The apparent overparameterization of the network NN2 (σ_{30}), however, must be attributed to the poor volatility estimate. In general, the differences between the models based on σ_{30} and σ_{VDAX} are more pronounced in the out-of-sample period, where σ_{VDAX} leads to far better results.

When looking at the magnitude of the improvement over the BS model, the gain through neural networks is considerable. The MAPE reduces from more than 12% to about 6.5%. Referring to market participants, the bid–ask spread for DAX options at the DTB is usually far smaller than 5% of the option price. Thus one can speak of an improvement which is also economically significant. It should be noted that such an improvement is quite easily obtained. As the results for the model BS + NN4 (σ_{VDAX}) show, it suffices to fit a network with as few as one hidden unit and two explanatory variables to the BS residuals. This achieves the same pricing accuracy as the more complex pure network NN3(σ_{VDAX}). Due to the simple network part such mixtures of theoretical pricing models and neural networks are very promising.

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 (Δ), gamma (Γ) and theta (Θ) of the neural network model NN3(σ_{VDAX}) are shown in Figures 8(a) to 8(d).[13] For computation of the derivatives, the volatility and interest rate were kept constant at $\sigma = 15\%$ and $r = 5\%$.

According to Cox and 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 fact' (Cox and Rubinstein, p. 156). 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 8(a), the delta-values fulfil 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.

In order to investigate whether this inconsistency comes from the data or from the network being unable to reproduce the derivatives appropriately, we fitted networks of similar complexity as model N3 to simulated BS prices.[14] As a result we obtained hedge parameters similar to those of the BS 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 dataset thus contains very few observations in this region which can be seen in Figure 8(d). As the most liquid options are those at-the-money with a short time-to-maturity they weigh heavily when the networks are fitted.[15]

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 8(c) 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. (Deep-in-the-money options with short time-to-maturity are again an exception).

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

[13] Note that the derivatives were taken with respect to the normalized index value S/X.
[14] We used 21,150 BS prices uniformly covering the area from $S/X = 0.85$, $r = 0\%$, $\sigma = 5\%$, $(T - t) = 0$ to $S/X = 1.15$, $r = 10\%$, $\sigma = 35\%$, $(T - t) = 0.75$.
[15] The same effect can be observed in Hutchinson, Lo and Poggio (1994, Figure 4, p. 865 and Figure 5, p. 867) where the authors also obtained a decreasing delta for high S/X values.

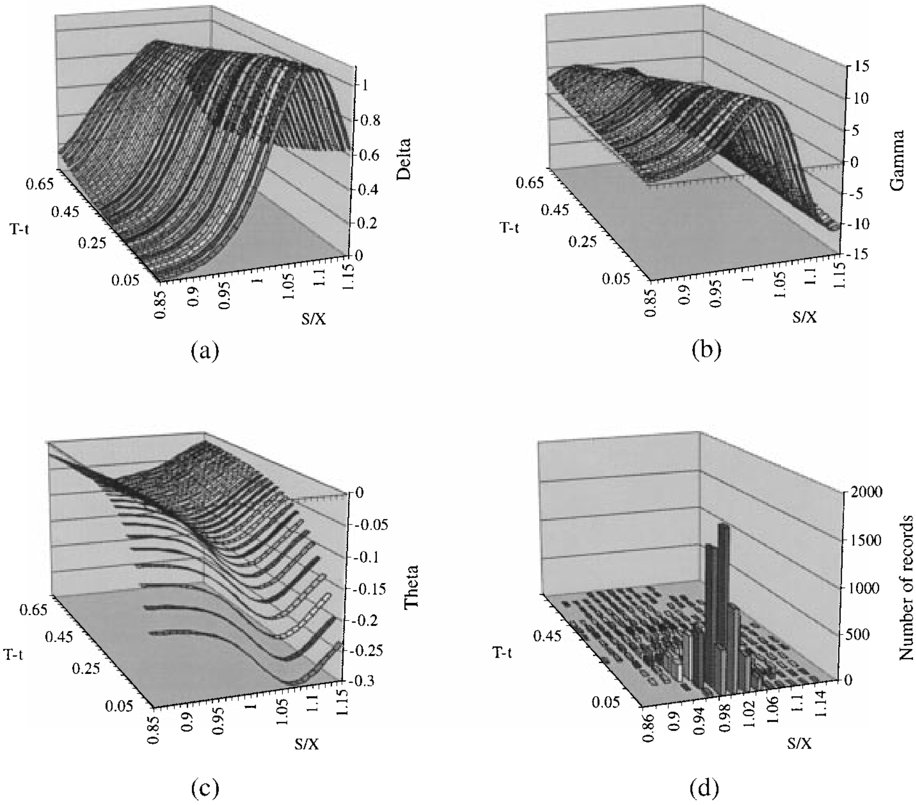


Figure 8 (a) Network Delta. (b) Network Gamma. (c) Network Theta. (d) Distribution of records

performance of actual hedging strategies based on neural network hedge parameters, however, needs to be investigated in future research.

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 (1989 b,c) and Teräsvirta, Lin and 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 variables suggested by theory, S/X, r, σ and $(T-t)$, show statistical significance.

Our statistical approach allows us to test for additional input variables in the network. It turns out that the index level S has some additional explanatory power when used in connection with a historical volatility estimate. When implied volatility is employed as the input variable, the index level shows no significance. Moreover, the use of implied volatility improves the out-of-sample results for both the BS and the network models.

The estimated networks show a higher pricing accuracy with respect to the performance measures R^2, RMSE, MAE and MAPE than the model of Black and Scholes both in-sample and out-of-sample. This result indicates that a restriction to significant hidden units and input connections helps both to avoid overfitting and to establish stable functional relationships. Fitting a network to the residuals of the BS model leads to additional significant contributions of S/X and $T-t$. The resulting pricing formula achieves a similar performance as the pure network NN3(σ_{VDAX}). This is particularly promising as the network part in the combined model is extremely simple. Therefore, the computational burden is reduced and the interpretation of the model facilitated.

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, 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.

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. We also like to thank two anonymous referees for valuable comments and suggestions.

REFERENCES

Allen, D. M., 'The prediction sum of squares as a criterion for selecting prediction variables', *Technical Report 23* (1971), University of Kentucky, Department of Statistics.

Amemiya, T., 'Selection of regressors', *International Economic Review*, **21** (1980), 331–354.

Anders, U. and Korn, O., 'Model selection in neural networks', ZEW Discussion Paper 96-21 (1996).

Black, F. and Scholes, M., 'The pricing of options and corporate liabilities', *Journal of Political Economy*, **81** (1973), 673–654.

Breiman, L., 'Better subset regression using the nonnegative garotte', *Technometrics*, **37** (1995), 373–384.

Cox, J. C. and Ross, S. A., 'The valuation of options for alternative stochastic processes', *Journal of Financial Economics*, **3** (1976), 145–166.

Cox, J. C. and Rubinstein, M., *Options Markets*, Englewood Cliffs, NJ: Prentice Hall, 1985.

Deutsche Börse, 'The DAX volatility index', Information Brochure, Deutsche Börse AG, (1995), Frankfurt/Main.

Duan, J. C., 'The GARCH option pricing model', *Mathematical Finance*, **5** (1995), 13–32.

Franke, J., 'Derivative products', in Walter, N. and von Rosen, R. (eds), *German Financial Markets*, Gresham Books, 1995, 186–215.

Geske, R., 'The valuation of compound options', *Journal of Financial Economics*, **7** (1979), 63–81.

Hornik, K., Stinchcombe, M. and White, H., 'Multilayer feedforward networks are universal approximators', *Neural Networks*, **2** (1989), 359–366.

Hornik, K., Stinchcombe, M. and White, H., 'Universal approximation of an unknown mapping and its derivatives using multilayer feedforward networks', *Neural Networks*, **3** (1990), 551–560.

Hull, J. C., *Options, Futures, and other Derivative Securities*, Englewood Cliffs, NJ: Prentice Hall, 1993.

Hull, J. C. and White, A., 'The pricing of options on assets with stochastic volatilities', *The Journal of Finance*, **38** (1987), 281–300.
Hutchinson, J. M., Lo, A. W. and Poggio, T., 'A nonparametric approach to pricing and hedging derivative securities via learning networks', *The Journal of Finance*, **49** (1994), 851–889.
Judge, G. G., Hill, R. C., Griffiths, W. E., Lütkepohl, H. and Lee, T. C., *Introduction to the Theory and Practice of Econometrics*, 2nd edition, New York: John Wiley, 1988.
Lajbcygier, P., Boek, C., Palaniswami, M. and Flitman, A., 'Neural network pricing of all ordinaries SPI options on futures', in Refenes, A.-P., Abu-Mostafa, Y., Moody, J., Weigend, A. (eds), *Neural Networks in Financial Engineering*, New York: World Scientific, 1995, 64–77.
Lee, T. H., White, H. and Granger, C. W. J., 'Testing for neglected nonlinearity in time series models', *Journal of Econometrics*, **56** (1993), 269–290.
Magee, L. and Veall, M. R., 'Selecting regressors for prediction using PRESS AND White *t* statistics', *Journal of Business & Economics Statistics*, **9** (1991), 91–96.
Malliaris, M. and Salchenberger, L., 'A neural network model for estimating option prices', *Journal of Applied Intelligence*, **3** (1993), 193–206.
Merton, R. C., 'Theory of rational option pricing', *Bell Journal of Economics and Management Science*, **4** (1973), 141–183.
Merton, R. C., 'Option pricing when underlying stock returns are discontinuous', *Journal of Financial Economics*, **3** (1976), 125–144.
Natenberg, S., *Option Volatility & Pricing: Advanced Trading Strategies and Techniques*, New York: Probus, 1994.
Poggio, T. and Girosi, F., 'Networks for approximation and learning', *Proceedings of the IEEE*, **78** (1990), No. 9, 91–106.
Resnick, B., Sheikh, A. M. and Song, Y.-S., 'Time varying volatilities and calculation of the weighted implied standard deviation', *Journal of Financial and Quantitative Analysis*, **28** (1993), No. 3, 417–430.
Rubinstein, M., 'Displaced diffusion option pricing', *The Journal of Finance*, **38** (1983), 213–217.
Rubinstein, M., '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*, **40** (1985), No. 2, 455–480.
Sheikh, A. M., 'Transaction data tests of S&P 100 call option pricing', *Journal of Financial and Quantitative Analysis*, **26** (1991), No. 4, 459–475.
Schmitt, C., 'Option pricing using EGARCH models', in Albrecht, P. (ed.), *Aktuarielle Ansätze für Finanz-Risiken*, Verlag Versicherungswirtschaft, 1996, 1311–1336.
Teräsvirta, T., Lin, C.-F. and Granger, C. W., 'Power of the neural network linearity test', *Journal of Time Series Analysis*, **14** (1993), No.^{2}, 209–220.
Tompkins, R., *Options Explained*, New York: Macmillan, 1994.
Walter, N. and von Rosen, R. (eds), *German Financial Markets*, Gresham Books, 1995.
White, H., 'Maximum likelihood estimation of misspecified models', *Econometrica*, **50** (1982), No. 1, 1–25.
White, H., 'Some asymptotic results for learning in single hidden-layer feedforward network models', *Journal of the American Statistical Association*, **84** (1989a), No. 404, 1003–1013.
White, H., 'Learning in neural networks: a statistical perspective', *Neural Computation*, **1** (1989b), 425–464.
White, H., An additional hidden unit test for neglected non-linearity in multilayer feedforward networks', *Proceedings of the International Joint Conference on Neural Networks* (1989c), Washington, DC. San Diego: SOS Printing, II, 451–455.
White, H., *Estimation, Inference and Specification Analysis*, New York: Cambridge University Press, 1994.
Xu, X. and Taylor, S. J., 'The term structure of volatility implied by foreign exchange options', *Journal of Financial and Quantitative Analysis*, **29** (1994), No. 1, 57–74.

Authors' biographies:
Dr Ulrich Anders studied Economic Engineering at the University of Karlsruhe, Germany and completed an MBA course at the University of Kent, UK. He received his PhD in the field of neural networks and statistics from the University of Karlsruhe. He has been working at the Centre for European Economic Research (ZEW) and the University of Otago, New Zealand. Currently, he is with Deutsche Bank Frankfurt.

Olaf Korn studied Statistics and Economics at the Universities of Dortmund, Sheffield and Basel. He is a Research Fellow at the Centre for European Economic Research (ZEW), Mannheim.

Christian Schmitt studied Financial Engineering at the University of Karlsruhe. He is a Research Fellow at the Centre for European Economic Research (ZEW), Mannheim.

Authors' address:
Ulrich Anders, Olaf Korn and **Christian Schmitt**, Centre for European Economic Research (ZEW), PO Box 10 34 43, 68034 Mannheim, Germany.

